

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Zentrum für Innere Medizin
III. Medizinische Klinik und Poliklinik
Direktor: Prof. Dr. med. Tobias B. Huber

Analyse der „Clinical Reasoning“-Kompetenz von fortgeschrittenen Medizinstudierenden in simulierten Anamnesegesprächen

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von

Tillmann Helm
aus Kiel

Hamburg 2020

Angenommen von der Medizinischen Fakultät der Universität Hamburg am: 22.06.2021

Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der Vorsitzende: Prof. Dr. Sven Anders

Prüfungsausschuss, zweite Gutachterin: Prof. Dr. Sigrid Harendza

Inhaltsverzeichnis

1. Einleitung	5
2. Arbeitshypothese und Fragestellung	10
3. Material und Methoden	11
3.1 360-Grad-Kompetenzprüfung	11
3.2 Videoaufzeichnungen und Patientenfälle	11
3.3 Studierendekollektiv	12
3.4 Entwicklung des Bewertungsbogens für „Clinical Reasoning“-Fähigkeiten	12
3.5 Pilottestung des Bewertungsbogens	13
3.6 Analyse der Anamnesevideos	14
3.7 Statistische Analyse	15
4. Ergebnisse	16
4.1 Ergebnisse der Pilottestung des Bewertungsbogens	16
4.2 Vergleich der „Clinical Reasoning“-Fähigkeiten nach Studienfortschritt auf Ebene aller Patientengespräche	17
4.3 Vergleich der „Clinical Reasoning“-Fähigkeiten zwischen den Universitäten auf Ebene aller Patientengespräche	18
4.4 Vergleich der „Clinical Reasoning“-Fähigkeiten auf Ebene einzelner Patientengespräche	20
4.4.1 Patientengespräch 1	20
4.4.2 Patientengespräch 2	21
4.4.3 Patientengespräch 3	21
4.4.4 Patientengespräch 4	22
4.4.5 Patientengespräch 5	23
5. Diskussion	25
6. Zusammenfassung	31
7. Summary	32
8. Abkürzungsverzeichnis	33
9. Tabellenverzeichnis	34
10. Literaturverzeichnis	35

11. Anhang.....	40
12. Lebenslauf	42
13. Danksagung.....	43

1. Einleitung

Die Ausbildung Medizinstudierender befindet sich aktuell im Wandel. An immer mehr medizinischen Fakultäten in Deutschland und in anderen europäischen Ländern werden vertikal-integrierte Studiengänge eingeführt. Das Ziel dieser neuen Curricula ist es, die angehenden Medizinerinnen und Mediziner möglichst gut auf ihre späteren klinischen Tätigkeiten vorzubereiten. Besonders durch die frühe Verbindung der vorklinischen, biomedizinischen Ausbildung mit klinischen Erfahrungen, sollen die Studierenden nach ihrem Abschluss bereits bessere klinische Fähigkeiten entwickelt haben (Wijnen-Meijer et al. 2015). Jeder Patientenkontakt einer Ärztin oder eines Arztes ist umfangreich und geht mit einer Vielzahl von unterbewussten Wahrnehmungen und kognitiven Prozessen einher (Charlin et al. 2007). Kahneman (2011) unterteilt diese Gedankenprozesse in zwei Wege, den intuitiven oder auch den schnellen Weg und den analytischen bzw. den langsamen Weg. In einem Anamnesegespräch gehören zu diesen Prozessen das Erkennen der Gesprächssituation, das schnelle Erstellen von einer ersten Hypothese und das Abgleichen dieser mit vom Patienten ausgehenden Zeichen und Symptomen. Anschließend kann die Hypothese dann möglicherweise bestätigt oder aber verworfen werden. Zusammen mit dem erlernten Wissen und der klinischen Erfahrung wird schlussendlich eine passende Diagnostik und Behandlung eingeleitet. Die Gesamtheit dieses beschriebenen kognitiven Prozesses bezeichnet man als „Clinical Reasoning“ (Charlin et al. 2007).

„Clinical Reasoning“ (klinisches Argumentieren) ist also der Gedanken- und Entscheidungsfindungsprozess, der während einer klinischen Interaktion mit Patientinnen und Patienten oder bei der Diskussion über Patientengeschichten durchlaufen wird (Thampy et al. 2019). Er hilft Ärztinnen und Ärzten zum Beispiel dabei, während einer fokussierten Anamnese Patientendaten zu erfassen und zu analysieren (Charlin et al. 2007). So ist es möglich, im klinischen Alltag argumentativ gut begründete Entscheidungen zu treffen und dabei die spezifische Situation, in deren Kontext die jeweilige Entscheidung getroffen wird, zu berücksichtigen. „Clinical Reasoning“ beinhaltet viele verschiedene Kompetenzen wie Entscheidungsvermögen, Verständnis und Erfahrung (Durning et al. 2011). Diese sind alle abhängig von dem Fachwissen der jeweiligen Person. Dabei ist es schwierig den Einfluss von Wissen, Erfahrung und eventuellen Schlussfolgerungen in einem Anamnesegespräch voneinander abzugrenzen. Zudem ist „Clinical Reasoning“ auch in Bezug auf einzelne Personen, die es anwenden, sehr variabel in unterschiedlichen Situationen und somit vom Kontext abhängig (Durning et al. 2011). Aus der Psychologie stammen ähnliche Untersuchungen, welche die Gedanken während eines Entscheidungsprozesses kategorisieren. Neben den von Kahneman definierten zwei Wegen des Denkens, gibt es die

so genannte „Dual Processing“-Theorie (Norman 2009). Diese beschreibt, dass Entscheidungen mit Hilfe von zwei Systemen getroffen werden können. Zum einen durch einen schnellen, unbewussten, kontextbezogenen Prozess (System 1), hier werden direkte Assoziationen zwischen einer neuen Information und bereits im Gedächtnis verankerten Beispielen hergestellt. Zum anderen durch einen langsamen, bewussten, konzeptionellen Prozess (System 2), bei welchem das Arbeitsgedächtnis aktiv wird und beispielsweise nach Mustern von bereits bekannten und als wahrscheinlich eingestuften Erkrankungen sucht (Norman et al 2017). Beim „Clinical Reasoning“ laufen beide Prozesse parallel bzw. nacheinander ab. Die Bedeutung des Zusammenspiels der beiden Systeme wird durch die Tatsache untermauert, dass klinisch tätige Ärztinnen und Ärzte zumeist die Anamnese bereits abschließen, nachdem sie nur 60-70% der wichtigen, zu erfassenden Daten gesammelt haben. Dies führt jedoch nicht dazu, dass später mit einer geringeren Wahrscheinlichkeit die richtige Diagnose gestellt wird (Norman 2009).

