

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

Entwicklung von klinischem Denken bei
Medizinstudierenden im Laufe des
klinischen Studienabschnittes

INAUGURAL-DISSERTATION

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

vorgelegt von

Insa Amelie de Temple, geb. Frisch

aus

Hattingen

Göttingen 2021

Dekan: Prof. Dr. W. Brück

Betreuungsausschuss

Betreuer: Prof. Dr. T. Raupach, MME

Ko-Betreuerin: PD Dr. S. Sennhenn Kirchner, MME

Prüfungskommission

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

Ko-Referent/in:

Drittreferent/in:

Datum der mündlichen Prüfung:

Hiermit erkläre ich, die Dissertation mit dem Titel "Entwicklung von klinischem Denken bei Medizinstudierenden im Laufe des klinischen Studienabschnittes" eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den

(Unterschrift)

Die Daten, auf denen die vorliegende Arbeit basiert, wurden teilweise publiziert und vorgestellt.

Raupach T, de Temple I, Middeke A, Anders S, Morton C, Schuelper N (2021): Effectiveness of a serious game addressing guideline adherence: cohort study with 1.5-year follow-up. BMC Med Educ. 21, 189

Effects of using a serious game on clinical reasoning in medical students in different years of undergraduate education. Vortrag im Rahmen der AMEE Konferenz, Basel, 25.-29.08.2018

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis.....	V
1 Einleitung	1
1.1 Lernen	1
1.1.1 <i>Clinical reasoning</i>	1
1.1.2 <i>Dual-process</i> -Theorie	2
1.1.3 Fallbasiertes Lernen	6
1.2 <i>Serious games</i> in der Medizindidaktik.....	8
1.2.1 <i>Serious games</i>	8
1.2.2 Wie lässt sich <i>clinical reasoning</i> messen?.....	9
1.2.3 Das <i>serious game</i> EMERGE.....	10
1.3 Aufbau des Medizinstudiums an der Universität Göttingen.....	14
1.3.1 Implementierung von EMERGE im Curriculum der UMG	14
1.4 Erarbeitung der Fragestellung	16
2 Material und Methoden	18
2.1 Studiendesign	19
2.2 Proband:innenrekrutierung.....	20
2.3 Organisation und Durchführung.....	21
2.4 Messung von <i>clinical reasoning</i>	24
2.4.1 <i>Logfiles</i>	24
2.4.2 Übergabeprotokolle	26
2.4.3 <i>Key-feature</i> -Prüfung	27
2.4.4 <i>Think-aloud</i> -Interviews.....	27
2.5 Statistische Analysen.....	29
3 Ergebnisse.....	31
3.1 Zusammensetzung der Kohorten.....	31
3.3 Prospektiver Vergleich zwischen Modul 3.1 und 6.x im Laufe des SoSe 2017.....	33
3.3.1 <i>Logfile</i> -Auswertung.....	33
3.3.2 Übergabeprotokolle	37
3.3.3 KF	39
3.3.4 <i>Think-aloud</i> -Interviews.....	40
3.4 <i>Follow-up</i> des Moduls 3.1 nach eineinhalb Jahren im WiSe 2018/19.....	40
3.5 Vergleich zwischen EMERGE-erfahrenen und EMERGE-naiven Studierenden im WiSe 2018/19	43

4	Diskussion.....	45
4.1	Wesentliche Ergebnisse.....	45
4.1.1	Analyse 1.....	45
4.1.2	Analyse 2.....	48
4.1.3	Analyse 3.....	49
4.2	Stärken und Schwächen der Studie.....	50
4.3	Potential und Limitationen von <i>serious games</i> in der Medizindidaktik.....	52
4.4	Bedeutung der Studie für Forschung und Lehre.....	54
4.5	Ausblick: Offene Fragen und Vorschläge für weitere Studien.....	54
5	Zusammenfassung.....	55
6	Anhang.....	57
7	Literaturverzeichnis.....	60

Abbildungsverzeichnis

Abbildung 1: Ein Modell des <i>clinical reasoning</i> nach Eva (2005).....	4
Abbildung 2: Die Sicht auf die virtuelle Notaufnahme EMERGE mit geöffneten Pager-Benachrichtigungen.	11
Abbildung 3: Anamnesefunktion im <i>serious game</i> EMERGE.....	12
Abbildung 4: Das Archiv zum Einsehen des Feedbacks zu jedem abgeschlossenen Fall.....	13
Abbildung 5: Veranschaulichung des Studiendesigns und der Forschungsfragen.....	19
Abbildung 6: Verteilung der Spieltermine in den beiden Modulen.....	21
Abbildung 7: Fallvorkommen pro Spieltermin.....	22
Abbildung 8: Methodische Auswertung pro Spieltermin.....	23
Abbildung 9: Beispiel eines <i>Logfiles</i> des Falles NSTEMI.....	24
Abbildung 10: Kriterien zur Auswertung der <i>Logfiles</i> für die drei Fälle.	25
Abbildung 11: Fallvignette „Aortenklappenstenose“ aus der <i>Key-feature</i> -Prüfung.....	27
Abbildung 12: Zusammensetzung der EMERGE-erfahrenen Kohorte im WiSe 18/19.....	31
Abbildung 13: Zusammensetzung der Kontrollgruppe „EMERGE-naiv“.....	32
Abbildung 14 a-c: Analyse 1: Veränderung der Spielleistung im Modul 3.1 im SoSe 2017 zwischen Termin zwei und Termin sechs in beiden Modulen.	35
Abbildung 15: Prozentpunktzahlen der in den Übergabeprotokollen erwähnten CR- <i>items</i> zu Termin zwei und sechs.	38
Abbildung 16: Die erreichten Punkte in der <i>Key-feature</i> -Prüfung	39
Abbildung 17: Erreichte Prozentpunktzahlen in der <i>Logfile</i> -Auswertung zu T2, T6 und T80 als Gesamtpunktzahl sowie aufgeteilt in Anamnese und Management.....	41
Abbildung 18 a; b: Vergleich der erreichten Prozentpunktzahlen der EMERGE-erfahrenen (Intervention) und EMERGE-naiven (Kontrollgruppe) Studierenden in der <i>Logfile</i> - Auswertung.....	44

Tabellenverzeichnis

Tabelle 1: Vergleich von intuitiven und analytischen Herangehensweisen an Entscheidungsprozesse nach Croskerry (2009).....	3
Tabelle 2: Charakteristika der intuitiven und analytischen Ansätze bei Entscheidungsprozessen nach Croskerry (2009).....	4
Tabelle 3: Erreichte Prozentpunktzahlen der Module 3.1 und 6.x zu Termin zwei und sechs.	33
Tabelle 4: Erreichte Prozentpunktzahlen beider Module zusammengefasst zu Termin zwei und sechs.....	36
Tabelle 5: Verlegungen beider Module zusammengefasst.	37
Tabelle 6: Bewertung der Übergabeprotokolle: Häufigkeiten der einzelnen Checklisten-Punkte aufgeteilt nach Modul und Spieltermin	38
Tabelle 7: Erreichte Punktzahlen, aufgeteilt nach Modul und Spieltermin.....	40
Tabelle 8: Ergebnisse der <i>Logfile</i> -Auswertung zu T2, T6 und T80	41
Tabelle 9: Erreichte Prozentpunktzahlen in der <i>Logfile</i> -Auswertung zu T2, T6 und T80 aufgeteilt nach Fällen sowie den Kategorien Anamnese und Management	42
Tabelle 10: Vergleich der EMERGE-erfahrenen (Intervention) und EMERGE-naiven Studierenden (Kontrollgruppe)	43

Abkürzungsverzeichnis

AVNRT	Atrioventrikuläre <i>Re-Entry</i> -Tachykardie
BGA	Blutgasanalyse
COPD	<i>Chronic obstructive pulmonary disease</i>
CR	<i>Clinical reasoning</i>
CT	Computertomographie
IC	<i>Informed consent</i>
KF	<i>Key-feature</i>
KPL	Kardiopulmonale Lehre
LAE	Lungenarterienembolie
M 3.1	Modul 3.1
M 6.x	Modul 6.x
NSTEMI	<i>Non-ST-Elevation Myocardial Infarction</i>
OSCE	<i>Objective structured clinical examination</i>
POL	Problemorientiertes Lernen
POLEMA	Problemorientiertes Lernen - elektronische Medien assistiert
SG	<i>Serious game</i>
SoSe	Sommersemester
SVT	Supraventrikuläre Tachykardie
STEMI	<i>ST-Elevation Myocardial Infarction</i>
T2	Termin 2
T6	Termin 6
T80	Termin 80
TA	<i>Think-aloud</i>
UaK	Unterricht am Krankenbett
UMG	Universitätsmedizin Göttingen
VH	Vorhof
VHF	Vorhofflimmern
WiSe	Wintersemester

1 Einleitung

1.1 Lernen

1.1.1 *Clinical reasoning*

Angehende Ärztinnen und Ärzte auf den klinischen Alltag vorzubereiten ist ein zentrales Ziel im Medizinstudium. Aufgrund von Wissen und Erfahrung Untersuchungen zu veranlassen, Diagnosen zu stellen und Therapien einzuleiten bedarf komplexer Kompetenzen, die *clinical reasoning* (CR) erfordern (Kassirer 2010). CR ist ein komplexes und multidimensionales Konstrukt (Connor et al. 2019). Norman (2005) fasst CR als „almost all of what doctors do“ zusammen. Higgs *et al.* (2008) bezeichnen es als das Fundament der klinischen Praxis. CR hat viele Synonyme, wie *problem-solving*, *judgement*, *decision-making* (Norman 2005) und klinisches Denken.

In der Literatur sind unterschiedliche Definitionen von CR zu finden, doch sie beinhalten meist zwei Aspekte. Ein Aspekt sind die kognitiven Vorgänge, die es den Behandelnden ermöglichen, Informationen wahrzunehmen, zu sammeln und zu analysieren. Der zweite Aspekt sind die daraus resultierenden Entscheidungen für die zu treffenden Maßnahmen, unter Berücksichtigung der Wünsche der Patient:innen (Eva *et al.* 2007; ten Cate and Durning 2018). Für diese Arbeit soll klinisches Denken als der kognitive Prozess, der von der Anamnese, der körperlichen Untersuchung über die laborchemische und apparative Diagnostik zur Verdachtsdiagnose, Differenzialdiagnosen und Therapie führt (Durning 2013) definiert sein.

Die Qualität und Genauigkeit ihres *clinical reasonings* unterscheidet Assistenzärzt:innen von klinischen ExpertInnen (Hobu *et al.* 1987; Pinnock *et al.* 2015). Deshalb ergeben sich für die Ausbildung von Ärzt:innen nun zwei wichtige Fragen: steht CR als essenzieller Bestandteil des medizinischen Sachverstandes und Fähigkeit für sich allein und sollte deshalb als Kompetenz isoliert beigebracht werden oder entsteht es eher als eine Konsequenz aus spezifischem Fachwissen? Wie entwickelt es sich und wie lässt es sich unterrichten und messen (Norman 2005)?

Bis in die 70er Jahre wurde Problemlösung als allgemeine Denkstrategie isoliert betrachtet und trainiert (Newell and Simon 1972). Ende der 70er Jahre verglichen Elstein *et al.* in einem Experiment Medizinstudierende und Ärzt:innen beim Lösen von klinischen Fällen. Dabei

stellte ein deutlich größerer Anteil der Ärzt:innen die richtige Diagnose als die Studierenden. In ihrer Herangehensweise an die Lösung der Fallbeispiele unterschieden sich die beiden Gruppen jedoch nicht voneinander. Der große Unterschied zwischen diesen beiden Gruppen schien das Wissen und insbesondere die Struktur ihres Wissens zu sein (Elstein *et al.* 1978). Ein isoliertes Betrachten und Üben von CR scheint dementsprechend nicht zielführend, da es wissensbasiert und an die Materie gekoppelt ist. Im Englischen wird dies *context specificity* genannt (Eva 2005). Um besser klinisch denken zu können, müssen sich Faktenwissen und klinische Erfahrung erweitern und ordnen (Eva 2005; Koens *et al.* 2005; Norman 2005; Kassirer 2010).

1.1.2 *Dual-process-Theorie*

In der Literatur wird Denken in zwei grundlegende Gruppen aufgeteilt: in intuitives und analytisches Denken. Diese Aufteilung ist als *dual process theory* (Kahnemann 2011) bekannt. Oft gebräuchliche Synonyme sind *fast and slow thinking* und *system 1 and system 2 reasoning*.

Die intuitive Herangehensweise basiert hauptsächlich auf den Erfahrungen der diagnostizierenden Person. Um diese Form des klinischen Denkens entwickeln und ausbauen zu können, ist es notwendig, viele klinische Fälle zu sehen und zu bearbeiten (Ericsson 2004; Norman 2005; Kassirer 2010). Dadurch können mit zunehmender klinischer Erfahrung Informationen unter der Überschrift einer bestimmten Diagnose zusammengefasst gespeichert werden. Dieser intuitive und automatisierte schnelle Vergleich erlaubt ein schnelles Erkennen von Mustern (*pattern recognition*) (Schmidt *et al.* 1990; Ericsson 2004). Er ist essenziell für das Erlangen von diagnostischer Expertise (Schmidt *et al.* 1990; Coderre *et al.* 2003; Norman 2005; Bowen 2006).

Besonders hilfreich ist dieses schnelle assoziative Denken bei Symptomkombinationen, die pathognomonisch für eine bestimmte Erkrankung sind. Ein Beispiel dafür ist die dermatombezogene Ausbreitung der charakteristischen Hauteffloreszenzen bei Herpes zoster (Croskerry 2009). Die meisten Ärzt:innen hätten vermutlich keinerlei Schwierigkeiten, bei dieser typischen Erscheinung die Blickdiagnose richtig und schnell zu stellen. Eine sehr bedeutsame und gefährliche Schwachstelle der intuitiven Herangehensweise wird jedoch deutlich, wenn sich ein Krankheitsbild atypisch präsentiert. In einer Studie zum akuten Koronarsyndrom konnte gezeigt werden, dass die diagnostische Fehlerrate um ein Zehnfaches zunahm, wenn die Patient:innen nicht das Kardinalsymptom „Brustschmerz“ aufwiesen (Brieger *et al.* 2004).

Fast thinking scheint demnach häufiger mit Fehlern assoziiert zu sein. Der erste Eindruck sei sogar bei erfahrenen Ärzt:innen oft inkorrekt, wenn auch hilfreich für die Diagnosefindung

(Eva 2002 und 2005). Dies ist auf verschiedene *bias* zurückzuführen. Der *availability bias* führt beispielsweise dazu, dass die Diagnosen die Ärztinnen und Ärzte stellen weniger präzise und richtig sind, wenn sie vorher eine:n andere:n Patient:in mit auf den ersten Blick ähnlichen Symptomen gesehen haben (Mamede *et al.* 2010). Auch Merkmale oder Symptome von Patient:innen, die nichts mit dem Krankheitsbild zu tun haben, können ausschlaggebend für eine fehlerhafte Verdachtsdiagnose sein (Mamede *et al.* 2014).

Es ist wichtig, einen Startpunkt für analytische Gedanken zu definieren. Diese Funktion kann eine initiale Verdachtsdiagnose einnehmen. Dabei muss sorgfältig reflektiert werden, dass diese Hypothese alternative, und vielleicht stimmigere, Erklärungen für die Symptome der erkrankten Person in den Hintergrund drängen könnte (Sibbald and Cavalcanti 2011).

Tabelle 1: Vergleich von intuitiven und analytischen Herangehensweisen an Entscheidungsprozesse nach Croskerry (2009). Die Verwendung erfolgt mit freundlicher Genehmigung von Wolters Kluwer.

Intuitiv / System 1	Analytisch / System 2
Aus Erfahrungen abgeleitet	Hypothetisch deduktiv
Nur begrenzt rational	Rational
Heuristisch	Normatives Denken
<i>Pattern recognition</i> / Wiedererkennung	Erworbenes, kritisches, logisches Denken
Unbewusste Komponenten	Bewusstes Hinterfragen

Die analytische Herangehensweise an klinisches Denken ist ein aktiver, rationaler Denkprozess, wie beispielsweise das hypothetisch-deduktive Modell des *clinical reasoning*, das Elstein *et al.* (1978) beschrieben. Das analytische Denken ist langsamer und bewusster als die intuitive Herangehensweise. Dafür ist es weniger fehleranfällig (Eva 2005; Croskerry 2009). Das analytische Denken ist schließlich ein aktiver Prozess, der auf Wissenschaft und Evidenz, Logik, pathophysiologischen Konzepten und Wahrscheinlichkeit beruht und sich selbst kritisch hinterfragt (Kassirer 2010). Somit entsteht weniger Raum für emotionale Fehlerquellen oder unterbewussten *bias*. Der Prozess des analytischen Denkens wird jedoch oft erst angestoßen, wenn ein bestimmtes Merkmal nicht zur typischen Lehrbuchmanifestation der Erkrankung passt (Mamede *et al.* 2007; Kassirer 2010).

Tabelle 2: Charakteristika der intuitiven und analytischen Ansätze bei Entscheidungsprozessen nach Croskerry (2009). Die Verwendung erfolgt mit freundlicher Genehmigung von Wolters Kluwer.

	Intuitiv / Typ 1	Analytisch / Typ 2
Prinzip	Assoziativ	Auf Regeln basierend
Bewusstsein/ Kontrolle	Wenig	Stark
Geschwindigkeit	Schnell	Langsam
Reliabilität	Niedrig	Hoch
Fehler	Häufig	Selten
Aufwand	Niedrig	Hoch
Emotionale Beeinflussbarkeit	Hoch	Niedrig
Wissenschaftlichkeit	Niedrig	Hoch

Aufgrund der einander ergänzenden Stärken und Schwächen der intuitiven und analytischen Herangehensweisen sollten sie nicht dichotom betrachtet werden. Sie sollten vielmehr als interaktive, einander ergänzende und sich gegenseitig beeinflussende Denkprozesse angesehen werden (Croskerry 2009). Für ein optimales *clinical reasoning* sollte, je nach Fall zu variierenden Anteilen, analytisches und nicht-analytisches Denken, kombiniert werden (Eva 2005; Bowen 2006). Die Forschung deutet an, dass einfache Fälle eher durch intuitives Denken (*pattern recognition*) und komplexere Fälle eher durch analytisches Denken gelöst werden (Eva 2005; Norman 2005; Kassirer 2010).

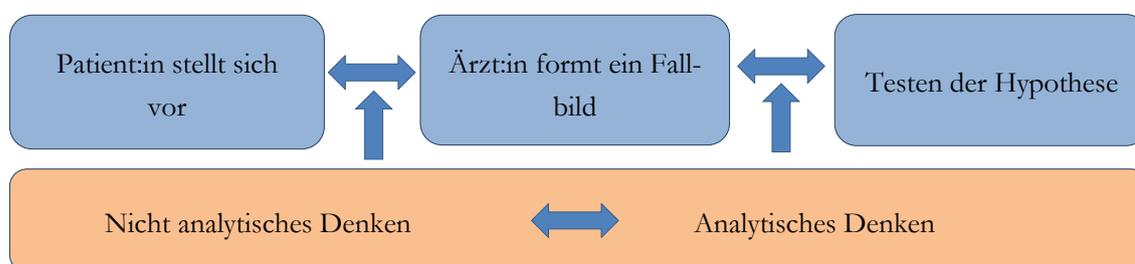


Abbildung 1: Ein Modell des *clinical reasoning* nach Eva (2005). Die Verwendung erfolgt mit freundlicher Genehmigung von John Wiley and Sons.

Ark *et al.* konnten anhand einer Studie den Vorteil dieses interaktiven, aus intuitiven und analytischen Komponenten bestehenden *Clinical-reasoning*-Modelles belegen. In dieser Studie ließen sie unerfahrene Studierende ein Elektrokardiogramm (EKG) befunden. Die Studierenden wurden in drei Gruppen aufgeteilt. Eine Gruppe wurde aufgefordert ein nicht-analytisches Denken anzuwenden (*pattern recognition*), aber eine vorschnelle Diagnosestellung zu vermeiden (also die Hypothese analytisch zu hinterfragen). Diese Gruppe wertete die EKGs mit einer signifikant größeren diagnostischen Genauigkeit aus, als die anderen beiden Gruppen, die jeweils zu einer nur analytischen oder nur nicht-analytischen Herangehensweise angewiesen worden waren. Die Beobachtungen dieser Versuche zeigen, dass eine Kombination aus beiden *Reasoning*-Strategien zu einer genaueren Diagnose führt, als eine rein analytische Herangehensweise (Ark *et al.* 2006).

Laut Feltovich und Barrows (1984) beinhaltet die mentale Abbildung eines Patientenfalles mehrere Merkmale, die mit der Krankheit assoziiert sind. Diese Abbildung nennen sie *illness scripts*. Diese gebündelten Informationen und Erfahrungen verknüpfen unterschiedliche Darstellungen einer Diagnose und ermöglichen das Lösen neuer Fälle durch den Vergleich mit bereits gesehenen Fällen aus der Vergangenheit (Bowen 2006). Wenn Studierende ihre ersten Patient:innen sehen, wenden sie bewusst ihr pathophysiologisches Wissen an, um den Fall zu lösen. Nachdem sie mehrere Patient:innen mit denselben oder ähnlichen Symptomen gesehen haben, beginnen sich diese pathophysiologischen Erklärungen in vereinfachte Modelle zu sortieren. Diese bieten eine für Diagnostik und Therapie ausreichende Erklärung für die beobachteten Symptome. Durch das Betrachten von vielen Patient:innen bekommen Studierende ein Gefühl dafür, auf welche Weisen eine Krankheit sich manifestieren kann. Die verschiedenen Merkmale, die das klinische Erscheinungsbild einer Krankheit charakterisieren, werden zu Fixpunkten, um die das Denken der Ärzt:innen sich entwickelt. Dies können beispielsweise Geschlecht, Alter oder Risikofaktoren sein. *Illness scripts* können somit einen nicht-analytischen Rahmen für weitere Diagnostik darstellen (Feltovich and Barrows 1984; Schmidt *et al.* 1990; Schmidt and Mamede 2015).

Es bleibt schwer vorstellbar, dass ein einzelnes Modell zusammenfassen kann, wie alle Menschen denken. Diese Modelle und Denkstrategien sollten daher nicht als definitive Prozesse, sondern eher als vorsichtige Theorien und Metaphern für Denkprozesse verstanden werden (Kassirer 2010).

Zusammenfassend ist im Rückblick auf die zurzeit vorhandene Literatur zum Thema *clinical reasoning* festzustellen, dass bisher wenig Evidenz für die Vorteile des Unterrichtens von

Reasoning-Strategien existiert. Stattdessen scheint die Kernvoraussetzung zum Erlangen (medizinischer) Expertise das aktive Bearbeiten klinischer Fälle mit Feedback zu sein (Ericsson 2004; Norman 2005; Kassirer 2010). Dabei ist es wichtig, dass diese Übung der späteren Anwendung so ähnlich wie möglich ist (Norman 2005).

1.1.3 Fallbasiertes Lernen

Das wiederholte Bearbeiten von klinischen Fällen ist eine valide Methode, Wissen zu trainieren und langfristig zu konsolidieren (Eva 2005; Norman 2005; Kassirer 2010). Studierende sollten eine große Sammlung an Fallbeispielen sehen, um flexible und an das jeweilige Problem angepasste Lösungen zu finden (Ark *et al.* 2006). So können *illness scripts* auf- und ausgebaut werden und ein stabiles Fundament für schnelles Denken bilden.

