

Aus dem
Institut für Didaktik und Ausbildungsforschung in der Medizin
Klinikum der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. Martin Fischer



Einfluss von adaptivem Feedback und kollaborativer Bearbeitung beim Lernen mit virtuellen Patientinnen und Patienten

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Sven-Markus Christian Lorenz Sarbu-Rothsching

aus
München

Jahr
2023

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erster Gutachter: Prof. Dr. Martin R. Fischer

Zweiter Gutachter: Prof. Dr. Dr. Christian Sommerhoff

Dritter Gutachter: Prof. Dr. Philipp Korber

Promovierter Mitbetreuer: PD Dr. phil. Jan Kiesewetter, Dr. phil. Jan Zottmann

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 13.11.2023

Danksagung

An dieser Stelle möchte ich allen danken, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Mein besonderer Dank gilt Dr. phil. Jan Zottmann, für die ausgezeichnete Betreuung und die enorme Unterstützung bei der Umsetzung der gesamten Arbeit.

Außerdem möchte ich mich bei meinem Doktorvater Prof. Dr. med. Martin Fischer für die stete Hilfsbereitschaft bedanken.

Mein Dank gilt auch PD Dr. phil. Jan Kiesewetter, Dr. med. Raphael Kunisch und Prof. Dr. med. Christian Sommerhoff, die mir im Rahmen des Projektes und der Fertigstellung meiner Arbeit zur Seite standen.

Nicht zuletzt gilt mein besonderer Dank Martin Adler, der mir bei der Umsetzung meiner Ideen und bei der Datenerhebung stets eine große Hilfe war.

Meinen Eltern und Freunden danke ich für ihre Geduld und ihre Ermutigungen während meines Studiums und während der Arbeit an dieser Dissertation.

Inhaltsverzeichnis

1. Problemstellung	6
1.1 Simulationsbasiertes Lernen im Medizinstudium.....	8
1.2 Feedback als Möglichkeit der Unterstützung simulationsbasierten Lernens	9
1.3 Kollaboratives Lernen mit Fallsimulationen.....	10
1.4 Zur Rolle des Vorwissens beim Lernen mit Fallsimulationen	11
1.5 Fragestellungen und Hypothesen.....	12
2. Methoden	14
2.1 Ethikvotum.....	14
2.2 Akquise.....	14
2.3 Beschreibung der Stichprobe	14
2.4 Studiendesign.....	15
2.5 Fallmaterial	18
2.6 Instrumente.....	19
2.6.1 Prätest	19
2.6.2 Lernphase	20
2.6.3 Testfall	20
2.6.4 Posttest.....	22
2.7 Vorbereitungen.....	24
2.7.1 Umsetzung als Feldstudie	24
2.7.2 Anpassungen an der Online-Oberfläche	24
2.8 Technischer Ablauf.....	27
2.9 Statistische Analyse	28
3. Ergebnisse	29
3.1 Untersuchung der Effekte von Feedbackart und Bearbeitungsform	29
3.1.1 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf das <i>konzeptuelle</i> und <i>strategische Wissen</i>	29
3.1.2 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die Lösung des Testfalls (Herr Schneider)	30
3.1.3 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die <i>Motivation</i> der Student:innen	31
3.1.4 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die <i>kognitive Belastung</i>	32
3.2 Untersuchung der Effekte von Feedbackart und Vorwissen	33
3.2.1 Voruntersuchung.....	33
3.2.2 Untersuchung des Einflusses der Feedbackart und des Vorwissens auf das <i>konzeptuelle und strategische Wissen</i>	33
3.2.3 Untersuchung des Einflusses der Feedbackart und des Vorwissens auf die Lösung des Testfalls (Herr Schneider).....	34
3.2.4 Untersuchung des Einflusses der Feedbackart auf die <i>Motivation</i> der Studierenden.....	36
3.2.5 Untersuchung des Einflusses des Vorwissens auf die <i>kognitive Belastung</i>	36
3.3 Deskriptive Auswertung der Interviewfragen aus dem Posttest	37
4. Diskussion der Ergebnisse	39

4.1 Zusammenfassung der Ergebnisse	39
4.2.1 Einfluss der Feedbackart bei Lernenden mit hohem Vorwissen.....	40
4.2.2 Einfluss der Bearbeitungsform.....	41
4.3 Ergebnisse in Bezug auf die Feedbackart und das Vorwissen.....	42
4.3.1 Einfluss der Feedbackart bei Lernenden mit unterschiedlichem Vorwissen.....	42
4.3.2 Einfluss des Vorwissens	43
4.4 Limitationen	44
4.5 Implikationen für Theorie und Praxis	45
5. Zusammenfassung.....	46
5.1 Zusammenfassung	46
5.2 Summary	48
6. Literatur.....	50
7. Anlagen.....	55
7.1 Anschreiben am Beispiel Standort Universität Würzburg	55
7.2 Begrüßungsmail.....	56
7.3 Leitfaden zum kollaborativen Arbeiten	56
8. Eidesstaatliche Versicherung.....	61
9. Lebenslauf.....	62

1. Problemstellung

„59-jährige Patientin mit diffusen Oberbauchschmerzen und Übelkeit“ – so könnte die Übergabe eines Rettungssanitäters an den diensthabenden Internisten der Notaufnahme lauten. Die Entscheidung über das weitere diagnostische Vorgehen muss nun innerhalb kürzester Zeit getroffen werden. Werden die Symptome der Patientin durch eine entzündete Gallenblase oder Bauchspeicheldrüse oder doch durch einen maskierten Herzinfarkt verursacht? Solche Differenzialdiagnosen gilt es abzuklären. Doch mit welcher Diagnostik sollte begonnen werden? Zuerst einmal körperlich untersuchen oder doch zuerst Blut abnehmen? Sofort eine Ultraschalluntersuchung veranlassen und ein EKG schreiben, um nichts zu übersehen? Der Weg zur richtigen Diagnose ist essenziell für das weitere Vorgehen und schlussendlich auch für das Wohl des Patienten. Daher gehört die *Diagnosekompetenz* zu den wichtigsten Kompetenzen praktizierender Mediziner:innen [1]. Klinische Fähigkeiten wie das Diagnostizieren können nicht durch bloßes Beobachten generiert werden [2]. Um zum Beispiel die Diagnose-richtigkeit zu verbessern und Fehler beim Diagnostizieren zu verringern, muss diese Kompetenz aktiv gelehrt werden [3], [4]. Diese Erkenntnis ist auch vermehrt explizit in das medizinische Curriculum integriert [5]. Auch der aktuelle nationale kompetenzbasierte Lernzielkatalog Medizin (NKLM), welcher diejenigen Kompetenzen, welche nach Abschluss des Medizinstudiums vorliegen sollten, definiert, sieht es vor, „zielgerichtet(e) und situationsangemessen(e) Indikationen unter Berücksichtigung der Priorisierung, Dringlichkeit und verfügbaren Ressourcen für diagnostische Verfahren stellen“ zu können [6].

Diagnosekompetenz subsumiert dabei die zielgerichtete Sammlung und Interpretation fallspezifischer Informationen, um so Unsicherheiten zu verringern und medizinische Entscheidungen zu treffen [7].

Neben der essenziellen Vermittlung von theoretischem Wissen liegt spätestens seit der Anpassung der ärztlichen Approbationsordnung im Jahr 2002 auch der Fokus auf der Förderung von Problemlösekompetenzen im Medizinstudium. Studierende sollen demnach in der Lage sein, ihr theoretisch erlangtes Wissen anzuwenden, um die ihnen präsentierten Probleme zu bewältigen. Die aktuelle Approbationsordnung sieht dazu ein problemorientiertes Arbeiten anhand der Bearbeitung von Fallbeispielen vor. Hierzu finden immer mehr digitale Angebote Einzug in den curricularen Kontext des Medizinstudiums und computerbasierte Fallsimulationen werden bereits seit dem Ende des letzten Jahrhunderts erfolgreich angewandt [8]. Diese digitale Präsentation virtueller Patienten, deren Krankheitsbild und deren Behandlung den Lerninhalt darstellen, ermöglichen den Lernenden, sich auf kreative Art und Weise mit dem vorgegebenen Material auseinanderzusetzen und gleichzeitig mit anderen Lernenden zu interagieren, was für ein erfolgreiches Lernen durch Simulationen wichtig erscheint [9]. Das digitale Lehrangebot in verschiedenen Ausführungen -ob in Form von interaktiven Simulationen oder geführten, interaktiven Theoriemodulen- ist dabei für Studierende sowohl vielseitig als auch unübersichtlich. Das sich daraus ergebende Ziel der medizinischen Lehrforschung ist es somit, zu eruieren, in welcher Form digitale Lehrangebote am besten einzusetzen sind und den größten Lernerfolg bieten.

Hierbei gilt es zu berücksichtigen, dass der Nutzen von Lerneinheiten mit Fallsimulationen für Studierende unterschiedlicher Vorwissensstufen unterschiedlich groß ausfällt. Es gibt Evidenzen dafür, dass vor allem Lernende mit viel Vorwissen von Fallsimulationen profitieren [10].

Im Einklang damit hat sich auch gezeigt, dass gerade Lernende mit wenig Vorwissen und Erfahrung spezielle Unterstützung (z.B. in Form eines spezifischen Feedbacks) benötigen, um komplexe Fähigkeiten zu erlernen [11].

Somit liegt der Fokus zunehmend auf der Entwicklung effizienter Feedback-Mechanismen.

Darüber hinaus wird auch -nachdem sich in der Vergangenheit häufig mit der Beforschung von individuellem Lernen auseinandergesetzt wurde- immer mehr das kollaborative Bearbeiten von Simulationen untersucht (z.B. [12], [13]).

Der Ansatz des kooperativen Falllösens spiegelt auch die Realität des medizinischen Arbeitslebens wider. Denn häufig ist der Arbeitsalltag von Problemen geprägt, die es - u.a. aufgrund der steigenden Multimorbidität immer älter werdender Patienten- gemeinsam und zumeist interdisziplinär zu lösen gilt [14].

Auch der NKLM sieht vor, Medizinstudierende darin zu schulen, „klinische Entscheidungsprozesse im Team(...) (zu) gestalten“ [6].

In einer Studie, bei der Medizinstudierende aus dem vierten Studienjahr entweder gemeinsam oder alleine insgesamt sechs Patientenfälle bearbeiten sollten, konnte bereits gezeigt werden, dass die Lernpaare den Einzel-Lernern in Bezug auf deren diagnostische Leistung überlegen waren [12]. Zur selben Erkenntnis gelangen auch Kühl et al. bei der Untersuchung einer E-learning Umgebung im Rahmen eines inverted classroom [15].

1.1 Simulationsbasiertes Lernen im Medizinstudium

Innerhalb der letzten Jahre -nicht zuletzt angetrieben durch die weltweite Covid-19-Pandemie und die damit verbundenen Einschränkungen auch in der Hochschullehre- entwickelten sich eine Vielzahl an vielversprechenden digitalen Lehrangeboten (siehe [16]).

Als eine der effektivsten digitalen Lernmethoden hat sich in der Vergangenheit das simulationsbasierte Lernen herauskristallisiert, um den Erwerb komplexer Fähigkeiten zu unterstützen [17]. Hierunter versteht man die Darbietung eines authentischen

Szenarios, welches von -in unsrem Fall- Medizinstudierenden durchlaufen wird [18]. Gerade zu Pandemiezeiten, einhergehend mit Kontaktbeschränkungen und dem Verbot von Präsenzlehre, bietet sich hierbei besonders die digitale Form der Darbietung eines solchen simulierten Szenarios an. Dass sich diese Art der Lehre gut in die klinische medizinische Ausbildung integrieren lässt, konnte bereits in vergangenen Studien gezeigt werden [19], [20]. Gerade dadurch, dass Studierende beim simulationsbasierten Lernen vor Szenarien gestellt werden, zu denen sie eigenständig eine Lösung finden müssen, werden ihre Problemlösekompetenzen gefördert [21]. Im Gesundheitswesen bietet diese Form des Lernens insgesamt einen Vorteil gegenüber nicht-simulationsbasiertem Lernen, wobei jedoch vor allem Lernende mit wenig Vorwissen ein gewisses Maß an Unterstützung und Feedback benötigen [22]. Aufgrund dieses Umstandes gilt es zu untersuchen, inwiefern man Lernenden durch anleitende Interventionen, wie zum Beispiel unterstützendes Feedback in Fallsimulationen, helfen kann [10].

1.2 Feedback als Möglichkeit der Unterstützung simulationsbasierten Lernens

Bei dem -sich in der Hochschullehre bereits etablierten- Lernen am Beispiel, kommen häufig Experten-validierte Lösungen zur Anwendung, die eine Musterlösung für den jeweiligen Fall oder die jeweilige Problemstellung bieten [23].

Insgesamt scheint Feedback, wenn es darum geht, komplexe Fähigkeiten wie das Diagnostizieren von Krankheiten zu erlernen, unabdingbar zu sein [22], [24]. Dabei ist die Präsentation eines statischen Feedbacks in Form einer Musterlösung am Ende der Fallbearbeitung und nach den eigenen Bemühungen der Lernenden ohne großen logistischen und personellen Aufwand umzusetzen, da es automatisch und direkt im Anschluss an die Bearbeitung digital angezeigt werden kann [25].

Im Gegensatz zu statischem Feedback, hat Feedback, welches individuell für jeden einzelnen Lernenden erzeugt wird, das Potential, auf spezifische Ansprüche von Lernenden je nach deren Leistung einzugehen [26]. Unter der Berücksichtigung, dass nicht jeder Studierende die gleichen Wissensvoraussetzungen mitbringt, kann ein solches Feedback dabei helfen, auf die individuellen Stärken und Schwächen eines jeden einzelnen einzugehen, da es Studierenden dabei aufzeigt, wie ihr aktueller Wissensstand einzuordnen ist und ihnen hilft, die Ansatzpunkte zu finden, den Wissensstand auszubauen und zu erweitern [27]; [28].

Bereits 2014 wurde ein erstes Projekt, welches sich mit der Realisierung adaptiven Feedbacks mittels Machine Learning beschäftigte, durchgeführt [29].

Am Beispiel Studierender, die Lösungsstrategien zum Thema Klimawandel erarbeiten sollten, wurde deutlich, dass diese durch automatisiertes, adaptives Feedback ihre Lösungen häufiger revidierten und somit die Qualität ihrer Begründungen steigerten [30]. In einer Folgestudie konnte gezeigt werden, dass kontextbezogenes Feedback einen Vorteil gegenüber statischem Feedback in der Lernunterstützung bietet [31].

Neben der Relevanz der Form des Feedbacks, gelangte Oliver Scheuer in einer Übersicht über verschiedene Techniken zur Schulung von Argumentationsfähigkeiten zu dem Schluss, dass auch die Lernsituation (d.h. ob allein gelernt wird oder kollaborativ mit einem oder mehreren Lernpartnern) beim Erwerb dieser Fertigkeit eine wichtige Rolle spielt [24]. Dem Schluss zur Folge, wurden bereits sowohl im Kontext der Lehrerbildung [32] als auch in der medizinischen Hochschullehre -beispielsweise in Form von klinischen Falldiskussionen (CCD)- Untersuchungen angestellt [33], [34].

1.3 Kollaboratives Lernen mit Fallsimulationen

In der Vergangenheit konnten Studien bereits zeigen, dass sich Kollaborationen von individuell Lernenden in der Art und Weise der Problemlösung unterscheiden:

Während gemeinsam mehr Hypothesen diskutiert werden, bevor sich auf eine Lösung geeinigt wird, zeigen individuell Lernende eine größere, effektivere Bereitschaft, Lösungen vorzuschlagen [35].