Die bei klinischen Expertinnen und Experten unterbewusst ablaufenden Gedankengänge sind bei Studierenden noch nicht so ausgeprägt. Bei ihnen kommt es eher zu einem bewussten Abgleichen von Symptomen und erlernten Krankheitsbildern. Mit vermehrtem Patientenkontakt erfolgt das Anwenden von „Clinical Reasoning“ immer intuitiver (Norman 2009), also auf dem schnellen Denkweg, und gipfelt in der Entwicklung von sogenannten „Illness scripts“ (Klein et al. 2019.). Bei diesen handelt es sich um spezifische Konzepte im Kopf von Medizinerinnen oder Medizinern, die es ihnen ermöglichen, generelles Wissen über ein Krankheitsbild mit persönlichen Erfahrungen und Erinnerungen an einzelne Patientenbegegnungen zu vereinen und diese jederzeit als innere Repräsentation eines Krankheitsbildes abrufen zu können (Custers 2015). Mit der Zeit bilden sich „Netzwerke“ zwischen bereits vorhandenen und neu gebildeten „Scripts“ (Thampy et al. 2019). „Clinical Reasoning“ lässt sich in zwei Phasen unterteilen, „Diagnostic Reasoning“ (diagnostisches Argumentieren) und „Therapeutic Reasoning“ (therapeutisches Argumentieren). Das eine beinhaltet die Entscheidung für ein spezifisches, auf den Patienten zugeschnittenes diagnostisches Vorgehen, das andere die individualisierte Therapie, die folgt (Durning 2011).

Bereits 1989 definierte Kassirer mehrere Reasoning-Strategien, die während eines diagnostischen Prozesses durchlaufen werden (Kassirer 1989). Zum einen findet das sogenannte „Probabilistic Reasoning“ statt, welches zur Einschätzung der Wahrscheinlichkeit einer bestimmten Diagnose beitragen soll. Es hilft dabei, mit der Prätestwahrscheinlichkeit eine diagnostische Hypothese zu entwickeln und die Bedeutung von klinischen Symptomen sowie Testergebnissen der Patientinnen und Patienten einzuordnen. Das „Causal Reasoning“ soll bei der Verifizierung der Hypothese behilflich sein, indem ein gedankliches

Modell erstellt wird, gegen welches die Symptome des Patienten in Bezug auf Zusammenhang und Vollständigkeit abgeglichen werden. Beim „Deterministic Reasoning“ handelt es sich um ein Repertoire an Regeln, die während einer jahrelangen klinischen Routine entstehen und ebenfalls zur Abgleichung mit der diagnostischen Hypothese verwendet werden (Kassierer 1989). Die beschriebenen Vorgänge laufen während einer Anamneseerhebung bei Experten automatisch ab. So zeigen Untersuchungen, dass bei der Stellung einer Diagnose durch erfahrene Ärztinnen und Ärzte intuitiv ein Prozess durchlaufen wird, in dem mehrere Hypothesen wahrscheinlicher Diagnosen aufgestellt werden (Brush et al. 2017). Anfängerinnen und Anfänger wiederum gleichen laufend bewusst die Symptome der zu behandelnden Personen mit erlernten Krankheitsbildern ab (Charlin et al. 2007). Die Wichtigkeit des intuitiven Generierens von möglichen Differenzialdiagnosen, wird dadurch unterstrichen, dass in 98% der Fälle letztendlich die richtige Diagnose gestellt wird, wenn die Ärztin oder der Arzt, diese Diagnose bereits in den ersten 5 Minuten des Patientenkontaktes als mögliche Differenzialdiagnose in Betracht gezogen hat. Ist dies nicht der Fall, ist die endgültige Diagnose nur zu 25% richtig (Brush et al 2017).

Das Stellen einer Diagnose nimmt eine zentrale Rolle in der täglichen ärztlichen Arbeit ein. Einen großen Anteil hierbei hat wiederum die Anamneseerhebung. Mehrere Studien zeigten, dass in 74-76% der Fälle die korrekte Diagnose bereits allein aus dem Anamnesegespräch und gegebenenfalls dem zusätzlichen Lesen des Einweisungsschreibens gestellt werden konnte (Hampton et al. 1975; Schmitt et al. 1986; Peterson et al. 1992). Aus diesem Grund sollte das Lehren und Lernen von „Clinical Reasoning“-Fähigkeiten ein entscheidender Teil in der medizinischen Ausbildung sein. In den letzten drei Jahrzehnten wurde „Clinical Reasoning“ von einer steigenden Anzahl medizinischer Ausbildungsstätten als eine wichtige zu erlernende Fähigkeit erkannt. Die konkrete Umsetzung allerdings, also die Entwicklung entsprechender Lehreinheiten und Messverfahren von „Clinical Reasoning“-Fähigkeiten, stellte sich komplexer dar (Yazdani et al. 2017). Ein Grund dafür ist die Tatsache, dass es sich beim „Clinical Reasoning“ um eine Fähigkeit handelt, die sich über jahrelanges klinisches Arbeiten hinweg entwickelt und schließlich eine Folge von umfangreichem und multidimensionalem Wissen ist (Norman 2005). In der Frage, wie „Clinical Reasoning“-Fähigkeiten am besten entwickelt und gefestigt werden können, unterscheiden sich die Meinungen. Während erfahrene Ärztinnen und Ärzte die Wichtigkeit, „Clinical Reasoning“ zu lehren, hervorheben, setzen die, sich noch in der Ausbildung befindlichen, Ärztinnen und Ärzte eher auf das Generieren von Wissen und das Nutzen von Online-Ressourcen (Durning et al. 2013). Eine optimale Art und Weise „Clinical Reasoning“-Fähigkeiten zu entwickeln wird in der aktuellen Literatur bisher nicht beschrieben. Die Basis sollte jedoch bereits früh im Studium während der vorklinischen Phase gelegt werden. Dies kann in Form von