Fallbasiertes Lernen bedeutet, dass Fallbeispiele vom Leitsymptom aus betrachtet und bearbeitet werden (beispielsweise „Schmerzen im linken Unterbauch“), ganz wie es Ärzt:innen in der Notaufnahme täglich erleben. Diese Herangehensweise steht der traditionellen Unterrichtsform gegenüber, bei der Vorlesungen zu einem bestimmten Krankheitsbild, beispielsweise Divertikulitis, gehalten werden. Fallbasiertes Lernen kann bereits während des Medizinstudiums einen Grundstein für diese klinische Erfahrung und somit den intuitiven *Reasoning*-Pfad legen (Barrows und Tamblyn). Diese Art von Unterricht findet häufig im Rahmen von ersten Begegnungen mit Patient:innen auf Station statt. Im praktischen Jahr, in Famulaturen, Blockpraktika oder beim Unterricht am Krankenbett haben Studierende die Möglichkeit, bei der Arbeit mit Patient:innen ihr Wissen anzuwenden und zu wiederholen.

Im klinischen Unterricht ist fallbasiertes Lernen meist patient:innenzentriert und daher schwierig standardisierbar und außerdem ressourcenintensiv (Srinivasan *et al.* 2007). Studien können allerdings seine Effektivität belegen. Durch ein klinisches Praktikum erlangen Studierende einen deutlichen Wissenszuwachs (Butterfield and Libertin 1993). Studierende mit einem ähnlichem Wissensstand vor der klinischen Rotation erzielen jedoch teilweise unterschiedliche Ergebnisse nach dem Praktikum (Schwartz *et al.* 1994). Dies weist darauf hin, dass bestimmte Faktoren eines Praktikums Konsequenzen für den Lernerfolg haben. Die Qualität von Unterricht am Krankenbett (UaK) sowie Supervision und das Angebot von Feedback variieren stark (Wimmers *et al.* 2006; Schmidt and Mamede 2015). Welche Faktoren schlussendlich jedoch zu dem unterschiedlich starken Lernzuwachs führen ist nicht abschließend erwiesen (Wimmers *et al.* 2006). Beispielsweise konnte keine direkte Korrelation zwischen der Anzahl der gesehenen Patient:innen und der Leistung am Ende des klinischen

Praktikums festgestellt werden, obwohl davon ausgegangen wird, dass zum Ausbau der klinischen Kompetenz eine gewisse Zahl an klinischen Fällen gesehen werden sollte (Leeuwen *et al.* 1997; Schmidt and Mamede 2015). Es ist nicht nur interessant wie viele, sondern auch welche Fallbeispiele in klinischen Praktika gesehen werden. In einer Studie von McLeod und Snell (1991) konnte gezeigt werden, dass Studierende während ihres Praktikums nicht immer die Krankheitsbilder sehen, die für besonders wichtig erachtet werden. Auch in einer großen amerikanischen Studie wurde deutlich, dass weniger als die Hälfte der teilnehmenden Medizinstudierenden während ihres klinischen Praktikums Patient:innen mit Krankheiten mit der höchsten Prävalenz sahen (Rattner 2001).

An einer Universitätsklinik wird mit hoher Wahrscheinlichkeit ein selektiertes Patientenkollektiv vorgestellt, das eher seltene Krankheitsbilder aufweist oder besonders komplikationsreiche und schwere Krankheitsverläufe zeigt. Außerdem sind bestimmte Patientenkollektive vom Studierendenunterricht ausgeschlossen, beispielsweise Patient:innen mit ansteckenden Infektionskrankheiten oder in akuter Lebensgefahr (Ziv *et al.* 2006), auch wenn die Kenntnis über diese Erkrankungen ebenso relevant für Studierende ist.

In der Medizin ist es wichtig zu verstehen, welche Krankheiten häufig auftreten und welche selten. Das fallbasierte Lernen sollte versuchen, dies so gut wie möglich abzubilden (Norman 2005). Hinzu kommt die Herausforderung, möglichst dieselben Bedingungen für Studierende im gleichen Semester zu schaffen. So sollten alle Studierenden eines Semesters bestenfalls Patienten mit den gleichen Krankheitsbildern kennenlernen. Doch es ist unwahrscheinlich, dass sich gleichzeitig genügend Patienten mit demselben Krankheitsbild in einer Universitätsklinik befinden, dass ein ganzes Semester sie in Kleingruppen untersuchen könnte. Die Studierenden eines Semesters würden also wahrscheinlich unterschiedliche Fälle sehen.

Zusammenfassend lässt sich also sagen, dass fallbasiertes Lernen hilft, durch Anwendung bereits gelerntes Wissen besser zu behalten (Kassirer 2010) und in Form von *illness scripts* zu sortieren und damit den intuitiven *Reasoning*-Pfad auszubilden. Jedoch ist es schwierig, sowohl die präsentierten Krankheitsbilder als auch die Umsetzung (Lehre, Feedback) im klinischen Alltag zu standardisieren. Darüber hinaus ist der Unterricht am Krankenbett recht ressourcenintensiv (Srinivasan *et al.* 2007).

Ein Weg der Standardisierung fallbasierten Lernens ist ein Kleingruppenunterricht ohne Patient:innen, der an der McMaster Universität in Hamilton, Kanada 1969 implementiert wurde (Neufeld *et al.* 1989) und seitdem an vielen Universitäten als alternative Lehrmöglichkeit angewendet wird (Cunningham *et al.* 2006). Im Curriculum der Humanmedizin an der Univer-

sität Göttingen findet diese Unterrichtsform beispielsweise als POLEMA („Problem-orientiertes Lernen, elektronische-Medien assistiert“) im Modul 6.x¹ statt. Hier werden in kleinen Gruppen von etwa sechs Studierenden in von TutorInnen moderierten Sitzungen klinische Fallbeispiele im *serial cue approach*, also schrittweise, gelöst. So können die Fallbeispiele vorher sorgfältig ausgewählt somit und eine realitätsnahe Abbildung erzeugt werden. Durch die Betreuung der einzelnen Kleingruppen ist diese Unterrichtsform jedoch weiterhin personal- und damit kostenaufwändig. In den letzten Jahren wurden daher an vielen medizinischen Fakultäten weltweit neue Ansätze des fallbasierten Lernens ausprobiert (Knight *et al.* 2010; Kleinert *et al.* 2015; Dankbaar Mary EW *et al.* 2016; Chon *et al.* 2018; Tsoy *et al.* 2019). Ein Ansatz ist der computergestützte Simulationsunterricht mit virtuellen Patient:innen.

1.2 *Serious games in der Medizindidaktik*

1.2.1 *Serious games*

Simulationstraining findet in vielen anderen Berufen, in denen es wenig Raum für Fehler gibt, längst Anwendung, wie beispielsweise in der Pilot:innenausbildung oder beim Militär (Ziv *et al.* 2006). Ziel ist es, eine Lernumgebung zu schaffen, in der Fehler passieren dürfen, ohne Menschen zu gefährden. Ziv *et al.* bezeichnen Simulation in der medizinischen Ausbildung sogar als einen „ethischen Imperativ“. Die Autor:innen sind der Meinung, dass Simulation eine gute Methode einer fallbasierten Ausbildung sein könnte, ohne Patient:innen einem unnötigen Risiko auszusetzen. *Serious games* (SG), aus dem Englischen übersetzt „ernsthafte Spiele“, sind Spiele, die nicht nur unterhaltsam, sondern vor allem lehrreich sein sollen (Gorbanev *et al.* 2018). Mit SG können Studierende auch höchst akute Situationen trainieren, in denen ihre Patient:innen in Lebensgefahr schweben; eine Situation, die aufgrund der Priorität der Patient:innensicherheit nur schwer im klinischen Alltag geübt werden kann (Knight *et al.* 2010). In den letzten Jahren haben sich Forschende zunehmend mit dem Einsatz von SG in der Medizindidaktik beschäftigt, was durch die wachsende Zahl der *case reports* und *reviews* in der Literatur erkennbar wird (Abdulmajed *et al.* 2015; Wang *et al.* 2016). Studierende bearbeiten einen Fall, doch wenn Fehler passieren, zieht dies keine fatalen Konsequenzen mit sich, sondern im besten Fall einen Lerneffekt (Gorbanev *et al.* 2018). Unterstützt wird dieser Lerneffekt durch *gamification*, also Spielelemente wie für die Spielleistung vergebene Punkte oder Sterne, die Motivation und Engagement der Spielenden fördern sollen (Rutledge

¹ Das Modul wurde zwischenzeitlich umbenannt und heißt nun „Modul 6.1 Klug entscheiden in der Medizin“

et al. 2018). Doch wie effektiv der tatsächliche Lerneffekt von SG ist, ist bisher nicht eindeutig mit Studien belegt. Oftmals bezieht sich die Literatur zum Thema SG auf die Evaluation des Spielerlebnisses, die Handhabung des SG oder die Meinung der Studierenden zum Spiel und nicht auf eine Messung des Lernerfolgs (Akl *et al.* 2010; Gorbanev *et al.* 2018). Da jedoch genau dies das Ziel von SG ist, soll in der hier vorgelegten Studie dieser Lerneffekt wissenschaftlich untersucht werden.

1.2.2 Wie lässt sich *clinical reasoning* messen?

Spannend ist nun, wie klinisches Denken getestet, bewertet, quantifiziert und erforscht werden kann. Aufgrund der Vielschichtigkeit des klinischen Denkens erfordert dies mehrere Methoden. Im Folgenden werden vier Methoden erläutert, die zum Erfassen von klinischem Denken herangezogen werden können.

1.2.2.1 *Logfiles*

Ein *Logfile* ist eine Protokolldatei, in der alle Ereignisse, die in einem Computersystem geschehen, gespeichert werden. In dieser Datei kann nachvollzogen werden, welche Aktion zu welchem Zeitpunkt ausgeführt wurde. Bei einer computerbasierten Übung, wie beispielsweise einer Simulation mit Patientenfällen, kann so im Nachhinein eine Bewertung bestimmter *items* erfolgen. So könnte durch das Auslesen der Protokolldatei dieser Computersimulation festgestellt werden, welche Aktionen unternommen wurden, also beispielsweise welche Anamnesefragen gestellt wurden, welche Diagnose gestellt wurde und welche Therapie eingeleitet wurde. Die Auswertung von *Logfiles* kann also als ergebnisorientierte Messmethode für CR genutzt werden.

1.2.2.2 Übergabeprotokolle

Eine weitere Möglichkeit zur Erfassung klinischen Denkens ist das Übergabeprotokoll (Durning *et al.* 2012). Wie die englische Bezeichnung *patient summary* ausdrückt, wird ein Fallbeispiel nach der Bearbeitung zusammengefasst wiedergegeben. Hier kann ein Einblick in das klinische Denken der präsentierenden Person erfolgen: was wird in der Übergabe genannt? Welche Schlussfolgerungen und Gedanken werden mitgeteilt? Welche Diagnose wurde warum gestellt, und welche Differenzialdiagnosen kamen infrage? Durning *et al.* implementierten diese Form der Messung von CR im Jahr 2012 in einem OSCE-Prüfungsformat (*objective structured clinical examination*) und zeigten die Validität und Umsetzbarkeit dieser Methode.

1.2.2.3 *Key-feature*-Prüfungen

Das *Key-feature*-Format wurde initial 1987 von Bordage und Page vorgestellt als ein effektiver Weg, klinische Entscheidungsfindung in schriftlichen Tests zu beurteilen (Page und Bordage 1995; Bordage und Page 2018). In jedem klinischen Fall gibt es essenzielle Schritte auf dem Weg zur erfolgreichen Lösung des Falles (Page und Bordage 1995; Farmer und Page 2005). Diese Punkte nennen sie *Key-features* (KF). In KF-Prüfungen wird ein kurzer klinischer Fall geschildert, zu dessen Auflösung longitudinal Fragen gestellt werden und bei denen mehr als nur eine Lösung richtig sein kann. Diese Fragen beziehen sich immer auf die Schlüsselmomente der Bearbeitung des Falles. Nach der Beantwortung eines KF wird die Auflösung verraten, um Folgefehler zu vermeiden (Page und Bordage 1995), ohne dass die gegebene Antwort nachträglich verbessert werden kann. KF-Tests können papier- oder computerbasiert gestellt werden. In dieser Studie wurden elektronische KF-Prüfungen verwendet, weshalb im Folgenden nur auf diese eingegangen wird.

1.2.2.4 *Think-aloud*-Interviews

Die grundlegende Annahme für Erforschung des kognitiven Verhaltens ist, dass Verhalten durch bewusste und unterbewusste Denkprozesse vermittelt wird (Davison *et al.* 1997). *Think aloud* (TA) ist eine etablierte wissenschaftliche Methode, um diese Denkprozesse zu untersuchen (Elstein *et al.* 1978; Ericsson and Simon 1984; Davison *et al.* 1997; Pinnock *et al.* 2015). Für ein TA-Interview verbalisieren Proband:innen ihre Gedanken, während sie eine Aufgabe bearbeiten (Duncker 1926; Burbach *et al.* 2015). So zeigen sie, wie sie Informationen nutzen, um eine Lösung für ein Problem zu finden (Ericsson and Simon 1984; Pinnock *et al.* 2015). Die TA-Interviews werden aufgenommen, transkribiert und später wissenschaftlich untersucht.

In *Logfiles* und Testergebnissen können die Ergebnisse des CR, also richtige oder falsche Diagnosen oder Therapievorschlüsse, quantifiziert werden. Mithilfe von TA-Interviews kann potenziell der Weg zu diesen Ergebnissen offengelegt werden. *Think aloud* ist damit eine Methode, die helfen kann, den Ablauf von Denkprozessen auf dem Weg zur richtigen Diagnose und Therapie zu verstehen (Pinnock *et al.* 2015).

1.2.3 **Das *serious game* EMERGE**

Das Spiel EMERGE ist eine von der Universitätsmedizin Göttingen in Zusammenarbeit mit dem Universitätsklinikum Hamburg Eppendorf und PatientZero Games GmbH[®] im Jahr 2012 konzipierte, bildschirmbasierte virtuelle Notaufnahme, in der Spielende klinische Fälle

selbstständig am Computer erarbeiten können. Es wurde für den Einsatz in der Lehre im Medizinstudium entwickelt. Die Spielenden schlüpfen in diesem *serious game* in die ärztliche Rolle in der Notaufnahme und behandeln bis zu zehn virtuelle Patient:innen gleichzeitig. Aufgabe ist, mit Hilfe von Anamnesefragen und körperlichen, apparativen und laborchemischen Untersuchungen die richtige Diagnose zu stellen, eine adäquate Therapie einzuleiten, und anschließend die Patient:innen auf die richtige Station zu verlegen oder nach Hause zu entlassen. Nach Abschluss des Falles kann ein Feedback zu der Behandlung und der gestellten Diagnose eingesehen werden.



Abbildung 2: Die Sicht auf die virtuelle Notaufnahme EMERGE mit geöffneten Pager- Benachrichtigungen

Studierende können zunächst eine Ärztin oder einen Arzt als Avatar auswählen und geben ihre Matrikelnummer zur Anmeldung ein. So erfolgt die Anwesenheitskontrolle im EMERGE-Seminar. Nach der Anmeldung öffnet sich die Sicht auf die virtuelle Notaufnahme, in der es eine Rezeption, einen Wartebereich sowie zwei Schockräume und fünf Behandlungszimmer gibt. In der oberen rechten Ecke des Bildschirms ist eine kleine Übersicht der Notaufnahme zu sehen, in der kleine Bilder anzeigen, welche Betten belegt sind und wo sich der Avatar gerade befindet. Navigiert wird durch ein Klicken auf einen Punkt in der virtuellen Notaufnahme. Wenn neue Simulationspatient:innen in die virtuelle Notaufnahme eingeliefert werden, erscheint im rechten unteren Bildrand eine Nachricht, in der Name und Leitsymptom der zu behandelnden Person angezeigt werden. Außerdem ist abzulesen, ob

die Person mit dem Rettungswagen eingeliefert wird oder selbstständig in die Notaufnahme kommt. Nun kann die Person in eines der fünf Behandlungszimmer oder einen der zwei Schockräume verwiesen werden. Die zu behandelnde Person ist automatisch an einen Monitor angeschlossen und die Vitalparameter Blutdruck, Puls und Sauerstoffsättigung sowie Name und Alter werden rechts im Bild angezeigt.

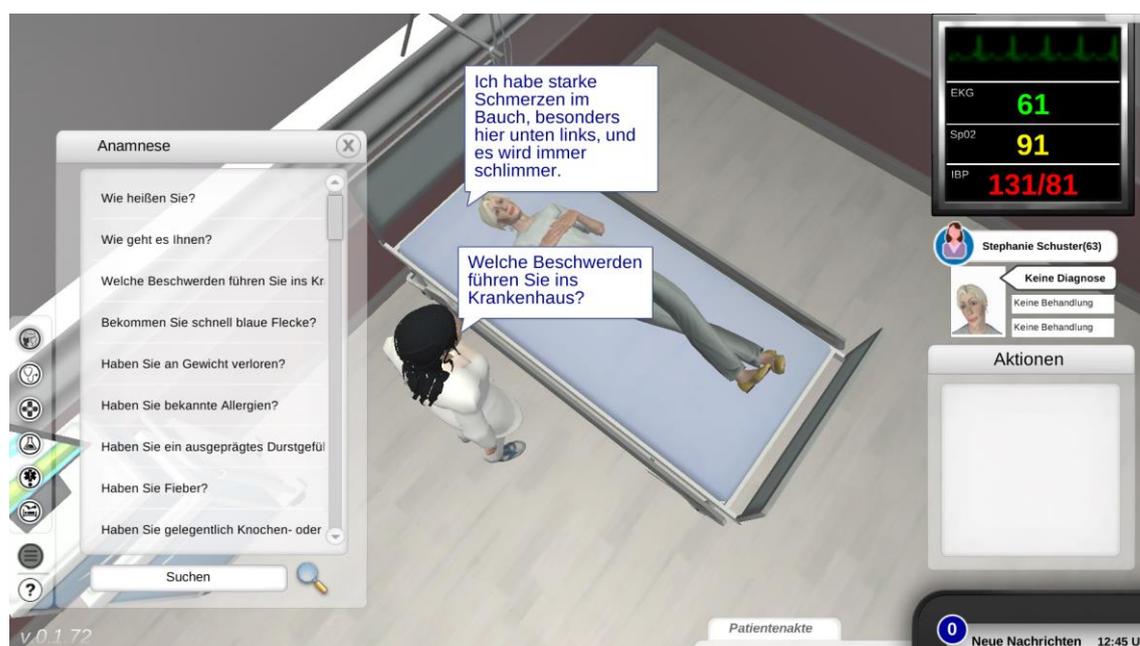


Abbildung 3: Anamnesefunktion im *serious game* EMERGE

Im Bedienungs Menü kann aus 73 vorgegebenen Anamnesefragen gewählt werden, auf die dann auf die einzelnen Fälle angepasste Antworten gegeben werden. Im Menü „Untersuchungen“ können unterschiedliche körperliche Untersuchungen wie beispielsweise des Abdomens oder des Thorax durchgeführt werden. Außerdem gibt es die Möglichkeit, apparative Untersuchungen anzufordern. Es finden sich hier beispielsweise ein Röntgen Thorax, EKG, Computertomographie (CT) oder Temperaturmessung. Außerdem gibt es die Möglichkeit, Konsile unterschiedlicher Fachrichtungen anzufordern. Im Labor können entweder vorgegebene Profile, wie zum Beispiel ein kleines oder großes Blutbild, Gerinnung oder Infarktprofil, oder auch einzelne Blutwerte angefordert werden. Unter dem Reiter „Behandlungen“ findet sich zum einen eine Medikamentenliste, zum anderen weitere Behandlungsmöglichkeiten wie beispielsweise Punktionen oder die Anlage von Kathetern. Dabei reagieren die virtuellen Patient:innen auf Behandlungen und Medikamente. Im Menüpunkt „Diagnose“ kann zwischen 227 Diagnosen ausgewählt werden. Um die Person zu verlegen oder zu entlassen, kann zwischen Intensivstation, Intermediate Care Station, Normalstation und

der Entlassung ausgewählt werden. Alle erhobenen Untersuchungsergebnisse sind in der Patientenakte nachzuvollziehen, die durch einen Klick aufzurufen ist. Darin werden alle durchgeführten Untersuchungen mit Ergebnissen und Anamnesefragen aufgelistet. Die bereits abgeschlossenen Fälle können im Archiv durch einen Klick auf einen der Computer in den virtuellen Behandlungszimmern eingesehen werden. Dort wird zu jedem Fall ein Oberarzt-kommentar angezeigt, der eine kurze Erklärung zum Krankheitsbild und zur richtigen Vorgehensweise bietet. Außerdem wird eine Bewertung der wichtigsten Aktionen angezeigt durch ein Punktesystem bestehend aus maximal drei Sternen, die in den Kategorien „Anamnese“, „Körperliche Untersuchung“, „Diagnose“, „Fall abgeschlossen“ und „Verlegung“ vergeben werden können.

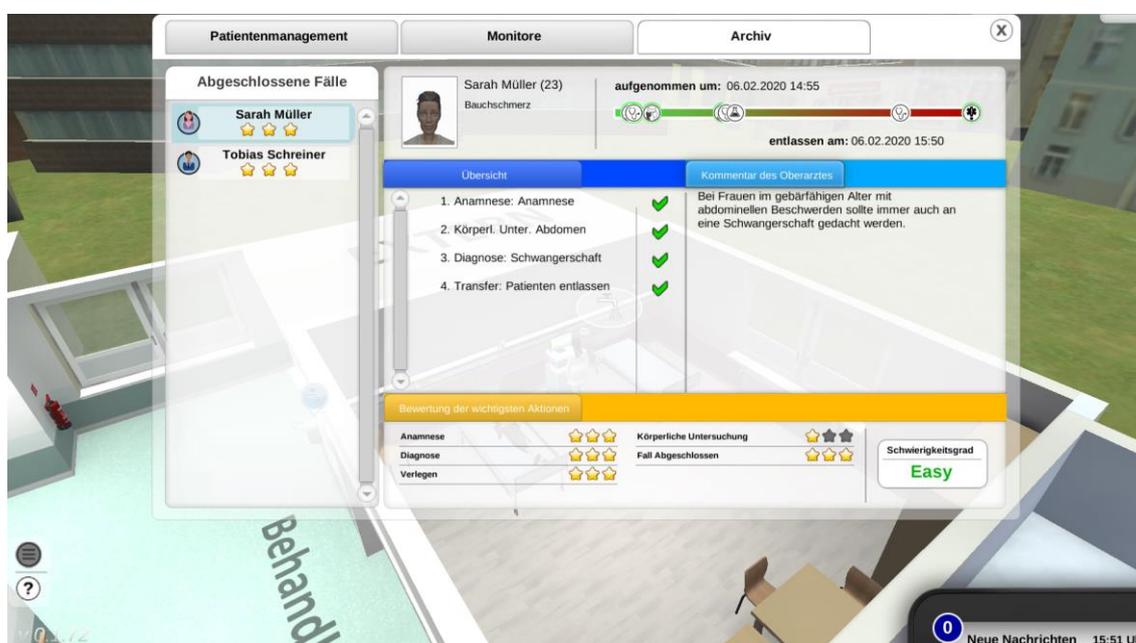


Abbildung 4: Das Archiv zum Einsehen des Feedbacks zu jedem abgeschlossenen Fall

In der virtuellen Notaufnahme werden, wie auch in einer echten Notaufnahme, häufig mehrere Personen kurz nacheinander oder gleichzeitig vorstellig. Hier ist es Aufgabe der Spielenden, zu triagieren, also zu entscheiden, wer vorrangig behandelt werden muss. Die Spielenden können mit einem Klick auf die Computer, die sich in jedem Behandlungsraum befinden, einen Überblick über ihre Notaufnahme verschaffen. Es wird ihnen dann ein Blick in der Vogelperspektive auf die virtuelle Notaufnahme mit ihren einzelnen Behandlungsräumen und Patient:innen und ihren Vitalparametern sowie eine Liste der im Wartebereich sitzenden Patient:innen angezeigt.