Im medizinischen Kontext konnten Tolsgaard et al. darüber hinaus erste, vielversprechende Resultate beim kollaborativen Lernen von klinischen Fähigkeiten in simulierten Umgebungen feststellen [36]

Trotzdem zeichnen Metaanalysen kein eindeutiges Bild, was die Überlegenheit von kollaborativem, simulationsbasiertem Lernen angeht [22], [37], [17]. So gilt es zu untersuchen, unter welchen Modalitäten Lernende wirklich von einer kollaborativen Fallbearbeitung profitieren können.

Darüber hinaus können gemeinsam Lernende ihren Partner zwar als Quelle von Feedback nutzen [38], es gibt jedoch Hinweise darauf, dass auch kollaborativ Lernende von adaptivem Feedback profitieren [39].

1.4 Zur Rolle des Vorwissens beim Lernen mit Fallsimulationen

Vorwissen spielt beim Erwerb neuen Wissens und neuer Fähigkeiten eine wichtige Rolle [40]. Aufbauend auf Ausubels Ausführungen, führte Bianca Simonsmeier eine Metaanalyse zu genau diesem Umstand durch [41].

Basierend auf ihren Erkenntnissen gibt es nun Überlegungen, zu welchem Zeitpunkt, unter welchen Umständen und ab welchem Wissensstand Fallsimulationen effektiv zum Einsatz kommen sollen. Kirschner et al. postulieren dabei, dass dies erst dann vielversprechend sei, sobald eine gewisse Basis in Form von theoretischem Wissen vorliegt. So werden Lernende mit der Simulation nicht überfordert und haben mehr kognitiven Raum, um die Problemstellung zu bewältigen [42]. Demgegenüber spricht für den früheren Einsatz, dass dadurch die Wissens-Umstrukturierung in höhere Ordnungskonzepte unterstützt wird und somit das neu gewonnene Wissen direkt zur

Bewältigung einer bestimmten Problemlösung angewandt werden kann [17]. So würde das theoretische Wissen direkt in Verbindung mit einem Fallbeispiel abgespeichert werden [43].

Lernende mit geringem Vorwissen profitieren dabei von einer ausführlichen Anleitung, da dadurch kognitive Kapazitäten gespart werden, welche dann zum Erlernen neuen Wissens zur Verfügung stehen [44]. Sobald die kognitive Belastung vorwissensschwächerer Lernenden steigt, sinkt gleichzeitig das Lernengagement und umgekehrt steigt es mit sinkender kognitiver Belastung und steigendem Vorwissen [45]. Somit ist es das Ziel, Studierende mit Lehrinterventionen kognitiv so wenig wie möglich zu belasten.

1.5 Fragestellungen und Hypothesen

Die vorliegende Arbeit beschäftigt sich explizit mit den Fragen nach der Auswahl der bestmöglichen Feedback-Art und der Frage, inwiefern das kollaborative Lösen von Simulationsfällen einen Vorteil gegenüber der individuellen Bearbeitung bietet. Hierbei ergeben sich die folgenden zentralen Fragestellungen und Hypothesen, zu deren Beantwortung eine experimentelle multizentrische Feldstudie in primärer Verantwortung des Autors dieser Arbeit durchgeführt wurde:

1. Welchen Einfluss haben unterschiedliche Feedback-Arten (adaptives Feedback versus statisches Feedback) und Bearbeitungsformen (kollaborativ versus individuell) auf die Diagnosekompetenz Medizinstudierender beim Lernen mit Fallsimulationen?

Hierbei vermuteten wir, dass das adaptive Feedback dem statischen Feedback im Hinblick auf die Förderung der Diagnosekompetenz Medizinstudierender überlegen ist (Hypothese H1a).

Darüber hinaus vermuteten wir, dass das kollaborative Lernen mit Fallsimulationen der individuellen Bearbeitungsform überlegen ist (Hypothese H1b).

Schließlich vermuteten wir, dass etwaige positive Einflüsse des Feedbacks nicht von der Bearbeitungsform abhängen (Hypothese H1c).

2. Welchen Einfluss haben unterschiedliche Feedback-Formen (adaptives Feedback versus statisches Feedback) und der Vorwissensstand (hoch versus niedrig) auf die Diagnosekompetenz Medizinstudierender beim Lernen mit Fallsimulationen?

Hierbei vermuteten wir, dass das adaptive Feedback dem statischen Feedback im Hinblick auf die Förderung der Diagnosekompetenz Medizinstudierender überlegen ist (Hypothese H2a).

Darüber hinaus vermuteten wir, dass Medizinstudierende mit hohem Vorwissen mehr vom Lernen mit Fallsimulationen profitieren als Medizinstudierende mit niedrigem Vorwissen (Hypothese H2b).

Schließlich vermuteten wir, dass etwaige positive Effekte des Feedbacks nicht vom Vorwissen der Lernenden abhängen (Hypothese H2c).

Die unabhängigen Variablen dieser Arbeit sind also die Feedbackart (unabhängige Variable 1, UV1), die Bearbeitungsform (unabhängige Variable 2, UV2) und das Vorwissen (unabhängige Variable 3, UV3).

2. Methoden

2.1 Ethikvotum

Das Ethikvotum zur Durchführung dieser Studie wurde am 15.05.2017 durch das Klinikum der Universität München erteilt. Die Durchführung des Projektes 17-250 wurde darin ohne Einwände genehmigt.

2.2 Akquise

Die Teilnehmenden für diese Feldstudie wurden allesamt per standardisierter E-Mail (siehe Anlage 7.1) mit Angaben zum zeitlichen Rahmen der Studie und zu den je Standort angebotenen Terminen auf die Studie aufmerksam gemacht und konnten sich unter Angabe ihres Namens und ihrer E-Mail-Adresse anmelden. Zusätzlich wurde in ausgewählten Lehrveranstaltungen persönlich für die Studie geworben.

2.3 Beschreibung der Stichprobe

Eingeschlossen wurden all diejenigen Proband:innen, welche alle drei Abschnitte der Studiendurchführung konsekutiv bearbeitet haben, bei denen die Lernphase wie vorgesehen abgeschlossen werden konnte, von denen die Erlaubnis zur Datenauswertung erteilt wurde und die im Testfall eine auswertbare und suffiziente Antwort gaben. Hierbei wurden all diejenigen Antworten, die aufgrund von Bezugslosigkeit, Unvollständigkeit oder Knappheit nicht kodierbar und somit nicht auswertbar waren, aus der Auswertung ausgeschlossen.

So ergab sich für die Studie, für die Proband:innen an den insgesamt vier verschiedenen medizinisch-universitären Standorten in Bayern (München: Medizinische Fakultäten der Ludwig-Maximilians-Universität und der Technischen Universität München, Würzburg: Medizinische Fakultät der Universität Würzburg, Erlangen: Medizinische

Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg) rekrutiert wurden, folgende Besetzung:

Die Kohorte aus München und Würzburg zeigte bei der Zellenbesetzung keine signifikanten Unterschiede bezüglich des Geschlechts der Student:innen. Insgesamt haben an diesen Standorten 121 weibliche und 51 männliche Studierende an der Studie teilgenommen. Der Anteil der weiblichen Teilnehmerinnen betrug somit 70 Prozent, was in etwa der Geschlechterverteilung im Medizinstudium entspricht [46]. In der Kohorte aus Erlangen entsprach der Anteil der Studentinnen 69 Prozent bei insgesamt 140 Teilnehmenden. Auch hierbei ergab sich kein signifikanter Unterschied in der Zellenbesetzung. Im Mittel betrug das Alter der Studierenden aus Erlangen 23 (23,34) Jahre mit einem Modus von 21 Jahren (37,4%), während es in München und Würzburg bei 24 (24,46) Jahren lag. Der Häufigkeitsgipfel fand sich hier bei 23 Jahren (41,9%). Die Spannweite lag dabei in München und Würzburg zwischen 21 und 35 Jahren, in Erlangen lag sie zwischen 19 und 39 Jahren.

2.4 Studiendesign

Es handelt sich bei dieser Studie um eine experimentelle randomisiert-kontrollierte Feldstudie, die vom 12.11.2019 bis zum 08.06.2020 durchgeführt wurde. Als Proband:innen kamen alle Medizinstudierenden infrage, die Grundwissen in allgemeiner oder innerer Medizin vorweisen konnten. Explizit wurden hierfür an den Standorten Würzburg (*Medizinische Fakultät Universität Würzburg*) und München (*TUM Fakultät für Medizin, Medizinische Fakultät der LMU München*) Studierende ab dem 8. Fachsemester rekrutiert. Diese wurden sodann randomisiert jeweils einer von zwei Erhebungsgruppen zugeteilt. Je eine der beiden Gruppen erhielt adaptives Feedback (UV1) auf die Bearbeitung der einzelnen Fallbeispiele, welches durch künstliche Intelligenz, der Verarbeitung natürlicher Sprache (natural language processing, NLP) [47] und eine

Feedback-Datenbank individuell Bezug auf das diagnostische Vorgehen der Studierenden nahm [48]. Die verwendete Künstliche-Intelligenz-Methode war dabei in der Lage, mithilfe von einem im Vorfeld durchgeführten Lernprozess, anhand von NLP zu erkennen, ob einzelne Diagnostika im Freitext als angewandt, nicht angewandt oder gar nicht erwähnt wurden. Daraufhin erfolgte ein Abgleich mit der Feedback-Datenbank (siehe Abbildung 1) [49].

Abbildung 1

Auszug aus der Feedback-Datenbank zur Veranschaulichung der Funktionsweise des adaptiven Feedbacks

Kategorie	Fachbegriff	Synonyme	Feedback
Echokardiographie	Echokardiographie	Echokardiografie; Herzecho; Herz-Echo; Echo; Mitralklappe; Perikarderguss; Klappenvegetationen; Vegetationen TEE; Mitralsuffizienz	Du hast gut erkannt, dass die Echokardiographie hier wichtig ist. Sie zeigt eine mittelgradige Mitralklappeninsuffizienz, einen Perikarderguss und eine Einflusstauung der Vena Cava inferior. Außerdem sind Klappenvegetationen nachweisbar.
Echokardiographie	NOT Echokardiographie	NOT Echokardiografie; NOT Herzecho; NOT Herz-Echo; NOT Echo; NOT Mitralklappe; NOT Perikarderguss; NOT Klappenvegetationen; NOT Vegetationen, NOT TEE; NOT Mitralsuffizienz	Du hättest erkennen sollen, dass die Echokardiographie hier wichtig ist. Sie zeigt eine mittelgradige Mitralklappeninsuffizienz, einen Perikarderguss und eine Einflusstauung der Vena Cava inferior. Außerdem sind Klappenvegetationen nachweisbar.

Das zutreffende Feedback wurde dann von der Künstlichen-Intelligenz-Methode ausgegeben (siehe Abbildung 2).

Abbildung 2

Beispiel eines adaptiven Feedbacks in der Lernumgebung

Freitext-Antwort

„75 jähriger Patient mit Fieber seit einem Tag. In Zusammenschau der Befunde kommt ein infektiöses Geschehen in Frage. Bei der körperlichen Untersuchung imponiert das Systolikum über der Mitralklappe welches in der **Echokardiographie** mit einer Klappenvegetation korreliert. Die Blutkulturen zeigen einen Nachweis von S. aureus. Daraus ergibt sich die Diagnose der bakteriellen Endokarditis.“

Adaptives Feedback

- ⬆ Du hast richtig erkannt, dass für diesen Fall eine körperliche Untersuchung wichtig ist. Einer Anamnese sollte sich immer eine gründliche körperliche Untersuchung anschließen.
- ⬆ Gut, dass dir die febrile Körpertemperatur mit 39,1°C aufgefallen ist. Die Herzfrequenz ist mit 135 Schlägen/Minute tachykard.
- Mit den Splinter-Blutungen unter den Nägeln und den alten Hämatomen an der Hüfte sind dir wichtige Hinweise zur richtigen Diagnose entgangen. Splinter Blutungen können z.B. bei bakterieller Endokarditis, Psoriasis, Trichinose und rheumatoider Arthritis vorkommen.
- ⬆ Wie du richtig erkannt hast, fällt bei der Auskultation des Herzens ein Systolikum mit p.m. im fünften ICR medioclavicular links mit Fortleitung in die Axilla auf. Dies deutet auf eine Mitralklappeninsuffizienz hin. Die Kombination Fieber und Herzgeräusch lassen wieder an eine bakterielle Endokarditis denken.
- Da hier Hinweise auf einen Infekt bestehen, hättest du dich für die Abnahme eines Labors entscheiden müssen. Hier zeigen sich eine normozytäre Anämie, eine Leukozytose mit Linksverschiebung der neutrophilen Granulozyten, eine Thrombozytopenie und ein erhöhter CRP-Wert. Auch diese Befundkonstellation passt wieder zur bakteriellen Endokarditis.
- Ein EKG zur Abklärung wäre sinnvoll gewesen. Es zeigt sich ein abnormes EKG mit Sinustachykardie und stattgehabtem Infarkt.
- ⬆ Du hast gut erkannt, dass die Echokardiographie hier wichtig ist. Sie zeigt eine mittelgradige Mitralklappeninsuffizienz, einen Perikarderguss und eine Einflusstauung der Vena Cava inferior. Außerdem sind Klappenvegetationen nachweisbar.
- ⬆ Blutkulturen sind in diesem Fall zur Diagnosefindung notwendig. Sie sind positiv auf S.aureus, dem häufigsten Erreger der Endokarditis acuta.
- Eine Thorax-Röntgenuntersuchung wäre hier obligat gewesen. Sie zeigt eine Stauung der Lunge sowie einen Pleuraerguss rechts.

Hinweis auf fehlendes Diagnostikum

Hervorheben des korrekt angewandten Diagnostikums

Die andere Gruppe erhielt statisches Feedback in Form einer vorformulierten Musterlösung. In beiden Gruppen wurden zunächst die Zellen für die kollaborative Bearbeitung (UV2) gefüllt, da dies die am kompliziertesten umzusetzende und fehleranfälligste Komponente darstellte. Kam es dabei aber bei einzelnen Gruppen zu technischen

Schwierigkeiten, sodass keine Kollaboration möglich war, wurden die Ergebnisse der Teilnehmenden individuell erhoben.

Somit ergab sich ein 2X2 Design mit insgesamt vier Gruppen: *individuell-adaptiv*, *individuell-statisch*, *kollaborativ-adaptiv*, *kollaborativ-statisch*. Die Zuteilung zu den kooperativen Dyaden erfolgte dabei ebenfalls randomisiert.

Tabelle 1

Visualisierung des 2X2-Designs Bearbeitungsform und Feedbackart

	Kollaborative Bearbeitung	Individuelle Bearbeitung
Adaptives Feedback	<i>kollaborativ-adaptiv</i>	<i>individuell-adaptiv</i>
Statisches Feedback	<i>kollaborativ-statisch</i>	<i>individuell-statisch</i>

Um die Umsetzung an den verschiedenen Standorten in einem weitestgehend identischen Rahmen zu ermöglichen, wurden diverse Bemühungen unternommen, um die Vorbereitung und Durchführung zu standardisieren. So wurde im Vorfeld sowohl ein standardisiertes Anschreiben verfasst, welches einen Überblick über die Ziele der Studie sowie deren Durchführung und die möglichen Termine geben sollte. Darüber hinaus wurde ein Begrüßungsschreiben aufgesetzt, das die genauen Voraussetzungen der Studie, sowie den individuellen Zugangscode eines jeden einzelnen Teilnehmenden beinhaltete.