theoriebasierten Lerneinheiten oder durch das Lernen an Fallbeispielen geschehen (Khin-Htun & Kushairi 2019). So wurde beispielsweise an der medizinischen Fakultät der Universität Hamburg ein „Clinical Reasoning“-Kurs entwickelt und durchgeführt. In insgesamt sechs Einheiten wurden hier, anhand von Fallbeispielen, Studierende im letzten Jahr ihrer Ausbildung unterrichtet. In der Selbstbeurteilung schätzten die Teilnehmenden ihre „Clinical Reasoning“-Fähigkeiten nach Beendigung des Kurses signifikant besser ein als zuvor (Harendza et al. 2017b). Hindernisse bei der Lehre von „Clinical Reasoning“ können auf mehreren Ebenen entstehen, welche in Form von Kontextfaktoren (die komplexe Natur von „Clinical Reasoning“ selbst), Umweltfaktoren (z.B. seltene Gelegenheiten zum Üben) und die Lehrenden betreffende Faktoren (zu wenig Wissen über „Clinical Reasoning“, zu wenig Zeit die Studierenden zu beobachten) beschrieben wurden (Khin-Htun & Kushairi 2019).

Studierende erlernen „Clinical Reasoning“-Fähigkeiten meistens während eines Unterrichts am Krankenbett oder wenn sie z.B. an einer Sprechstunde teilnehmen. „Clinical Reasoning“ wird dabei jedoch meist nicht explizit gelehrt (Bowen 2006). In der Regel sind die Lehrenden häufig nicht anwesend, während Studierende beispielsweise Anamnesen erheben oder eine körperliche Untersuchung durchführen. Dabei ist die direkte Beobachtung von Studierenden in Interaktion mit Patientinnen und Patienten ein essenzieller Teil der Entwicklung solider klinischer Fähigkeiten, wie das erwähnte Erheben einer Anamnese und die Durchführung einer körperlichen Untersuchung (Howley & Wilson 2004). Generell spielt bei der Entwicklung von Fähigkeiten Feedback eine große Rolle und dieses sollte im besten Fall direkt am Arbeitsplatz gegeben werden (Norcini & Burch 2007). Dieses Feedback sollte auf die Bedürfnisse der Lernenden zugeschnitten und auf die wichtigen Aspekte der zu erlernenden Fähigkeit fokussiert sein. Zudem ist von großer Bedeutung, dass die Feedbackgebenden gut ausgebildet sind und über die entsprechenden Werkzeuge wie z.B. Bewertungskriterien verfügen (Norcini & Burch 2007). Oft sind Bewertungen von der Interaktion Studierender mit Patientinnen und Patienten unzuverlässig, weil den Beobachtenden eine gemeinsame Basis dafür fehlt, was und wie sie beurteilen sollen (Pulito et al. 2006). Von welcher Wichtigkeit die interaktive Teilnahme von Studierenden während Unterrichteinheiten zum „Clinical Reasoning“ ist, zeigt eine Studie von Weidenbusch et al. (2019). Einheiten, die in Form von sogenannten Live-CCDs („Clinical Case Discussions“) durchgeführt wurden, waren effektiver und nachhaltiger als solche, in denen nur mit Patientenfällen in Papierform gearbeitet wurde (Weidenbusch et al 2019). Klein et al. (2019) verfolgten den Ansatz, durch das Lernen aus Fehlern im diagnostischen Prozess, „Clinical Reasoning“-Fähigkeiten bei jungen Medizinstudierenden zu festigen. Die Überlegung ist, dass das Wissen über typische kognitive Fehler im „Clinical Reasoning“-Prozess erweitert

wird und so die Studierenden dabei unterstützt werden, später falsche Entscheidungen und Handlungen zu vermeiden (Klein et al. 2019).

Im Laufe des Medizinstudiums werden die klinischen Fähigkeiten von Studierenden häufig durch Ärztinnen und Ärzte beurteilt. Vielfach findet dies aber informal und basierend auf den aktuellen Eindrücken der Bewertenden statt. Zudem ist es meistens der Endpunkt eines Prozesses, der beurteilt wird, und nicht der Prozess an sich (Thampy et al. 2019). Ein Ansatz, die Leistung von Studierenden zu objektivieren, ist die Durchführung von OSCEs (Objective Structured Clinical Examinations). Hier werden sowohl festgelegte Bewertungskriterien, als auch stets die gleichen bzw. ähnliche klinische Szenarios verwendet (Zayyan 2011). Um die Anwendung von „Clinical Reasoning“ bei Studierenden bewerten zu können, werden also einheitliche und möglichst objektive Kriterien benötigt, an denen sich die Prüfenden orientieren können. Am Anfang eines jeden „Clinical Reasoning“-Prozesses steht die Anamnese. Sie hat daher, wie oben ausgeführt, eine besonders große Bedeutung für das Stellen einer Diagnose und das Nennen möglicher Differenzialdiagnosen. Ist dieser erste Teil fehlerhaft, wird es schwer, dies im weiteren Prozess wieder zu korrigieren, wie in einer qualitativen Studie gezeigt werden konnte (Haring et al. 2017). Haring et al. (2017) zeichneten Arzt-Patientengespräche von Studierenden am Ende ihrer Rotation Innere Medizin auf. Erfahrene Ärztinnen und Ärzte sahen sich diese Videos an und kommentierten währenddessen laut, wann die Studierenden ihrer Meinung nach Anzeichen von „Clinical Reasoning“ zeigten. Die Kommentare wurden anschließend analysiert und daraus auf der Basis von Übereinstimmungen grundlegende „Clinical Reasoning“-Indikatoren formuliert. Von den Expertinnen und Experten selbst wurden die Faktoren „Kontext“, „das Betrachten von sich selbst als Referenz“ und „Emotionen/Gefühle“ als Variablen bei der Beurteilung identifiziert (Haring et al. 2017). Diese Einschätzungen stimmen größtenteils mit dem überein, was auch Kogan et al. 2011 beschrieben. Zusätzlich nannten diese jedoch noch das „Übertragen von der subjektiven Beurteilung auf eine numerische Skala“ und das unterschiedliche „Interpretieren von gleichen Handlungen“ als Gründe für das Zustandekommen verschiedener Bewertungen (Kogan et al. 2011).