Ein Fallbeispiel, beispielsweise „NSTEMI“, präsentiert sich nicht jedes Mal absolut gleich. Durch geringe Veränderungen der Fallpräsentation wird zu vermeiden versucht, dass sich wiederholende Fälle direkt als solche erkannt werden. Das soll dazu führen, dass die Studierenden nicht die 83-jährige Dame mit dem Blutdruck von 163/95 mmHg erkennen, und beim dritten Spieltermin gleich wissen, dass NSTEMI hier die richtige Diagnose sein wird. Bei passenden Fällen entscheidet ein Zufallsgenerator über das Geschlecht der erkrankten Person. Auch Alter und Vitalparameter werden in einem definierten Referenzbereich zufällig angezeigt. Außerdem sind bei den meisten Fällen in der Anamneseliste mehrere Antwortmöglichkeiten hinterlegt. Diese in der Grundaussage zwar identischen Antworten sind leicht unterschiedlich formuliert, um einen möglichen Wiedererkennungswert zu minimieren.

1.3 Aufbau des Medizinstudiums an der Universität Göttingen

An der Universität Göttingen ist das Medizinstudium in Form eines Regelstudiengangs organisiert. In den ersten vier Semestern werden naturwissenschaftliche Grundlagen unterrichtet und daraufhin mit dem ersten Abschnitt der ärztlichen Prüfung geprüft. Im darauffolgenden klinischen Studienabschnitt werden die einzelnen klinischen Spezialisierungen in Modulform unterrichtet (Humanmedizin in Göttingen 2020). In einem Modul finden üblicherweise drei einander ergänzende Lehrformen statt: Vorlesungen, Seminare und Unterricht am Krankenbett. Im dritten klinischen Semester, also im siebten Semester des Medizinstudiums, findet das Modul 3.1 Kardiopulmonale Lehre statt. Dies stellt das erste Modul nach Organ-schwerpunkten dar.

Im sechsten klinischen Semester findet ein Repetitionsmodul statt, genannt Modul 6.x: Klug entscheiden in der Medizin: Vom Symptom zur Diagnose und weiter zur individuell angepassten Therapie. Hier wird vom Leitsymptom aus interdisziplinär und differenzialdiagnostisch gedacht, um vor dem praktischen Jahr auf das klinische Arbeiten vorzubereiten (M6.1 - Klug entscheiden in der Medizin 2020).

Im Curriculum der Universität werden kardiologische und pulmologische Krankheitsbilder dementsprechend zu zwei Zeitpunkten gelehrt; im Modul 3.1 im dritten, und im Modul 6.x im sechsten klinischen Semester.

1.3.1 Implementierung von EMERGE im Curriculum der UMG

In einer vorangehenden Studie der Abteilung für Medizindidaktik und Ausbildungsforschung der Universität Göttingen im Sommersemester 2016 wurde das Spiel EMERGE mit dem *Problem-based-learning*-Lehrformat POLEMA, einem seit 2010 im Curriculum der UMG

etabliertem Lehrformat, verglichen (Middeke *et al.* 2018). Ziel der Studie war, die Lernerfolge von Studierenden, die mit POLEMA lernen, mit dem der EMERGE spielenden Studierenden zu vergleichen und somit das *serious game* als gleichwertige Lernmethode zu etablieren.

Das Semester teilte sich in zwei Gruppen, von denen eine an POLEMA und die andere an EMERGE-Sitzungen teilnahm. Am Ende des Moduls wurde in einer formativen *Key-feature*-Prüfung der Lernerfolg beider Gruppen getestet und miteinander verglichen.

Beide Gruppen nahmen an zehn 90-minütigen Veranstaltungen der von ihnen gewählten Methode teil. Die POLEMA-Studierenden besprachen in Gruppen von sechs bis acht Studierenden und studentischen TutorInnen insgesamt fünf Fälle im Detail. Die EMERGE-Gruppe bearbeitete insgesamt mehr als 40 Fälle in der virtuellen Notaufnahme in Einzelarbeit. In der formativen KF-Prüfung am Ende der Intervention erzielte die EMERGE-Kohorte gleiche oder bessere Ergebnisse als die POLEMA-Gruppe. Damit wurde EMERGE im Sommersemester 2017 als gleichwertiges Lehrformat in die curriculare Lehre des Moduls 6.x aufgenommen. Studierende können seitdem wählen, ob sie im Repetitionsmodul die klinischen Fälle in Kleingruppen im POLEMA-Seminar oder in Einzelarbeit mit dem *serious game* EMERGE bearbeiten möchten.

1.4 Erarbeitung der Fragestellung

Clinical reasoning ist ein komplexes Konstrukt, das für diese Arbeit als die kognitiven Schritte von der Fallpräsentation zu Diagnose und Therapie definiert sein soll (Connor *et al.* 2019). Damit ist es eine ärztliche Kernkompetenz (Norman 2005; Ajjawi und Higgs 2008; Dumas *et al.* 2018). Die Güte und Genauigkeit ihres klinischen Denkens unterscheidet Medizinstudierende von erfahrenen Ärzt:innen (van de Wiel *et al.* 1998). Obwohl CR essenziell für eine optimale medizinische Versorgung (Kassirer 2010) und damit auch die Sicherheit von Patient:innen ist, ist es selten ein Schwerpunkt des medizinischen Curriculums (Amey *et al.* 2017).

Das klinische Denken von Mediziner:innen entwickelt sich im Laufe ihres Werdegangs (Bowen 2006). Initial geschieht medizinische Problemlösung durch analytisches Betrachten der Pathophysiologie. Zunehmende klinische Exposition erlaubt Studierenden, Probleme mit ähnlichen zu vergleichen, die zuvor gesehen wurden (Schmidt *et al.* 1990) und *illness scripts* zu formen. Diese klinische Erfahrung kann in Famulaturen oder UaK gesammelt werden. In diesen Lehrformaten variieren jedoch sowohl Anzahl und Art der gesehenen Krankheitsbilder als auch die Qualität der Supervision (Wimmers *et al.* 2006). Eine standardisierte und kosteneffektive Möglichkeit des fallbasierten Lernens ist der Simulationsunterricht mit einem *serious game* (Gentry *et al.* 2019).

In den meisten Studien zum Thema *serious games* in der Medizindidaktik werden bisher hauptsächlich das Spielerlebnis, die Handhabung und der Einfluss des SG auf die Motivation der Studierenden evaluiert (Akl *et al.* 2010; Gorbanev *et al.* 2018). Ob SG eine gute Lehr- und Lernmethode zum Training von klinischem Denken sein können ist bisher nicht ausreichend erforscht (Connolly *et al.* 2012). In dieser Studie soll dies nun anhand der virtuellen Notaufnahme EMERGE untersucht werden. Es konnte bereits gezeigt werden, dass Studierende mit dem SG EMERGE im letzten klinischen Semester vergleichbare kurzfristige Lernerfolge erzielen wie mit dem Kleingruppenunterricht POLEMA (Middeke *et al.* 2018). Weiterführend stellt sich die Frage, ob dieses Ergebnis auch auf Studierende in anderen Semestern übertragbar ist, und ob diese Lernerfolge über einen längeren Zeitraum anhalten.

Es ergeben sich folgende Forschungsfragen:

1. Inwiefern unterscheiden sich Studierende des dritten und sechsten klinischen Semesters hinsichtlich ihrer Befähigung zum klinischen Denken? Zu welchem Zeitpunkt im Medizinstudium erzeugt die Nutzung einer digitalen Notaufnahme-Simulation den größten Lernerfolg?
2. Ist der vermutete Lernzuwachs auch nach eineinhalb Jahren noch festzustellen?
3. Wie unterscheidet sich die Spielleistung von Studierenden des sechsten klinischen Semesters, die bereits zu Beginn des klinischen Studiums mit EMERGE trainiert haben, von Studierenden desselben Semesters, die bisher nicht EMERGE gespielt haben?

Zur Beantwortung der drei Forschungsfragen wurden in dieser Studie drei Analysen durchgeführt, um das klinische Denken von Medizinstudierenden des dritten und sechsten klinischen Semesters der UMG in der virtuellen Notaufnahme EMERGE zu zwei Zeitpunkten verglichen. Dafür wurden *Logfiles*, Übergabeprotokolle, eine *Key-feature*-Prüfung und *Think-aloud*-Interviews erhoben und ausgewertet.

Ziel dieser Studie ist es somit, die Entwicklung des *clinical reasonings* Medizinstudierender im Laufe ihres Studiums näher zu beleuchten. Außerdem soll mit Hilfe dieser Studie untersucht werden, ob die virtuelle Notaufnahme EMERGE als Lehr- und Lernmethode für Studierende verschiedener Semester einen Vorteil bringt.

2 Material und Methoden

In dieser Arbeit wird eine prospektive, nicht-randomisierte Interventionsstudie beschrieben, die im Sommersemester 2017 stattfand und an der alle Studierenden des dritten und sechsten klinischen Semesters der Universitätsmedizin Göttingen (UMG) teilnehmen konnten. Die Studie wurde um ein longitudinales *follow-up* nach drei Semestern erweitert. Dafür fand die Datenerhebung zu Beginn des Wintersemesters 2018 statt. Im Folgenden werden zunächst Aufbau und Ablauf der Studie beschrieben. Zur Beantwortung der Studienfragen kamen qualitative und quantitative Erhebungsinstrumente zum Einsatz, die im Anschluss erläutert werden. Zusätzlich wird auf die explorative Datenanalyse eingegangen.

Zur Beantwortung der drei Forschungsfragen wurden folgende drei Analysen durchgeführt:

Analyse 1: Vergleich des Zuwachses an *clinical reasoning* von Medizinstudierenden im dritten klinischen Semester mit dem Zuwachs von *clinical reasoning* von Medizinstudierenden im sechsten klinischen Semester im Verlauf eines Moduls

Analyse 2: Prüfung des langanhaltenden Lernerfolgs mit einem *follow-up* nach eineinhalb Jahren

Analyse 3: Vergleich von Studierenden im sechsten klinischen Semester, die zuvor im Modul 3.1 mit EMERGE gearbeitet hatten, mit EMERGE unerfahrenen Kommilitonen

Gemäß des Entscheids der Ethik-Kommission der medizinischen Fakultät der Universität Göttingen vom 08.03.2017 bestanden keine ethischen und rechtlichen Bedenken gegen die Durchführung des vorliegenden Forschungsvorhabens (Antragsnummer 15/3/17).

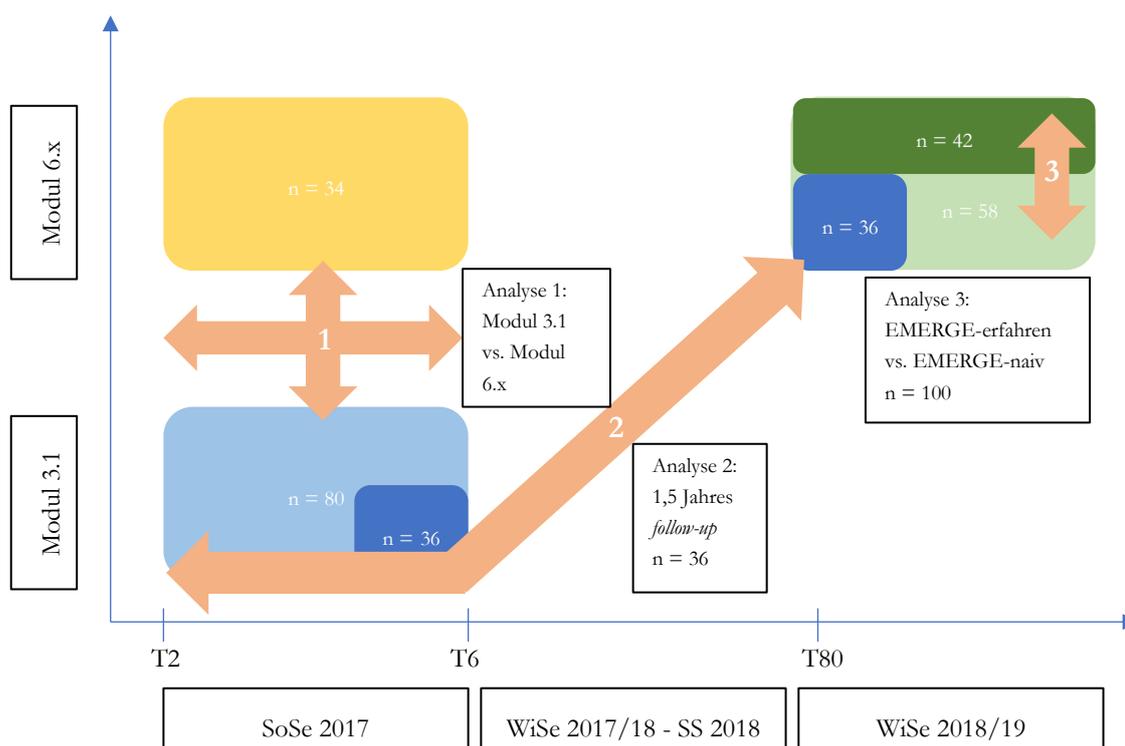


Abbildung 5: Veranschaulichung des Studiendesigns und der Forschungsfragen

2.1 Studiendesign

Diese prospektive Interventionsstudie wurde im Sommersemester 2017 und im Wintersemester 2018/19 an der UMG durchgeführt. Die Studierenden aus dem dritten und sechsten klinischen Semester im Sommersemester 2017 sowie im Wintersemester 2018/19 nahmen an der Studie teil.

Das Modul 3.1 dauert sechs Wochen und beinhaltet Vorlesungen, Seminare, Fallbesprechungen im Kleingruppenunterricht, pathologische Demonstrationen, Auskultationsübungen, EMERGE-Seminare und Unterricht am Krankenbett. Die Studierenden in Regelstudienzeit nehmen eineinhalb Jahre später am sechseinhalbwöchigen Modul 6.x teil. In diesem Repetitionsmodul werden, zur Vorbereitung auf das Staatsexamen und das praktische Jahr, wichtige Krankheitsbilder der vorangehenden Module wiederholt. Dabei werden wochenweise thematische Schwerpunkte gesetzt, beispielsweise steht in der ersten Modulwoche die kardiopulmonale Lehre im Fokus. Die Krankheitsbilder werden vom Leitsymptom ausgehend betrachtet. Als Lehrveranstaltungen werden Vorlesungen, Seminare und fallbasiertes Lernen, entweder mit POLEMA oder der virtuellen Notaufnahme EMERGE, angeboten.

Zuvor wurde nur im Repetitionsmodul im sechsten klinischen Semester mit dem *serious game* gearbeitet. Zum Zeitpunkt dieser Studie wurde EMERGE zusätzlich im Modul 3.1 implementiert.

Bisher ist EMERGE für die Göttinger Studierenden nur in Form dieser Seminare verfügbar und nicht für den privaten Gebrauch.

2.2 Proband:innenrekrutierung

Zum Zeitpunkt dieser Studie besuchten Teilnehmende der Module 3.1 (148 Studierende) und 6.x (118 Studierende) EMERGE-Seminare.

Die Studierenden, die im Sommersemester 2017 für die Module 3.1 und 6.x angemeldet waren, erhielten eine Woche vor Modulbeginn eine E-Mail, in der sie über diese Studie informiert wurden. Die E-Mail enthielt Informationen zum Hintergrund und dem Ablauf der Studie und die Einverständniserklärung zur Datenauswertung. Am ersten Tag der Module 3.1 und 6.x wurden die Studierenden zudem mündlich über die Studie aufgeklärt. In der Einführungsveranstaltung wurde den Studierenden Raum gegeben, Fragen zur Studie zu stellen. Neben den Studienzielen wurde dabei auf die genutzten Datenquellen und insbesondere Aspekte des Datenschutzes eingegangen. Die Studierenden wurden darüber informiert, dass sie den Durchführenden der Studie durch Unterzeichnen der Einverständniserklärung die Genehmigung erteilen, ihre persönlichen Daten in anonymisierter Form auszuwerten. Die unterzeichneten Einverständniserklärungen wurden vor Beginn des ersten Spieltermins eingesammelt.

Die Teilnahme an den Lehrveranstaltungen war unabhängig von diesem Forschungsprojekt verpflichtend im Rahmen der modularen Lehre mit Anwesenheitspflicht zu mindestens 80 % der Lehrveranstaltungen.

Bei dem jeweils ersten EMERGE-Termin wurde darüber hinaus das Konzept der *Think-aloud*-Interviews erläutert und die Studierenden konnten Fragen hierzu stellen. Studierende, die sich grundsätzlich eine Teilnahme an den Interviews vorstellen konnten, erhielten dazu ein ergänzendes Informationsblatt und Einverständniserklärung. Ein Studieneinschluss für die TA-Interviews erfolgte mit der Abgabe der unterschriebenen Einverständniserklärung zu Beginn des jeweils zweiten EMERGE-Termins. TA-Interviews erfolgten dann ab dem zweiten EMERGE-Termin.

2.3 Organisation und Durchführung

Die EMERGE-Seminare fanden im Computerraum DiPS 3 in der UMG statt. Es bestanden technische Kapazitäten von bis zu 50 Teilnehmenden, je nach Größe der Seminargruppen. Die Lehrveranstaltung wurde von einem Arzt geleitet. Zudem war zur technischen Unterstützung eine studentische Hilfskraft sowie im Hintergrund die Lehr-IT anwesend. Die Studierenden loggten sich mit ihrer Matrikelnummer im Computer ein. Somit konnte eine Anwesenheitskontrolle zu allen Terminen erfolgen. Sämtliche Daten, die während des Spiels erzeugt wurden, konnten als *Logfiles* in Form einer Excel[®]-Datei eingesehen werden. Die *Logfiles* wurden anonymisiert und danach ausgewertet. Die Studierenden des Moduls 3.1 nahmen an sechs EMERGE-Terminen von jeweils 90 Minuten teil; die Studierenden des Moduls 6.x an elf Terminen von jeweils 90 Minuten. In den ersten sechs EMERGE-Sitzungen lösten beide Kohorten dieselben Fälle, die inhaltlich alle aus der Kardiologie und Pneumologie stammten. Die letzten fünf Sitzungen des Moduls 6.x beinhalteten auch Patientenfälle aus weiteren internistischen Fachrichtungen, wie beispielsweise Hämatologie oder Gastroenterologie. Deshalb wurden diese Sitzungen im Rahmen der Studie zugunsten der Vergleichbarkeit beider Kohorten nicht berücksichtigt.

	Woche 1		Woche 2		Woche 3		Woche 4		Woche 5		Woche 6	
Modul 3.1	T1		T2		T3		T4		T5		T6	
Modul 6.x	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	

Abbildung 6: Verteilung der Spieltermine in den beiden Modulen.

Die in die Studie mit einbezogenen Spielsitzungen sind gelb unterlegt.

Der erste Spieltermin diente der Einarbeitung der Studierenden in die Handhabung des *serious games*. Die gewonnenen Daten aus dieser Sitzung beider Kohorten wurden deshalb nicht in die Studie einbezogen. Am zweiten und sechsten EMERGE-Termin wurden den Studierenden in der virtuellen Notaufnahme dieselben Krankheitsbilder präsentiert, um einen direkten longitudinalen Vergleich zu Beginn und zum Ende des Moduls zu ermöglichen. Pro Termin wurden zwischen sechs und acht Fälle bearbeitet. Für diese Studie wurde beim zweiten bis sechsten Spieltermin festgelegt, welcher Patient sich bereits in einem Behandlungszimmer in der virtuellen Notaufnahme befinden sollte.

Termin 1	Termin 2	Termin 3	Termin 4	Termin 5	Termin 6
Hypoglykämie	Stabile LAE	Vorhofflimmern	AVNRT	Aortenstenose	Stabile LAE
Hyperthyreose	NSTEMI	Hyponatriämie	Asthma	Pneumonie	NSTEMI
Gastroenteritis	Hypertonus	Diastolische Herzinsuffizienz	NSTEMI	Mitralinsuffizienz	Hypertonus
Hyponatriämie	Pneumothorax	Pneumonie	VHF	COPD	VHF
Alkohol-intoxikation	STEMI	Aortendissektion	Stabile LAE	STEMI	
Harnwegsinfekt	Hyperthyreose	Bronchitis	Synkope	Herzinsuffizienz	
	Hypoglykämie	Systolische Herzinsuffizienz	Hypertonus	VH-Septumdefekt	

Abbildung 7: Fallvorkommen pro Spieltermin

Die in die Studie mit einbezogenen Spielsitzungen sind gelb unterlegt.

Zu jedem der Spieltermine lösten die Studierenden 90 Minuten lang EMERGE-Fälle. Das klinische Denken der Studierenden beider Kohorten wurde am jeweils zweiten und sechsten Spieltermin anhand ihrer Spielleistung in der virtuellen Notaufnahme mit folgenden vier Methoden verglichen. Die *Logfiles* von Termin zwei (T2) und Termin sechs (T6) wurden ausgewertet und als Beginn und Ende des Moduls miteinander verglichen, um den potenziellen Lernzuwachs während des Semesters zu messen. Zu Termin zwei und sechs nahmen außerdem acht Studierende pro Kohorte an *Think-aloud*-Interviews teil. Zu jedem der Spieltermine zwei bis sechs fertigten die Studierenden ein Übergabeprotokoll zu einem festgelegten EMERGE-Fall an. Zu Termin sechs nahmen alle Studierenden außerdem an einer *Key-feature*-Prüfung teil.

Dieser Abschnitt der Studie dient der Beantwortung der ersten Forschungsfrage:

1. Inwiefern unterscheiden sich Studierende des dritten und sechsten klinischen Semesters hinsichtlich ihrer Befähigung zum klinischen Denken? Zu welchem Zeitpunkt im Medizinstudium erzeugt die Nutzung einer digitalen Notaufnahme-Simulation den größten Lernerfolg?

Termin 2	Termin 3	Termin 4	Termin 5	Termin 6
Übergabeprotokolle				
<i>Logfiles</i>				<i>Logfiles</i>
<i>Think-aloud-Interviews</i>				<i>Think-aloud-Interviews</i>
				<i>Key-feature-Prüfung</i>

Abbildung 8: Methodische Auswertung pro Spieltermin

Eineinhalb Jahre nach diesem Zeitpunkt der Datenerhebung bestanden im Wintersemester 2018/19 die Teilnehmenden des Moduls 6.x etwa zur Hälfte aus Regelzeitstudierenden und zur anderen Hälfte aus Studierenden außerhalb der Regelstudienzeit. Die Studierenden im Modul 6.x im WiSe 2018/19 in Regelstudienzeit hatten im SoSe 2017 EMERGE im Rahmen ihres Moduls 3.1 gespielt. Die nicht in Regelstudienzeit Studierenden hatten zu einem früheren Zeitpunkt das Modul 3.1 besucht. Da das *serious game* im Sommersemester 2017 erst für das Modul 3.1 implementiert wurde und bisher außerhalb des Unterrichts nicht verfügbar ist, hatten diese Studierenden bisher keine Berührungspunkte mit dem *serious game* und waren somit EMERGE-naiv. Aufgrund dieser Umstände konnten nun zum einen die Teilnehmenden des Moduls 6.x im Wintersemester 2018/19 in EMERGE-erfahrene und EMERGE-naive Studierende unterteilt und die beiden Kohorten untereinander verglichen werden. Zum anderen konnten die Studierenden, die bereits im Modul 3.1 EMERGE als Teil ihres Curriculums gespielt hatten, nun eineinhalb Jahre später mit einem *follow-up*-Termin im Modul 6.x im Wintersemester 2018 zu drei Zeitpunkten verglichen werden. Da seit Beginn des Moduls 3.1 im Sommersemester 2017 80 Wochen vergangen waren, wurde der Spieltermin T80 genannt. Somit konnten die zweite und dritte Forschungsfrage untersucht werden:

2. Ist der vermutete Lernzuwachs auch nach 1.5 Jahren noch festzustellen?
3. Wie unterscheidet sich die Spielleistung von Studierenden des 6. klinischen Semesters, die bereits zu Beginn des klinischen Studiums mit EMERGE trainiert haben, von Studierenden desselben Semesters, die bisher nicht EMERGE gespielt haben?