Ein weiteres 2X2-Design ergab sich durch die Teilnahme Studierender am Universitätsstandort Erlangen (*Medizinische Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg*). Im Unterschied zu den anderen Standorten wurde die Erhebung hier an die Vorlesungsreihe im Fach Allgemeinmedizin angegliedert und ersetzte einen kompletten Vorlesungstermin. Da hier davon auszugehen war, dass viele Studierende das Angebot in Anspruch nehmen und sich eine Vielzahl von ihnen zur Teilnahme im Hörsaal aufhalten würden, wurden an diesem Standort lediglich die beiden Kohorten

individuell-adaptiv und individuell-statisch erhoben. Ein weiterer Unterschied im Vergleich zu den Standorten München und Würzburg lag darin, dass sich die Studierenden allesamt im fünften Fachsemester der humanmedizinischen Hochschulausbildung befanden und hierbei somit davon ausgegangen werden musste, dass sie über ein niedrigeres Vorwissen (UV3) verfügten. Somit dient die Kohorte aus Erlangen zum Vergleich zu den Kohorten der anderen bayerischen Universitätsstandorte (Würzburg, München).

Tabelle 2

Visualisierung des 2X2-Designs Vorwissen und Feedbackart

	Niedriges Vorwissen	Hohes Vorwissen
Adaptives Feedback	<i>niedriges Vorwissen- adaptiv</i>	<i>hohes Vorwissen- adaptiv</i>
Statisches Feedback	<i>niedriges Vorwissen- statisch</i>	<i>hohes Vorwissen- statisch</i>

2.5 Fallmaterial

Bei FAMULUS 3 wurde das Fallmaterial aus den vergangenen FAMULUS-Projektphasen eingesetzt [50]. Um jedoch eine Erhebung an möglichst vielen verschiedenen Standorten ermöglichen zu können und bereits bekannte Kritik zu beachten, wurden sowohl die Patientenfälle zum Leitsymptom „Fieber unklarer Genese“, deren adaptives Feedback, sowie die in den Prä- und Posttests verwendeten Key-Feature-Fälle inhaltlich gemeinsam mit Fachexpert:innen aus Würzburg und München überarbeitet. Hierfür wurden Fachärzt:innen für Innere Medizin, Kardiologie, Endokrinologie und Infektiologie zu Rate gezogen. Ziel der experimentellen Feldstudie war es, die Erhebung an eine curriculare Veranstaltung der Partneruniversitäten zu knüpfen. Um sowohl den

Fachbereich Allgemeinmedizin als auch den der inneren Medizin ansprechen zu können, wurde sich auf das Leitsymptom „Fieber unklarer Genese“ beschränkt. Eine Freischaltung zu den Fällen zum Leitsymptom „Rückenschmerz“ konnte jedoch von den Proband:innen fakultativ im Anschluss an die Erhebung unter Angabe ihrer E-Mail-Adresse beantragt werden. So wurden aus Prä- und Posttest fachliche Fragen, sowie Key-Feature Fälle zum Thema „Rückenschmerz“ entfernt.

Die Beschränkung auf nur ein Leitsymptom hatte darüber hinaus zur Folge, dass die insgesamt antizipierte Bearbeitungszeit reduziert werden konnte.

2.6 Instrumente

2.6.1 Prätest

Der Prätest diente zum einen dazu, die jeweiligen Proband:innen studienfortschrittlich einzuordnen, ihren Wissensstand abzufragen, ihre Diagnosekompetenz zu evaluieren und zu guter Letzt dazu -vor allem unter dem Aspekt der Feldstudie- sie mit der online Umgebung in *CASUS* vertraut zu machen. So wurden zu Beginn nach der Begrüßung und der Einverständnigabe zur anonymen Datenverarbeitung zunächst Angaben zu Geschlecht, Alter, Muttersprache, Abiturnote, Physikumsnote, Semesterzahl, abgeschlossenen Berufsausbildungen, bisherigen Famulaturen und angestrebter Fachrichtung gemacht. Anschließend folgten vier fachliche MC-Fragen zur Erhebung des *konzeptuellen Wissens*, sowie zwei Key-Feature-Fälle aus je drei Karten (Differentialdiagnose, Untersuchungen, Weiteres Vorgehen) zum Themenkomplex „*Fieber unklarer Genese*“ zur Erhebung des *strategischen Wissens*. Zudem erfolgte die Erhebung der *Motivation* mittels drei Items und jeweils einer Likert-Skala von 1 bis 5 (gekürzte Version nach [51]) und zum Abschluss die Erhebung der zusammengefassten intrinsischen und extrinsischen kognitiven Belastung mittels der vier Items umfassenden Opfermann-Skala [52]. Analog zu den Vorstudien (siehe [10]) und in Anlehnung an die

Empfehlung Opfermanns wurden sowohl die intrinsische als auch die extrinsische kognitive Belastung erfasst und als Variable *Kognitive Belastung* zusammengefasst.

Der Prätest wurde von der eigentlichen Erhebung abgekoppelt und von den Studierenden bereits im Vorfeld zur Bearbeitung der Lernfälle bearbeitet. Für die Bearbeitung des Prätests wurden 20 Minuten angedacht.

2.6.2 Lernphase

Die Lernphase bestand aus drei aufeinanderfolgenden virtuellen Patientenfällen, welche bereits in Studie 1 und 2 erstellt und erprobt wurden.

So startete diese Phase mit der virtuellen Patientin „Frau Hoffmann“ (Hepatitis A), hierauf folgte „Frau Klein“ (Hyperthyreosis factitia) und den Abschluss bildete „Herr Schwarz“ (EBV-Infektion).

Aus den Erkenntnissen vorheriger Studienabschnitte konnte die Bearbeitungszeit eines einzelnen Lernfalls auf 20 Minuten festgelegt werden. Für die Lernphase ergab sich somit eine insgesamt Bearbeitungszeit von 60 Minuten.

2.6.3 Testfall

Der Testfall bestand aus der individuellen Bearbeitung eines weiteren virtuellen Patienten „Herr Schneider“ (bakterielle Endokarditis).

Zur Bewertung der Leistung bei der Bearbeitung des Testfalls und um die Diagnosekompetenz eines jeden Teilnehmenden quantifizieren zu können, wurde interprofessionell ein Score entwickelt, welcher bei der Bewertung der einzelnen Gesichtspunkte der Fallbearbeitung zur Anwendung kam.

Jeder Fall bietet grundsätzlich drei Aufgabenstellungen, die bei der Bewertung der Diagnosekompetenz zum Tragen kommen (siehe Tabelle 3).

Zum einen geht es darum, die richtige Diagnose zu stellen, aber auch darum, den diagnostischen Weg bis zur Diagnose logisch und nachvollziehbar zu beschreiben und zu guter Letzt alle relevanten und zielführenden Diagnostika anzuwenden. Die Diagnoserichtigkeit wird als Variable *Diagnoseakkuratheit* festgehalten. Hierbei wurde eine Bewertungseinheit (BE) vergeben, wenn die Antwort mit der zu erwartenden Diagnose übereinstimmte oder es sich dabei um ein Synonym handelte. Neben der Nachvollziehbarkeit der Begründung, welche mit maximal *drei* BE bewertet wurde (jeweils eine BE für die Einhaltung der Struktur, die Nachvollziehbarkeit und die Begründung über das Verwerfen falscher Differentialdiagnosen) und als Variable *Bewertung der Argumentation* festgelegt wurde, bot zu guter Letzt die schriftliche Ausführung über das diagnostische Vorgehen eines jeden Teilnehmers die Grundlage für die letzten zu erreichenden BE für das diagnostische Vorgehen. Hierbei wurde sich auf insgesamt

Tabelle 3
Unterpunkte des Testfall-Scores

Variable	Max.
Diagnoseakkuratheit	1
Qualität der Begründung	6
Bewertung der Argumentation (Struktur, Nachvollziehbarkeit, Verwerfen falscher DDs)	3

wurde sich auf insgesamt

sechs Punkte verständigt, die bei dem Vorliegen einer bakteriellen Endokarditis unerlässlich zur Diagnosestellung sind. Diese Bewertungseinheiten wurden sowohl in Abstimmung mit der aktuellen Leitlinie zur Diagnostik der Endokarditis als auch mit einem kardiologischen Fachexperten festgelegt und ergeben die Variable *Qualität der Begründung*.

Eine Stichprobe aller Studierenden (10% der Gesamtkohorte) wurde von zwei unabhängigen Experten anhand des erarbeiteten Scores bewertet. Bei einer hierbei erhobenen Interrater-Reliabilität von Cohen`s $k = 0,85$ wurde der manuell erhobene Score von einem einzelnen Experten auf die restlichen Studierenden angewandt.

Zusätzlich wurde die Bearbeitungszeit des Testfalls als Variable *Bearbeitungszeit* in die Auswertungen aufgenommen.

2.6.4 Posttest

Im Anschluss an den Testfall folgte analog zum Prätest ein Fragebogen. Dieser Fragebogen bestand ebenfalls aus einer Abfrage von *konzeptuellem Wissen* in Form von vier Multiple-Choice-Fragen und einer Einschätzung des *strategischen Wissens* durch zwei Key-Feature-Fälle zum Themenkomplex „Fieber“. Darüber hinaus beinhaltete der Posttest analog zum Prätest erneut eine Abfrage zu *kognitiver Belastung* und *Motivation*, sowie drei Freitext-Fragen zur Evaluation des Projektes. Hierbei konnten frei formulierte Antworten auf die Fragen nach nötigen Veränderungen zur Hochschulintegration, den Stärken und Schwächen der Simulation gegeben werden.

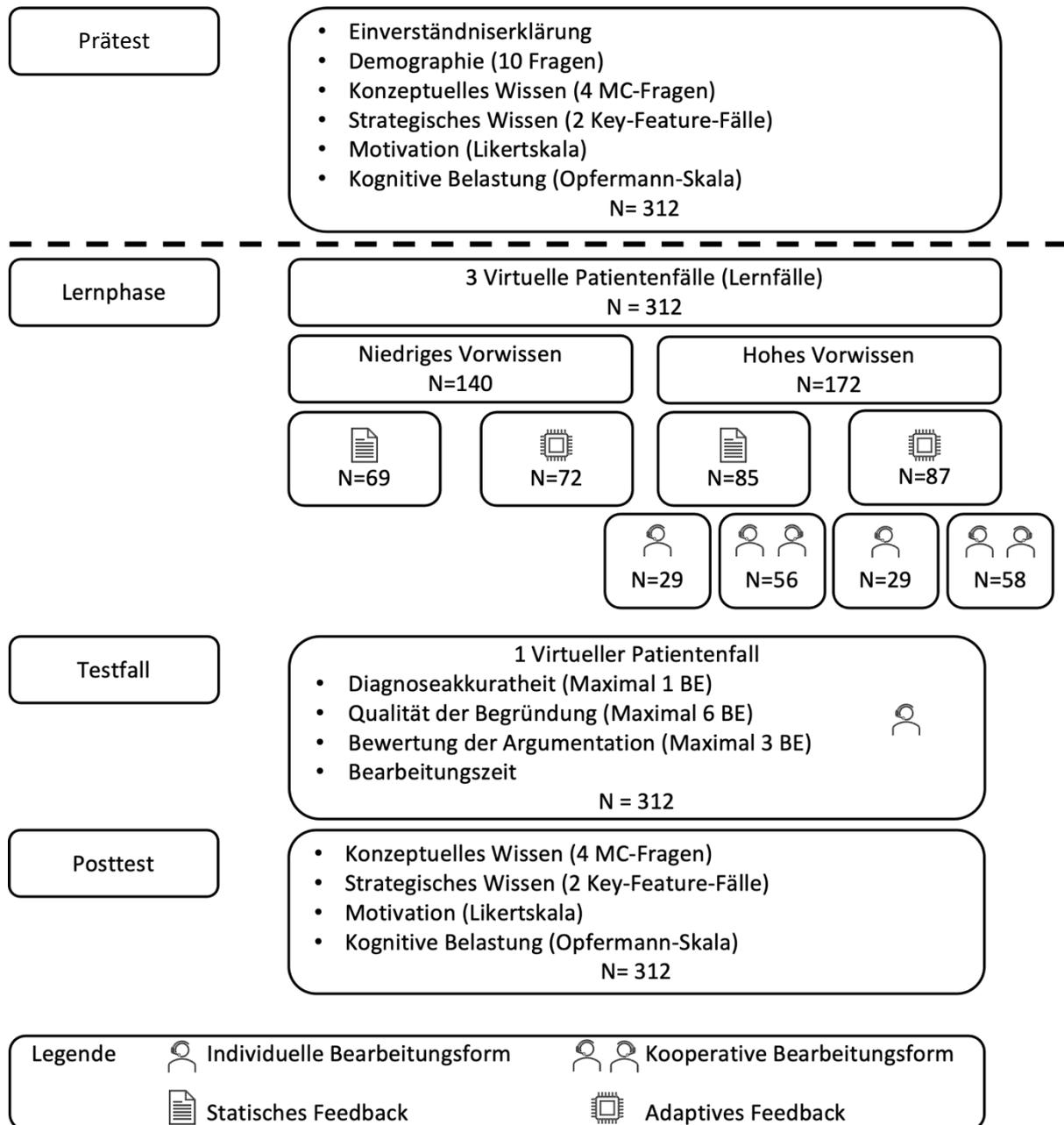
Die letzte Seite des Tests beinhaltete ein Freitext-Feld zur Angabe der E-Mail-Adresse und fakultativen Freischaltung für die vier virtuellen Patienten zum Leitsymptom „Rückenschmerz“. Hierbei konnte keine Verknüpfung der Studienleistung mit der jeweiligen E-Mail-Adresse hergestellt werden. Die Anonymität der Studienteilnehmer:innen blieb gewahrt. Somit diente die Zahl der Anmeldungen zum fakultativen Angebot als Maß für die Akzeptanz der Lehrintervention bei den Studierenden.

Für die Bearbeitung von Testfall und Posttest wurden insgesamt 30 Minuten angedacht. Somit ergab sich für die Studienteilnahme am Erhebungstag eine Gesamtdauer von 90 Minuten, was der Länge einer Seminareinheit entspricht.

Zur Veranschaulichung des Studienablaufs und der dabei erhobenen Variablen siehe Abbildung 3.

Abbildung 3

Veranschaulichung des Studiendesigns inklusive der erhobenen Variablen



2.7 Vorbereitungen

2.7.1 Umsetzung als Feldstudie

Die bisher unter Laborbedingungen -bei denen die Studiendurchführung unter Aufsicht und in Präsenz durchgeführt wurde [50]- erprobten Instrumente sollten im Rahmen dieser Studie im Feld eingesetzt werden. Hierfür waren einige Vorkehrungen vonnöten, um eine stabile und standardisierte Durchführung an verschiedenen Standorten zu ermöglichen. Das oberste Ziel bei der Planung der Feldstudie war es, auf keinen wesentlichen Aspekt der vorangegangenen Projektphasen verzichten zu müssen. So galt es, insbesondere die Kollaboration für die Feldbedingungen zu optimieren und den zeitlichen Rahmen -welcher ursprünglich bei vier Stunden Bearbeitungszeit lag- anzupassen bzw. zu verkürzen.

2.7.2 Anpassungen an der Online-Oberfläche

Die Feldstudie wurde in der Online-Umgebung CASUS [8] realisiert. Um dabei einen reibungslosen Ablauf ohne physisch präsente Aufsicht zu ermöglichen, wurden im Vergleich zu den bisherigen Erhebungen einige Anpassungen und Ergänzungen vorgenommen. Ziel war es dabei, die Durchführung der Studie für die Studienteilnehmenden zu vereinfachen und sicherzustellen, dass jede einzelne Erhebung standardisiert ablaufen kann. Die im Rahmen der Studie implementierten Anpassungen und Ergänzungen der Benutzeroberfläche sind seither Bestandteil der Lernplattform und stehen allen Nutzern zur Verfügung. Die Umsetzung erfolgte in enger Zusammenarbeit mit dem Geschäftsführer der INSTRUCT gGmbH.