Operationalisierte Kriterien, mit denen sich „Clinical Reasoning“ standardisiert bewerten lässt, sind bisher nicht bekannt. Mit der Entwicklung und Testung eines Bogens zur standardisierten Bewertung der Ausprägung von „Clinical Reasoning“ in Anamnesegesprächen beschäftigt sich diese Arbeit.

2. Arbeitshypothese und Fragestellung

„Clinical Reasoning“ stellt eine sehr wichtige Fähigkeit in der alltäglichen Arbeit praktisch tätiger Ärztinnen und Ärzte dar. Die Qualität von „Clinical Reasoning“ ist jedoch, aufgrund der Komplexität der zugrundeliegenden Denkprozesse schwer objektiv zu beurteilen. Es ist bisher wenig bekannt, ob Studierende im Rahmen des Medizinstudiums „Clinical Reasoning“-Fähigkeiten erwerben und in welchem Ausbildungsstadium dies stattfindet. Ziel dieser Arbeit war es daher, einen Bewertungsbogen basierend auf qualitativ identifizierten „Clinical Reasoning“-Indikatoren zu entwickeln und mit dessen Hilfe die „Clinical Reasoning“-Fähigkeiten von Studierenden in simulierten Anamnesegesprächen zu beurteilen.

Mit Hilfe dieser Analyse sollten folgende Fragen beantwortet werden:

- Lassen sich „Clinical Reasoning“-Fähigkeiten objektiv bewerten und vergleichen?
- Gibt es Unterschiede in Bezug auf die „Clinical Reasoning“-Fähigkeiten von Studierenden aus einem vertikal-integrierten Studiengang im Vergleich mit Studierenden aus einem klassischen Medizinstudiengang?
- Weisen Medizinstudierende im Praktischen Jahr im Vergleich mit Medizinstudierenden am Ende des 10. Semesters unterschiedliche Ausprägungen von „Clinical Reasoning“-Fähigkeiten auf?

3. Material und Methoden

3.1 360-Grad-Kompetenzprüfung

Im Jahr 2017 wurde am Universitätsklinikum Hamburg-Eppendorf eine neu entwickelte 360-Grad-Kompetenzprüfung zur Bewertung wichtiger Kompetenzen von ärztlichen Berufsanfängern durchgeführt. Diese sollte für Studierende einen ersten Arbeitstag an einem Krankenhaus simulieren und war in drei Phasen unterteilt. Zunächst führten die Studierenden eine Sprechstunde durch, in der sie mehrere Anamnesegespräche führten, welche auf Video aufgezeichnet wurden. Anschließend konnten sie in 2,5 Stunden ein Behandlungskonzept für ihre jeweiligen Patientinnen und Patienten entwickeln und dieses im folgenden Abschnitt einer ärztlichen „Kollegin“ bzw. einem ärztlichen „Kollegen“ in Form einer Übergabe präsentieren. Abschließend fand eine 30-minütige Nachbesprechung statt (Harendza et al. 2017a).

3.2 Videoaufzeichnungen und Patientenfälle

Für die Analyse der „Clinical Reasoning“-Fähigkeiten wurden die Videoaufzeichnungen der Anamnesegespräche des oben beschriebenen Assessments verwendet. Es handelt sich um 10-minütige Anamnesegespräche von Studierenden mit Schauspielpatientinnen und -patienten. Alle Studierenden führten in unterschiedlicher Reihenfolge je fünf Gespräche mit denselben fünf Schauspielpatientinnen und -patienten durch.

- Fall 1: Eine 42-jährige Frau mit Schwindelanfällen seit mehreren Wochen. In stressigen Situationen hat sie gelegentlich Herzrasen und dabei das Gefühl, das Herz schlage ihr bis zum Hals. (Diagnose: Intermittierendes Vorhofflimmern)
- Fall 2: Ein 53-jähriger Mann mit starker Müdigkeit seit Längerem und Bluthusten am Abend zuvor. Der Patient arbeitet bereits seit 15 Jahren als Missionar im Kongo. (Diagnose: Granulomatose mit Polyangiitis)
- Fall 3: Eine 58-jährige Frau mit seit drei Tagen anhaltenden Bauchschmerzen. Aufgrund der starken Schmerzen sitzt die Patientin im Rollstuhl. (Diagnose: gedeckt perforierte Sigmadivertikulitis)
- Fall 4: Ein 54-jähriger Mann mit seit drei Tagen anhaltenden Flankenschmerzen links. Trotz Schmerzmedikation mit Paracetamol kann der Patient vor Schmerzen nicht mehr schlafen. (Diagnose: gedeckt perforiertes Aortenaneurysma)

- Fall 5: Eine 36-jährige Frau mit Schmerzen in der linken Brust- und Schulterregion. Zusätzlich klagt sie über Erkältungssymptomatik ohne Husten und Fieber bis zu 38,5° Celsius. Bei der Patientin ist bereits eine rheumatoide Arthritis vorbekannt. (Diagnose: Herpes zoster unter Immunsuppression)

3.3 Studierendenkollektiv

Bei dem Studierendenkollektiv handelte es sich um insgesamt 69 Teilnehmende von drei deutschen Universitäten mit jeweils unterschiedlichen Curricula (Hamburg, Technische Universität München, Oldenburg). Von diesen kamen 37 Studierende aus Hamburg (integrierter Modellstudiengang), 27 aus München (klassisches Curriculum) und 5 aus Oldenburg (Modellstudiengang mit Adaptation an das Groninger Modell). Alle Teilnehmenden befanden sich bereits in einer fortgeschrittenen Phase ihres Studiums, größtenteils entweder im 10. Studiensemester oder im Praktischen Jahr (10. Semester: n=25; PJ: n=41; vor dem 10. Semester: n=3). Die Kompetenzprüfung fand während der vorlesungsfreien Zeit im Juli 2017 statt und die Studierenden hatten sich freiwillig angemeldet.