Zur Auswertung von T80 wurde nur eine Methode verwendet, nämlich die Auswertung der *Logfiles*. Zu T2, T6 und T80 wurden die gleichen Fallbeispiele präsentiert.

2.4 Messung von *clinical reasoning*

Im Folgenden werden die Messmethoden für klinisches Denken aufgeführt. Dabei wurden zur Analyse 1 *Logfiles*, Übergabeprotokolle, *Key-feature*-Prüfungen und *Think-aloud*-Interviews angewendet, für Analyse 2 und 3 die Auswertung der *Logfiles*.

2.4.1 *Logfiles*

Während die Studierenden an einem SG-Seminar teilnahmen, wurden alle Aktivitäten im Computerspiel in *Logfiles* als Microsoft Excel® Dateien aufgezeichnet.

Events:								
48	8:51 AM	NSTEMI_MN_exam.exc	Spawn	NSTEMI_MN_exam.exc				
138	8:51 AM	NSTEMI_MN_exam.exc	AssignToRoom	Behandlungsraum 3,Bed 2				
319	8:53 AM	NSTEMI_MN_exam.exc	Anamnesis	Wie heißen Sie?				
323	8:53 AM	NSTEMI_MN_exam.exc	Anamnesis	Wie geht es Ihnen?				
332	8:53 AM	NSTEMI_MN_exam.exc	Anamnesis	Welche Beschwerden führen Sie ins Krankenhaus?				
340	8:53 AM	NSTEMI_MN_exam.exc	Anamnesis	Wie alt sind Sie?				
348	8:53 AM	NSTEMI_MN_exam.exc	Anamnesis	Sind bei Ihnen irgendwelche Erkrankungen bekannt?				
357	8:54 AM	NSTEMI_MN_exam.exc	Anamnesis	Nehmen Sie regelmäßig Medikamente?				
367	8:54 AM	NSTEMI_MN_exam.exc	Anamnesis	Haben Sie Probleme beim Atmen?				
380	8:54 AM	NSTEMI_MN_exam.exc	Anamnesis	Haben Sie Schmerzen? Wenn ja, können Sie diese beschreiben?				
386	8:54 AM	NSTEMI_MN_exam.exc	Anamnesis	Haben Sie manchmal Brustschmerzen?				
401	8:54 AM	NSTEMI_MN_exam.exc	PerformAction	Körperl. Unter. Thorax				
570	8:55 AM	NSTEMI_MN_exam.exc	PerformAction	Labor => Labs: Profile: Blutbild,BGA,Infarkt-Profil,				
613	8:55 AM	NSTEMI_MN_exam.exc	PerformAction	Elektrokardiographie (EKG)				
1549	9:02 AM	NSTEMI_MN_exam.exc	PerformAction	Morphin 10 mg i.v.				
2074	9:06 AM	NSTEMI_MN_exam.exc	Diagnosis	NSTEMI				
2124	9:07 AM	NSTEMI_MN_exam.exc	PerformAction	Heparin 5000 IE i.v.				
2313	9:08 AM	NSTEMI_MN_exam.exc	AbortAction	Dabigatran 110 mg p.o.				
2318	9:08 AM	NSTEMI_MN_exam.exc	PerformAction	Acetylsalicylsäure 500 mg i.v.				
2383	9:09 AM	NSTEMI_MN_exam.exc	PerformAction	Ticagrelor 90 mg p.o.				
2449	9:10 AM	NSTEMI_MN_exam.exc	PerformAction	Glyceroltrinitrat akut Spray 0,4 mg p.o.				
2560	9:10 AM	NSTEMI_MN_exam.exc	PerformAction	Herzkatheteruntersuchung				
4216	9:31 AM	NSTEMI_MN_exam.exc	Transfer	Intermediate Care Station				

Abbildung 9: Beispiel eines *Logfiles* des Falles NSTEMI

Die *Logfiles* der Termine zwei und sechs, die im direkten Vergleich zueinander gestellt werden können, da an diesen beiden Tagen dieselben Fälle gespielt wurden, wurden aus der Excel®-Tabelle ausgelesen und in eine SPSS®-Datei kodiert. Somit konnte jede einzelne Handlung jedes Studierenden in der virtuellen Notaufnahme statistisch auswertbar gemacht werden. Hierbei handelt es sich um die Fälle „stabile Lungenarterienembolie“, „arterielle Hypertonie“ und „NSTEMI“. Von den Studienleitern wurden die wichtigsten Schlüsselmomente für die Bearbeitung dieser drei Fälle festgelegt, anhand derer das klinische Denken beurteilt werden

sollte. Diese wurden zu Checklisten für den jeweiligen Fall zusammengefasst, die in Abbildung 10 nachzuvollziehen sind. Diese Punkte waren den Studierenden nicht bekannt. Wurde die richtige Frage gestellt beziehungsweise die richtige Maßnahme unternommen, wurde „1“ codiert, wenn nicht, „0“. So konnten die Studierenden für jeden der drei virtuellen Patient:innen jeweils maximal 14 (LAE), 10 (Hypertensive Krise) und 19 (NSTEMI) Punkte erreichen. Die Gesamtpunktzahlen wurden in die Kategorien Anamnese und Management aufgeteilt. Um eine Vergleichbarkeit zwischen den Fällen herzustellen, wurden die Punktzahlen in Prozent umgerechnet.

LAE	Hypertensive Krise	NSTEMI
Anamnese		
Husten	Schwindel	Dyspnoe
Schmerzen	Anamnese des Leitsymptoms	Beginn des Leitsymptoms
Anamnese des Leitsymptoms	Vorerkrankungen	Anamnese des Leitsymptoms
Rauchen	Medikamentenanamnese	Rauchen
Vorerkrankungen		Allergien
Medikamentenanamnese		Vorerkrankungen
		Familienanamnese
		Medikamentenanamnese
Anamnese: max. 6 Punkte	Anamnese: max. 4 Punkte	Anamnese: max. 8 Punkte
Management		
BGA	Körperliche Untersuchung gesamt	Körperliche Untersuchung des Thorax
Troponin	Kleines Blutbild/ Routineprofil	Labor: Troponin T
D-Dimere	EKG	Labor: CKMB
EKG	Richtige Diagnose	Labor: TSH
Echokardiographie	Nitrat	EKG
CT Thorax	Antihypertensive Behandlung	Koronarangiographie
Korrekte Diagnose	(Amlodipin, Urapidil, Metoprolol, Enalapril, Verapamil)	Korrekte Diagnose
Antikoagulation		Medikamentöse Behandlung: ASS
		Medikamentöse Behandlung: Heparin
		Medikamentöse Behandlung: Nitrat
Management: max. 8 Punkte	Management: max. 6 Punkte	Management: max. 11 Punkte
Insgesamt: max. 14 Punkte	Insgesamt: max. 10 Punkte	Insgesamt: max. 19 Punkte

Abbildung 10: Kriterien zur Auswertung der *Logfiles* für die drei Fälle

Pro Fallbeispiel konnten Studierende maximal 19 (NSTEMI), 14 (Stabile LAE) und 10 (Hypertonus) Punkte bekommen, also insgesamt maximal 43 Punkte. Die Gesamtpunktzahl wurde außerdem in „Anamnese“ und „Management“ aufgeteilt.

2.4.2 Übergabeprotokolle

Zum Spieltermin eins bis sechs verfassten die Studierenden nach der jeweiligen EMERGE-Sitzung eine Übergabe für die jeweiligen festgelegten ersten Patient:innen, die ihnen in der virtuellen Notaufnahme begegneten. Dabei war der erste Spieltermin für die Studierenden als Übung gedacht und wurde nicht für diese Studie ausgewertet. Am zweiten und sechsten Termin wurde derselbe klinische Fall (LAE) als erstes gezeigt und musste dementsprechend beschrieben werden. Folgende Aufgabenstellung wurde standardisiert in allen Seminargruppen zu Beginn vorgelesen:

„Fast geschafft, die Schicht auf der Notaufnahme geht zu Ende. Können Sie sich noch an den Patienten erinnern, der ganz am Anfang der Schicht schon in einem der Zimmer lag? Bitte fassen Sie seine Anamnese, ihre diagnostischen Entscheidungen und den klinischen Verlauf in folgendem Freitextfeld kurz zusammen (maximal 10 Zeilen).“

Für die Erfassung der Übergabetexte wurde das Programm EvaSys® benutzt. Die Studierenden gaben wie bei der Anmeldung für das Spiel EMERGE auch bei EvaSys® ihre Matrikelnummer an. Daher konnten SG-Seminar und Übergabetext miteinander korreliert und danach zur Auswertung anonymisiert werden.

Die Übergabetexte wurden anhand der an den jeweiligen Fall angepassten inhaltlichen Kriterien, die den Studierenden nicht bekannt waren, bewertet. Pro Fall wurden sieben bis neun Inhalte zur Messung von *clinical reasoning* festgelegt und die Übergaben somit dichotom quantifiziert (1 = Kriterium wurde in der Übergabe erwähnt; 0 = wurde nicht erwähnt) und damit vergleichbar gemacht.

Für die Übergaben wurden vom Leiter dieser Studie vier EMERGE Fälle ausgesucht, die in inhaltlichem Bezug zu den gelehrtten Krankheitsbildern in der jeweiligen Modulwoche des Moduls 3.1 standen (stabile Lungenarterienembolie, AVNRT, VHF, Aortenklappenstenose). Im Repetitionsmodul 6.x waren die Krankheitsbilder der Kardiologie und Pneumologie bereits in der ersten Modulwoche wiederholt worden.

Der Fall „stabile Lungenembolie“ wurde zu den Terminen zwei und sechs, also den beiden Messzeitpunkten der Analyse 1, gezeigt. Für diesen Fall wurden acht Inhalte festgelegt, die

im Übergabetext erwähnt werden sollten. Anhand dieser Schlüsselbegriffe wurde gemessen, inwiefern die Studierenden die Essenz dieses Fallbeispiels verstanden haben, und wie ihr klinisches Denken auf dem Weg vom Leitsymptom zu Diagnose und Therapie funktioniert. Es wurde jeweils ein Punkt für das Erwähnen des Leitsymptoms „Atemnot“, der Sauerstoffsättigung, der Tachykardie, des Wells Scores, der D-Dimere, der echokardiographisch diagnostizierten Rechtsherzbelastung, eines CTs sowie der richtigen Diagnose vergeben. Somit konnten insgesamt maximal acht Punkte erreicht werden. Die Punktzahlen wurden zur besseren Vergleichbarkeit in Prozent umgerechnet.

2.4.3 *Key-feature-Prüfung*

Im Rahmen dieser Studie fand zum Zeitpunkt T6 eine formative (unbenotete) *Key-feature-Prüfung* statt. Für jedes richtig beantwortete KF erhielten die Studierenden einen Punkt. Es wurden fünf Fallbeispiele aus dem Themengebiet kardiopulmonale Lehre gestellt (Dilatative Kardiomyopathie, Pneumonie, AVNRT, VHF, Aortenstenose).

Der 75-jährige Herr Droschus ist noch gut zu Fuß – in diesem Jahr wollte er mit seinen Freunden vom Kegelclub zum Bismarckturm wandern, doch auf halber Strecke wurde ihm plötzlich schwindelig.
Er konnte sich gerade noch hinsetzen, bevor er für kurze Zeit das Bewußtsein verlor.
Als die sofort verständigte Notärztin eintrifft, geht es Herrn Droschus schon wieder gut.

Er kann ihr berichten, daß er früher einmal Rhythmusstörungen hatte; mit dem Betablocker und dem Medikament „Cordarex“ habe er damit aber schon seit längerem keine Last mehr.
Die Ärztin untersucht Herrn Droschus und tastet einen sehr langsam ansteigenden, eher schwachen Puls.
Bei der Herz-Auskultation hört sie ein spindelförmiges 4/6-Systolikum mit Punctum maximum im 2. Intercostalraum rechts parasternal und Fortleitung in die Carotiden.

Frage ID: 936609, Fragetyp: Long Menu, Antworten gefordert: 1, Punkte: 1.0

Welche Verdachtsdiagnose lässt sich aus dem Untersuchungsbefund ableiten?
(Bitte kreuzen Sie **eine** Antwort an!)

Abbildung 11: Fallvignette „Aortenklappenstenose“ aus der *Key-feature-Prüfung*

2.4.4 *Think-aloud-Interviews*

Am zweiten und sechsten EMERGE Spieltermin nahmen pro Seminargruppe zwei Studierende an einem TA-Interview teil. Zu Beginn des zweiten EMERGE Termins wurden dafür pro Seminargruppe zwei Freiwillige gesucht. Bei jeweils vier Seminargruppen à circa 50 Studierenden der Module 3.1 und 6.x nahmen pro Kohorte also acht Studierende an den TA-Interviews teil. Während der Rest der Kohorte wie gewohnt 90 Minuten lang im selben Raum mit dem *serious game* arbeitete, spielten die TA-Interviewten einzeln in einem anderen Raum

jeweils 30 Minuten lang. In dieser Zeit bearbeiteten sie die drei Fälle „stabile Lungenembolie“, „hypertensive Krise“ und „NSTEMI“. Dabei wurden sie von einem studentischen Mitglied der Arbeitsgruppe Medizindidaktik und Ausbildungsforschung beaufsichtigt. Dieselben Freiwilligen nahmen am zweiten und sechsten Termin an einem TA-Interview teil. Die Reihenfolge des Erscheinens der Patient:innen im Spiel war zufällig. Die Studierenden, die nicht am TA-Interview teilnahmen, behandelten dieselben Fälle, doch aufgrund der längeren Spielzeit wurden ihnen darüber hinaus noch weitere Fallbeispiele präsentiert. An allen weiteren Spielterminen besuchten die TA-Interviewten regulär das EMERGE Seminar ihrer Gruppe. Als Anerkennung für die besondere Teilnahme an diesem Teil der Studie erhielten die TA-Interviewten einen Büchergutschein.

Für das TA-Interview sollten die Studierenden alle Gedanken, die ihnen während der halben Stunde in der virtuellen Notaufnahme in den Sinn kamen, laut aussprechen.

Jedes TA-Interview begann mit folgender Aufgabenstellung:

„Vielen Dank, dass Sie sich dazu bereit erklärt haben, an den *Think-aloud*-Interviews teilzunehmen. Was bedeutet das eigentlich? Tatsächlich geht es um Lautes Denken. Wir bitten Sie hierfür, während Sie im Spiel EMERGE Patienten behandeln, Ihre Gedanken zu formulieren. Beschreiben Sie, was Sie tun und weshalb Sie welche Entscheidung treffen, vom Eintreffen des Patient:innen in Ihrer Notaufnahme bis zur gestellten Diagnose und der Verlegung des Patienten.“

Daraufhin wurde das Spiel gestartet und die 30-minütige EMERGE Sitzung mit einem Diktiergerät aufgenommen. Die Studierenden wurden bei zu langen Schweigepausen daran erinnert, laut zu denken. Außerdem wurden beim zweiten Spieltermin eventuelle kurze Zwischenfragen zur Spielführung gestattet und beantwortet. Ansonsten schwieg die beaufsichtigende Person und verzichtete auf Hinweise und Wertung in jeglicher Form. Die Aufnahmen wurden im Anschluss transkribiert.

Bewertet wurden diese 16 anonymisierten Transkripte von zwei unabhängigen Oberärzten und Medizindidaktikern mit Expertise sowohl zum Thema *serious games* als auch *clinical reasoning*. Es war für die Bewertenden nicht erkenntlich, welches Transkript welchem Messzeitpunkt oder welcher Kohorte zuzuordnen war.

Die beiden Oberärzte markierten die Textpassagen, in denen sie klinisches Denken entdeckten. Die Bewertung der TA-Transkripte der beiden Auswertenden wurde dann miteinander verglichen. Im Folgenden wurden die bewerteten Transkripte von den Studienleitenden eingesehen und die markierten Textpassagen kategorisiert und somit zu 22 *items* des klinischen

Denkens zusammengefasst. Beispielsweise wurden „Hypoglykämie kann Kopfschmerzen verursachen“, „Bluthochdruck kann Kopfschmerzen verursachen“ und „Migräne kann Kopfschmerzen verursachen“ aus unterschiedlichen Transkripten zu einem *item* „potential causes of headache“ vereint. Die Anzahl der *items* pro Modul und Zeitpunkt wurden miteinander verglichen. Dafür wurden die Mittelwerte der *item*-Punktzahlen gebildet, die beide Oberärzte und die einer der beiden Oberärzte markiert hatte.

2.5 Statistische Analysen

Es wurde zunächst eine deskriptive Statistik der *Logfiles*, Übergabeprotokolle, TA-Interviews und der KF-Prüfung zur Beschreibung der Häufigkeiten mit Hilfe von Mittelwert, Standardabweichung, Median, Minimum und Maximum erstellt. Als Fehlerindikator wurde der Standardfehler angegeben. Das Studienkollektiv wurde in Bezug auf Alter und Geschlecht beschrieben. Um die Leistung der Studierenden zu erfassen, wurde die dichotome Auswertung der *Logfiles* zu einer Punktzahl pro Fall addiert. Es wurde sowohl mit dieser Gesamtpunktzahl gearbeitet als auch mit den einzelnen Punktzahlen für den Bereich „Anamnese“ und „Patienten Management“ (leitliniengerechte diagnostische Maßnahmen, Diagnosestellung und Therapie).

Für die erste Analyse wurde die Leistungsdifferenz der Studierenden in der *Logfile*-Auswertung beider Module zwischen dem zweiten und sechsten Spieltermin im Sommersemester 2017 anhand von Varianzanalysen (*analyses of variance*, ANOVAs) erhoben. Außerdem wurden die Anteile von richtig gestellten Diagnosen pro Fall zwischen den Spielterminen zwei und sechs mit McNemar-Tests verglichen. Die Punktsummen der dichotomisierten Übergabeprotokolle wurden in Prozent umgerechnet und die Ergebnisse von Termin zwei mit Termin sechs verglichen. Darüber hinaus wurde eine Variable definiert, um die *Logfile*-Ergebnisse mit den Übergabe-Ergebnissen zu vergleichen. Anhand dieser Variable wurde eine Korrelation zwischen beiden Methoden zu den jeweiligen Messzeitpunkten berechnet. Die KF-Prüfung zu Termin sechs wurde ausgewertet und die Ergebnisse der Kohorten 3.1 und 6.x mit Hilfe von unabhängigen T-Tests verglichen. Die Ergebnisse der Auswertung der TA-Interviews wurden anhand von Häufigkeiten beschrieben.

Für die zweite Analyse wurden die Leistungsveränderungen in den *Logfiles* der longitudinalen Kohorte zu T2, T6 und T80 mittels Friedman-Test erhoben. Es wurden die Gesamtpunktzahlen, sowie die Punktzahlen aufgeteilt auf „Anamnese“ und „Management“ aufgeteilt betrachtet. Zur Signifikanztestung wurden gepaarte T-Tests durchgeführt, um Unterschiede zu

den drei Messzeitpunkten zu eruieren. Außerdem wurden die Übergabeprotokolle des Spieltermins T80 in die Analyse mit einbezogen.

Für die dritte Analyse wurden zum Zeitpunkt T80 die drei zuvor beschriebenen Punktzahlen in der *Logfile*-Auswertung der EMERGE-naiven Kohorte mit denen der EMERGE-erfahrenen Kohorte verglichen. Außerdem wurden die erreichten Punktzahlen des Übergabeprotokolls von T80 untereinander verglichen. Dies geschah jeweils mit Hilfe von unabhängigen T-Tests.

Die Daten wurden als Mittelwert \pm Standardabweichung aufgeführt. Für die statistische Analyse wurde das Statistikprogramm SPSS für Windows, Version 14.0.1 (SPSS Inc. Chicago, Illinois, USA) verwendet. Das Signifikanzniveau wurde auf 5 % festgelegt.

3 Ergebnisse

3.1 Zusammensetzung der Kohorten

Im Sommersemester 2017, also zum ersten Messpunkt dieser zweizeitigen Studie, spielten zwei Kohorten, Medizinstudierende des Moduls 3.1 (Kardiopulmonale Lehre) und 6.x (Repetitionsmodul), im Rahmen ihrer curricularen Lehre das *serious game* EMERGE. Zu diesem Zeitpunkt waren im Modul 3.1 147 Studierende eingeschrieben, von denen alle die Einverständniserklärung zur Auswertung ihrer Daten unterschrieben (*informed consent* = IC). Davon nahmen acht Studierende an den TA-Interviews zu Termin 2 und Termin 6 teil. Die *Logfiles* der TA-Interviewten wurden nicht in die Studie mit einbezogen, da die Interviews mit 30 Minuten Dauer im Vergleich zu einer normalen Spielsitzung (90 Minuten) kürzer und somit nicht vergleichbar waren. Außerdem gaben die TA-Interviewten keine Übergabeprotokolle ab. Somit ergab sich für die Auswertung der *Logfiles* und Übergabeprotokolle eine Kohorte von 139 Studierenden ($n = 139$). Im selben Semester entschieden sich 57 Studierende ($n = 57$) des Moduls 6.x für das Lehrformat EMERGE. Aus dieser Kohorte nahmen ebenfalls acht Studierende an TA-Interviews teil. Drei Semester später, im Wintersemester 2018/19, waren die Studierenden, die im Sommersemester 2017 das Modul 3.1 besucht und in Regelstudienzeit weiterstudiert hatten, im letzten klinischen Semester und somit im Modul 6.x angekommen. Im Modul 6.x hatten sie die Wahl zwischen den zwei Lehrformaten EMERGE und POL. In diesem Semester wählten 58 Studierende das Format EMERGE. Davon waren bei 36 Studierenden die Daten aller drei Messzeitpunkte (T2, T6 und T80) vollständig und das Einverständnis zur Datenauswertung lag vor ($n = 36$).