2.7.2.1 *Live-Chat*

Trotz der erprobten Online-Umsetzung ließ sich nicht ausschließen, dass es bei einzelnen Proband:innen während der Studienteilnahme zu technischen Problemen kommt, die Verbindung zu ihren Kollaborationspartnern abbricht oder ein Kollaborationspartner der Studie kurzfristig fernbleibt. Da es bei diesem Studiendesign nicht

umsetzbar war, eine physische Aufsicht für jeden Erhebungstermin bereitzustellen und bei der Studie ohnehin eine Teilnahme von zuhause aus möglich sein sollte, musste eine Lösung gefunden werden, dennoch eine Art technische Unterstützung in Realzeit anzubieten. So wurde für FAMULUS ein Live-Chat-Support eingerichtet, der von jedem Teilnehmenden verwendet und zentral bedient werden konnte. Hierdurch wurde sichergestellt, dass jeder Proband die Studie vollständig durchlaufen konnte, auch wenn es zwischenzeitlich zu Problemen kam.

2.7.2.2 Anpassungen in der Handhabung

Da diese Studie multizentrisch konzipiert wurde und nicht jede medizinische Universität bei E-Learning Angeboten auf *CASUS* setzt, musste die Handhabung der Oberfläche so intuitiv und eindeutig wie möglich gestaltet werden. Das beste Beispiel hierfür ist die neue Verbindung von einem Patientenfall zum nächsten. Ursprünglich hätte es mehrere Klicks benötigt, um von einem beendeten Hauptfall zum nächsten zu gelangen. Hierbei bestand das Risiko, dass man sich aus Versehen abmeldet und die Studie somit vorzeitig beendet. So wurde eine Lösung implementiert, die es ermöglichte, durch den Klick auf einen -in einem Textfeld auf der letzten Seite eines jeden Falles angezeigten Link- direkt zum darauffolgenden Fall zu gelangen.

Darüber hinaus wurden einige Anpassungen am User-Interface vorgenommen. So wurde(n)

- die Notizfunktion in der Art überarbeitet, eine Notiz anzufertigen ohne dafür eine neue Notiz anlegen zu müssen
- die Pfeiltasten zur Navigation so gestaltet, dass immer klar war, welche es zu drücken galt, um fortzufahren
- Hinweise zum Vergrößern von Bildern ergänzt
- der Logout-Prozess auf der letzten Seite des letzten Hauptfalls ausführlich beschrieben

- das Longmenu bei Eingabe der Diagnose um eine freie Eingabemöglichkeit ergänzt.

2.7.2.3 Einführungsvideo und Leitfaden

Um die komplexeste Herausforderung dieses Experimentes -nämlich die kollaborative Verschaltung von zwei Proband:innen, während diese räumlich getrennt voneinander sind- zu meistern, wurde sowohl ein PDF-Dokument mit einem *Leitfaden zum kollaborativen Arbeiten* als auch ein knapp zweiminütiges Einführungsvideo erstellt. Ziel des Videos war es, eine visuelle Schritt-für-Schritt-Anleitung zu liefern, welche nicht nur den Verschaltungsprozess, sondern auch den Live-Chat-Support vorstellte. Wichtig bei der Erstellung des Videos war zum einen, es nicht mit Informationen zu überladen und zum anderen, genügend Informationen zu liefern, sodass eine reibungslose Kollaboration in Kombination mit dem zur Verfügung gestellten Leitfaden gewährleistet war. Realisiert wurde dies durch Bildschirmaufnahmen, welche anschließend mittels Voice-over kommentiert wurden. Das Endprodukt wurde in die Begrüßungsseite des Prätests integriert, um sicherzustellen, dass jeder der Teilnehmenden sich bereits vor Antritt zur Hauptsitzung mit der Handhabung vertraut macht. Der Leitfaden wurde als Anhang an die Begrüßungsmail zur Verfügung gestellt (siehe Anlage 7.2).

2.7.2.4 PIN

Da die Erhebung unter Laborbedingungen mit fünf individuellen PINs durchgeführt wurde, musste das Zugangssystem zur Feldstudie vereinfacht werden. So erhielt jeder Proband nunmehr einen einzigen PIN-Code, welcher ihn für alle drei Erhebungs-Abschnitte (Prätest, Trainingsphase, Posttest) freischaltete. Jeder Teilnehmende wurde in *CASUS* jedem Erhebungsabschnitt mit demselben Zugang hinzugefügt. Die Kurse wurden dann zeitlich versetzt freigeschaltet, sodass es nicht möglich war, die Reihenfolge individuell abzuändern.

2.7.2.5 Pilotierung der Anpassungen

Um diese Anpassungen unter Realbedingungen zu erproben, wurde bereits mehrere Monate vor Studienstart eine erfolgreiche Pilotierung mit einer kleinen Gruppe Studierender durchgeführt.

2.8 Technischer Ablauf

Nach erfolgter Anmeldung empfing jeder der Teilnehmenden eine standardisierte Begrüßungsmail mit genaueren Angaben zu technischen Voraussetzungen, Ablauf und zur Funktion der Kollaboration während der Erhebung (siehe Anlage 7.3).

Jede/r Proband:in besaß dazu einen individuellen Code durch welchen sie/er Zugang zu den Kursen auf <https://talky.casus.net> erhielt. Für jeden Erhebungstag wurde ein neues Kurspaket in CASUS erstellt. Jedem dieser Kurse wurden die Proband:innen mit ihrem individuellen PIN zugeteilt, sodass nicht mehrere PIN-Codes nötig waren. Zugleich wurden die kollaborativen Dyaden zufällig eingeteilt. Die Kurse wurden dabei jeweils kopiert, deren externe IDs übernommen und aus den abgelaufenen Kursen gelöscht, sodass die Fall-Verlinkung nach wie vor funktionierte. Damit nicht alle Kurse zum Zeitpunkt des Prätests zur Auswahl standen, wurden diese nach einem individuellen Zeitplan je Erhebungstag freigeschaltet. So war der Prätest bis fünf Minuten vor der eigentlichen Erhebung anwählbar, die Trainingsphase zum ausgeschriebenen Termin und der Posttest jeweils bis 45 Minuten nach Beginn der Erhebung.

Um sicherzustellen, dass der Prätest auch wirklich im Vorfeld bearbeitet wurde, erhielten die Teilnehmenden zwei Tage vor angemeldetem Termin eine Erinnerungs-Mail mit dem wiederholten Hinweis auf die technischen Voraussetzungen der Studierendurchführung und der verpflichtenden Bearbeitung des Prätests.

Die technische Unterstützung wurde jeweils eine halbe Stunde vor offiziellem Start der Erhebung aufgenommen. Zum Start wählten sich die Proband:innen mit ihrem Zugangscod ein, warteten auf ihre Kollaborationspartner:in, verbanden sich mit dieser

und bearbeiteten die drei Hauptfälle gemeinsam. Im Anschluss daran wurde die Verbindung getrennt und jeder Teilnehmende bearbeitete den Testfall sowie den Posttest individuell.

2.9 Statistische Analyse

Die erhobenen Daten wurden mit SPSS 28.0 (IBM Inc.) ausgewertet. Hierbei wurden Kovarianzanalysen (englisch analysis of covariance, kurz ANCOVA) gerechnet. Die Aufschlüsselung der Werte der dabei berichteten Ergebnisse kann Tabelle 4 entnommen werden [53, 54].

Tabelle 4

Aufschlüsselung der berichteten Werte und Abkürzungen

$F(df_{\text{Zähler}}, df_{\text{Nenner}}) = F\text{-Wert}, p = \text{Signifikanz}, \text{partial } \eta^2 = \text{Effektstärke}$

F	Das Testverfahren verwendet eine F-Statistik. Dieser liegt eine F-Verteilung zugrunde.
(df _{Zähler} , df _{Nenner})	Diese beiden Parameter der F-Verteilung beeinflussen deren Aussehen und die Grenzen der Signifikanz
F-Wert	Dieser Wert wird in der F-Verteilung nachgeschlagen, um den p-Wert zu berechnen
p	Dies ist der Wert, nach dem sich die Signifikanz richtet
partial η^2	Gibt die Effektstärke an

Ein p-Wert kleiner 0,05 wird als signifikant angesehen. Die Einteilung der Effektstärke des partial η^2 wurde anhand Cohens Empfehlungen vorgenommen:

0,01 < $\eta^2 \leq 0,06$ entspricht einem kleinen Effekt. 0,06 < $\eta^2 \leq 0,14$ entspricht einem Effekt mittlerer Größe und $\eta^2 > 0,14$ einem großen Effekt [55].

3. Ergebnisse

3.1 Untersuchung der Effekte von Feedbackart und Bearbeitungsform

3.1.1 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf das *konzeptuelle* und *strategische Wissen*

Kovarianzanalysen (mit den jeweiligen Werten aus dem Prätest und der *Bearbeitungszeit* als Kovariate, um deren Einfluss auf das Ergebnis zu kontrollieren) zeigten keinen Effekt der Feedbackart auf das *konzeptuelle Wissen* im Post-Test, $F(1,168)=0,204$; $p=0,652$. Auch die Bearbeitungsart hatte keinen Effekt auf das *konzeptuelle Wissen* im Post-Test, $F(1,168)=.072$; $p=0,788$. Die Mittelwerte der richtigen Antworten in der Einzelbedingung lagen bei der Gruppe mit statischem Feedback bei 3,86 (SD 1,36) und bei der Gruppe mit adaptivem Feedback bei 3,86 (SD 0,88). In der Kollaboration wurden mit statischem Feedback im Mittel 3,89 (SD 0,97) und mit adaptivem Feedback 3,86 (SD 0,98) Punkte erreicht (siehe Tabelle 5).

Ebenfalls fand sich kein Effekt der Feedbackart auf das *strategische Wissen* im Post-Test, $F(1,168)=0,100$; $p=0,752$. Die Bearbeitungsart hatte auch keinen Effekt auf das *strategische Wissen* im Post-Test, $F(1,168)=.257$; $p=0,613$. Die Mittelwerte der bei der Bearbeitung der Key-Feature-Fälle erreichten Punkte lagen in der Einzelbedingung bei der Gruppe mit statischem Feedback bei 4,67 (SD 0,31) und bei der Gruppe mit adaptivem Feedback bei 4,74 (SD 0,27). In der Kollaboration wurden mit statischem Feedback im Mittel 4,72 (0,38) und mit adaptivem Feedback 4,69 (0,40) Punkte erreicht (siehe Tabelle 5).

3.1.2 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die Lösung des Testfalls (Herr Schneider)

Um den Einfluss vorbestehender Unterschiede in *strategischem Wissen* im Prätest und den Einfluss der *Bearbeitungszeit* zu kontrollieren, dienten diese beiden Variablen als Kovariate.

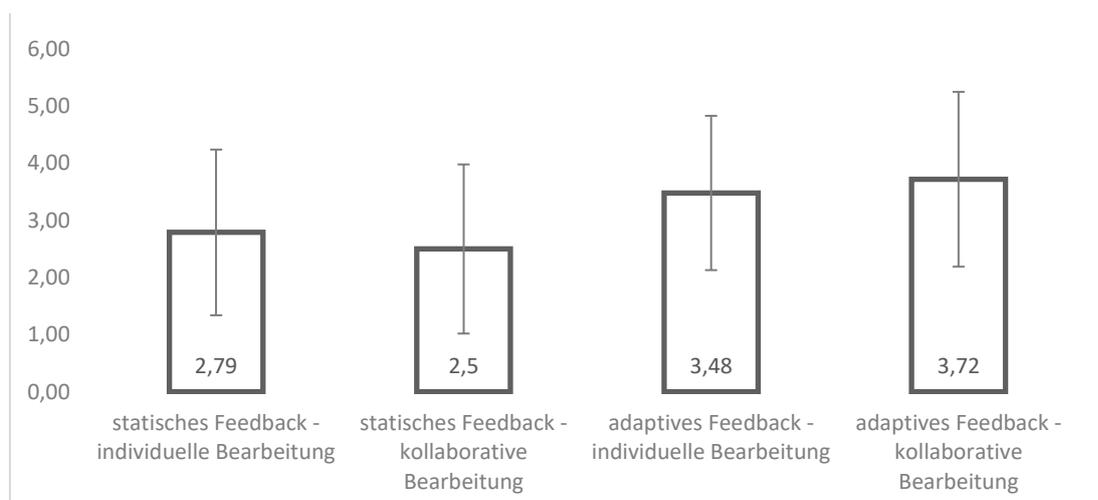
Ein Effekt des adaptiven Feedbacks von mittlerer Stärke zeigte sich beim Einfluss auf die *Qualität der Begründung*, $F(1,168)=15,624$; $p=0,001$, $\text{partial } \eta^2=0,086$.

Die Bearbeitungsform zeigte keinen signifikanten Einfluss, $F(1,168)=1,508$; $p=0,221$.

Die Mittelwerte für die *Qualität der Begründung* lagen in der Einzelbedingung bei der Gruppe mit statischem Feedback bei 2,79 (SD 1,45) und bei der Gruppe mit adaptivem Feedback bei 3,48 (SD 1,35). In der Kollaboration wurde die *Qualität der Begründung* der Gruppe mit statischem Feedback im Mittel mit 2,50 (SD 1,48) und bei der Gruppe mit adaptivem Feedback mit 3,72 (SD 1,53) Punkten bewertet (siehe Tabelle 5). Abbildung 4 zeigt eine graphische Darstellung der Mittelwerte der einzelnen Gruppen (siehe Abbildung 4).

Abbildung 4

Qualität der Begründung in Bezug auf Feedbackart und Bearbeitungsform



Bei der *Diagnoseakkuratheit* wurde kein Effekt der Feedbackart gefunden, $F(1,168)=2,019$; $p=0,157$. Die Bearbeitungsform hatte ebenso keinen signifikanten Einfluss, $F(1,168)=0,074$; $p=0,786$.

Die Mittelwerte für die *Diagnoseakkuratheit* lagen in der Einzelbedingung bei der Gruppe mit statischem Feedback bei 0,45 (SD 0,51) und bei der Gruppe mit adaptivem Feedback bei 0,62 (SD 0,49). In der Kollaboration lag die *Diagnoseakkuratheit* der Gruppe mit statischem Feedback im Mittel bei 0,54 (SD 0,50) und bei der Gruppe mit adaptivem Feedback bei 0,60 (SD 0,49) (siehe Tabelle 5).

Die *Bewertung der Argumentation* fiel darüber hinaus bei Proband:innen mit adaptivem Feedback mit einer kleinen Effektstärke signifikant besser aus, $F(1,168)=6,872$; $p=0,010$; partial $\eta^2=0,040$. Hier lagen die Mittelwerte in der Einzelbedingung bei der Gruppe mit statischem Feedback bei 0,72 (SD 0,59) und bei der Gruppe mit adaptivem Feedback bei 0,97 (SD 0,57). In der Kollaboration wurden im Mittel in der Gruppe mit statischem Feedback 0,77 (SD 0,54) und in der Gruppe mit adaptivem Feedback 1,02 (SD 0,66) Punkte erreicht (siehe Tabelle 5).

Es wurde weder in Bezug auf die *Qualität der Begründung* ($F(1,168)=0,818$; $p=0,367$), noch auf die *Diagnoseakkuratheit* ($F(1,168)=0,443$; $p=0,507$), noch auf die *Bewertung der Argumentation* ($F(1,168)=0,001$; $p=0,980$), ein Interaktionseffekt zwischen Feedbackart und Bearbeitungsform gefunden.