3.4 Entwicklung des Bewertungsbogens für „Clinical Reasoning“-Fähigkeiten

Zur Bewertung und zum Vergleich der „Clinical Reasoning“-Fähigkeiten der Studierenden wurde auf eine Studie zurückgegriffen, in der beobachtbare Aspekte, mit denen „Clinical Reasoning“ operationalisiert werden kann, erarbeitet wurden (Haring et al. 2017). In dieser Studie wurden erfahrenen Ärztinnen und Ärzten aufgezeichnete Anamnesevideos von Studierenden mit Schauspielpatientinnen und -patienten gezeigt. Die Ärzte waren angehalten immer laut auszusprechen, wenn sie das Vorhandensein oder das Fehlen von „Clinical Reasoning“ erkannten. Zusätzlich sollten sie erläutern, woran sie dies festmachten. Anhand dieser Kommentare konnten verschiedene „Clinical Reasoning“-Indikatoren identifiziert werden. Insgesamt wurden die folgenden acht der von Haring et al. (2017) ermittelten Bewertungskriterien in den für diese Studie etablierten Bewertungsbogen aufgenommen:

1. Kontrolle über das Gespräch übernehmen
2. Rückversichern bei den Patienten
3. Stellen der Fragen in einer logischen Reihenfolge
4. Zusammenfassen
5. Zeichen für pathophysiologisches Denken
6. Erkennen von und Reagieren auf relevante Informationen

7. Konkretisieren von Symptomen
8. Gesammelte Daten und Effektivität

Diese acht Bewertungskriterien erfüllten die Auswahlbedingungen, dass „Clinical Reasoning“-Aspekte sich verbal bemerkbar machen und in ihrer Ausprägung für die Bewertung graduierbar sind. Aspekte wie „Körpersprache“, die bei Haring et al. (2017) auch herausgearbeitet worden waren, wurden nicht in den Bewertungsbogen für diese Arbeit aufgenommen. Jedes der ausgewählten Kriterien wurde zudem mit einer kurzen Definition versehen, welche einer freien Übersetzung der Beschreibungen von Haring et al. (2017) entsprach. Zur Beurteilung der acht „Clinical Reasoning“-Indikatoren wurde eine Likert-Skala mit ein bis fünf Punkten verwendet (1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft teilweise zu, 4 = trifft eher zu, 5 = trifft voll zu). Alle Studierenden konnten so je Patientenfall und Bewertungskategorie minimal einen Punkt (trifft nicht zu) und maximal fünf Punkte (trifft voll zu) erhalten. Die minimale Punktzahl für den gesamten Bewertungsbogen pro Gespräch beträgt somit 8 Punkte und die maximale Punktzahl 40 Punkte. Der Bewertungsbogen befindet sich im Anhang 1.

3.5 Pilottestung des Bewertungsbogens

Zur Pilottestung des Bewertungsbogens wurden zehn Anamnesevideos von zwei unabhängigen Personen, einer Ärztin (Betreuerin der Promotion) und einem Medizinstudenten (Promovierender) mit dem Bogen bewertet. Die zehn ausgewählten Videos verteilten sich auf fünf Studierende (je zwei Videos). Bei der Auswahl wurde ebenfalls darauf geachtet, dass jeder Patientinnen- bzw. Patientenfall je zweimal vorkam. Die Ergebnisse der Beurteilungen wurden ausgewertet und verglichen. Gab es bei den Bewertungen eine Übereinstimmung von mindestens 80%, wurde dies als Zeichen angesehen, dass es bei unabhängigen Bewertungen unterschiedlicher Personen zu vergleichbaren Ergebnissen kommt. Die Berechnung zur Übereinstimmung wurde sowohl auf der Ebene der einzelnen Anamnesegespräche als auch auf der Ebene der einzelnen Bewertungskriterien durchgeführt. Wenn sich in einzelnen Fällen eine Übereinstimmung von unter 80% ergab, wurden diese nochmals gesondert besprochen um zu klären, wie es zu dieser unterschiedlichen Bewertung kam. Anschließend an die vergleichenden Bewertungen wurde der Bewertungsbogen nicht mehr verändert, sondern in seiner ursprünglichen Form für alle weiteren Analysen verwendet.

3.6 Analyse der Anamnesevideos

Insgesamt wurde Videomaterial von 68 Studierenden analysiert. Von einer Teilnehmerin (PJ) aus München waren aus technischen Gründen die Videos nicht auswertbar. Da alle Studierenden jeweils fünf etwa 10-minütige Anamnesegespräche geführt haben, wurden 340 Videoaufzeichnungen ausgewertet. Die Videodateien wurden in Bezug auf die Studierenden randomisiert analysiert. Es war vor der Auswertung nicht bekannt, von welcher Universität die Studierenden im jeweiligen Video kamen und in welchem Semester sie waren. Die Reihenfolge der Patientenfälle, in der die einzelnen Videos der Studierenden angesehen wurden, war stets die gleiche. Sie begann mit dem Patientenfall 1 und endete chronologisch mit dem fünften.

Für die Analyse des Videomaterials wurden für die jeweiligen Bewertungskriterien unterschiedliche Methoden angewandt, um festzustellen, welche Punktzahl zwischen eins und fünf die Teilnehmenden für das einzelne Gespräch erhielten. Diese Methoden des Kennzeichnens, ob oder inwiefern die Vorgaben für die einzelnen Kriterien erfüllt wurden, dienten dazu, eine möglichst große Vergleichbarkeit und Konstanz über die Zeit der Videoanalysen zu erreichen. Die Bewertungskriterien konnten hierfür in drei Gruppen unterteilt werden. Bei „Kontrolle über das Gespräch übernehmen“, „Stellen der Fragen in einer logischen Reihenfolge“ und „Erkennen von und Reagieren auf relevante Informationen“ ging es darum, einen Gesamteindruck vom Gesprächsverlauf bezüglich des jeweiligen Kriteriums zu gewinnen und diesen dann zu dokumentieren. Hierfür wurden Pfeile verwendet, die eine Tendenz anzeigen sollten. Wurden zum Beispiel bei der „Gesprächskontrolle“ zwei Pfeile nach oben ($\uparrow\uparrow$) notiert, wurde das als „trifft voll zu“ und damit mit der maximalen Anzahl von fünf Punkten bewertet. Ein Pfeil nach oben (\uparrow) wurde als „trifft eher zu“ und damit mit vier Punkten bewertet. Dementsprechend wurde ein waagerechter Pfeil (\rightarrow) mit drei Punkten („trifft teilweise zu“), ein Pfeil nach unten (\downarrow) zwei Punkten („trifft eher nicht zu“) und zwei Pfeile nach unten ($\downarrow\downarrow$) mit der minimalen Punktzahl von eins („trifft nicht zu“) bewertet. Bei den vier Bewertungskriterien „Rückversichern bei den Patienten“, „Zusammenfassen“, „Zeichen für pathophysiologisches Denken“ und „Konkretisieren von Symptomen“ wurde immer, wenn sich beispielsweise ein Studierender bei einem Patienten rückversicherte, dies mit einem Haken (\checkmark) vermerkt. Je mehr Haken Teilnehmende in einem Gespräch erhalten hatten, desto mehr Punkte bekamen sie. Die Kategorie „Zusammenfassen“ war unter den oben genannten Kriterien dabei eine Besonderheit. Fasste beispielsweise ein Teilnehmer den Gesprächsverlauf am Ende komplett zusammen, so wurde dies mit genauso vielen Haken vermerkt, wie bei jemandem, der immer wieder zwischendurch kurz zusammenfasste. Beim Bewertungskriterium