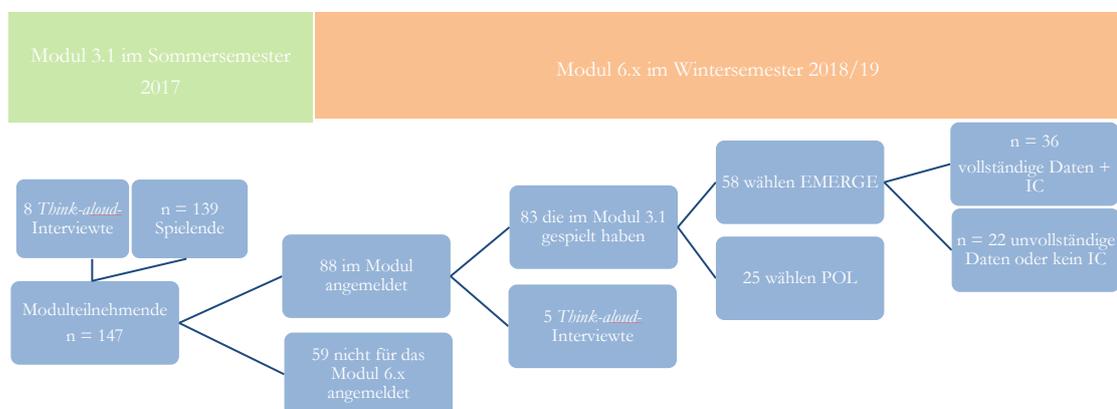


Abbildung 12: Zusammensetzung der EMERGE-erfahrenen Kohorte im WiSe 18/19

Im Wintersemester 2018/19 gab es 67 EMERGE-naive Studierende. Davon entschieden sich 21 für POL und 46 für EMERGE. Von diesen 46 konnten die Daten von 42 in die Studie einbezogen werden, da sie die Einverständniserklärung unterschrieben haben und am einfließenden Spieltermin (T80) anwesend waren. Dies ergibt $n = 42$ für die EMERGE-naiven Studierenden und, wie im vorigen Abschnitt beschrieben, $n = 58$ für die EMERGE-erfahrenen Studierenden.

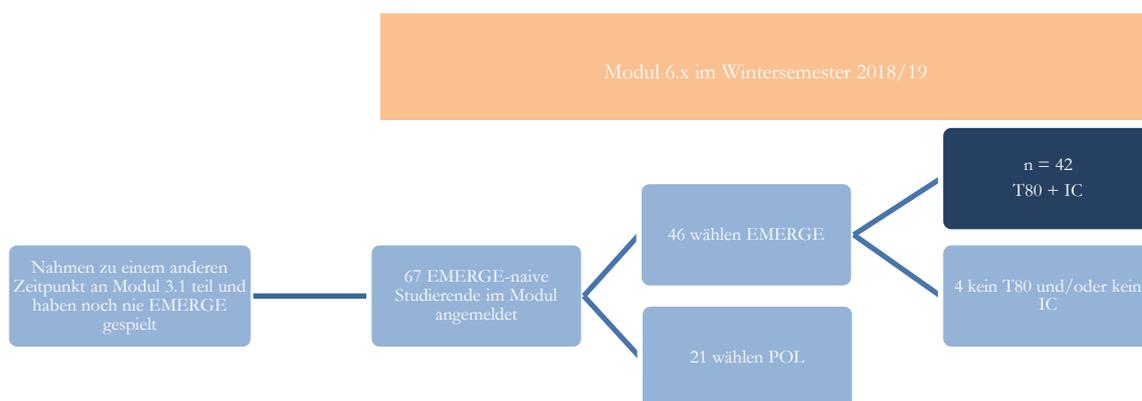


Abbildung 13: Zusammensetzung der Kontrollgruppe „EMERGE-naiv“

Insgesamt unterschrieben 238 Studierende die Einverständniserklärung zur Studienteilnahme (92,2 % der Studierenden). Es wurden zu drei Messzeitpunkten Daten zur Auswertung erhoben. Von den Personen, die zufällig einen der Messzeitpunkte nicht wahrnehmen konnten, flossen keine Daten in die Analyse ein. Die Daten von 178 Studierenden konnten in den drei Analysen berücksichtigt werden (74,8 %). Wie in Abbildung 5 nachzuvollziehen, setzte sich die Gruppe aus Kohorte 1 mit $n = 80$, Kohorte 2 mit $n = 34$ und Kohorte 3 mit $n = 100$ zusammen.

Das Durchschnittsalter der Studierenden lag während der Analyse 1 bei $25,3 \pm 0,4$; während der Analyse 2 bei $24,9 \pm 0,6$ und während der Analyse 3 bei $26,3 \pm 0,4$ Jahren. Im Sommersemester 2017 waren die Studienteilnehmenden der Kohorte 3.1 zu 37,7 % männlich und 62,3 % weiblich, die der Kohorte 6.x zu 50,0 % männlich und 50,0 % weiblich. Die Kohorte 6.x im Wintersemester 2018/19 bestand aus 39,0 % Männern und 61,0 % Frauen.

3.2 Prospektiver Vergleich zwischen Modul 3.1 und 6.x im Laufe des Sommersemesters 2017

3.2.1 *Logfile*-Auswertung

Die *Logfiles* und somit die Spielaktivität der Studierenden beider Module wurde codiert und anhand einer individuellen Checkliste für jeden der drei Fälle (stabile LAE, hypertensive Krise und NSTEMI) ausgewertet. Dabei wurden die Punkte auf der Checkliste in die beiden Kategorien „Anamnese“ und „Management“ aufgeteilt. Die erreichten Prozentpunktzahlen der beiden Kohorten wurden in Prozent miteinander und von Spieltermin zwei zu sechs verglichen. Die Reliabilitätsanalyse der Variablen ergab ein Cronbachs α von 0,708.

Tabelle 3: Erreichte Prozentpunktzahlen der Module 3.1 und 6.x zu Termin zwei und sechs

Kriterien	Modul 3.1 [%]		Modul 6.x [%]	
	Termin 2	Termin 6	Termin 2	Termin 6
Gesamt	55,9 \pm 1,0	63,3 \pm 1,3	57,8 \pm 1,8	65,5 \pm 1,8
Anamnese	55,0 \pm 1,7	57,8 \pm 1,8	59,8 \pm 2,3	60,5 \pm 2,0
Management	56,6 \pm 1,1	67,2 \pm 1,5	56,2 \pm 2,1	69,2 \pm 2,2

Die erreichten Prozentpunktzahlen der beiden Studierendengruppen unterscheiden sich zu Termin 2 (Modul 3.1: 55,9 % \pm 1,0 und Modul 6.x: 57,8 % \pm 1,8) nicht signifikant voneinander ($p = 0,845$). Zu Termin 6 zeigt sich ebenfalls kein signifikanter Unterschied zwischen den Gruppen (Modul 3.1: 63,3 % \pm 1,3 und Modul 6.x: 65,5 % \pm 1,8, $p = 0,583$). Aufgeteilt in die Bereiche „Anamnese“ und „Management“ sind ebenfalls keine signifikanten Unterschiede zwischen den erreichten Prozentpunktzahlen der beiden Gruppen festzustellen.

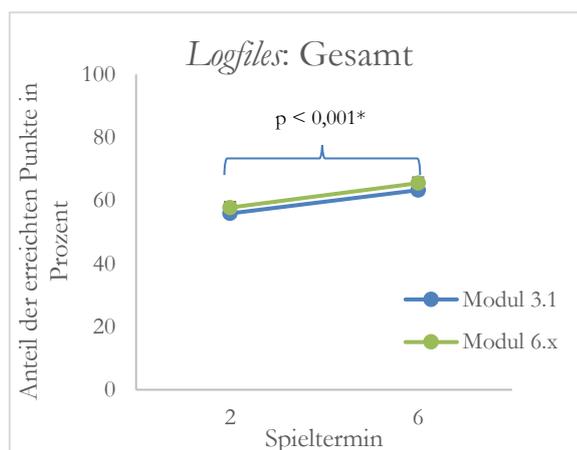
Werden die Prozentpunktzahlen an Termin zwei mit denen an Termin sechs verglichen, ist jedoch ein Unterschied abzulesen. Für Modul 3.1 ist von Termin 2 (55,9 % \pm 1,0) zu Termin sechs (63,3 % \pm 1,3) ein signifikanter Punktzuwachs festzustellen ($p < 0,001^*$), der sich hauptsächlich auf einen Unterschied im Bereich „Management“ zurückführen lässt (56,6 % \pm 1,1 zu Termin zwei, 67,2 % \pm 1,5 zu Termin 6, $p < 0,001^*$). In der Kategorie Anamnese unterscheiden sich die erreichten Prozentpunktzahlen der Studierenden des Modul 3.1 zu

Termin zwei und sechs nicht signifikant voneinander. Im Modul 6.x zeichnet sich ein ähnliches Bild ab. Zu Termin zwei werden $57,8 \% \pm 1,8$ der Punkte erzielt, zu Termin sechs sind es mit $65,5 \% \pm 1,8$ signifikant mehr ($p = 0,001^*$). Auch in diesem Modul lässt sich dies hauptsächlich durch einen Unterschied im Bereich „Management“ erklären. Mit $69,2 \% \pm 2,2$ erreichen die Studierenden zu Termin sechs signifikant mehr Punkte als zu Termin zwei mit $56,2 \% \pm 2,1$, $p < 0,001^*$. Im Modul 6.x unterscheiden sich die Prozentpunktzahlen im Bereich „Anamnese“ zu Termin zwei und sechs ebenfalls nicht signifikant voneinander.

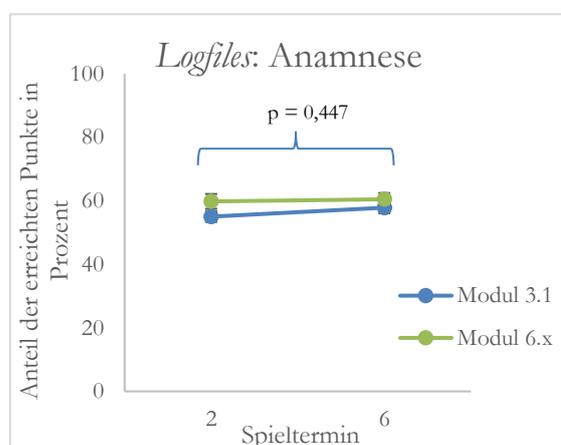
Zusammenfassend ist festzustellen, dass die statistischen Auswertungen der Analyse 1 in beiden Vergleichsgruppen einen signifikanten Punktzuwachs von Termin zwei zu Termin sechs verzeichnen. Die Module 3.1 und 6.x unterschieden sich dabei zu keinem der Zeitpunkte signifikant voneinander.

Kombiniert man die Punktzahlen beider Module, steigen die Gesamtpunktzahlen in Prozent von $57,6 \% \pm 1,1$ auf $65,5 \% \pm 1,2$ ($p < 0,001^*$; $d = 0,656$) an. Der Punktzuwachs ist hauptsächlich auf den Bereich „Management“ zurückzuführen ($57,4 \% \pm 1,2$ auf $69,8 \% \pm 1,3$; $p < 0,001^*$), während im Bereich „Anamnese“ in der *Logfile*-Auswertung kaum Veränderung im Punktescore zwischen Termin zwei und Termin sechs zu sehen ist ($57,9 \% \pm 1,6$ auf $59,6 \% \pm 1,6$; $p = 0,447$). Dieses Muster zeigt sich bei allen drei virtuellen Patient:innen. Dies ist in Abbildung 14 a-c, in der die Ergebnisse beider Kohorten zu Spieltermin zwei und sechs aufgetragen sind, nachzuvollziehen.

a)



b)



c)

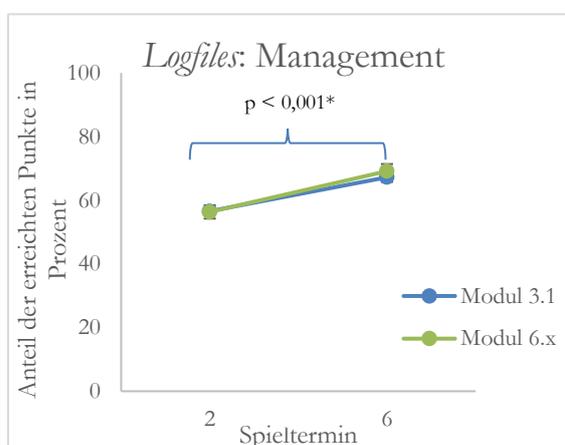


Abbildung 14 a-c: Analyse 1: Veränderung der Spielleistung im Modul 3.1 im SoSe 2017 zwischen Termin zwei und Termin sechs in beiden Modulen

Die p-Werte wurden anhand ANOVAs berechnet und beziehen sich auf den Zeitvergleich.

Aufgrund des fehlenden Gruppenunterschiedes zwischen den beiden Kohorten wurden für die nachfolgenden Berechnungen der *Logfiles* in Analyse 1 Modul 3.1 und Modul 6.x zusammengefasst.

Tabelle 4: Erreichte Prozentpunktzahlen beider Module zusammengefasst zu Termin zwei und sechs

Variable		Termin 2 [%]	Termin 6 [%]	p-Wert	Cohen's d
Lungenembolie, stabil	Anamnese	70,5 ± 1,9	70,5 ± 2,0	1,000	—
	Management	53,9 ± 1,9	73,8 ± 1,7	<0,001	1,02
Hypertensive Krise	Anamnese	55,7 ± 2,4	58,2 ± 2,2	0,343	—
	Management	57,4 ± 2,5	64,8 ± 2,2	0,007	0,34
NSTEMI	Anamnese	46,3 ± 2,6	50,5 ± 2,3	0,110	—
	Management	60,6 ± 1,9	69,0 ± 2,0	<0,001	0,44

Betrachtet man den Anteil an richtigen Diagnosen beider Kohorten gemeinsam, so sieht man einen Zuwachs von 57,9 % auf 80,7 % für den Fall „stabile LAE“ ($p < 0,001^*$), von 50,0 % auf 66,3 % für den Fall „NSTEMI“ ($p < 0,014^*$) und von 80,7 % auf 94,3 % für den Fall „hypertensive Krise“ ($p = 0,002^*$).

In Tabelle 5 kann nachvollzogen werden, auf welche Station die jeweiligen Patient:innen nach Diagnostik und Therapie verlegt wurden, oder ob sie nach Hause entlassen wurden. Die Stichprobe n der Studierenden variiert von Fall zu Fall, da nur die Daten der Studierenden berücksichtigt wurden, die den jeweiligen Fall von Aufnahme bis hin zur Verlegung bearbeitet haben.

Am Ende des Moduls verlegten die Studierenden die virtuellen Patient:innen häufiger auf die richtige Station als zu Beginn des Semesters. Der Trend in allen drei Fällen ging dahin, dass Patient:innen am Ende des Moduls eher auf weniger intensive Stationen verlegt wurden als zu Beginn.

Tabelle 5: Verlegungen beider Module zusammengefasst
Die richtige Lösung ist jeweils fett gedruckt.

Stabile Lungenembolie [%] (n = 51)		
Verlegung	T2	T6
Entlassung	0,0 (0)	2,0 (1)
Normalstation	11,8 (6)	19,6 (10)
Intermediate Care Unit (IMC)	35,3 (18)	58,8 (30)
Intensivstation	49,0 (25)	17,6 (9)
OP	3,9 (2)	2,0 (1)
Hypertensive Krise (n = 58)		
Verlegung	T2	T6
Entlassung	19,0 (11)	75,9 (44)
Normalstation	56,9 (33)	22,4 (13)
IMC	20,7 (12)	1,7 (1)
Intensivstation	3,4 (2)	0,0 (0)
OP	0,0 (0)	0,0 (0)
NSTEMI (n = 22)		
Verlegung	T2	T6
Normalstation	13,6 (3)	27,3 (6)
IMC	54,5 (12)	68,2 (15)
Intensivstation	31,8 (7)	4,5 (1)

3.2.2 Übergabeprotokolle

Zum zweiten und sechsten Spieltermin verfassten die Studierenden der Module 3.1 und 6.x ein Übergabeprotokoll über virtuelle Patient:innen mit einer Lungenarterienembolie. Anhand einer Checkliste bestehend aus acht essenziellen Schritten zur Lösung des Falles wurden die Übergabeprotokolle bewertet. Cronbachs α der Variablen der Übergabeprotokolle beträgt 0,465. Die Studierenden der Module 3.1 und 6.x unterscheiden sich anhand der Gesamtpunktzahlen, die sie in der Auswertung der Übergabeprotokolle erreichten, nicht signifikant voneinander (Tabelle 6). Ungepaarte T-Tests ergeben $p = 0,683$ zu Termin zwei und $p = 0,769$ zu Termin sechs. Daher wurden beide Kohorten für die nach folgenden Analysen zusammengefasst. Beide Kohorten zusammengefasst erwähnten in ihrem Übergabetext am

zweiten Spieltermin 32,2 % \pm 1,7 und am sechsten Spieltermin 54,0 % \pm 1,7 der Punkte auf der Checkliste.

Tabelle 6: Bewertung der Übergabeprotokolle: Häufigkeiten der einzelnen Checklisten-Punkte aufgeteilt nach Modul und Spieltermin

Kriterien	Modul 3.1 (n = 80) [%] (n)		Modul 6.x (n = 34) [%] (n)	
	Termin 2	Termin 6	Termin 2	Termin 6
Leitsymptom Atemnot	83,8 (67)	96,3 (77)	88,2 (30)	100 (34)
Sauerstoffsättigung	25,0 (20)	37,5 (30)	14,7 (5)	35,3 (12)
Tachykardie	36,3 (29)	67,5 (54)	32,4 (11)	58,8 (20)
D-Dimer Testergebnis	36,3 (29)	55,0 (44)	20,6 (7)	50 (17)
CT-Thorax	23,8 (19)	76,3 (61)	50,0 (17)	67,6 (23)
Diagnose LAE	53,5 (42)	90,0 (72)	61,8 (21)	76,5 (26)
Insgesamt	31,8 \pm 1,6	52,4 \pm 1,8	33,3 \pm 3,4	51,3 \pm 3,5

Von Termin zwei zu Termin sechs ist ein signifikanter Zuwachs an CR-*items* in den Übergabeprotokollen festzustellen (beide Module zusammengefasst: $p < 0,001^*$, $d = 0,95$), dies ist in Abbildung 15 graphisch dargestellt. Auch wenn kein signifikanter Gruppenunterschied bezüglich der erreichten Gesamtpunktzahlen festzustellen ist, fällt auf, dass zu Termin sechs die richtige Diagnose „Lungenarterienembolie“ von 90 % der Studierenden im Modul 3.1, doch nur von 76,5 % der Studierenden im Modul 6.x gestellt wird ($p = 0,064$).

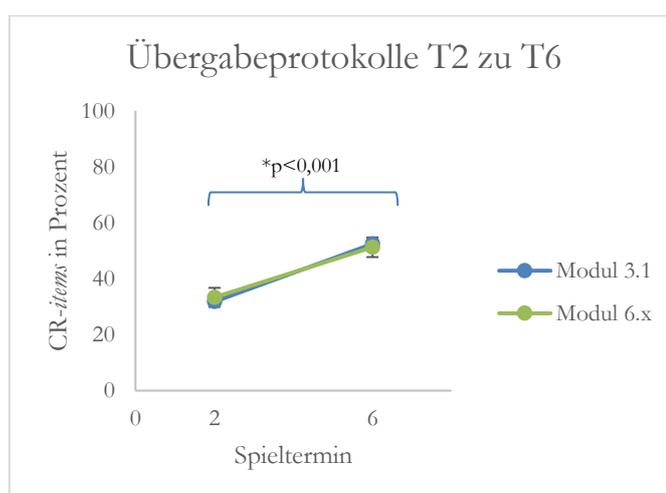


Abbildung 15: Prozentpunktzahlen der in den Übergabeprotokollen erwähnten CR-*items* zu Termin zwei und sechs. Die p-Werte wurden anhand ANOVAs berechnet und beziehen sich auf den Zeitvergleich.

Zwischen den erreichten Prozentpunkten der Übergabeprotokolle und der *Logfiles* für den Fall „stabile LAE“ besteht eine signifikante positive Korrelation. Zu Termin zwei beträgt der Korrelationskoeffizient nach Pearson 0,856 mit einer Signifikanz von $p < 0,001^*$. Zu Termin sechs beträgt der Korrelationskoeffizient nach Pearson 0,718 mit einer Signifikanz von $p < 0,001^*$.

3.2.3 *Key-feature*

Die KF-Prüfung, die zum sechsten Spieltermin stattfand, beinhaltete fünf klinische Fälle, zu denen jeweils fünf KF-Fragen gestellt wurden. Die Reliabilitätsanalyse der Variablen der KF-Prüfung ergibt ein Cronbachs α von 0,796. Das Gesamtergebnis von Modul 6.x liegt mit einem Mittelwert von 64,5 % über dem Ergebnis von Modul 3.1 mit einem Mittelwert von 61,5 %. Dieser Unterschied ist nicht signifikant ($p = 0,393$). Die größten Leistungsunterschiede der beiden Kohorten zeigen sich in den Fällen AVNRT (Modul 3.1 59,8 % \pm 0,2, Modul 6.x 48,8 % \pm 0,3, $p = 0,092$), Pneumonie (Modul 3.1 67,2 % \pm 0,2, Modul 6.x 75,8 % \pm 0,1, $p = 0,040^*$) und dilatative Kardiomyopathie (Modul 3.1 60,0 % \pm 0,1, Modul 6.x 75,0 % \pm 0,2, $p = 0,006^*$). Für die Fälle Aortenstenose (Modul 3.1 73,6 % \pm 0,1, Modul 6.x 74,8 % \pm 0,2, $p = 0,780$) und VHF (Modul 3.1 46,8 % \pm 0,2, Modul 6.x 48,2 % \pm 0,3, $p = 0,804$) ist kein signifikanter Unterschied zwischen den beiden Modulen bezüglich ihrer Leistung in der KF-Prüfung festzustellen.

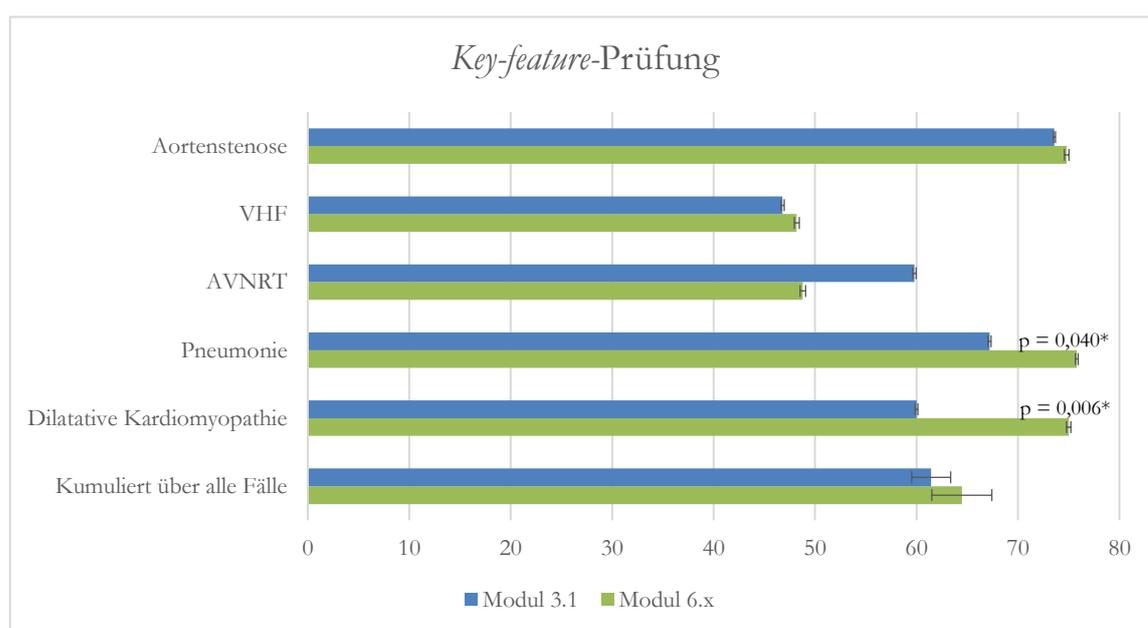


Abbildung 16: Die erreichten Punkte in der KF-Prüfung

Modul 3.1 n = 80, Modul 6.x n = 34; p-Werte durch unabhängige T-Tests

3.2.4 *Think-aloud-Interviews*

Die 16 TA-Transkripte wurden von zwei unabhängigen Oberärzten gelesen und bewertet. Die Textpassagen, in denen die beiden Oberärzte klinisches Denken entdeckten, wurden markiert und später zu 22 *items*, die klinisches Denken zeigen, zusammengefasst. Es wurden pro Kohorte und Spieltermin die Mittelwerte der *item*-Punktzahlen gebildet, die beide Oberärzte markiert hatten, und die, die einer der beiden Oberärzte markiert hatte. In der Tabelle sind diese Mittelwerte der erreichten Punkte aufgeführt. Beide Kohorten scheinen zu Zeitpunkt sechs etwas häufiger klinisches Denken zu zeigen als zu Zeitpunkt zwei. Es besteht ein großer Unterschied zwischen den Punktzahlen, die nur einer der beiden Ärzte markiert hat und denen, die beide Ärzte markiert haben.