3.1.3 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die *Motivation* der Student:innen

Um den Einfluss der vorbestehenden *Motivation* aus dem Prätest und den Einfluss der *Bearbeitungszeit* zu kontrollieren, dienten diese beiden Variablen als Kovariate. Es zeigte sich kein Effekt der Feedbackart auf die *Motivation* der Studierenden, $F(1,166)=0,106$; $p=0,746$. Ebenfalls zeigte sich kein Effekt der Bearbeitungsform auf

die *Motivation* im Posttest, $F(1,166)=2,045$; $p=0,155$. Die Mittelwerte der *Motivation* der einzelnen Gruppen lassen sich aus Tabelle 5 entnehmen.

3.1.4 Untersuchung des Einflusses der Feedbackart und der Bearbeitungsform auf die *kognitive Belastung*

Betrachtet man die *kognitive Belastung*, so zeigte sich bei kleiner Effektstärke ein signifikant höherer Wert in Bezug auf das adaptive Feedback, $F(1,168)=4,105$; $p=0,044$; $\text{partial } \eta^2=0,024$; . Die Bearbeitungsform zeigte keinen Einfluss auf die extrinsische *kognitive Belastung*, $F(1,168)=0,689$; $p=0,408$. Tabelle 5 zeigt die Mittelwerte der einzelnen Gruppen auf.

Tabelle 5

Ergebnisse zu den Effekten von Feedbackart und Bearbeitungsform

	Statisches Feedback		Adaptives Feedback	
	Einzelbedingung	Kollaboration	Einzelbedingung	Kollaboration
	M (SD)	M (SD)	M (SD)	M (SD)
Konzeptuelles Wissen (MC-Fragen)	3,86 (1,36)	3,89 (0,97)	3,86 (0,88)	3,86 (0,98)
	N=29	N=56	N=29	N=58
Strategisches Wissen (Key Features)	4,67 (0,31)	4,72 (0,38)	4,74 (0,27)	4,69 (0,40)
	N=29	N=56	N=29	N=58
Wissensanwendung				
(Post-Test-Fallbearbeitung)				
Qualität der	2,79 (1,45)	2,50 (1,48)	3,48(1,35)	3,72 (1,53)
Begründung	N=29	N=56	N=29	N=58
Diagnoseakkuratheit	0,45 (0,51)	0,54(0,50)	0,62 (0,49)	0,60 (0,49)
	N=29	N=56	N=29	N=58
Bewertung der	0,72 (0,59)	0,77 (0,54)	0,97 (0,57)	1,02(0,66)
Argumentation	N=29	N=56	N=29	N=58
Kognitive Belastung	8,79(2,65)	8,34 (2,18)	9,34 (2,66)	9,55 (2,60)
	N=29	N=56	N=29	N=58
Motivation	13,31 (2,39)	14,24 (1,83)	14,00 (1,54)	13,98 (1,34)
	N=29	N=54	N=29	N=58

3.2 Untersuchung der Effekte von Feedbackart und Vorwissen

3.2.1 Voruntersuchung

Die Semesterzahl allein ist nicht zwangsläufig mit einem gewissen Vorwissenstand gleichzusetzen [56]. Um die Einteilung der Gruppen anhand der Semesterzahl in hohes und niedriges Vorwissen zu validieren, wurden die Ergebnisse aus dem Prätest all jener Teilnehmenden, die die Patientenfälle der Lernphase individuell bearbeitet haben, miteinander verglichen. Hierbei fiel ein signifikanter Unterschied in Bezug auf das *strategische Wissen* auf ($F(1,196)=34,696$; $p<0,001$). Darüber hinaus zeigte sich ein deskriptiver Unterschied bei den Mittelwerten des *konzeptuellen Wissens* (5.Semester: 2,24 (SD: 1,15); ab 8. Semester: 2,52 (SD: 1,01)).

3.2.2 Untersuchung des Einflusses der Feedbackart und des Vorwissens auf das *konzeptuelle* und *strategische Wissen*

Die Varianzanalyse ergab, dass auch in dieser Kohorte die Feedbackart keinen signifikanten Einfluss auf das *konzeptuelle Wissen* ($F(1,194)=0,795$; $p=0,374$) und das *strategische Wissen* ($F(1,194)=1,970$; $p=0,162$) hatte. Hierbei dienten wieder die jeweiligen Ergebnisse aus dem Prätest sowie die *Bearbeitungszeit* als Kovariate. Es zeigte sich ein mittelstarker Effekt im Zusammenhang zwischen Vorwissen und dem Punktwert im *konzeptuellen Wissen*, $F(1,194)=22,177$; $p<0,001$; $\text{partial } \eta^2=0,104$, sowie ein kleiner Effekt auf das *strategische Wissen*, $F(1,187)=5,235$; $p=0,023$; $\text{partial } \eta^2=0,028$.

Die Mittelwerte für das *konzeptuelle Wissen* lagen in der vorwissensstarken Gruppe mit statischem Feedback bei 3,86 (SD 1,36) und bei der vorwissensstarken Gruppe mit adaptivem Feedback bei 3,86 (SD 0,88). In der vorwissensschwachen Gruppe mit statischem Feedback lag der Punktwert im Mittel bei 3,13 (SD 1,21) und bei der Gruppe mit adaptivem Feedback bei 2,81 (SD 1,24) (siehe Tabelle 6). Das *strategische Wissen* wurde in der vorwissensstarken Gruppe mit statischem Feedback im

Mittel mit 4,67 (SD 0,30) und in der vorwissensstarken Gruppe mit adaptivem Feedback mit 4,74 (SD 0,27) Punkten bewertet. Die vorwissensschwache Gruppe erzielte im Mittel mit statischem Feedback 4,48 (SD 0,35) und mit adaptivem Feedback 4,57 (SD 0,34) Punkte (siehe Tabelle 6).

3.2.3 Untersuchung des Einflusses der Feedbackart und des Vorwissens auf die Lösung des Testfalls (Herr Schneider)

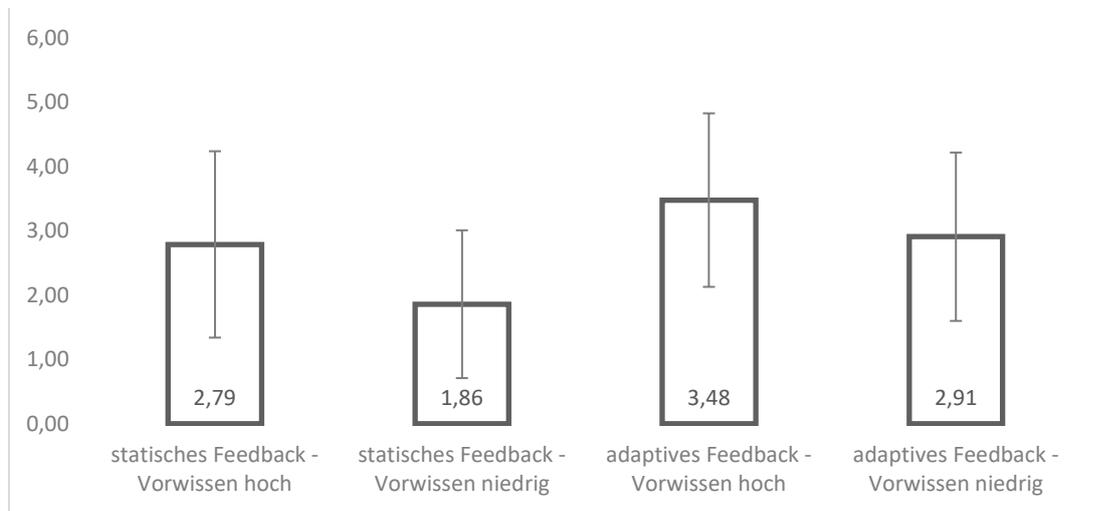
Um den Einfluss vorbestehender Unterschiede in *strategischem Wissen* im Prätest und den Einfluss der *Bearbeitungszeit* zu kontrollieren, dienten diese beiden Variablen als Kovariate.

Ein Haupteffekt des Vorwissens, $F(1,194)=15,676$; $p<0,001$; partial $\eta^2=0,075$ sowie ein Haupteffekt des adaptiven Feedbacks auf die *Qualität der Begründung*, $F(1,194)=18,708$; $p<0,001$; partial $\eta^2=0,089$ konnte festgestellt werden.

Die Mittelwerte für die *Qualität der Begründung* lagen in der vorwissensstarken Gruppe mit statischem Feedback bei 2,79 (SD 1,45) und bei der Gruppe mit adaptivem Feedback bei 3,48 (SD 1,35). In der vorwissensschwachen Gruppe wurde die *Qualität der Begründung* der Gruppe mit statischem Feedback im Mittel mit 1,86 (SD 1,15) und bei der Gruppe mit adaptivem Feedback mit 2,91 (SD 1,31) Punkten bewertet (siehe Tabelle 6). Abbildung 5 zeigt eine graphische Darstellung der Mittelwerte der einzelnen Gruppen (siehe Abbildung 5).

Abbildung 5

Qualität der Begründung in Bezug auf Feedbackart und Vorwissen



Für die *Diagnoseakkuratheit* zeigt sich ein signifikanter Haupteffekt des Vorwissens, $F(1,194)=6,539$; $p=0,011$; $\text{partial } \eta^2=0,033$, der Effekt des adaptiven Feedbacks wurde gefunden, dieser war aber nicht signifikant, $F(1,194)=2,704$; $p=0,102$; $\text{partial } \eta^2=0,014$. Die Mittelwerte für die *Diagnoseakkuratheit* lagen in der vorwissensstarken Gruppe bei der Gruppe mit statischem Feedback bei 0,45 (SD 0,51) und bei der Gruppe mit adaptivem Feedback bei 0,62 (SD 0,49). In der vorwissensschwachen Gruppe lag die *Diagnoseakkuratheit* der Gruppe mit statischem Feedback im Mittel bei 0,27 (SD 0,45) und bei der Gruppe mit adaptivem Feedback bei 0,36 (SD 0,48) (siehe Tabelle 6).

Für die *Bewertung der Argumentation* fand sich lediglich ein Haupteffekt des Vorwissens, $F(1,194)=7,902$; $p=0,005$; $\text{partial } \eta^2=0,040$, die Feedbackart hat hier keinen signifikanten Einfluss auf die *Bewertung der Argumentation* ($F(1,194)=0,336$; $p=0,324$).

Hier lagen die Mittelwerte in der vorwissensstarken Gruppe mit statischem Feedback bei 0,72 (SD 0,59) und mit adaptivem Feedback bei 0,97 (SD 0,57). In der vorwissensschwachen Gruppe wurden im Mittel mit statischem Feedback 0,62 (SD 0,59) und mit adaptivem Feedback 0,57 (SD 0,58) Punkte erreicht (siehe Tabelle 6).

Auch bei dieser Stichprobe wurde weder in Bezug auf die *Qualität der Begründung* ($F(1,194)=0,388$; $p=0,534$), noch auf die *Diagnoseakkuratheit* ($F(1,168)=0,601$; $p=0,439$), noch auf die *Bewertung der Argumentation* ($F(1,168)=7,902$; $p=0,092$) ein Interaktionseffekt zwischen Feedbackart und dem Vorwissen gefunden.

3.2.4 Untersuchung des Einflusses der Feedbackart auf die *Motivation* der Studierenden

Bei der *Motivation* zeigte sich in der ANCOVA, dass das adaptive Feedback einen kleinen positiven Effekt auf diese Variable aufweist, $F(1,183)=4,153$; $p=0,043$; partial $\eta^2=0,022$. Die Mittelwerte lagen hier mit statischem Feedback bei 13,31 (vorwissensstark, SD 2,39) und 12,79 (vorwissensschwach, SD 2,37). Mit adaptivem Feedback lag die *Motivation* im Mittel bei 14,00 (vorwissensstark, SD 1,54) und 13,63 (vorwissensschwach, SD 2,02) (siehe Tabelle 6).

3.2.5 Untersuchung des Einflusses des Vorwissens auf die *kognitive Belastung*

Es fand sich ein Zusammenhang zwischen Vorwissen und kognitiver Belastung. Dabei lag diese bei den vorwissensschwächeren Teilnehmenden aus Erlangen (5. Fachsemester) signifikant über dem der vorwissensstärkeren Teilnehmenden aus München und Würzburg (8. Fachsemester und höher), $F(1,189)=12,154$; $p<0,001$; partial $\eta^2=0,061$.

Die Mittelwerte der einzelnen Gruppen lassen sich aus Tabelle 6 entnehmen.

Die *kognitive Belastung* korrelierte negativ und signifikant mit der *Qualität der Begründung* ($r= -0,251$; $p<0,001$) und der *Diagnoseakkuratheit* ($r= -0,219$; $p=0,002$).

Tabelle 6*Ergebnisse zu den Effekten von Feedbackart und Vorwissensstand (Vergleichsgruppe Erlangen)*

	Statisches Feedback		Adaptives Feedback	
	Vorwissen hoch	Vorwissen niedrig	Vorwissen hoch	Vorwissen niedrig
	M (SD)	M (SD)	M (SD)	M (SD)
Konzeptuelles Wissen (MC-Fragen)	3,86 (1,36)	3,13 (1,21)	3,86 (0,88)	2,81 (1,24)
	N=29	N=71	N=29	N=69
Strategisches Wissen (Key Features)	4,67 (0,30)	4,48 (0,35)	4,74 (0,27)	4,57 (0,34)
	N=29	N=68	N=29	N=65
Wissensanwendung				
(Post-Test-Fallbearbeitung)				
Qualität der	2,79 (1,45)	1,86 (1,15)	3,48 (1,35)	2,91 (1,31)
Begründung	N=29	N=71	N=29	N=69
Diagnoseakkuratheit	0,45 (0,51)	0,27 (0,45)	0,62 (0,49)	0,36 (0,48)
	N=29	N=71	N=29	N=69
Bewertung der	0,72 (0,59)	0,62 (0,59)	0,97 (0,57)	0,57 (0,58)
Argumentation	N=29	N=71	N=29	N=69
Kognitive Belastung	8,79 (2,65)	10,23 (2,37)	9,34 (2,66)	10,71 (2,45)
	N=29	N=69	N=29	N=66
Motivation	13,31 (2,39)	12,79 (2,37)	14,00 (1,54)	13,63 (2,02)
	N=29	N=66	N=29	N=63

3.3 Deskriptive Auswertung der Interviewfragen aus dem Posttest

Die deskriptive Inhaltsanalyse der Interviewfragen am Ende des Posttests erfolgte durch Filterung und Sortierung der Antworten nach den Kategorien „Benutzeroberfläche“, „Virtuelle Patienten“, „Diagnoseprozess“ und „Feedback“. Die Auswertung erbrachte Hinweise darauf, dass das kollaborative Arbeiten sehr großen Anklang fand („Durch das Sprechen fällt es einem auch leichter sich zu orientieren und sich Sachen zu merken, man kann sich gegenseitig auf bestimmte Sachen hinweisen.“; „(Die) Idee mit der Konferenz fand ich super, man konnte gut zusammenarbeiten und diskutieren. Da konnte man wirklich viel mitnehmen“; „Dadurch bekommt man nochmal neuen

Input und wir konnten gut an die Fälle herangehen.“). Des Weiteren wurde die Realitätsnähe der Fallbeispiele („Die Fälle waren realitätsnah, nicht zu komplex aber auch nicht langweilig“), die Möglichkeit des freien Diagnostizierens und die individuelle Rückmeldung in Form des adaptiven Feedbacks am Ende eines jeden Falles („Die Analyse der Antworten war sehr gut. Dass einem gesagt wurde was gut war und genannt wurde und auf was man hätte achten können hat sehr geholfen. Die extra Informationen bei den Analysen hat auch viel Klarheit geschaffen und auch die Beschreibung der letztendlichen Diagnose.“) positiv bewertet.