„gesammelte Daten und Effektivität“ dienten die Fallbeschreibungen der einzelnen Patienten und Patientinnen als Bewertungsgrundlage. Aus diesen Angaben wurde eine bestimmte Anzahl von Daten ermittelt, an die die Teilnehmenden pro Fall maximal für ihr weiteres differentialdiagnostisches Denken gelangen konnten. Die Anzahl der vergebenen Punkte für dieses Kriterium pro Anamnesegegespräch (von eins bis fünf) konnte so genau an der Menge der erlangten Daten abgelesen werden (Tabelle 1).

	1 Punkt	2 Punkte	3 Punkte	4 Punkte	5 Punkte
Fall 1	0-1	2-3	4-5	6-7	8-9
Fall 2	0-3	4-7	8-12	13-16	17-20
Fall 3	0-4	5-9	10-15	16-20	21-25
Fall 4	0-3	4-7	8-11	12-15	16-19
Fall 5	0-2	3-6	7-10	11-14	15-18

Tabelle 1. Anzahl zu erlangender Daten für bestimmte Punktzahlen pro Patientenfall

Am Ende der Analyse einer Gesprächsaufzeichnung wurden alle vermerkten Zeichen bzw. Symbole in Punktzahlen übersetzt und auf dem Bewertungsbogen festgehalten.

3.7 Statistische Analyse

Für die statistische Auswertung wurden zunächst alle Daten von den Bewertungsbögen in zwei umfassende Datentabellen übertragen. Anschließend wurden die Codes der Studierenden aufgelöst und die Studierenden nach Universität bzw. Studienfortschritt zugeordnet. Dies diente als Ausgangspunkt für die weiteren Berechnungen. Als Programm wurde die Statistiksoftware SPSS Version 23 verwendet. Hiermit wurden dann mehrere zweifaktorielle Varianzanalysen (ANOVA) durchgeführt, wobei es sich bei dem ersten Faktor jeweils um die Universität, von der die Studierenden kamen (Universität Hamburg, TU München, Oldenburg), und bei dem zweiten Faktor um den Studienfortschritt (PJ vs. 10. Semester) handelte. Zunächst wurden so Vergleiche über alle Patientenfälle hinweg angefertigt und anschließend Vergleiche für jeden Patientenfall separat. Innerhalb der Varianzanalysen wurden die zuvor ermittelten Mittelwerte der verschiedenen Gruppen verglichen. Als signifikant wurden p-Werte von $< 0,05$ angesehen.

4. Ergebnisse

4.1 Ergebnisse der Pilottestung des Bewertungsbogens

Für den Vergleich der ersten Anwendung des Bewertungsbogens durch den Promovierenden und die Betreuerin der Promotion wurden zehn Testgespräche ausgewählt. Auf der Ebene der einzelnen Items ergab sich durchgehend eine Übereinstimmung von >80%. Die minimale Übereinstimmung lag bei 85,4% (Item 8) und maximale bei 100% (Item 4 und 5). Verglichen wurden hier jeweils die Mittelwerte der Bewertungen einzelner Items, wobei keines der Items pro Teilnehmendem im Mittelwert mehr als 1 Punkt voneinander abweichen durfte. Über alle Items ergab die Übereinstimmung sowohl des Mittelwerts als auch der Summe 95,8% (Tabelle 2).

Items	Promovierender MW ± SD	Betreuerin MW ± SD	Übereinstimmung %
1	4,30 ± 0,64	3,90 ± 0,83	90,70
2	3,40 ± 1,01	3,20 ± 0,97	94,10
3	4,00 ± 0,63	4,50 ± 0,67	88,90
4	2,20 ± 0,97	2,20 ± 0,74	100
5	4,20 ± 0,74	4,20 ± 0,74	100
6	3,70 ± 0,90	4,20 ± 0,60	88,10
7	4,10 ± 0,83	4,40 ± 0,66	93,20
8	3,50 ± 0,80	4,10 ± 0,70	85,40
Mittelwert über alle Items	3,70 ± 0,82	3,80 ± 0,74	95,80
Summenwert über alle Items	29,40 ± 2,34	30,70 ± 2,11	95,80

Tabelle 2. Vergleich der Bewertungen auf Itemebene

SD = Standardabweichung, MW = Mittelwert, Item 1 = Kontrolle über das Gespräch übernehmen, Item 2 = Rückversichern bei den Patienten, Item 3 = Stellen der Fragen in einer logischen Reihenfolge, Item 4 = Zusammenfassen, Item 5 = Zeichen für pathophysiologisches Denken, Item 6 = Erkennen von und Reagieren auf relevante Informationen, Item 7 = Konkretisieren von Symptomen, Item 8 = Gesammelte Daten und Effektivität

Beim Vergleich der Bewertungen auf Ebene der ausgewählten Gespräche kam es bei acht von zehn ausgewerteten Anamnesegesprächen zu einer Übereinstimmung von >80%. Die maximale Übereinstimmung lag bei 91,9% (Testgespräch 4). In zwei Fällen jedoch, lag die Übereinstimmung der Bewertung bei weniger als 80%. Es handelte sich hierbei um die Testgespräche 7 (68,8%) und 5 (71,9%). Die Bewertung dieser beiden Gespräche wurde nach der Pilottestung vom Promovierenden und der Betreuerin noch mal besprochen. Der Mittelwert der Übereinstimmungen lag bei 86,07% (Tabelle 3).