Tabelle 7: Erreichte Punktzahlen (von 22), aufgeteilt nach Modul und Spieltermin

	Termin 2		Termin 6	
	Beide Rater	Ein Rater	Beide Rater	Ein Rater
Modul 3.1 (n = 8)	4,0 ± 1,0	9,3 ± 0,8	6,5 ± 0,9	12,4 ± 0,9
Modul 6.x (n = 8)	5,0 ± 1,0	11,1 ± 0,9	5,5 ± 1,1	10,1 ± 1,2
Beide (n = 16)	4,5 ± 0,7	10,2 ± 0,6	6,0 ± 0,7	11,3 ± 0,8

3.3 *Follow-up* des Moduls 3.1 nach eineinhalb Jahren im WiSe 2018/19

Um zu erforschen, ob der in Analyse 1 zwischen T2 und T6 festgestellte Punktezuwachs auch über einen langen Zeitraum andauert, wurde die Spielleistung von den 36 Studierenden, die im SoSe 2017 an Modul 3.1 teilgenommen hatten und eineinhalb Jahre später im WiSe 2018/19 am Modul 6.x teilnahmen, zu T2, T6 und T80 verglichen. T80 stellt dabei ein ein- einhalbjähriges *follow-up* dar.

Mit Hilfe von Friedmann-Tests wurde die Entwicklung der *Logfile*-Ergebnisse in Prozent von T2, T6 bis T80 berechnet. Werden die Ergebnisse der drei Fälle „stabile LAE“, „hypertensive Krise“ und „NSTEMI“ zusammengerechnet betrachtet, ist im Friedmann-Test eine signifi-

kante Veränderung zwischen diesen drei Zeitpunkten festzustellen ($p < 0,001^*$). Die Aufteilung dieser Ergebnisse in die Kategorien „Anamnese“ und „Management“ zeigt, dass diese Signifikanz sich erneut auf die Punktekategorie „Management“ zurückführen lässt ($p < 0,001^*$), während die „Anamnese“-Punktzahlen der Studierenden sich nicht signifikant verändern ($p = 0,376$). P-Werte wurden mit dem Friedmann Test berechnet.

Tabelle 8: Ergebnisse der *Logfile*-Auswertung zu T2, T6 und T80.

Variable		Termin 2 [%]	Termin 6 [%]	Termin 80 [%]	p-Wert
Alle Fälle	Insgesamt	55,7 ± 2,0	65,4 ± 2,3	66,0 ± 1,5	< 0,001*
	Anamnese	54,0 ± 3,1	60,0 ± 3,3	56,5 ± 2,6	0,376
	Management	56,9 ± 1,9	69,4 ± 2,3	72,8 ± 1,5	< 0,001*

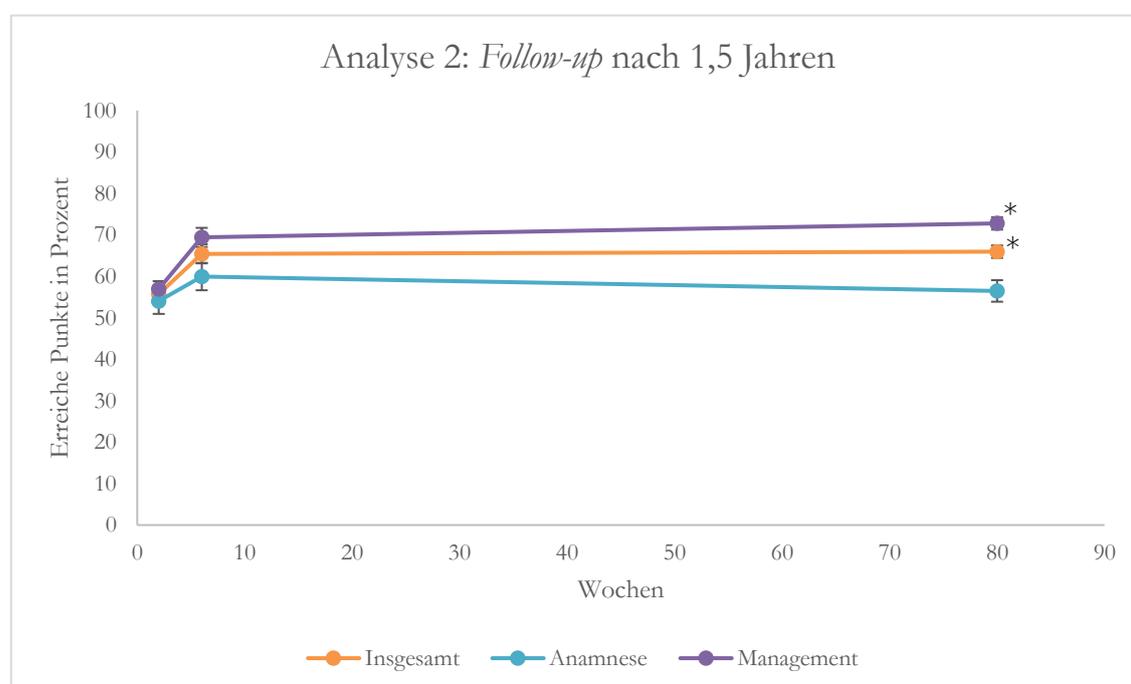


Abbildung 17: Erreichte Prozentpunktzahlen in der *Logfile*-Auswertung zu T2, T6 und T80 als Gesamtpunktzahl sowie aufgeteilt in Anamnese und Management

Werden die Gesamtpunktzahlen der Studierenden zu den drei Zeitpunkten nicht nur in die Kategorien „Anamnese“ und „Management“, sondern auch noch in die drei Fälle aufgeteilt, so zeigt sich im Management der Fälle „stabile LAE“ und „NSTEMI“ eine signifikante Veränderung der erreichten Punkte im Laufe der drei Zeitpunkte. Auch im Management des Falles „hypertensive Krise“ erreichen die Studierenden am Ende des Moduls mehr Punkte, dies ist jedoch kein signifikantes Ergebnis.

Tabelle 9: Erreichte Prozentpunktzahlen in der *Logfile*-Auswertung zu T2, T6 und T80 aufgeteilt nach Fällen sowie den Kategorien Anamnese und Management

Variable		Termin 2 [%]	Termin 6 [%]	Termin 80 [%]	p - Wert
Lungenembolie, stabil (n = 36)	Anamnese	68,5 ± 3,6	72,2 ± 3,3	65,7 ± 3,4	0,299
	Management	55,9 ± 3,4	74,7 ± 3,0	72,6 ± 2,8	< 0,001*
Hypertensive Krise (n = 32)	Anamnese	51,6 ± 4,0	56,3 ± 4,5	64,1 ± 4,0	0,076
	Management	57,3 ± 4,0	64,1 ± 3,5	66,7 ± 2,9	0,218
NSTEMI (n = 32)	Anamnese	44,9 ± 4,9	51,6 ± 5,0	45,7 ± 3,5	0,558
	Management	58,5 ± 3,0	65,9 ± 3,9	76,4 ± 2,3	< 0,001*

Um herauszufinden, zu welchem der drei Messzeitpunkte es zu einer signifikanten Veränderung der erreichten Punkte kam, wurden bei den Kategorien, bei denen der Friedman-Test signifikant ausfiel, gepaarte T-Tests zwischen den einzelnen Zeitpunkten durchgeführt. Die p-Werte der gepaarten T-Tests zwischen Woche zwei und Woche sechs bezüglich des Management-Scores und der Gesamtpunktzahl betragen < 0,05*. Daraus lässt sich schließen, dass es zwischen diesen beiden Messzeitpunkten zu dem signifikanten Punktezuwachs kommt. Zwischen T6 und T80 hingegen kommt es zu keiner signifikanten Veränderung der erreichten Punkte.

Bei der Auswertung der Übergabeprotolle zeigt sich ein ähnliches Muster. Betrachtet man die Ergebnisse der Auswertung der Übergabeprotokolle zu den drei Zeitpunkten, erwähnten die Studierenden (n = 36) zu T2 32,9 % ± 2,7, zu T6 57,9 % ± 2,1 und zu T80 50,4 % ± 2,4 der acht CR-items. Dies ergibt im Friedman-Test einen p-Wert von < 0,001*; die Veränderung im Laufe der Zeit ist somit signifikant. Im gepaarten T-Test ist der Punktezuwachs in

den Übergabeprotokollen zwischen T2 und T6 ist mit $p < 0,001^*$ signifikant; zwischen T6 und T80 kommt es mit $p = 0,020^*$ zu einem signifikanten Punkteverlust.

Auch zum Zeitpunkt T80 besteht zwischen den Prozentscores der Übergabeprotokolle und der *Logfiles* für den Fall „stabile LAE“ eine signifikante positive Korrelation. Der Korrelationskoeffizient nach Pearson beträgt 0,782 mit einer Signifikanz von $p < 0,001^*$.

3.4 Vergleich zwischen EMERGE-erfahrenen und EMERGE-naiven Studierenden im WiSe 2018/19

Im Wintersemester 2018/19 bestand das Modul 6.x aus 58 EMERGE-erfahrenen Studierenden (Interventionsgruppe) und 42 EMERGE-naiven Studierenden (Kontrollgruppe). Diese beiden Gruppen wurden anhand ihrer *Logfiles* miteinander verglichen. Studierende mit vorheriger EMERGE-Exposition erreichten signifikant bessere *Logfile*-Ergebnisse als Studierende, die vorher nicht mit EMERGE gearbeitet hatten ($p = 0,003^*$, $d = 0,618$). Auch hier lässt sich das signifikante Ergebnis besonders auf den Unterschied der „Management“-*items* zurückführen ($p < 0,001^*$, $d = 0,811$), während für die Ergebnisse im Bereich „Anamnese“ keine signifikanten Unterschiede festzustellen sind.

Tabelle 10: Vergleich der EMERGE-erfahrenen (Intervention) und EMERGE-naiven Studierenden (Kontrollgruppe)

Termin 80		EMERGE naiv (n = 42) [%]	EMERGE erfahren (n = 58) [%]	p-Wert	Cohen's d
Alle Fälle zusammengefasst	Anamnese	55,4 ± 2,1	55,9 ± 1,9	0,875	-
	Management	63,5 ± 2,1	72,6 ± 1,2	< 0,001*	0,811
	Gesamtprozentzahl	60,1 ± 1,4	65,6 ± 1,2	0,003*	0,618
Lungenembolie, stabil	Anamnese (%)	63,9 ± 3,6	63,2 ± 2,6	0,878	-
	Management (%)	58,0 ± 3,5	73,7 ± 2,2	< 0,001*	0,808
Hypertensive Krise	Anamnese (%)	63,4 ± 3,5	61,6 ± 3,2	0,709	-
	Management (%)	69,5 ± 3,1	68,5 ± 2,1	0,769	-
NSTEMI	Anamnese (%)	44,9 ± 3,1	47,6 ± 2,3	0,479	-
	Management (%)	64,3 ± 2,2	74,0 ± 1,8	< 0,001*	0,713

Beim einzelnen Betrachten der *Logfile*-Ergebnisse der drei Fälle wird deutlich, dass sich besonders im „Management“ der Fälle „stabile LAE“ und „NSTEMI“ Unterschiede zwischen den beiden Kohorten aufzeigen. Bei der Behandlung der Patient:innen mit einer stabilen Lungenembolie erreichten die EMERGE-naiven Studierenden durchschnittlich $58,0\% \pm 3,5$ im Bereich „Management“; ihre EMERGE-erfahrenen Kommiliton:innen $73,7\% \pm 2,2$. Sie unterscheiden sich signifikant ($p = 0,001^*$). Beim Management von virtuellen Patient:innen mit einem NSTEMI erreichten die EMERGE-naiven Studierenden $64,3\% \pm 2,2$. Ihre EMERGE-erfahrenen Kommiliton:innen waren ihnen mit $74,0\% \pm 1,8$ signifikant überlegen ($p = 0,001^*$). Im Anhang sind in Tabelle 9a-c die Häufigkeiten der einzelnen CR-*items* für die drei Fälle und beide Kohorten im Vergleich aufgeführt.

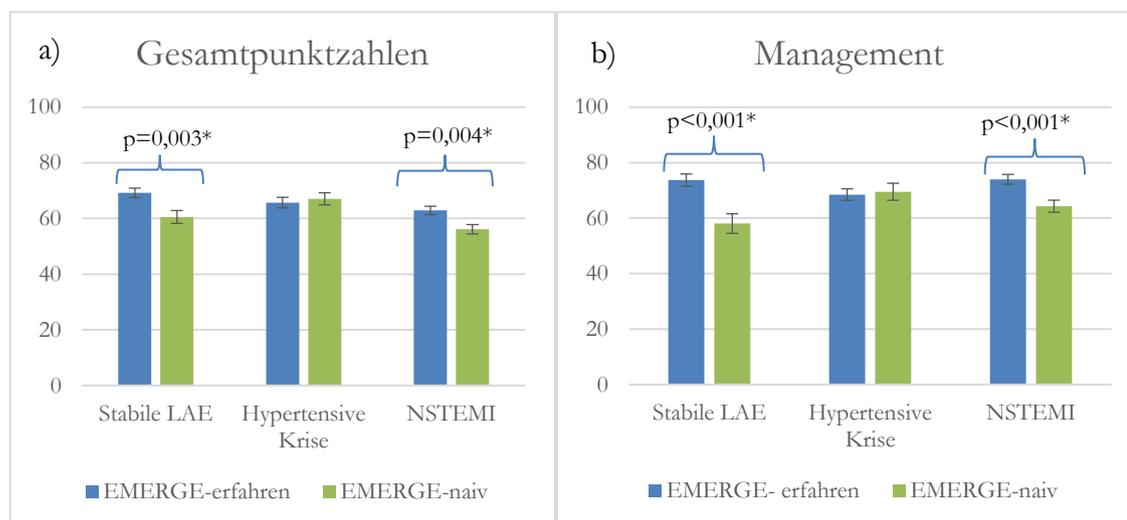


Abbildung 18a; b: Vergleich der erreichten Prozentpunktzahlen der EMERGE-erfahrenen (Intervention) und EMERGE-naiven (Kontrollgruppe) Studierenden in der *Logfile*-Auswertung, nach Fällen aufgeteilt

4 Diskussion

Wenn es um die Vermittlung von Faktenwissen geht, so erscheinen Lernprozess und Lehrmethoden recht geradlinig. Lehrbücher und wissenschaftliche Publikationen enthalten alle wichtigen Fakten, die man zum Bestehen der Prüfungen im Medizinstudium lernen muss. *Clinical reasoning* hingegen ist eine Fähigkeit, die am besten durch wiederholte Exposition gegenüber klinischen Fällen und Fallbeispielen trainiert werden kann (Eva 2005; Bowen 2006) und von zeitnahe Feedback begleitet werden sollte (Squire 2007; Kassirer 2010). Studien suggerieren, dass computerbasiertes (Simulations-)lernen eine effektive Möglichkeit dieses fallbasierten Lernens sein kann (Knight *et al.* 2010; Buijs-Spanjers *et al.* 2018; Chon *et al.* 2018 und 2019; Middeke *et al.* 2018; Mohan *et al.* 2018).

4.1 Wesentliche Ergebnisse

Simulationstraining ist eine weit verbreitete Lehrmethode in vielen Berufsausbildungen, unter anderem in der Medizin (Ziv *et al.* 2006). Im Rahmen dieser Arbeit wurde der kurz- und langfristige Lernerfolg durch ein *serious game* zu verschiedenen Zeitpunkten im Medizinstudium untersucht. Außerdem wurde die Nutzung der virtuellen Notaufnahme damit als neue Lehrform im Modul 3.1 etabliert.

Die vorliegende Studie zeigt erstmals einen positiven und nachhaltigen Effekt eines *serious games* als Lehrmethode für das diagnostische und therapeutische Management von Patient:innen mit kardiopulmonalen Krankheitsbildern. Die Nutzung des *serious games* war kurzfristig und langfristig mit besserem klinischem Denken assoziiert.

Bezugnehmend auf die unter 1.4 genannten Forschungsfragen kann von folgenden Beobachtungen berichtet werden:

4.1.1 Analyse 1

Die Studierenden des dritten und sechsten klinischen Semesters unterscheiden sich entgegen der Erwartungen weder zu Beginn noch am Ende des SoSe2017 voneinander. Im Laufe des Semesters stiegen die Gesamtpunktzahlen der *Logfiles* der aufgrund ihrer Ähnlichkeit zusammengefassten Semestern von $57,6 \pm 1,1$ % auf $65,5 \pm 1,2$ % ($p < 0,001^*$; $d = 0,656$). Diese Verbesserung der Leistung ist auf einen deutlichen Punktezuwachs im Bereich „Management“ zurückzuführen ($57,4 \pm 1,2$ % auf $69,8 \pm 1,3$ %; $p < 0,001^*$), während im Bereich „Anamnese“ keine Veränderung der Leistung zu erkennen war. Dies weist darauf hin, dass

mit dem Spiel EMERGE (differenzial-) diagnostisches Denken, Diagnosefindung und sinnvolles Triagieren geübt werden kann, EMERGE aber keinen nennenswerten Effekt auf die Qualität der Anamneseerhebung der Spielenden hat. Dies könnte an der eher statischen Anamnese-Funktion von EMERGE liegen, die wenig Individualität der Anamnesefragen und somit keine gute Trennschärfe bietet. Demnach eignen sich andere Lehrformate wie Unterricht am Krankenbett oder Praktika vermutlich besser, um Gesprächskompetenz und Anamneseerhebung zu üben. Eine neuseeländischen Pilotstudie konnte kürzlich den Hinweis geben, dass ein SG auch eine effektive und motivierende Lehrmethode für Anamnese-fähigkeiten sein kann (Alyami *et al.* 2019). Das SG dieser Studie ist allerdings ausschließlich ein Anamnese-Trainer, während der Fokus von EMERGE auf dem diagnostischen und therapeutischen Prozedere liegt.

Dass die Studierenden des dritten und sechsten klinischen Semesters sich kaum in ihrer Spielleistung unterscheiden ist verwunderlich. Das Modul „kardiopulmonale Lehre“ ist das erste klinische Modul im Medizinstudium. Über sechs Wochen lernen die Studierenden in Vorlesungen, Seminaren und im Kleingruppenunterricht wichtige Aspekte der Kardiologie. Dieses Wissen wird in den nächsten Semestern angewendet, beispielsweise im Rahmen von Famulaturen und Blockpraktika. Dem gegenübergestellt wurden Studierende des Repetitionsmoduls 6.x im letzten Semester ihres Studiums. Deshalb wäre es zu erwarten, dass diese die virtuellen Patient:innen besser diagnostizieren und behandeln würden als ihre Kommiliton:innen des Moduls 3.1. Im Gegensatz zu dieser Annahme konnte im Querschnittsvergleich weder zu Beginn noch am Ende des jeweiligen Moduls einen Leistungsunterschied der beiden Kohorten festgestellt werden (T2: Modul 3.1 55,9 % \pm 1,0, Modul 6.x 57,8 % \pm 1,8; T6: Modul 3.1 63,3 % \pm 1,3, Modul 6.x 65,5 % \pm 1,8). Die Studierenden beider Module schienen gleichsam von der Nutzung des *serious games* zu profitieren, unabhängig von ihrer Semesterzugehörigkeit.

Eine mögliche Erklärung für die ähnliche Leistung kann das unterschiedliche Motivationslevel der beiden Kohorten sein. Studien belegen, dass Motivation die akademische Leistung positiv beeinflusst (Artino 2007; Doménech-Betoret *et al.* 2017). Nach drei Jahren der präklinischen Vorbereitung beginnt im Modul 3.1 der Einblick in den klinischen Alltag mit UaK und praktischen Seminaren wie Auskultations- und Untersuchungskursen. Viele Studierende freuen sich auf das klinische Arbeiten und sind deshalb in diesem Modul besonders motiviert. Die Teilnehmenden des Modul 6.x hingegen sind am Ende ihres Studiums angelangt. Sie haben bereits 19 Module im klinischen Studienabschnitt abgeschlossen und absolvieren nun ihre letzten Pflichtveranstaltungen vor dem schriftlichen Examen und dem praktischen Jahr.

Es wäre möglich, dass sie mit weniger Begeisterung und Motivation an die EMERGE-Seminare herangehen als ihre Kommiliton:innen aus dem Modul 3.1.

Es ist bekannt, dass summative Prüfungen Lernverhalten und Retention wesentlich beeinflussen (Raupach *et al.* 2013 und 2016; Raupach und Schuelper 2018). Während die Studierenden des Modul 3.1 dieses Modul mit einer Prüfung in kardiopulmonaler Lehre abschließen, haben die Studierenden des Modul 6.x bereits alle Prüfungen der inneren Medizin abgelegt. Sie müssen nach dem Modul 6.x keine weitere summative Prüfung mit kardiologischen Fragen bestehen. Der Ausblick auf eine summative Prüfung kann eine Lernmotivation für die Studierenden des Modul 3.1 sein, die sie von den fortgeschritteneren Studierenden unterscheidet.

Eine weitere Erklärung für diese Beobachtung ist der unterschiedliche Aufbau beider Module. Während im Modul 3.1 sechs Wochen lang Kardiologie und Pulmologie unterrichtet werden, dient das Repetitionsmodul 6.x der Vorbereitung auf das Examen und praktische Jahr und bietet eine Wiederholung mehrerer klinischer Fächer. Dabei ist die erste Woche der Wiederholung der Kardiologie und Pulmologie gewidmet. Studierende des Repetitionsmoduls erwarten also auch in der virtuellen Notaufnahme Patient:innen aus allen Fachbereichen, während die Kommiliton:innen zu Beginn ihres klinischen Studiums nur die Kardiologie in dieser Lehreinheit erwarten. Dies kann den Schwierigkeitsgrad des differenzialdiagnostischen Arbeitens für die erfahreneren Studierenden erhöhen. So zeigte sich in der Auswertung der *Logfiles* und der Übergabeprotokolle, dass die erfahreneren Studierenden auch nicht kardiologische Differenzialdiagnosen in Erwägung zogen. Für den Fall „hypertensive Krise“ mit dem Leitsymptom „Kopfschmerz“ dachten die 6.x-Studierenden neben der hypertensiven Krise auch an Differenzialdiagnosen aus der Neurologie wie beispielsweise die Migräne. Im Fall „stabile Lungenarterienembolie“ präsentierten sich die Patient:innen mit stabilen Vitalparametern, ohne Vorerkrankungen und mit dem Leitsymptom Luftnot. Hier wurde von den Teilnehmenden des Moduls 6.x, die im Semester zuvor Psychiatrie gelernt hatten, differenzialdiagnostisch eine Panikattacke gedacht, während die 3.1-Studierenden für die Luftnot nur Erklärungen aus dem kardiologischen Formenkreis suchen konnten.