Nachbesserungsbedarf sahen die Studierenden bei der Gestaltung der Benutzeroberfläche („Vielleicht ein ansprechenderes Design finden, nicht ganz so grau und leer.“) und vereinzelt auch bei der Stabilität der Verbindung mit dem Kollaborationspartner. Darüber hinaus bestand der Wunsch nach einem Ausbau der visuellen Attribute der Falloberfläche („Vielleicht würden Bilder die Fälle noch anschaulicher machen.“).

Bei der Analyse der Antworten der Kohorte der vorwissensschwächeren Studierenden auf die Interviewfragen ließ sich darüber hinaus der Wunsch nach Unterstützung bei der Falllösung erheben („Studenten am Anfang der Klinik bräuchten mehr Information, was einzelne Laborparameter, Medikamente oder Untersuchungen bedeuten. Viele Krankheiten kennt man zu dem Zeitpunkt noch gar nicht, weshalb Thesen zur Diagnose anhand der Symptome sehr schwer zu treffen sind. Die Punkte, welche weiteren Untersuchungen man macht oder wie man weiter vorgeht bzgl. der Behandlung können auch erst in späteren Studienabschnitten beantwortet werden.“).

Als weiterer Hinweis für die hohe Akzeptanz lässt sich außerdem berichten, dass sich an den Standorten München und Würzburg etwa die Hälfte der Studierenden für die fakultative Bearbeitung weiterer Fallbeispiele zum Leitsymptom Rückenschmerz angemeldet haben.

4. Diskussion der Ergebnisse

4.1 Zusammenfassung der Ergebnisse

Die Ergebnisse zeigen, dass sich der erwartete positive Effekt des adaptiven Feedbacks in der Kohorte mit höherem Vorwissen auf die Diagnosekompetenz bzw. deren Teilaspekte *Qualität der Begründung* und *Bewertung der Argumentation* nachweisen lässt. Ebenso zeigte sich ein positiver Haupteffekt des adaptiven Feedbacks in der Kohorte der individuell Lernenden auf die *Qualität der Begründung*.

Außerdem konnte gezeigt werden, dass der positive Effekt des adaptiven Feedbacks unabhängig von der Bearbeitungsform und vom Vorwissen der Lernenden besteht.

In der Kohorte der vorwissensstärkeren Lernenden ging das adaptive Feedback mit einer erhöhten kognitiven Belastung einher. Dies war auch in der Kohorte der individuell Lernenden unterschiedlicher Vorwissenstufen aus München, Würzburg und Erlangen der Fall, in der sich darüber hinaus auch ein positiver Effekt des adaptiven Feedbacks auf die *Motivation* fand.

Gleichzeitig ging in dieser Kohorte ein niedriges Vorwissen erwartungskonform mit einer erhöhten kognitiven Belastung einher.

Entgegen den Erwartungen hatte das adaptive Feedback in den beiden untersuchten Kohorten allerdings weder einen Einfluss auf die Entwicklung des *konzeptuellen* und *strategischen Wissens* noch auf die *Diagnoseakkuratheit*.

Die Erwartung, dass die kollaborative Bearbeitungsform einen positiven Einfluss auf die Diagnosekompetenz der Lernenden haben würde, konnte in dieser Studie ebenfalls nicht bestätigt werden.

4.2 Ergebnisse in Bezug auf die Feedbackart und die Bearbeitungsform

4.2.1 Einfluss der Feedbackart bei Lernenden mit hohem Vorwissen

Das adaptive Feedback hatte -wie erwartet- einen positiven Effekt auf die *Qualität der Begründung*. Dies bedeutet, dass die Herangehensweise an die Lösung des Testfalls bei Proband:innen, die mit adaptivem Feedback gelernt hatten, signifikant besser bewertet wurde. Um die fundierte Herangehensweise an einen Patientenfall zu schulen, ist es unabdingbar, den Studierenden eine inhaltliche Rückmeldung zu ihrer Leistung zu präsentieren, um so ihre Begründungsqualität zu steigern. Adaptivem Feedback gelingt es dabei, alle drei Feedbackaspekte nach Hattie und Timperley (2007) zu vereinen: So wird neben der richtigen Lösung und der Nennung der Ziele (Feed-up) auch Bezug auf die Stärken und Schwächen der einzelnen Lernenden genommen (Feedback) und ein Vorschlag zur Verbesserung der Herangehensweise (Feed-forward) gemacht [57]. Der positive Effekt des adaptiven Feedbacks zeigte sich dabei unabhängig davon, ob kollaborativ oder individuell gelernt wurde.

Auch konnte ein positiver Einfluss des adaptiven Feedbacks auf die *Bewertung der Argumentation* festgestellt werden. Durch die Individualisierung des Feedbacks kann somit auch dieser Aspekt der Diagnosekompetenz gefördert werden.

Auf das *konzeptuelle* und *strategische Wissen* sowie auf die *Diagnoseakkuratheit* scheint die Feedbackart keinen Einfluss zu nehmen. Das Erlernen konzeptuellen Wissens ist ein Prozess, der der Überführung von Informationen aus dem Arbeitsgedächtnis in das Langzeitgedächtnis bedarf. Hierfür sind Wiederholungen unabdingbar. Möglicherweise genügte der zeitliche Rahmen der vorliegenden Studie in Seminarlänge nicht, um diesen Anforderungen gerecht zu werden [58].

In früheren Studien in der gleichen Lernumgebung mit Fällen zu Lernstörungen im Kontext der Lehrer:innenbildung [25], wurde bereits beobachtet, dass adaptives Feedback keinen Effekt auf das *strategische Wissen* hatte. Gruber und Kollegen fanden

schon im Jahr 1998, dass das Lösen von Key-Feature-Fällen vor allem inhalts-spezifisches strategisches Wissen benötigt [59].

Dass sich keine Unterschiede in der Förderung der *Diagnoseakkuratheit* ergeben, könnte darauf zurückzuführen sein, dass dieser Aspekt der Lösung sowohl im statischen als auch im adaptiven Setting als statische Antwort präsentiert wurde. Mit Hilfe von adaptivem Feedback gelingt es zwar, die Hinführung und den Weg zur richtigen Diagnose zu schulen, die Diagnose selbst wird dadurch aber nicht signifikant häufiger richtig gestellt. In der Medizin ist es neben der Stellung der richtigen Diagnose mindestens genauso wichtig, eine Struktur für die Herangehensweise an einen beliebigen Fall zu besitzen. Nur so gelingt es, im Rahmen der Diagnostik nichts zu übersehen. Dadurch wird auch die Fähigkeit geschult, Fälle zu lösen und Patienten zu diagnostizieren, die nicht den Lehrbuchfällen entsprechen, wie es in vorliegender Studie der Fall war.

Unsere Ergebnisse zeigen außerdem, dass das adaptive Feedback mit einer erhöhten *kognitiven Belastung* einhergeht. Es scheint, als ob die Auseinandersetzung mit einer individuellen Rückmeldung kognitiv herausfordernder als eine allgemeine Musterlösung ist.

4.2.2 Einfluss der Bearbeitungsform

In Bezug auf die Bearbeitungsform konnte weder ein Effekt auf das *konzeptuelle* oder *strategische Wissen* noch auf die Teilfertigkeiten der Diagnosekompetenz festgestellt werden.

Da diese Studie als Feldstudie konzipiert war, konnte die Bearbeitung der einzelnen Abschnitte nicht wie im Labor kontrolliert werden. Somit ist unklar, wie die jeweiligen Dyaden mit der Bearbeitung der Lernfälle umgegangen sind und wie die jeweilige Kollaboration genau aussah. Da die Zusammensetzung der Dyaden nach dem Zufallsprinzip geschah, ist davon auszugehen, dass die Kollaborationspartner zum ersten Mal

mit der/dem jeweiligen Gegenüber zusammenarbeiteten und dadurch nicht der erwünschte Wissensaustausch zu einem gemeinsamen Wissenspool, aus dem beide Kollaborationspartner gleichermaßen schöpfen können, erfolgte [60], wodurch der Vorteil der kollaborativen Bearbeitung nicht zur Gänze zum Tragen kam. Des Weiteren lässt sich keine Aussage über den Umgang der Dyaden mit dem adaptiven Feedback treffen. Vor allem ist in dieser Hinsicht nicht eindeutig, inwiefern das Feedback für beide Kollaborationspartner in gleicher Weise adaptiv war. Je nachdem, wie heterolog die Zusammensetzung einer Dyade hinsichtlich des Vorwissens aussah, wurde eine Dysbalance im Profit des Feedbacks begünstigt. Es soll an dieser Stelle aber festgehalten werden, dass auch kein Nachteil der kollaborativen Bearbeitung festgestellt werden konnte. Die Kollaboration wurde zudem in den Freitextantworten der Studierenden positiv bewertet.

Vor diesem Hintergrund sollten zukünftige Studien Prozessdaten während des Trainings sammeln und analysieren, inwiefern die Zusammenarbeit der Dyaden effektiver gestaltet werden könnte, um die Diagnosekompetenz von kollaborativ lernenden Studierenden zu fördern. Hierfür könnte sich etwa eine instruktionale Unterstützung mit Kollaborationsskripts eignen, die den Lernenden bei der kollaborativen Fallbearbeitung beispielsweise Rollen zuweist [61].

Kollaborationsskripts steigern die Leistung beim simulationsbasierten Lernen [62]. Dies zeigt auch ihr erfolgreicher Einsatz in der Notfallmedizin [63].

4.3 Ergebnisse in Bezug auf die Feedbackart und das Vorwissen

4.3.1 Einfluss der Feedbackart bei Lernenden mit unterschiedlichem Vorwissen

Das adaptive Feedback hatte einen positiven Einfluss auf die *Qualität der Begründung*. Dieser Effekt zeigte sich unabhängig vom Vorwissen der Studierenden, was den

Schluss nahelegt, dass dieser Haupteffekt über verschiedene Vorwissenstadien hinweg reproduzierbar ist.

Statisches Feedback erfordert das selbstständige Auseinandersetzen mit der Lösung und den Abgleich der Musterlösung mit der eigenen, während das adaptive Feedback diese Aufgabe übernimmt. Der sich beim Lernen mit statischem Feedback daraus ergebende notwendige Eigenanteil wird nach Krause [64] nur selten erbracht.

Darüber hinaus ließ sich beim Vergleich der individuell Lernenden kein Effekt der Feedbackart auf die *Bewertung der Argumentation*, die *Diagnoseakkuratheit*, sowie die Entwicklung des *strategischen* oder *konzeptuellen Wissens* erkennen.

Die *Motivation* der mit adaptivem Feedback Lernenden lag signifikant über der *Motivation* der Lernenden, welche statisches Feedback erhielten. Individuelle konstruktive Anreize, sowohl durch positives Feedback zur selbstbestimmten Handlung, als auch durch negatives Feedback zur Autonomieunterstützung sind in der Lage, die Lernmotivation zu steigern ([65]; [58]).

4.3.2 Einfluss des Vorwissens

Proband:innen mit niedrigerem Vorwissen wiesen eine signifikant höhere extrinsische und intrinsische *kognitive Belastung* als die vorwissensstärkeren Lernenden auf. Das zeigt, dass es für diese Kohorte kognitiv anstrengender war, sich mit den Fällen und deren Lösung auseinander zu setzen. Eine mögliche Erklärung dafür ist, dass Studierende über die klinische Studienlaufbahn vermehrt mit Fallbeispielen in Berührung kommen und ihnen das Lernen anhand von Simulationen zunehmend vertraut ist. Gleichzeitig fällt eine Falllösung umso leichter, wenn man über ein breites Vorwissen verfügt, da dadurch die Unterscheidung und Wahrnehmung der Informationen eher gelingt [66]. Diese Auswirkung des Vorwissens auf die *kognitive Belastung* konnte bereits von Kiesewetter und Kollegen [10] im Rahmen ihrer Studie zur Untersuchung, wie

virtuelle Patientenfall-Formate und Vorwissen in Bezug auf das Erlernen von klinischer Begründungsqualität interagieren, festgestellt werden.

4.4 Limitationen

Die Studie wurde als Feldstudie konzipiert. Daher konnte nicht alles wie in den vorangegangenen Studienabschnitten im Labor kontrolliert werden. Beispielsweise konnte nicht überprüft werden, wie die jeweilige Kollaboration in den Dyaden genau aussah oder ob bei der Fallbearbeitung eventuell Hilfsmittel zu Rate gezogen wurden oder wie lange die Interaktion in der Lernphase jeweils andauerte. Die vorliegende Studie hat dennoch gezeigt, dass die vorher unter Laborbedingungen erprobte Lehrmethode auch unter Feldbedingungen einsetzbar ist. Außerdem ergibt sich durch die erfolgreiche Durchführung der Studie unter Feldbedingungen eine hohe externe Validität.

Der Feldbedingung geschuldet, kam es bei einzelnen wenigen Kollaborations-Dyaden zu technischen Problemen, die aus der Ferne nicht zu beheben waren. Hierbei konnte zum Beispiel der Bildschirm nicht geteilt werden, die Tonübertragung war nicht ausreichend oder die Internet-Verbindung ist während der Lernintervention abgebrochen. Auch wenn solche Probleme nur vereinzelt auftraten, kann nicht gänzlich ausgeschlossen werden, dass sie einen Einfluss auf die Lernenden in dieser Bedingung hatten.

Des Weiteren war nicht vorgesehen, Prozessdaten in der Lernphase zu erheben. Somit ist im Nachhinein nicht ersichtlich, wie sich mit dem einzelnen Fall auseinandergesetzt wurde, welche Diagnostika wie lange betrachtet wurden und wie lange es bis zur jeweiligen Diagnosestellung gedauert hat.

Gerade die Realisierung der kollaborativen Bearbeitungsform bedarf eines großen technischen Aufwands. Um dies zu verbessern, müsste die Stabilität der Verbindung zwischen den Kollaborationspartnern optimiert werden. In dieser Form der Studierendurchführung war der personell besetzte Chat-Support unabdingbar.

4.5 Implikationen für Theorie und Praxis

Basierend auf unseren Ergebnissen sollte in Zukunft beim simulationsbasierten Lernen vermehrt auf die Form des adaptiven Feedbacks gesetzt werden. Hierbei gilt zu berücksichtigen, dass das Arbeiten mit Fallsimulationen bei vorwissensschwächeren Studierenden mit einer erhöhten kognitiven Belastung einhergeht.

Um vorwissensschwächere Studierende kognitiv zu entlasten, könnte es förderlich sein, bei Bedarf Hinweise bereitzustellen, die die Falllösung unterstützen und lenken [67].

Allerdings fand sich auch für diese Gruppe ein positiver Effekt des adaptiven Feedbacks auf die Diagnosekompetenz bzw. den Teilaspekt der *Qualität der Begründung*. Der Effekt bestand dabei unabhängig und ohne Interaktion zwischen Feedbackart und Vorwissen. Vor diesem Hintergrund erscheint es sinnvoll, auch schon in früheren Semestern der medizinischen Ausbildung auf eine adaptive Form des Feedbacks zu setzen, wenn mit klinischen Fällen gelernt wird.

In der vorliegenden Studie kamen all diejenigen Studierenden als Teilnehmende in Frage, welche sich im klinischen Studienabschnitt befinden. Die Kohorte umfasste also keine Novizen aus den ersten vorklinischen Semestern. Zukünftiger Forschungsgegenstand könnte es somit auch sein, zu untersuchen, ab welchem Zeitpunkt im Medizinstudium der Einsatz solcher Lehrmethoden geeignet ist.

Die Ergebnisse unserer Studie erlauben zum aktuellen Zeitpunkt keine Empfehlung der kollaborativen Bearbeitungsform als bessere Alternative zum individuellen Lernen mit Fällen. Künftige Forschungsvorhaben sollten sich der Frage widmen, wie Lernende bei der kollaborativen Fallbearbeitung instruktional unterstützt werden könnten [68].