Testgespräche	Promovierender Gesamtpunktzahl	Betreuerin Gesamtpunktzahl	Übereinstimmung %
1	36	32	88,90
2	30	34	88,20
3	31	34	91,20
4	34	37	91,90
5	23	32	71,90
6	31	28	90,30
7	22	32	68,80
8	23	21	91,30
9	32	28	90,30
10	33	29	87,90
Summe aller Gespräche	295	307	
Mittelwert der Übereinstimmung			86,07

Tabelle 3. Vergleich der Gesamtpunktzahlen der Bewertenden auf Gesprächsebene

4.2 Vergleich der „Clinical Reasoning“-Fähigkeiten nach Studienfortschritt auf Ebene aller Patientengespräche

Insgesamt war das Kriterium „Kontrolle über das Gespräch übernehmen“ das am höchsten ($4,33 \pm 0,44$) und das Kriterium „Zusammenfassen“ das am niedrigsten bewertete Kriterium ($2,25 \pm 0,91$). Alle Teilnehmenden erreichten im Mittel einen Gesamtscore von $143,93 \pm$

14,67 Punkten. Der Mittelwert der PJ-Studierenden für den Gesamtscore lag dabei bei $147,53 \pm 12,50$ und der Mittelwert der Studierenden im 10. Semester bei $140,44 \pm 15,88$ Punkten. Dieser Unterschied erwies sich als statistisch nicht signifikant. Zudem erzielten die PJ-Studierenden höhere Punktzahlen in allen Bewertungskategorien mit Ausnahme des Kriteriums „Erkennen und Reagieren auf relevante Informationen“. Diese Unterschiede waren ebenfalls statistisch nicht signifikant (Tabelle 4).

Items	Gesamt (n = 68) MW \pm SD	PJ (n = 40) MW \pm SD	Semester 10 (n = 25) MW \pm SD	p-Wert	Cohens d
Kontrolle	4,33 \pm 0,44	4,39 \pm 0,39	4,31 \pm 0,44	0,67	0,19
Informationen	3,64 \pm 0,45	3,64 \pm 0,43	3,74 \pm 0,42	0,19	-0,24
Konkretisieren	3,85 \pm 0,52	3,92 \pm 0,46	3,78 \pm 0,62	0,83	0,26
Pathophysiologie	3,83 \pm 0,57	3,94 \pm 0,45	3,77 \pm 0,60	0,90	0,33
Reihenfolge	3,99 \pm 0,46	4,07 \pm 0,45	3,89 \pm 0,47	0,51	0,38
Rückversichern	3,30 \pm 0,72	3,46 \pm 0,69	3,06 \pm 0,70	0,45	0,56
Zusammenfassen	2,25 \pm 0,91	2,46 \pm 0,93	1,98 \pm 0,83	0,32	0,55
Daten	3,61 \pm 0,34	3,65 \pm 0,29	3,57 \pm 0,39	0,94	0,22
Gesamtscore	143,93 \pm 14,67	147,53 \pm 12,5	140,44 \pm 15,88	0,82	0,50

Tabelle 4. Vergleich der Studierenden nach Studienzeit auf Itemebene für alle Patientenfälle

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d, SD = Standardabweichung, MW = Mittelwert, PJ = Praktisches Jahr

4.3 Vergleich der „Clinical Reasoning“-Fähigkeiten zwischen den Universitäten auf Ebene aller Patientengespräche

Beim Vergleich der Studierenden aus Hamburg (vertikal-integriertes Curriculum) und der Studierenden aus München (nicht vertikal-integriertes Curriculum) zeigten sich mehrere signifikante Unterschiede. Die Teilnehmenden aus Hamburg ($3,58 \pm 0,44$) versicherten sich signifikant ($p = 0,007$) häufiger bei den Schauspielpatienten rück als es die Teilnehmenden

aus München ($3,02 \pm 0,67$) taten. Für diesen Unterschied zeigte sich ein starker Effekt ($d = 0,84$). Auch wurden die Fragen der Hamburger Studierenden ($4,10 \pm 0,44$) signifikant ($p = 0,03$) öfter in einer logischen Reihenfolge gestellt im Vergleich zu den Studierenden aus München ($3,33 \pm 0,49$). Für diesen Unterschied zeigte sich ein kleiner Effekt ($d = 0,48$). Zudem wurden signifikant ($p < 0,001$) häufiger Zeichen für pathophysiologisches Denken in den Fragen und Aussagen der Studierenden aus Hamburg ($3,92 \pm 0,49$) erkannt als bei den Studierenden aus München ($3,85 \pm 0,59$). In Bezug auf die anderen Bewertungskriterien wurde kein signifikanter Unterschied in die eine oder andere Richtung festgestellt. Im Gesamtscore erreichten die Hamburger Studierenden mit $148,65 \pm 14,26$ signifikant ($p = 0,003$) mehr Punkte als die Münchner Studierenden ($140,19 \pm 13,16$). Bei diesem Unterschied zeigte sich ein mittlerer Effekt ($d = 0,62$) (Tabelle 5).

Items	Gesamt (n = 68) MW \pm SD	UHH (n = 37) MW \pm SD	TUM (n = 26) MW \pm SD	p-Wert	Cohens d
Kontrolle	4,33 \pm 0,44	4,38 \pm 0,43	4,25 \pm 0,45	0,43	0,31
Informationen	3,64 \pm 0,45	3,62 \pm 0,49	3,72 \pm 0,39	0,19	-0,24
Konkretisieren	3,85 \pm 0,52	3,94 \pm 0,49	3,81 \pm 0,52	0,20	0,26
Pathophysiologie	3,83 \pm 0,57	3,92 \pm 0,49	3,85 \pm 0,59	<0,001	0,12
Reihenfolge	3,99 \pm 0,46	4,10 \pm 0,44	3,88 \pm 0,49	0,03	0,48
Rückversichern	3,30 \pm 0,72	3,58 \pm 0,65	3,02 \pm 0,67	0,007	0,84
Zusammenfassen	2,25 \pm 0,91	2,55 \pm 0,91	1,91 \pm 0,71	0,07	0,78
Daten	3,61 \pm 0,34	3,65 \pm 0,35	3,60 \pm 0,28	0,20	0,15
Gesamtscore	143,93 \pm 14,67	148,65 \pm 14,26	140,19 \pm 13,16	0,003	0,62