In dieser Studie profitierten Anfänger:innen genauso wie Erfahrene vom Spiel EMERGE. In einer Studie an der Universität Köln von Chon *et al.* übten Studierende unterschiedlicher Semester in einer virtuellen Notaufnahme. Vor und nach der Trainingsphase mit dem *serious game* nahmen sie an einem Test teil. In dieser Studie profitierten weniger fortgeschrittene Studierende (drittes klinisches Semester) signifikant mehr von der Intervention als fortgeschrittene Studierende (sechstes klinisches Semester) (Chon *et al.* 2019).

Diese Ergebnisse unterscheiden sich von einer Studie zum Notfall-Simulationsspiel abcdeSIM (Dankbaar *et al.* 2016). In dieser niederländischen Studie wurden Assistenzärzt:innen in zwei Gruppen geteilt. Die Kontrollgruppe lernte anhand von Texten, die Interventionsgruppe nahm zusätzlich am Spiel abcdeSIM teil. Am Ende war die Leistung der Interventionsgruppe signifikant besser als die der Kontrollgruppe. Die gleiche Studie wurde an Medizinstudierenden durchgeführt. Hier schnitt die Interventionsgruppe nicht besser ab als die Kontrollgruppe. In der Studie zeigte sich also ein „*expertise reversal effect*“. Dies bedeutet, dass Expert:innen von einer abwechslungsreichen Lernumgebung profitieren, während Anfänger:innen keinen Vorteil aus der Lernmethode ziehen können (Rey and Buchwald 2011). Dieser Effekt war in der vorliegenden EMERGE-Studie nicht festzustellen. Das könnte zum einen daran liegen, dass die Studierenden des Moduls 3.1 und 6.x sich bezüglich ihrer Expertise weniger unterscheiden als Studierende von Assistenzärzt:innen. Auch der Schwierigkeitsgrad der ausgesuchten Fallbeispiele könnte hier eine Rolle spielen. In der vorliegenden Studie wurden drei häufige kardiopulmonale Krankheitsbilder behandelt mit einer jeweils klassischen Präsentation. Es wäre interessant herauszufinden, ob sich bei komplizierteren EMERGE-Fällen ein *expertise reversal effect* wie in der Studie von Dankbaar *et al.* feststellen lassen würde.

4.1.2 Analyse 2

Die Ergebnisse der Longitudinalkohorte zu den Zeitpunkten T6 und T80 unterschieden sich nicht voneinander. Die *Logfile*-Auswertung ergab einen signifikanten Anstieg der Gesamtpunktzahlen von T2 zu T6 von $55,7 \pm 2,0$ % auf $65,4 \pm 2,3$ % ($p < 0,001^*$) und betrug zum *follow-up* nach eineinhalb Jahren unveränderte $66,0 \pm 1,5$ %. Die mit Hilfe von EMERGE im Modul 3.1 erworbenen *Clinical-reasoning*-Fähigkeiten waren eineinhalb Jahre nach der Intervention also unverändert vorhanden. Diese Beobachtung kann bedeuten, dass der mit EMERGE unterstützte Wissenszuwachs nachhaltig ist. Obwohl Medizinstudierende hervorragend kurzfristig für Tests lernen können (Larsen *et al.* 2009) fehlt die nachhaltige Retention des Wissens im Medizinstudium häufig (Sullivan *et al.* 2013). Vorangehende Studien zeigen, dass summative Prüfungen einen erheblichen Einfluss auf die Leistung der Studierenden haben (Newble and Jaeger 1983; Raupach *et al.* 2013). Aktuelle Erkenntnisse legen jedoch nahe, dass wiederholte formative (also unbenotete) Tests ebenfalls einen günstigeren Effekt auf das nachhaltige Lernen der Studierenden haben könnten (Raupach *et al.* 2016). Die Fallbearbeitung in der virtuellen Notaufnahme ist durch den Oberarztkommentar und die Bewertung der Fallbearbeitung in Form von Sternen in unterschiedlichen Kategorien eine formative Prüfung. Der dadurch erlebte *testing effect* (Raupach and Schuelper 2018; Sennhenn-

Kirchner *et al.* 2018) könnte zu einem nachhaltigen Lernen durch EMERGE beitragen. Es wäre interessant zu sehen, ob dieser langanhaltende Lerneffekt auch in EMERGE-unabhängigen Testungen sichtbar ist. Hier wäre beispielsweise eine erneute KF-Prüfung möglich.

4.1.3 Analyse 3

Diese EMERGE-erfahrenen Studierenden schnitten mit $65,6 \% \pm 1,2$ deutlich besser ab als ihre EMERGE-naiven Kommiliton:innen ($60,1 \% \pm 1,4$; $p=0,003^*$), auch hier lag der Effekt im Unterschied des Managements begründet ($72,6 \% \pm 1,2$ im Vergleich zu $63,5 \% \pm 2,1$, $p < 0,001^*$). Am Spieltermin T80 stellten EMERGE-erfahrene Studierende zweimal so häufig die richtige Diagnose bei Patient:innen mit Lungenembolie und dreimal so häufig bei Patient:innen mit NSTEMI wie EMERGE-naive Studierende.

Die Leistung der Studierenden der beiden Gruppen unterscheidet sich also erheblich. Der einzige strukturelle Unterschied zwischen beiden Gruppen ist, dass die einen bereits EMERGE-erfahren sind, während die anderen noch nie EMERGE gespielt haben. T80 ist der zweite Spieltermin im Modul 6.x im Wintersemester 2018/19. Beide Kohorten konnten sich am ersten Spieltermin des Semesters mit der Spielführung und den diagnostischen und therapeutischen Möglichkeiten des Spiels vertraut machen. Daher ist diesbezüglich kein Vorteil der EMERGE-erfahrenen Kohorte zu erwarten. EMERGE versucht, durch die zufällige Anzeige von Name, Alter und Geschlecht der virtuellen Patient:innen, eine Wiedererkennung der Fälle zu vermeiden. Trotzdem ist es möglich, dass Studierende durch häufiges Spielen gewisse Merkmale oder Antworten einem Krankheitsbild zuordnen können und somit einen Vorteil haben, der sich nicht nur auf einen Lernzuwachs im klinischen Denken zurückführen lässt.

Die EMERGE-erfahrenen Studierenden haben in Regelstudienzeit studiert. Es könnte die These aufgestellt werden, die Regelstudienzeit-Studierenden seien fleißiger, während die nicht in Regelstudienzeit Studierenden Urlaubssemester nehmen oder Prüfungen nicht bestehen würden. Doch die Regelstudienzeit könnte auch für Erasmus- oder andere Stipendien verlassen werden oder Studierende könnten für eine aufwändige Doktorarbeit ein Semester aussetzen. Dies sind spekulative Überlegungen und es ist nicht ersichtlich, dass die beiden Gruppen nicht dieselben Voraussetzungen mitbringen. In einer zukünftigen Studie wäre es möglich, die Prüfungsergebnisse der Studierenden miteinander zu vergleichen, um die Gleichheit beider Kohorten zu beweisen.

Die Unterschiede der Punktzahlen konnten in jeder Analyse auf Veränderungen der Management-Ergebnisse zurückgeführt werden. Dies weist erneut darauf hin, dass die Simulation

EMERGE das klinische Denken fördert, jedoch keinen Einfluss auf die Fähigkeit der Anamneseerhebung hat.

Für alle drei getesteten Fallbeispiele zeigt sich ein ähnliches Muster, weshalb ein verallgemeinerbarer Effekt möglich ist. Aufgrund des Prinzips der *case specificity* bleibt dieser verallgemeinerbare Effekt allerdings für andere Fälle zu prüfen. In einer vorangehenden Studie an der Universitätsmedizin Göttingen wurde dies untersucht. Studierende übten sechs Wochen lang mit dem *serious game* EMERGE klinisches Denken. Die Studierenden wurden in drei Gruppen aufgeteilt, die Patient:innen mit unterschiedlichen Krankheitsbildern in der virtuellen Notaufnahme sahen. Alle Studierenden bearbeiteten eine zufällige Fallauswahl. Zusätzlich sah Gruppe A Patient:innen mit einer Pankreatitis und einem NSTEMI. Gruppe B beschäftigte sich zusätzlich mit einem Asthmaanfall und einer gastrointestinalen Blutung. Gruppe AB sah alle vier Fälle zusätzlich in der virtuellen Notaufnahme. In der letzten EMERGE-Sitzung wurden die *Logfiles* der Studierenden der drei Gruppen ausgewertet und verglichen. Es gab keinen signifikanten Unterschied zwischen den Gruppen A, B und AB bezüglich ihrer Spielleistung. Somit konnte gezeigt werden, dass das klinische Denken, das anhand des *serious games* geübt wurde, nicht zwingend fallspezifisch ist, sondern möglicherweise auf andere klinische Fälle angewendet werden kann (Middeke *et al.* 2020).

4.2 Stärken und Schwächen der Studie

Eine Stärke der Studie ist die praxisnahe Simulation des Spiels EMERGE. Die Informationen zu den virtuellen Patient:innen werden mit der *Serial-cue*-Methode vorgestellt. Das bedeutet, die Informationen zum Fall sind den Studierenden nicht von Anfang an bekannt, sondern werden auf Nachfrage hin gegeben (Schmidt and Mamede 2015). Zu Beginn kennen die Spielenden nur das Leitsymptom. Dies bildet eine echte klinische Konsultation realitätsnah ab.

Eine bedeutende Stärke dieser Studie ist, dass sie nicht in einem *learning lab*, sondern im Rahmen des Curriculums einer medizinischen Fakultät stattfand. Dadurch konnte der Effekt des Spiels als Erweiterung des regulären Unterrichts beobachtet und somit realistischer eingeschätzt werden. Trotzdem konnte durch das Curriculum eine kontrollierte Studienumgebung geschaffen werden.

Eine weitere Stärke der Studie ist die Betrachtung von CR anhand mehrerer Methoden. Ein großer Teil der Forschung zu CR bezieht sich nur auf den Weg von der Anamnese und Untersuchung zur richtigen Diagnose (McGuire 1985). Eine richtige Diagnose impliziert aber nicht automatisch, ob suffizientes CR angewendet wurde (McGuire 1985; Eva 2005;

Norman 2005). In der vorliegenden Arbeit hingegen wurde versucht, die Mehrdimensionalität von klinischem Denken durch eine für jeden Fall individuelle Checkliste sowie eine KF-Prüfung und TA-Interviews zu erfassen.

Eine große Limitation dieser Studie ist, dass drei der vier verwendeten Methoden mit dem Spiel EMERGE in Verbindung stehen. Die Leistung der Studierenden wurde aus ihrer Aktivität während des Spiels abgeleitet. Diese Methode ist objektiv, reliabel und valide. Die Gültigkeit der Messkriterien sollte jedoch anhand einer vom Spiel unabhängigen Methode getestet werden. T6 war jedoch der einzige Spieltermin, an dem eine KF-Prüfung durchgeführt wurde. Folgende Studien müssen zeigen, ob sich der hier gezeigte Effekt auf die Praxis anwenden lässt. Dafür würde sich beispielsweise eine OSCE-Prüfung anbieten.

Eine weitere Limitation der Studie ist, dass das Modul 3.1 durch eine summative Prüfung im Fach Kardiopulmonale Lehre beendet wird, wohingegen die summative Abschlussklausur nach dem Modul 6.x ausschließlich Fragen zu den Fächern „Klinische Pharmakologie“ und „Geschichte, Theorie und Ethik der Medizin“ beinhaltet. Somit arbeiten die Studierenden im Modul 6.x nicht auf eine summative Klausur mit kardiologischen Inhalten hin. Es ist also möglich, dass die Studierenden des Modul 3.1 sich im Hinblick auf die Klausur im Laufe des Moduls stärker auf kardiopulmonale Inhalte vorbereiten als die Studierenden aus dem sechsten klinischen Semester. Es ist möglich, dass es in einem vergleichbareren Kontext einen Unterschied zwischen den beiden Modulen gegeben hätte.

Eine weitere Schwäche der vorliegenden Studie ist die recht kleine *follow-up*-Kohorte ($n = 36$), die als EMERGE-Erfahrene an Modul 6.x teilnahmen. Dies ergab sich durch Fehlzeiten an einem der drei Messtermine und durch Abweichungen mancher Studierenden von der Regelstudienzeit. Eventuell hätte die Stichprobe vergrößert werden können, hätte man Studierenden mitgeteilt, an welchen Spielterminen ihre Daten ausgewertet werden würden, doch dies hätte die Spielleistung gegebenenfalls verfälschen können.

Auch das Fehlen einer Kontrollgruppe im Modul 3.1 ist eine wichtige Limitation dieser Studie. Das Einführen einer Gruppe, die nicht am *serious game* teilnimmt, wäre im Curriculum nicht umsetzbar gewesen.

Bezüglich der TA-Interviews weist diese Studie mehrere Limitationen auf. Mit acht Studierenden pro Kohorte wurde eine recht kleine Stichprobe untersucht. Die TA-Interviews wurden außerdem nur von zwei Oberärzten hinsichtlich des CR beurteilt. Es besteht ein großer Unterschied zwischen den Punktzahlen, die nur einer der beiden Ärzte markiert hat und denen, die beide Ärzte markiert haben. Die Ergebnisse dieser Bewertung weisen demnach

keine gute *Inter-rater*-Reliabilität auf. Hier hätte eine größere Gruppe an Bewertenden eine aussagekräftigere Auswertung bieten können. Die erreichten Punktzahlen sind in beiden Kohorten und zu beiden Zeitpunkten niedrig. Mit $12,4 \pm 0,9$ von 22 CR-*items* erreichte das Modul 3.1 zu Termin sechs die höchste vergebene Punktzahl. Dies entspricht 56,3 % der CR-*items*. Da sich die Studierenden freiwillig melden durften, um am TA-Interview teilzunehmen, ist es außerdem möglich, dass hier ein Selektions-*bias* stattfand. Auch kannten die Studierenden diese Methode bisher noch nicht, und in manchen Interviews entstanden Pausen und Studierende mussten erinnert werden, ihre Gedanken laut auszusprechen.

Zwar konnte eine signifikante Steigerung im Bereich Management beobachtet werden, doch die Leistung der Studierenden war zu Termin 80 bestenfalls moderat ($72,6 \pm 1,2$ % in der Interventionsgruppe). Dies könnte damit zu tun haben, dass es sich bei den EMERGE-Seminaren zwar um Pflichtveranstaltungen, aber nicht um summative Prüfungen handelte.

4.3 Potential und Limitationen von *serious games* in der Medizindidaktik

Patientenmanagement wie individuelle Diagnostik und Therapie erfordern *clinical reasoning* und sind ein essenzieller Bestandteil des klinischen Alltags. Ärzt:innen haben besonders am Anfang ihrer Karriere oft Schwierigkeiten, ihr im traditionellen Medizinstudium erworbenes Faktenwissen auf Station richtig und sicher anzuwenden, da ihr *clinical reasoning* und ihre *illness scripts* noch sehr lückenhaft ausgebildet sind (Coderre *et al.* 2003). Durch häufiges Lösen von klinischen Fällen kann CR am besten trainiert werden (Eva 2005; Norman 2005; Kassirer 2010). Fallbasiertes Lernen im Medizinstudium kann allerdings Herausforderungen mit sich bringen. Unterricht am Krankenbett oder Untersuchungskurse mit Patient:innen müssen in kleinen Gruppen stattfinden und sind deshalb personal-, zeit-, und kostenaufwändig (Srinivasan *et al.* 2007; Kassirer 2010) und variabel in ihrer Qualität (Wimmers *et al.* 2006; Schmidt and Mamede 2015).

Eine Möglichkeit des fallbasierten Lernens ist die Simulation, beispielsweise in Form eines *serious games*. SG können eine kosteneffiziente, standardisierte Methode sein, junge Ärztinnen und Ärzte in *clinical reasoning* zu trainieren. Mit einem SG wie der virtuellen Notaufnahme EMERGE ist es möglich, große Gruppen von Studierenden gleichzeitig mit wenig personellem Aufwand viele klinische Fälle bearbeiten zu lassen. In der virtuellen Notaufnahme übernimmt jede:r einzelne Studierende die Rolle der behandelnden Ärzt:in. Somit haben alle die Möglichkeit, ihr Wissen anzuwenden. Dies treibt den Lernprozess voran (Wentzell *et al.*

2018). Die Studierenden behandeln außerdem selbstständig Patient:innen mit lebensbedrohlichen Erkrankungen, ohne dabei die Sicherheit von Menschen zu gefährden (Knight *et al.* 2010). Simulationsunterricht ist am effektivsten, wenn ein Raum geschaffen wird, in dem Fehler passieren dürfen und aus diesen gelernt werden kann (Bearman *et al.* 2019). Studien konnten zeigen, dass mit Hilfe von digitalen Lernmethoden sowohl diagnostische Fähigkeiten von Medizinstudierenden ausgebaut werden können (Wentzell *et al.* 2018) als auch Fähigkeiten von Ärzt:innen positiv beeinflussen (Cullinan *et al.* 2017; Maertens *et al.* 2017). Dies könnte unter anderem darauf zurückzuführen sein, dass Lernerfolg positiv mit dem Engagement der Lernenden korreliert ist (Maheu-Cadotte *et al.* 2018) und *serious games* als neues Lehrformat Studierende motiviert (Dankbaar *et al.* 2016; Dankbaar 2017; Chon *et al.* 2018).

Ein *serious game* hat Limitationen, besonders bezüglich des sozialen Aspektes der Medizin. Empathische Kommunikation, das Überbringen schlechter Nachrichten oder das Besprechen verschiedener Therapieoptionen sind Situationen, in denen Übung durch Interaktion mit echten Menschen unersetzlich ist. Auch bietet eine *gaming session*, in der 50 Studierende in einem Computerraum von ein oder zwei Lehrenden beaufsichtigt werden, eventuell weniger Möglichkeiten zum Austausch unter Studierenden oder individuellen Fragen an die Lehrenden, als ein Kleingruppenunterricht.

Eine Stärke von einem *serious game* wie EMERGE ist die Anzahl und die Möglichkeit der sorgfältigen Auswahl der Fälle. In der virtuellen Notaufnahme müssen mehrere, oft vier bis sechs, Patient:innen gleichzeitig behandelt werden. Während im UaK oder POL pro Studiengruppe und Stunde ein Fall besprochen wird, können im *serious game* in der gleichen Zeit deutlich mehr Krankheitsbilder wiederholt werden. Somit haben die Studierenden die Möglichkeit, *illness scripts* zu mehr klinischen Fällen aufzubauen. Dabei bleibt zu bedenken, dass nicht nur die Anzahl der präsentierten Fälle wichtig ist, um einen großen Lerneffekt zu erzielen. Studien, die den Lerneffekt von klinischen Rotationen erforschen, bestätigen einen deutlichen Wissenszuwachs der Studierenden während dieser Praktika. Die Studierenden weisen am Ende der Rotation jedoch sehr unterschiedlichen Wissenszuwachs auf. Es gibt dabei keinen Hinweis auf einen direkten Zusammenhang zwischen der Zahl der gesehenen Patient:innen und der klinischen Kompetenz (Wimmers *et al.* 2006).

4.4 Bedeutung der Studie für Forschung und Lehre

Die vorliegende Studie zeigt, dass dieses *serious game* als ein kurz- und langfristig effektives Lehrformat zum Erlernen des Managements von kardiologischen Krankheitsbildern angewendet werden kann und sowohl erfahrene als auch unerfahrene Medizinstudierende gleichsam davon profitieren können.

Die hohe Korrelation zwischen Übergabefällen und den *Logfile*-Ergebnissen und die Ähnlichkeit zu den Ergebnissen der KF-Prüfung als etablierter Prüfungsform für klinisches Denken eröffnet außerdem die Diskussion, ob ein *serious game* in Zukunft als Prüfungsformat angewendet werden könnte.

4.5 Ausblick: Offene Fragen und Vorschläge für weitere Studien

Der Einfluss des Spiels EMERGE auf das klinische Denken als komplexe Fähigkeit bestehend aus analytischen und intuitiven Komponenten bleibt ein spannendes Thema.

Es bleibt offen, ob sich die erhobenen Ergebnisse auf die postgraduale medizinische Lehre übertragen lassen. Außerdem wäre interessant, ob sich Studierende, die EMERGE gespielt haben, besser auf ihr praktisches Jahr vorbereitet fühlen und dann tatsächlich auch im praktischen Jahr besser klinisch denken.

Weiterführend zu dieser Studie muss erforscht werden, welche Merkmale eines *serious games* zum Lernerfolg der Studierenden beitragen. Zu dieser Frage wurde bisher beispielsweise der Einfluss der Realitätsnähe eines Simulationsspiels auf Motivation und Lernerfolg der Studierenden erforscht (Dankbaar *et al.* 2016). Es wäre interessant, weitere Studien zu möglichen Faktoren anzuschließen.

5 Zusammenfassung

Clinical reasoning ist eine ärztliche Kernkompetenz, die eine Grundvoraussetzung für präzise Diagnostik und Therapie von Patient:innen darstellt. Diese wird am besten durch wiederholtes Bearbeiten von Fallbeispielen geübt. Im klinischen Alltag kann es aufgrund von Mangel an Personal, Ressourcen und Patient:innen mit bestimmten Krankheitsbildern eine Herausforderung sein, allen Medizinstudierenden einen vergleichbaren und guten klinischen Unterricht zu gewährleisten, der Inzidenz und Prävalenz realistisch abbildet und außerdem Patient:innen keiner unnötigen Gefahr aussetzt. Seit ein paar Jahren werden Studien zum Einsatz von *serious games* in der Medizindidaktik durchgeführt; bisher wurden jedoch häufiger Spielführung oder der Einfluss der Spiele auf studentische Motivation untersucht als Evidenz für die Wirksamkeit dieser Lehrmethode geschaffen.

Diese Studie möchte daher die kurz- und langfristige Effektivität von der virtuellen Notaufnahme EMERGE als Lehrmethode des klinischen Denkens untersuchen und herausfinden, zu welchem Zeitpunkt im Studium die Studierenden am Meisten von ihr profitieren können.

In dieser Studie wurden vier Methoden angewendet, um CR zu messen; *Logfiles*, Übergabeprotokolle, eine *Key-feature*-Prüfung und *Think-aloud*-Interviews. Im Sommersemester 2017 wurden die Ergebnisse der Studierenden des dritten und sechsten klinischen Semesters miteinander verglichen. Außerdem wurde die Entwicklung des klinischen Denkens der Studierenden im Laufe des Semesters beurteilt. Im Wintersemester 2018/19 wurden die Ergebnisse der Studierenden, die bereits im Semester der ersten Datenerhebung an Modul 3.1 teilgenommen hatten und nun im Modul 6.x angekommen waren, ausgewertet, und mit ihren vorherigen Ergebnissen verglichen. Die Ergebnisse der EMERGE-erfahrenen 6.x-Studierenden wurde dann zusätzlich mit denen ihrer EMERGE-naiven Kommiliton:innen im selben Semester verglichen.