5. Zusammenfassung

5.1 Zusammenfassung

Einleitung

Im Medizinstudium hat sich das simulationsbasierte Lernen als eine der effektivsten Lehrmethoden etabliert, um die Diagnosekompetenz Studierender zu fördern [17]. Individualisiertes, adaptives Feedback kann das simulationsbasierte Lernen unterstützen [26], ist aber für Dozierende nur mit großem Ressourcenaufwand realisierbar. Zudem ist die Befundlage zum Vergleich kollaborativer und individueller Fallbearbeitungen derzeit noch unklar [37]. Daher wurde untersucht, inwiefern verschiedene Arten von Feedback (adaptiv vs. statisch) und Formen der Bearbeitung (kollaborativ vs. individuell) einzelner Fallsimulationen Einfluss auf die Diagnosekompetenz Medizinstudierender haben. Darüber hinaus sollte untersucht werden, inwiefern auch vorwissenschwächere Studierende von dem Einsatz simulationsbasierten Lernens profitieren können.

Methoden

Eine experimentelle Feldstudie mit 2x2-faktoriellem Design wurde online mit Medizinstudierenden ab dem 8. Fachsemester an den drei bayerischen Universitäten LMU, TU München und der Universität Würzburg durchgeführt. Alle 172 (121 weiblich, 51 männlich) Teilnehmenden bearbeiteten drei Fallbeispiele mit dem Leitsymptom Fieber in einer simulationsbasierten Lernumgebung. Ein Teil der Lernenden erhielt im Anschluss an jede Fallbearbeitung ein mittels einer Künstlichen-Intelligenz-Methode automatisiertes adaptives Feedback auf ihr diagnostisches Vorgehen; der andere Teil erhielt statisches Feedback in Form einer Musterlösung. Ein Teil der Lernenden bearbeitete die Fälle individuell, der andere Teil mit einem Kollaborationspartner via Screen-Sharing und Sprachchat. Nach Bearbeitung der Fallbeispiele lösten alle

Studierenden einen Testfall individuell und ohne Feedback. Die Lösung des Testfalls wurde herangezogen, um die Diagnosekompetenz der Lernenden anhand der Aspekte (1) Qualität der Begründung, (2) Bewertung der Argumentation und (3) Diagnoseakkuratheit einzuschätzen. Die Studienteilnahme dauerte insgesamt ca. 90 Minuten.

Ein weiteres 2x2-faktorielles Design ergab sich durch die Durchführung der Studie am Universitätsstandort Erlangen. Hier nahmen insgesamt 140 Studierende teil, die sich allesamt im 5. Fachsemester befanden und die Fallbeispiele lediglich individuell bearbeiteten. Aufgrund dieser Tatsache ergab sich eine Vergleichsgruppe mit –der geringeren Semesterzahl geschuldet- geringerem Vorwissen zur Kohorte derjenigen Studierenden, welche sich bereits mindestens im 8. Fachsemester befanden und welche die Patientenfälle ebenfalls individuell bearbeiteten.

Ergebnisse

Das adaptive Feedback hatte einen signifikanten positiven Effekt mittlerer Stärke auf die Qualität der Begründung ($F(1,168)=15,624$; $p<0,001$, $\eta^2p=0,086$) und einen signifikanten positiven Effekt kleiner Stärke auf die Bewertung der Argumentation ($F(1,168)=6,872$; $p=0,010$; $\eta^2p=0,040$), nicht jedoch auf die Diagnoseakkuratheit ($F(1,168)=20,019$; $p=0,157$). Die Bearbeitungsform hatte keinen Effekt auf die drei Teilaspekte der Diagnosekompetenz und es zeigten sich keine Interaktionseffekte zwischen Bearbeitungsform und Feedbackart.

Bei der Kohorte der individuell Lernenden konnte ein signifikanter negativer Effekt des Vorwissens sowohl auf die Qualität der Begründung ($F(1,194)=15,676$; $p<0,001$; $\eta^2p=0,075$), als auch auf die Bewertung der Argumentation ($F(1,194)=7,902$; $p=0,005$; $\eta^2p=0,040$) und die Diagnoseakkuratheit ($F(1,194)=6,539$; $p=0,011$; $\eta^2p=0,033$) erhoben werden. Darüber hinaus zeigte sich auch in dieser Kohorte ein signifikanter positiver Effekt großer Stärke auf die Qualität der Begründung ($F(1,194)=18,708$; $p<0,001$;

$\eta^2p=0,089$). Zudem zeigte sich kein Interaktionseffekt zwischen der Feedbackart und dem Vorwissen.

Diskussion

Die Ergebnisse machen deutlich, dass adaptives Feedback durch das Aufzeigen individueller Stärken und Schwächen beim Diagnostizieren das Potenzial hat, eine fundierte Herangehensweise an Patientenfälle zu fördern. Es ist dabei in der Lage, die drei Aspekte von Feedback, nämlich Feed-up, Feed-back und Feed-forward [57] zu vereinen. Erfolgt das Feedback automatisiert, ist es zugleich kompetenzfördernd und ressourcenschonend. Der positive Effekt der Adaptivität zeigte sich unabhängig von der Bearbeitungsform, die keinen Einfluss auf die Diagnosekompetenz hatte. Außerdem bleibt der Effekt auch bei vorwissenschwächeren Studierenden bestehen, wobei das Vorwissen selbst sehr wohl Einfluss auf die Leistung im Testfall hatte.

In künftigen Studien sollte beforscht werden, inwiefern das kollaborative Lernen mit Fällen instruktional unterstützt werden kann. Dass der Einsatz von Kollaborationskripts im Kontext des simulationsbasierten Lernens vielversprechend sein kann, konnte bereits gezeigt werden [63].

5.2 Summary

Introduction

In medical studies, simulation-based learning has established itself as one of the most effective teaching methods to promote students' diagnostic reasoning skills [17]. Individualized, adaptive feedback can support simulation-based learning [26], but can only be implemented by lecturers with a great amount of resources. In addition, the findings on the comparison of collaborative and individual case-work are currently still unclear [37]. Therefore, it was examined to what extent different types of feedback (adaptive vs. static) and learning mode (collaborative vs. individual) of case simulations influence

the diagnostic competence of medical students. In addition, it should be examined to what extent students with low prior knowledge can benefit from the use of simulation-based learning.

Methods

An experimental field study with a 2x2 factorial design was conducted online with medical students from the 8th semester onwards at the three Bavarian universities LMU, TU Munich and the University of Würzburg. All 172 (121 female, 51 male) participants worked on three virtual patients with the leading symptom fever in a simulation-based learning environment. After each virtual patient case, one part of the learners received automated adaptive feedback on their diagnostic procedure using an artificial intelligence method; the other part received static feedback in the form of a sample solution. One part of the learners worked on the cases individually, others with a collaboration partner via screen sharing and voice chat. After processing the case, all students solved a test case individually and without feedback. The solution of the test case was used to assess the learners' diagnostic competence based on the aspects (1) quality of justification, (2) evaluation of argumentation and (3) diagnostic accuracy. Participation in the study lasted about 90 minutes in total. Another 2x2-factorial design resulted from conducting the study at the University of Erlangen. A total of 140 students took part, all of whom were in their 5th semester and only worked on the virtual patients individually. This resulted in a comparison group with - due to the lower number of semesters - less prior knowledge to the cohort of those students who were at least in their 8th semester and who also worked on the patient cases individually.

Results

The adaptive feedback had a significant positive effect of medium strength on the quality of justification ($F(1,168)=15.624$; $p<.001$, $\eta^2p=.086$) and a significant positive effect of small strength on the evaluation of reasoning ($F(1,168)=6.872$; $p=.010$; $\eta^2p=.040$),

but not on diagnostic accuracy ($F(1,168)=2.019$; $p=.157$). The form of processing had no effect on the three aspects of diagnostic competence and there were no interaction effects between the learning mode and the type of feedback. In the cohort of individual learners, a significant negative effect of prior knowledge was found both on the quality of the justification ($F(1,194)=15,676$; $p<.001$; $\eta^2p=.075$), on the evaluation of the argumentation ($F(1,194) =7.902$; $p=.005$; $\eta^2p=.040$) and diagnostic accuracy ($F(1,194)=6.539$; $p=.011$; $\eta^2p=.033$). In addition, this cohort also showed a significant positive effect of large strength on the quality of the justification ($F(1,194)=18.708$; $p<.001$; $\eta^2p=.089$). In addition, there was no interaction effect between the type of feedback and prior knowledge.

Discussion

The results clearly demonstrate that adaptive feedback has the potential to promote an informed approach to patient cases by highlighting individual strengths and weaknesses in diagnosis. It combines the three aspects of feedback, namely feed-up, feedback and feed-forward [57]. If the feedback is automated, it promotes competence and conserves resources at the same time. The positive effect of the adaptivity was shown independently of the form of processing, which had no influence on the diagnostic competence. In addition, the effect also persists for students with weaker prior knowledge, although the prior knowledge itself did have an influence on the performance in the test case. Future studies should investigate the extent to which collaborative learning can be instructively supported with cases. It has already been shown that the use of collaboration scripts in the context of simulation-based learning can be promising [63].

6. Literatur

1. Bansal, A., et al., *Developing Medical Students' Broad Clinical Diagnostic Reasoning Through GP-Facilitated Teaching in Hospital Placements*. Adv Med Educ Pract, 2020. **11**: p. 379-388.
2. Williams, R.G., et al., *Variations in senior medical student diagnostic justification ability*. Acad Med, 2014. **89**(5): p. 790-8.
3. Ark, T.K., L.R. Brooks, and K.W. Eva, *The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies*. Med Educ, 2007. **41**(3): p. 281-7.
4. Mamede, S., et al., *How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases*. Acad Med, 2014. **89**(1): p. 121-7.
5. Eva, K.W., *What every teacher needs to know about clinical reasoning*. Med Educ, 2005. **39**(1): p. 98-106.
6. Fakultätentag, M., *Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM)*. 2021.
7. Heitzmann, N., et al., *Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education*. FRONTLINE LEARNING RESEARCH, 2019. **7**: p. 1-24.
8. Fischer, M.R., *CASUS - An authoring and learning tool supporting diagnostic reasoning*. Zeitschrift für Hochschuldidaktik, 2000. **1**(1): p. 87-98.
9. Stegmann, K., Fischer, F., *Kurzbericht „Auswirkungen digitaler Medien auf den Wissens- und Kompetenzerwerb an der Hochschule“*, in *Beitrag im Rahmen eines Expertenhearings des Wissenschaftsrates*. 2016.
10. Kiesewetter, J., et al., *Learning clinical reasoning: how virtual patient case format and prior knowledge interact*. BMC Medical Education, 2020. **20**(1): p. 73.
11. Wisniewski, B., K. Zierer, and J. Hattie, *The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research*. Frontiers in Psychology, 2020. **10**.
12. Hautz, W.E., et al., *Diagnostic performance by medical students working individually or in teams*. Jama, 2015. **313**(3): p. 303-4.
13. Kiesewetter, J., F. Fischer, and M.R. Fischer, *Collaborative Clinical Reasoning-A Systematic Review of Empirical Studies*. J Contin Educ Health Prof, 2017. **37**(2): p. 123-128.
14. Radkowsch, A., et al., *Learning to diagnose collaboratively: validating a simulation for medical students*. GMS J Med Educ, 2020. **37**(5): p. Doc51.
15. Kühl, S.J., et al., *Investigating the self-study phase of an inverted biochemistry classroom - collaborative dyadic learning makes the difference*. BMC Med Educ, 2019. **19**(1): p. 64.
16. Tolks, D., Kuhn S., Kaap-Fröhlich, S., *Lehre in Zeiten von COVID-19. Herausforderungen und Chancen für die digitale Lehre*. GMS J Med Educ, 2020. **37**(7): p. Doc 103.
17. Chernikova, O., et al., *Simulation-Based Learning in Higher Education: A Meta-Analysis*. Review of Educational Research, 2020. **90**(4): p. 499-541.
18. Codreanu, E., et al., *Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education*. Teaching and Teacher Education, 2020. **95**: p. 103146.
19. Fischer, M., et al., *Virtuelle Patienten in der medizinischen Ausbildung: Vergleich verschiedener Strategien zur curricularen Integration*. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2008. **102**: p. 648-653.

20. Hege, I., et al., *Experiences with different integration strategies of case-based e-learning*. Med Teach, 2007. **29**(8): p. 791-7.
21. Ziv, A., et al., *Simulation-based medical education: an ethical imperative*. Acad Med, 2003. **78**(8): p. 783-8.
22. Cook, D.A., et al., *Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis*. Simul Healthc, 2012. **7**(5): p. 308-20.
23. Renkl, A., *Toward an Instructionally Oriented Theory of Example-Based Learning*. Cognitive science, 2014. **38** 1: p. 1-37.
24. Oliver Scheuer, B.M.M., Frank Loll, Niels Pinkwart, *Automated Analysis and Feedback Techniques to Support and Teach Argumentation: A Survey*, in *Educational Technologies for Teaching Argumentation Skills*. 2012, Bentham Science Publisher. p. 71-124.
25. Sailer, M., et al., *Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning*. Learning and Instruction, 2022: p. 101620.
26. Plass, J. and S. Pawar, *Toward a taxonomy of adaptivity for learning*. Journal of Research on Technology in Education, 2020. **52**: p. 275-300.
27. Bimba, A.T., et al., *Adaptive feedback in computer-based learning environments: a review*. Adaptive Behavior, 2017. **25**(5): p. 217-234.
28. Narciss, S., et al., *Exploring feedback and student characteristics relevant for personalizing feedback strategies*. Computers & Education, 2014. **71**: p. 56-76.
29. Hughes, S., Hastings, P., Britt, M. A., Wallace, P., Blaum, D. *Machine Learning for Holistic Evaluation of Scientific Essays*. in *17th International Conference on Artificial Intelligence in Education (AIED'15)*. 2015.
30. Zhu, M., et al., *Investigating the impact of automated feedback on students' scientific argumentation*. International Journal of Science Education, 2017. **39**(12): p. 1648-1668.
31. Zhu, M., O.L. Liu, and H.-S. Lee, *The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing*. Computers & Education, 2020. **143**: p. 103668.
32. Zottmann, J., et al., *Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching experience on knowledge convergence*. Computers in Human Behavior, 2013. **29**: p. 2100–2108.
33. Weidenbusch, M., et al., *Can clinical case discussions foster clinical reasoning skills in undergraduate medical education? A randomised controlled trial*. BMJ Open, 2019. **9**(9): p. e025973.
34. Koenemann, N., et al., *Clinical Case Discussions - a novel, supervised peer-teaching format to promote clinical reasoning in medical students*. GMS J Med Educ, 2020. **37**(5): p. Doc48.
35. Csanadi, A., I. Kollar, and F. Fischer, *Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: better together?* European Journal of Psychology of Education, 2021. **36**(1): p. 147-168.
36. Tolsgaard, M.G., K.M. Kulasegaram, and C.V. Ringsted, *Collaborative learning of clinical skills in health professions education: the why, how, when and for whom*. Med Educ, 2016. **50**(1): p. 69-78.