Die Studierenden des dritten und sechsten klinischen Semesters unterscheiden sich weder zu Beginn noch am Ende des SoSe 2017 voneinander. Im Laufe des Semesters stiegen die Gesamtpunktzahlen der *Logfiles* der aufgrund ihrer Ähnlichkeit zusammengefassten Module von $57,6 \% \pm 1,1$ auf $65,5 \% \pm 1,2$ ($p < 0,001^*$; $d = 0,656$). Diese Verbesserung der Leistung ist auf einen deutlichen Punktezuwachs im Bereich „Management“ zurückzuführen ($57,4 \% \pm 1,2$ auf $69,8 \% \pm 1,3$; $p < 0,001^*$), während im Bereich „Anamnese“ keine Veränderung stattfand. Die EMERGE-erfahrenen Studierenden im Modul 6.x im WiSe 2018/19 zeigten, dass ihr Lernerfolg aus dem Modul 3.1 eineinhalb Jahre zuvor anhielt. Ihre Gesamtpunktzahlen in der *Logfile*-Auswertung waren während ihres Moduls 3.1 von $55,7 \% \pm 2,0$ signifikant auf

65,4 % \pm 2,3 angestiegen ($p < 0,001^*$) und betragen zum *follow-up* nach eineinhalb Jahren unveränderte 65,6 % \pm 1,2. Diese EMERGE-erfahrenen Studierenden schnitten somit deutlich besser ab als ihre EMERGE-naiven Kommiliton:innen (60,1 % \pm 1,4; $p = 0,003^*$), auch hier lag der Effekt im Unterschied des Managements begründet (72,6 % \pm 1,2 im Vergleich zu 63,5 % \pm 2,1, $p < 0,001^*$).

In dieser prospektiven Studie mit einem *follow-up* von eineinhalb Jahren ist die Nutzung der virtuellen Notaufnahme EMERGE mit einer anhaltenden Steigerung der Befähigung zum klinischen Denken bei Medizinstudierenden sowohl zu Beginn als auch am Ende ihres klinischen Studiums assoziiert.

6 Anhang

Tabelle A1: Analyse 3: Prozentanteil der *Logfiles*, in denen das jeweilige *clinical-reasoning-item* genannt wurde.

Stabile Lungenembolie	EMERGE-naiv [%] (n = 42)	EMERGE-erfahren [%] (n = 58)	p - Wert
Anamnese			
Husten	26,2 (11)	32,8 (19)	0,479
Schmerzen	52,4 (22)	62,1 (36)	0,333
Anamnese des Leitsymptoms	78,6 (33)	70,7 (41)	0,375
Rauchen	54,8 (23)	50,0 (29)	0,638
Vorerkrankungen	95,2 (40)	84,5 (49)	0,090
Medikamentenanamnese	76,2 (32)	79,3 (46)	0,710
Anamnese: max. 6 Punkte	3,8 ± 0,2	3,8 ± 0,2	0,878
Management			
BGA	85,7 (36)	87,9 (51)	0,745
Troponin	71,4 (30)	91,4 (53)	0,009*
D-dimere	57,1 (24)	79,3 (46)	0,017*
EKG	83,3 (35)	98,3 (57)	0,007*
Echokardiographie	19,0 (8)	29,3 (17)	0,242
CT-Thorax	52,4 (22)	69,0 (40)	0,092
Richtige Diagnose	59,5 (25)	79,3 (46)	0,031*
Behandlung: Antikoagulation	35,7 (15)	55,2 (32)	0,054*
Management: max. 8 Punkte	4,6 ± 0,3	5,9 ± 0,2	<0,001*
Gesamtpunktzahl: max. 14 Punkte	8,5 ± 0,3	9,7 ± 0,2	0,003*

Hypertensive Krise	EMERGE-naiv [%] (n = 41)	EMERGE-erfahren [%] (n = 56)	p - Wert
Anamnese			
Schwindel	17,1 (7)	19,6 (11)	0,748
Anamnese des Leitsymptoms	65,9 (27)	66,1 (37)	0,982
Vorerkrankungen	80,5 (33)	71,4 (40)	0,307
Medikamentenanamnese	90,2 (37)	89,3 (50)	0,878
Anamnese: max. 4 Punkte	2,5 ± 0.1	2,4 ± 0.1	0,641
Management			
Körperliche Untersuchung	73,2 (30)	73,2 (41)	0,996
Kleines Blutbild	70,7 (29)	66,1 (37)	0,627
EKG	48,8 (20)	57,1 (32)	0,415
Richtige Diagnose	97,6 (40)	94,6 (53)	0,475
Behandlung: Nitrate	56,1 (23)	60,7 (34)	0,648
Behandlung: Antihypertensiva	70,7 (29)	58,9 (33)	0,232
Management: max. 6 Punkte	4,1 ± 0.2	4,0 ± 0,2	0,678
Gesamtpunktzahl: max. 10 Punkte	6,6 ± 0.3	6,3 ± 0,2	0,578

NSTEMI	EMERGE-naiv [%] (n = 42)	EMERGE- erfahren [%] (n = 58)	p - Wert
Anamnese			
Atemnot	42,9 (18)	43,1 (25)	0,980
Beginn des Leitsymptoms	61,9 (26)	55,2 (32)	0,501
Anamnese des Leitsymptoms	16,7 (7)	32,8 (19)	0,070
Rauchen	28,6 (12)	36,2 (21)	0,423
Allergien	26,2 (11)	36,2 (21)	0,289
Vorerkrankungen	71,4 (30)	69,0 (40)	0,791
Familienanamnese	40,5 (17)	29,3 (17)	0,245
Medikamentenanamnese	71,4 (30)	79,3 (46)	0,362
Anamnese: max. 8 Punkte	3,6 ± 0,2	3,8 ± 0,2	0,479
Management			
Körperliche Untersuchung des Thorax	88,1 (37)	94,8 (55)	0,221
Troponin	100,0 (42)	100,0 (58)	—
CK	92,9 (39)	96,6 (56)	0,403
CKMB	92,9 (39)	98,3 (57)	0,172
TSH	2,4 (1)	27,6 (16)	0,001*
EKG	97,6 (41)	98,3 (57)	0,817
Koronarangiographie	78,6 (33)	89,7 (52)	0,126
Richtige Diagnose	69,0 (29)	87,9 (51)	0,020*
Behandlung: ASS	26,2 (11)	34,5 (20)	0,376
Behandlung: Heparin	21,4 (9)	36,2 (21)	0,111
Behandlung: Nitrate	38,1 (16)	50,0 (29)	0,238
Management: max. 11 Punkte	7,0 ± 0,2	8,1 ± 0,2	0,001*
Total score: max. 19 Punkte	10,7 ± 0,3	12,0 ± 0,3	0,004*

7 Literaturverzeichnis

- Abdulmajed H, Park YS, Tekian A (2015): Assessment of educational games for health professions: A systematic review of trends and outcomes. *Med Teach* 37, 27–32
- Ajjawi R, Higgs J (2008): Learning to reason: A journey of professional socialisation. *Adv Health Sci Educ* 13, 133–150
- Akl E, Pretorius R, Sackett K, Erdley S, Bhoopathi P, Alfarah Z, Schünemann H (2010): The effect of educational games on medical students' learning outcomes: A systematic review: BEME Guide No 14. *Med Teach* 32, 16–27
- Alyami H, Alawami M, Lyndon M, Alyami M, Coomarasamy C, Henning M, Hill A, Sundram F (2019): Impact of using a 3D visual metaphor serious game to teach history-taking content to medical students: Longitudinal mixed methods pilot study. *JMIR Serious Games* 7, e13748
- Amey L, Donald KJ, Teodorczuk A (2017): Teaching clinical reasoning to medical students. *Br J Hosp Med* 78, 399–401
- Ark TK, Brooks LR, Eva KW (2006): Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med* 81, 405–409
- Artino AR (2007): Motivational beliefs and perceptions of instructional quality: predicting satisfaction with online training: Predicting satisfaction with online training. *J Comput Assist Learn* 24, 260–270
- Barrows HS, Tamblyn RM: *Problem-based learning: An approach to medical education*. Springer Publishing Company, New York 1980
- Bearman M, Greenhill J, Nestel D (2019): The power of simulation: a large-scale narrative analysis of learners' experiences. *Med Educ* 53, 369–379
- Bordage G, Page G (2018): The key-features approach to assess clinical decisions: validity evidence to date. *Adv Health Sci Educ* 23, 1005–1036
- Bowen JL (2006): Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 355, 2217–2225
- Brieger D, Eagle K, Goodman S et al. (2004): Acute coronary syndromes without chest pain, an underdiagnosed and undertreated high-risk group: Insights from the Global Registry of Acute Coronary Events. *Chest* 126, 461–469

- Buijs-Spanjers KR, Hegge HH, Jansen CJ, Hoogendoorn E, de Rooij SE (2018): A web-based serious game on delirium as an educational intervention for medical students: Randomized controlled trial. *JMIR Serious Games* 6, 17
- Burbach B, Barnason S, Thompson SA (2015): Using “think aloud” to capture clinical reasoning during patient simulation. *Int J Nurs Educ Scholarsh* 12, 1–7
- Butterfield PS, Libertin AG (1993): Learning outcomes of an ambulatory care rotation in internal medicine for junior medical students. *J Gen Intern Med* 8, 189–192
- Chon SH, Hilgers S, Timmermann F, Dratsch T, Plum PS, Berlth F, Datta R, Alakus H, Schlößer HA, Schramm C, et al. (2018): Web-based immersive patient simulator as a curricular tool for objective structured clinical examination preparation in surgery: Development and evaluation. *JMIR Serious Games* 6, e10693
- Chon SH, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, Datta RR, Schramm C, Haneder S, Späth MR et al. (2019): Serious games in surgical medical education: A virtual emergency department as a tool for teaching clinical reasoning to medical students. *JMIR Serious Games* 7, e13028
- Coderre S, Mandin H, Harasym PH, Fick GH (2003): Diagnostic reasoning strategies and diagnostic success. *Med Educ* 37, 695–703
- Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM (2012): A systematic literature review of empirical evidence on computer games and serious games. *Comput Educ* 59, 661–686
- Connor DM, Durning SJ, Rencic JJ (2019): Clinical reasoning as a core competency: *Acad Med* 95, 1166–1171
- Croskerry P (2009): A universal model of diagnostic reasoning: *Acad Med* 84, 1022–1028
- Cullinan S, O’Mahony D, Byrne S (2017): Use of an e-learning educational module to better equip doctors to prescribe for older patients: A randomized controlled trial. *Drugs Aging* 34, 367–374
- Cunningham CE, Deal K, Neville A, Rimas H, Lohfeld L (2006): Modeling the problem-based learning preferences of McMaster University undergraduate medical students using a discrete choice conjoint experiment. *Adv Health Sci Educ* 11, 245–266
- Dankbaar M (2017): Serious games and blended learning; effects on performance and motivation in medical education. *Perspect Med Educ* 6, 58–60

- Dankbaar MEW, Alsmas J, Jansen EEH, van Merriënboer JJG, van Saase JLJM, Schuit SCE (2016): An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ* 21, 505–521
- Dankbaar MEW, Roozeboom MB, Oprins EAPB, Rutten F, van Merriënboer JJG, van Saase JLJM, Schuit SCE (2016): Preparing residents effectively in emergency skills training with a serious game: *Simul Healthc J Soc Simul Healthc* 12, 9–16
- Davison GC, Vogel RS, Coffman SG (1997): Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *J Consult Clin Psychol* 65, 950–958
- Doménech-Betoret F, Abellán-Roselló L, Gómez-Artiga A (2017): Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Front Psychol* 8, 1193
- Dumas D, Torre DM, Durning SJ (2018): Using relational reasoning strategies to help improve clinical reasoning practice: *Acad Med* 93, 709–714
- Duncker K (1926): A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). *Pedagog Semin J Genet Psychol* 33, 642–708
- Durning SJ (2013): Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Acad Med* 88, 442–448
- Durning SJ, Artino A, Boulet J, La Rochelle J, Van Der Vleuten C, Arze B, Schuwirth L (2012): The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach* 34, 30–37
- Elstein AS, Shulman LS, Sprafka SA: *Medical problem solving: an analysis of clinical reasoning*. Harvard University Press, Cambridge, Mass 1978
- Ericsson KA (2004): Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains: *Acad Med* 79, 70–81
- Ericsson KA, Simon HA: *Protocol analysis: Verbal reports as data*. The MIT Press, Cambridge 1984
- Eva KW (2002): The aging physician: Changes in cognitive processing and their impact on medical practice. *Acad Med* 77, S1–S6
- Eva KW (2005): What every teacher needs to know about clinical reasoning. *Med Educ* 39, 98–106

- Eva KW, Hatala RM, LeBlanc VR, Brooks LR (2007): Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information: clinical expertise. *Med Educ* 41, 1152–1158
- Farmer EA, Page G (2005): A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 39, 1188–1194
- Feltovich P, Barrows H: Issues of generality in medical problem solving. In: Schmidt HG (Hrsg): *Tutorials in Problem-Based Learning: A New Direction in Teaching the Health Professions*. Van Gorcum 1984, 128–42
- Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, Dauwels-Okutsu S, Nikolaou CK, Zary N, Campbell J, Car J (2019): Serious gaming and gamification education in health professions: Systematic review. *J Med Internet Res* 21, e12994
- Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, Yepes FJ, Muñoz Ó (2018): A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online* 23, 1438718
- Higgs J, Jones M, Loftus S, Christensen N: *Clinical reasoning in the health professions*. 3. Auflage; Elsevier Butterworth Heinemann, Edinburgh 2008
- Hobu PPM, Schmidt HG, Boshuizen HPA, Patel VL (1987): Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. *Med Educ* 21, 471–476
- Humanmedizin in Göttingen. <https://www.umg.eu/studium-lehre/studieninteressierte/studienangebot/humanmedizin-in-goettingen/>; Zugriff am 24.08.2020
- Kahnemann D: *Thinking fast and slow*. Farrar, Straus and Giroux, New York 2011
- Kassirer JP (2010): Teaching clinical reasoning: Case-based and coached: *Acad Med* 85, 1118–1124
- Kleinert R, Wahba R, Chang D-H, Plum P, Hölscher AH, Stippel DL (2015): 3D Immersive patient simulators and their impact on learning success: A thematic review. *J Med Internet Res* 17, e91
- Knight JF, Carley S, Tregunna B, Jarvis S, Smithies R, de Freitas S, Dunwell I, Mackway-Jones K (2010): Serious gaming technology in major incident triage training: A pragmatic controlled trial. *Resuscitation* 81, 1175–1179
- Koens F, Mann KV, Custers EJFM, Ten Cate OTJ (2005): Analysing the concept of context in medical education. *Med Educ* 39, 1243–1249

- Larsen DP, Butler AC, Roediger III HL (2009): Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 43, 1174–1181
- Leeuwen YDV, Pollemans MC, Drop MJ (1997): Factors influencing the growth in knowledge of trainees in general practice. *Adv Health Sci Educ Theory Pr* 2, 61–70
- M6.1 - Klug entscheiden in der Medizin. <https://www.umg.eu/studium-lehre/studiengaenge/humanmedizin/klinik-humanmedizin/module/m61-klug-entscheiden-in-der-medizin/>; Zugriff am 24.08.2020
- Maertens H, Aggarwal R, Moreels N, Vermassen F, Van Herzele I (2017): A proficiency based stepwise endovascular curricular training (PROSPECT) program enhances operative performance in real life: A randomized controlled trial. *Eur J Vasc Endovasc Surg* 54, 387–396
- Maheu-Cadotte M-A, Cossette S, Dubé V, Fontaine G, Mailhot T, Lavoie P, Cournoyer A, Balli F, Mathieu-Dupuis G (2018): Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: A systematic review and meta-analysis protocol. *BMJ Open* 8, e019871
- Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM (2007): Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning: Clinical expertise. *Med Educ* 41, 1185–1192
- Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C, Schmidt HG (2010): Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 304, 1198
- Mamede S, van Gog T, van den Berge K, van Saase JLCM, Schmidt HG (2014): Why do doctors make mistakes? A study of the role of salient distracting clinical features: *Acad Med* 89, 114–120
- McGuire C (1985): Medical problem-solving: a critique of the literature. *J Med Educ* 60, 587–595
- McLeod PJ, Snell L (1991): Casemix in an internal medicine clerkship: Educational value of the clinical problems seen. *J Gen Intern Med* 6, 455–459
- Middeke A, Anders S, Schuelper M, Raupach T, Schuelper N (2018): Training of clinical reasoning with a serious game versus small-group problem-based learning: A prospective study. *PLOS ONE* 13, e0203851

- Middeke A, Anders S, Raupach T, Schuelper N (2020): Transfer of clinical reasoning trained with a serious game to comparable clinical problems: A prospective randomized study. *Simul Healthc J Soc Simul Healthc* 15, 75–81
- Mohan D, Fischhoff B, Angus DC, Rosengart MR, Wallace DJ, Yealy DM, Farris C, Chang C-CH, Kerti S, Barnato AE (2018): Serious games may improve physician heuristics in trauma triage. *Proc Natl Acad Sci* 115, 9204–9209
- Neufeld V, Woodward C, MacLeod S (1989): The McMaster M.D. program: a case study of renewal in medical education. *Acad Med* 64, 423–32
- Newble DI, Jaeger K (1983): The effect of assessments and examinations on the learning of medical students. *Med Educ* 17, 165–171
- Newell A, Simon HA: *Human problem solving*. Prentice-Hall, Englewood Cliffs 1972
- Norman G (2005): Research in clinical reasoning: past history and current trends. *Med Educ* 39, 418–427
- 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
- Pinnock R, Young L, Spence F, Henning M, Hazell W (2015): Can think aloud be used to teach and assess clinical reasoning in graduate medical education? *J Grad Med Educ* 7, 334–337
- Rattner SL (2001): Documenting and comparing medical students' clinical experiences. *JAMA* 286, 1035
- Raupach T, Schuelper N (2018): Reconsidering the role of assessments in undergraduate medical education. *Med Educ* 52, 464–466
- Raupach T, Brown J, Anders S, Hasenfuss G, Harendza S (2013): Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *BMC Med* 11, 61
- Raupach T, Andresen JC, Meyer K, Strobel L, Koziolok M, Jung W, Brown J, Anders S (2016): Test-enhanced learning of clinical reasoning: a crossover randomised trial. *Med Educ* 50, 711–720
- Rey GD, Buchwald F (2011): The expertise reversal effect: Cognitive load and motivational explanations. *J Exp Psychol Appl* 17, 33–48

- Rutledge C, Walsh CM, Swinger N, Auerbach M, Castro D, Dewan M, Khattab M, Rake A, Harwayne-Gidansky I, Raymond TT, et al. (2018): Gamification in action: theoretical and practical considerations for medical educators. *Acad Med* 93, 1014–1020
- Schmidt H, Norman G, Boshuizen H (1990): A cognitive perspective on medical expertise: Theory implications. *Acad Med* 67, 287
- Schmidt HG, Mamede S (2015): How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Med Educ* 49, 961–973
- Schwartz RW, Donnelly MB, Sloan DA, Young B (1994): Knowledge gain in a problem-based surgery clerkship: *Acad Med* 69, 148–51
- Sennhenn-Kirchner S, Goerlich Y, Kirchner B, Notbohm M, Schiekirka S, Simmenroth A, Raupach T (2018): The effect of repeated testing vs repeated practice on skills learning in undergraduate dental education. *Eur J Dent Educ* 22, e42–e47
- Sibbald M, Cavalcanti RB (2011): The biasing effect of clinical history on physical examination diagnostic accuracy: Effect of clinical history on diagnostic accuracy. *Med Educ* 45, 827–834
- Squire KD (2007): Games, learning, and society: Building a field. *Educ Technol* 47, 51–55
- Srinivasan M, Wilkes M, Stevenson F, Nguyen T, Slavin S (2007): Comparing problem-based learning with case-based learning: Effects of a major curricular shift at two institutions: *Acad Med* 82, 74–82
- Sullivan PB, Gregg N, Adams E, Rodgers C, Hull J (2013): How much of the paediatric core curriculum do medical students remember? *Adv Health Sci Educ* 18, 365–373
- ten Cate O, Durning SJ: Understanding clinical reasoning from multiple perspectives: A conceptual and theoretical overview. In: ten Cate O, Custers EJFM, Durning SJ (Hrsg.): *Principles and Practice of case-based clinical reasoning education*. Springer International Publishing, Cham 2018, 35–46
- Tsoy D, Sneath P, Rempel J, Huang S, Bodnariuc N, Mercuri M, Pardhan A, Chan TM (2019): Creating GridlockED: A serious game for teaching about multipatient environments. *Acad Med* 94, 66–70
- van de Wiel MWJ, Schmidt HG, Boshuizen HPA (1998): A failure to reproduce the intermediate effect. *Acad Med* 73, 894–900

- Wang R, DeMaria S, Goldberg A, Katz D (2016): A systematic review of serious games in training health care professionals: *Simul Healthc J Soc Simul Healthc* 11, 41–51
- Wentzell S, Moran L, Dobranowski J, Levinson A, Hannigan A, Dunne CP, McGrath D (2018): E-learning for chest x-ray interpretation improves medical student skills and confidence levels. *BMC Med Educ* 18, 256
- Wimmers PF, Schmidt HG, Splinter TAW (2006): Influence of clerkship experiences on clinical competence. *Med Educ* 40, 450–458
- Ziv A, Wolpe PR, Small SD, Glick S (2006): Simulation-based medical education: An ethical imperative: *Simul Healthc J Soc Simul Healthc* 1, 252–256

Danksagung

Ich möchte mich herzlich bei allen Mitarbeitenden und Studierenden der Universitätsmedizin Göttingen bedanken, die diese Studie und meine Dissertation möglich gemacht haben.

Ein außerordentlicher Dank gebührt Herrn Prof. Dr. Tobias Raupach für die Bereitstellung des Themas, sowie seine sehr intensive und engagierte Betreuung während der Datenerhebung und des Erstellungsprozesses dieser Dissertation. Vielen Dank für Deine immer prompten und geduldigen Antworten auf meine Fragen und Deine hilfreichen Anregungen.

Ganz besonders danke ich Dr. Nikolai Schülper für wertvolle Hinweise während der Datenerhebung und -auswertung und nicht zuletzt für das Korrekturlesen der Dissertation.

Außerdem möchte ich mich bei den Mitarbeitern der Lehr-IT, besonders Christian Müncher und Fred Viezens, für die technische Hilfe bei den EMERGE-Seminaren sowie bei der Erfassung der *Logfiles* bedanken.

Lebenslauf

Mein Name ist Insa Amelie de Temple, geb. Frisch. Ich wurde am 03.04.1992 in Hattingen als Tochter von Andreas und Susanna Frisch geboren.

An der Auguste-Viktoria-Schule in Flensburg erreichte ich im Juni 2011 die allgemeine Hochschulreife. Das elfte Schuljahr verbrachte ich zuvor an der Ponderosa High School in Parker, Colorado, USA. Von November 2011 bis März 2012 absolvierte ich einen Bundesfreiwilligendienst im Universitätskrankenhaus Hamburg Eppendorf auf der neuropädiatrischen Station. Im April 2012 begann ich an der Georg-August-Universität in Göttingen mit dem Medizinstudium. Im Sommersemester 2016 studierte ich im Rahmen des ERASMUS-Programmes an der Università di Bologna in Italien. Im April 2017 begann ich mit der Arbeit an dieser Dissertation und stellte im August 2018 erste Ergebnisse bei der Association for Medical Education in Europe (AMEE) Konferenz in Basel vor. In meinem Praktischen Jahr arbeitete ich in der Inneren Medizin am Evangelischen Krankenhaus Weende in Göttingen, in der Chirurgie am Königin-Elisabeth-Herzberge Krankenhaus in Berlin, sowie in der Pädiatrie am Royal Devon and Exeter University Hospital in England und im Waldkrankenhaus in Berlin. Mein Studium schloss ich im November 2019 ab und erlangte im Dezember 2019 die Approbation als Ärztin.

Von Mai 2020 bis April 2021 arbeitete ich als Assistenzärztin in der Viszeralchirurgie am Stadtspital Triemli in Zürich. Seit Mai 2021 bin ich als Assistenzärztin in der Kinderchirurgie am Universitäts-Kinderspital in Zürich tätig.