37. Cook, D.A., et al., *Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis*. Med Teach, 2013. **35**(1): p. e867-98.
38. Weinberger, A., K. Stegmann, and F. Fischer, *Learning to argue online: Scripted groups surpass individuals (unscripted groups do not)*. Computers in Human Behavior, 2010. **26**(4): p. 506-515.
39. Harold F. O'Neil, R.S.P., *Role of task-specific adapted feedback on a computer-based collaborative problem-solving task*, in *Web-based learning: Theory, research, and practice*. 2012, New York: Routledge. p. 239-254.
40. Ausubel, D.P., *Educational psychology: a cognitive view*. Educational psychology: a cognitive view. 1968, Holt, Rinehart and Winston: New York.
41. Simonsmeier, B., et al., *Domain-Specific Prior Knowledge and Learning: A Meta-Analysis*. Educational Psychologist, 2021.
42. Kirschner, P.A., J. Sweller, and R.E. Clark, *Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching*. Educational Psychologist, 2006. **41**(2): p. 75-86.
43. Kolodner, J.L., *An introduction to case-based reasoning*. Artificial Intelligence Review, 1992. **6**(1): p. 3-34.
44. Schmidt, H.G., et al., *Problem-Based Learning is Compatible with Human Cognitive Architecture: Commentary on Kirschner, Sweller, and*. Educational Psychologist, 2007. **42**(2): p. 91-97.
45. Dong, A., M.S.-Y. Jong, and R.B. King, *How Does Prior Knowledge Influence Learning Engagement? The Mediating Roles of Cognitive Load and Help-Seeking*. Frontiers in Psychology, 2020. **11**.
46. Bundesamt, S. *Studierende insgesamt und Studierende Deutsche im Studienfach Medizin (Allgemein-Medizin) nach Geschlecht*. 2021; Available from: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/lrbil05.html>.
47. Bauer, E., et al., *Learning to Diagnose Students' Behavioral, Developmental, and Learning Disorders in a Simulation-Based Learning Environment for Pre-Service Teachers*, in *Learning to Diagnose with Simulations : Examples from Teacher Education and Medical Education*, F. Fischer and A. Opitz, Editors. 2022, Springer International Publishing: Cham. p. 97-107.
48. Pfeiffer, J., et al., *FAMULUS: Interactive Annotation and Feedback Generation for Teaching Diagnostic Reasoning*. ArXiv, 2019. **abs/1908.11254**.
49. Sailer, M., et al., *Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning*. Learning and Instruction, 2023. **83**: p. 101620.
50. Jung, V.M., *Die Auswirkung des Falldarbietungsformats und des Vorwissens auf diagnostische Kompetenzen bei Medizinstudierenden*, in *Medizinischen Fakultät*. 2019, Ludwig-Maximilians-Universität München.
51. Wigfield, A., *Expectancy-value theory of achievement motivation: A developmental perspective*. Educational Psychology Review, 1994. **6**(1): p. 49-78.
52. Opfermann, M., *There's more to it than instructional design. The role of individual learner characteristics for hypermedia learning*. Wissensprozesse und digitale Medien. 11. 2008, Berlin: Logos-Verl. 219, [64] S.

53. Hemmerich, W.A. *Einfaktorielle ANCOVA: Ergebnisse interpretieren und berichten*. 2023; Available from: <https://statistikguru.de/spss/einfaktorielle-ancova/ergebnisse-interpretieren-berichten.html>.
54. Jürgen Bortz, C.S., *Statistik für Human- und Sozialwissenschaftler*. 7 ed. Springer-Lehrbuch. 2010.
55. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. Hoboken: Taylor and Francis. 1988.
56. Zottmann, J.M., et al., *Fostering the analytical competency of pre-service teachers in a computer-supported case-based learning environment: A matter of perspective?* Interactive Learning Environments, 2012. **20**(6): p. 513-532.
57. Hattie, J. and H. Timperley, *The Power of Feedback*. Review of Educational Research, 2007. **77**(1): p. 81-112.
58. Mareike Kunter, U.T., *Psychologie des Unterrichts*. Vol. 1. 2013: Brill | Schöningh.
59. Gruber, T.R., *The acquisition of strategic knowledge*. 1989: Academic Press Professional, Inc.
60. Janssen, J. and P.A. Kirschner, *Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda*. Educational Technology Research and Development, 2020. **68**(2): p. 783-805.
61. Fischer, F., et al., *Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning*. Educational psychologist, 2013. **48**(1): p. 56-66.
62. Radkowitz, A., et al., *Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations*. Learning and Instruction, 2021. **75**: p. 101487.
63. Zottmann, J.M., et al., *Just watching is not enough: Fostering simulation-based learning with collaboration scripts*. GMS journal for medical education, 2018. **35**(3): p. Doc35-Doc35.
64. Krause, U.M., *Feedback und kooperatives Lernen*. 2007: Waxmann Verlag.
65. Deci, E.L.R., Richard M., *Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik*, in *Zeitschrift für Pädagogik*. 1993. p. 223-238.
66. Regina Schönberger, V.J., Michael Sailer, Frank Fischer, Martin Fischer, Jan Kiesewetter. *Evaluation of virtual case simulations considering the extraneous cognitive load and emotions of medical students*. in *AMEE*. 2018. Basel.
67. Chernikova, O., et al., *Facilitating Diagnostic Competences in Higher Education—a Meta-Analysis in Medical and Teacher Education*. Educational Psychology Review, 2020. **32**(1): p. 157-196.
68. Goeze, A., et al., *Getting immersed in teacher and student perspectives? Facilitating analytical competence using video cases in teacher education*. Instructional Science, 2014. **42**.

7. Anlagen

7.1 Anschreiben am Beispiel Standort Universität Würzburg

Liebe Medizinstudierende,

gerne möchte ich Sie heute auf eine spannende Studie hinweisen, die wir im Wintersemester 2019/20 mit den KollegInnen der LMU München durchführen: das Projekt FAMULUS geht der Frage nach, wie Diagnosekompetenzen von Studierenden mit Hilfe von Online-Simulationen gefördert werden können. Für diese Untersuchung sind wir derzeit auf der Suche nach TeilnehmerInnen.

Sie möchten Ihre eigenen Diagnose-Fähigkeiten schärfen? Sie wollen Ihr internistisches Wissen mit authentischen Patientenfällen vertiefen? Dann würden wir uns sehr freuen, wenn Sie sich über folgende eMail-Adresse für einen der nachstehenden Termine anmelden:

Beispieltermin 1

oder

Beispieltermin 2

Bereits im Vorfeld zu dem von Ihnen gewählten Termin werden Sie den ersten Abschnitt der Studie (ca. 20 Minuten) bearbeiten, um sich mit der online-Umgebung vertraut zu machen. Die eigentliche Fallsimulation findet dann zum gewählten Zeitpunkt statt. Sie werden hierbei die Fälle gemeinsam mit einem Kommilitonen über Sprach-Chat bearbeiten. Im Anschluss daran folgt noch ein letzter Abschnitt, welchen Sie dann wieder einzeln durchlaufen werden

Die Teilnahme an der Studie nimmt insgesamt ca. 90 Minuten in Anspruch – benötigt wird ein Computer mit Internetzugang, ein Headset oder Mikrofon am Endgerät, sowie der Browser Mozilla Firefox.

Die bei dieser Studie erhobenen Daten werden anonym ausgewertet, es können keine Rückschlüsse auf Ihre Person gezogen werden. Die erhobenen Daten werden außerdem ausschließlich für Forschungszwecke erhoben und in keinem Fall an Dritte weitergegeben.

Bei dieser Studie werden die Vorschriften über den Datenschutz eingehalten. Es werden die Daten nur in verschlüsselter Form gespeichert, das heißt weder Ihr Name, noch Ihre Initialen oder das exakte Geburtsdatum erscheinen im Verschlüsselungscode.

Durch Ihre Teilnahme erklären Sie sich mit dieser Art der Datenverwertung einverstanden.

Dieses Einverständnis können Sie jederzeit widerrufen. Aus dem Widerruf entsteht Ihnen kein Nachteil.

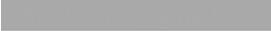
Bitte unterstützen Sie das FAMULUS-Projekt und helfen Sie uns dabei, die medizinische Hochschullehre zu verbessern.

Mit freundlichen Grüßen,

7.2 Begrüßungsmail

Herzlich Willkommen bei FAMULUS!

Sie haben sich für die Erhebung am **Beispieldatum** angemeldet. Nachfolgend finden Sie alle wichtigen Informationen dazu.

Sollten nach dem Durchlesen dieser Mail dennoch Fragen offen bleiben, so stehen wir Ihnen gerne unter  mit Rat und Tat zur Seite.

Ziel dieser Studie ist es, Ihre Diagnosekompetenz Leitsymptom-orientiert anhand von Online-Fallsimulationen zu schulen. Um die Durchführung so reibungslos wie möglich zu gestalten folgen anschließend ein paar wichtige Hinweise:

Für die Studienteilnahme benötigen Sie zwingend den Browser **Mozilla Firefox (R)**! Bitte achten Sie darauf, dass dieser auf Ihrem Computer installiert ist.

Bitte bearbeiten Sie bereits jetzt oder zumindest im **Vorfeld** zu dem von Ihnen gewählten Termin den ersten der drei Studienabschnitte (Abschnitt **A**) um sich mit der Online- Umgebung vertraut zu machen und die Voraussetzung für die Hauptsimulation zu erfüllen (Dauer ca. 15 Minuten).

Geben Sie dazu bitte die unten stehende PIN auf <https://talky.casus.net> (Das `https://` ist wichtig!) ein.

Für die Abschnitte B und C finden Sie sich bitte dann am **Beispieldatum** erneut an einem internetfähigen Computer ein und melden sich unter oben genannter URL erneut mit Ihrer PIN an.

Abschnitt **B** werden Sie gemeinsam mit einem Kommilitonen bearbeiten. Die Kommunikation erfolgt dabei über Sprach-Chat und Sie teilen sich über Screen-Sharing einen Bildschirm.

Sollte Ihre PIN **fett** gedruckt sein, so werden Sie Ihren Bildschirm freigeben und die Navigation übernehmen. Um Ihnen das Prinzip der Vernetzung Ihrer beiden Computer zu erläutern, finden Sie dieser eMail einen „Leitfaden zum kooperativen Arbeiten“ als PDF angehängt.

Wichtig: Falls Ihr Kommilitone zum vereinbarten Termin nicht innerhalb von **7** Minuten erscheint, müssen Sie nicht länger auf ihn warten. Loggen Sie sich aus und kontaktieren Sie uns einfach unter Angabe Ihrer alten PIN über das in die Online-Umgebung eingebaute Live-Chat-Tool. Sie erhalten umgehend eine neue PIN und können die Falle alleine bearbeiten.

Das gleiche gilt, falls Sie sich erst nach dieser Frist einloggen wollen. Ihre PIN verfällt und Sie erhalten von uns ebenfalls umgehend nach Kontaktaufnahme eine Ersatz-PIN zur individuellen Sitzung. Leider ist es nicht möglich, Wünsche zu Kooperationspartnern zu berücksichtigen. Die Zuteilung zu den Gruppen erfolgt randomisiert.

Ihre PIN lautet: **XXXXXX**

Nun wünschen wir Ihnen ganz viel Spaß und einen großen Lernerfolg beim Bearbeiten der Online-Fälle!

Das gesamte FAMULUS-Team

7.3 Leitfaden zum kollaborativen Arbeiten



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN



Herzlich Willkommen bei



Dieses Dokument soll Ihnen als
Leitfaden für das kooperative
Bearbeiten der Fälle in Abschnitt B
dienen.

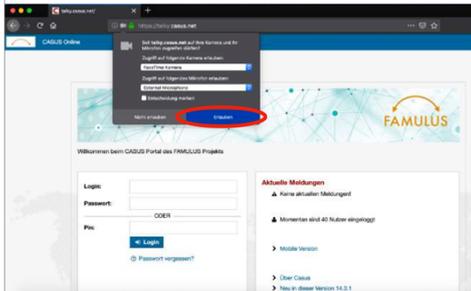
Öffnen Sie
Mozilla Firefox (R)



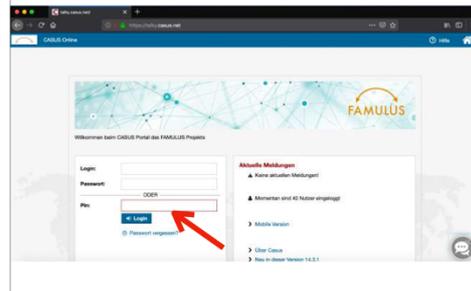
Und begeben Sie sich auf

<https://talky.casus.net>

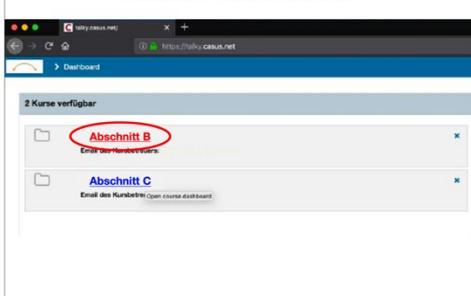
Um zusammen arbeiten zu können benötigt CASUS die Erlaubnis, auf Ihr Mikrofon zugreifen zu dürfen.



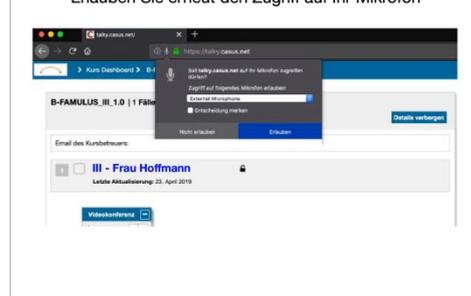
Geben Sie anschließend Ihren persönlichen Zugangscode ein und loggen Sie sich ein



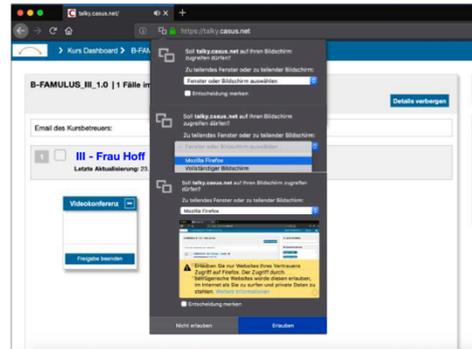
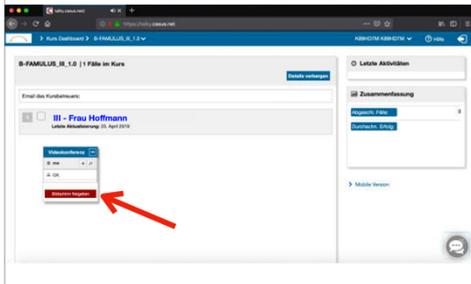
Wählen Sie den Hauptkurs aus



Erlauben Sie erneut den Zugriff auf Ihr Mikrofon

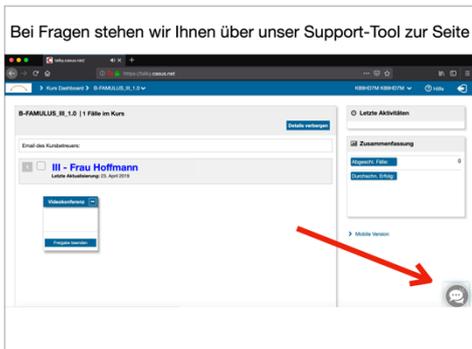


Und geben Sie abschließend - sobald Ihr Kooperationspartner online ist - Ihren Bildschirm frei



Sie sind nun mit Ihrem Kooperationspartner verbunden und können die Fälle gemeinsam bearbeiten.

Dabei teilen Sie sich einen Bildschirm.



Viel Spaß!

8. Eidesstaatliche Versicherung

Sarbu-Rothsching, Sven-Markus Christian Lorenz

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema **„Einfluss von adaptivem Feedback und kollaborativer Bearbeitung beim Lernen mit virtuellen Patientinnen und Patienten“** selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 05.02.2024

Ort, Datum

Sarbu-Rothsching Sven-Markus

Unterschrift Doktorandin/Doktorand