Tabelle 5. Vergleich der Studierenden nach Universität auf Itemebene für alle Patientenfälle

Die Bewertungskriterien sind hier in gekürzter Form dargestellt, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d , SD = Standardabweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München

4.4 Vergleich der „Clinical Reasoning“-Fähigkeiten auf Ebene einzelner Patientengespräche

4.4.1 Patientengespräch 1

In Bezug auf das erste Patientengespräch („42-jährige Frau mit Schwindelanfällen“) war das am höchsten ($4,28 \pm 0,56$) bewertete Kriterium „Gesammelte Daten und Effektivität“ und das am niedrigsten bewertete ($1,93 \pm 1,10$) „Zusammenfassen“. Der Gesamtscore aller Teilnehmenden lag bei $26,94 \pm 4,16$. Insgesamt wurden die PJ-Studierenden ($27,78 \pm 3,64$) bei diesem Patientengespräch besser bewertet als die Studierenden aus dem 10. Semester ($26,16 \pm 4,64$). Hierfür zeigte sich ein kleiner Effekt ($d = 0,39$). Im Vergleich der Universitäten wurden die Hamburger Studierenden ($28,08 \pm 4,03$) insgesamt signifikant ($p = 0,02$) besser bewertet als die Münchener Studierenden ($25,61 \pm 4,23$). Hier zeigte sich ein mittlerer Effekt ($d = 0,60$) (Tabelle 6).

Items	Gesamt (n=68) MW \pm SD	PJ (n=40) MW \pm SD	Semester 10 (n=25) MW \pm SD	d	UHH (n=37) MW \pm SD	TUM (n=26) MW \pm SD	d
Kontrolle	3,43 \pm 0,86	3,50 \pm 0,74	3,48 \pm 0,94	0,02	3,57 \pm 0,92	3,27 \pm 0,81	0,28
Informationen	3,65 \pm 0,92	3,65 \pm 0,94	3,76 \pm 0,81	-0,13	3,54 \pm 1,00	3,88 \pm 0,75	-0,39
Konkretisieren	3,47 \pm 0,92	3,60 \pm 0,86	3,28 \pm 1,00	0,34	3,70 \pm 0,83	3,19 \pm 0,96	0,57
Pathophysiologie	3,65 \pm 0,90	3,80 \pm 0,90	3,48 \pm 0,90	0,36	3,81 \pm 0,81	3,50 \pm 0,97	0,34
Reihenfolge	3,65 \pm 0,72	3,68 \pm 0,69	3,68 \pm 0,79	-0,01	3,76 \pm 0,79	3,46 \pm 0,63	0,42
Rückversichern	2,90 \pm 1,14	3,25 \pm 1,09	2,32 \pm 1,01	0,89	3,38 \pm 0,94	2,35 \pm 1,17	0,98
Zusammenfassen	1,93 \pm 1,10	2,03 \pm 1,13	1,84 \pm 1,08	0,17	2,19 \pm 1,06	1,58 \pm 1,04	0,58
Daten	4,28 \pm 0,56	4,28 \pm 0,59	4,32 \pm 0,55	-0,08	4,19 \pm 0,56	4,38 \pm 0,56	-0,35
Gesamtscore	26,94 \pm 4,16	27,78 \pm 3,64	26,16 \pm 4,64	0,39	28,08 \pm 4,03	25,61 \pm 4,23	0,60

Tabelle 6. Vergleich der Studierenden anhand des Patientengesprächs 1

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d , SD = Standardabweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München, PJ = Praktisches Jahr

4.4.2 Patientengespräch 2

Beim zweiten Patientengespräch („53-jähriger Mann mit Müdigkeit und Bluthusten“) war „Kontrolle über das Gespräch übernehmen“ das insgesamt am höchsten ($4,21 \pm 0,72$) und „Zusammenfassen“ ($2,09 \pm 1,13$) das insgesamt am niedrigsten bewertete Kriterium. Der Gesamtscore aller Studierenden lag bei $28,03 \pm 3,65$. Die PJ-Studierenden ($29,08 \pm 2,96$) wurden auch bei diesem Patientengespräch, über alle Bewertungskriterien betrachtet, besser bewertet als die Studierenden aus dem 10. Semester ($26,72 \pm 4,12$). Hierbei zeigte sich ein mittlerer Effekt ($d = 0,66$). Im Vergleich der Universitäten wurden die Hamburger Studierenden ($29,24 \pm 3,28$) signifikant ($p = 0,01$) besser als die Münchener Studierenden ($27,19 \pm 3,2$) bewertet. Es zeigte sich hierbei ebenfalls ein mittlerer Effekt ($d = 0,63$) (Tabelle 7).

Items	Gesamt (n=68) MW \pm SD	PJ (n=40) MW \pm SD	Semester 10 (n=25) MW \pm SD	d	UHH (n=37) MW \pm SD	TUM (n=26) MW \pm SD	d
Kontrolle	$4,21 \pm 0,72$	$4,35 \pm 0,65$	$4,04 \pm 0,77$	0,43	$4,30 \pm 0,73$	$4,19 \pm 0,62$	0,16
Informationen	$4,06 \pm 0,76$	$4,15 \pm 0,61$	$4,00 \pm 0,94$	0,19	$4,22 \pm 0,70$	$3,96 \pm 0,76$	0,35
Konkretisieren	$3,59 \pm 0,90$	$3,60 \pm 0,80$	$3,52 \pm 1,06$	0,09	$3,76 \pm 0,82$	$3,42 \pm 0,97$	0,37
Pathophysiologie	$3,99 \pm 0,85$	$4,15 \pm 0,69$	$3,80 \pm 0,94$	0,43	$4,11 \pm 0,83$	$4,04 \pm 0,71$	0,09
Reihenfolge	$3,96 \pm 1,08$	$3,68 \pm 0,69$	$3,68 \pm 0,79$	-0,01	$4,08 \pm 0,54$	$3,88 \pm 0,60$	0,34
Rückversichern	$3,09 \pm 0,98$	$3,13 \pm 0,98$	$3,04 \pm 1,00$	0,09	$3,27 \pm 0,98$	$2,85 \pm 0,82$	0,47
Zusammenfassen	$2,09 \pm 1,13$	$2,50 \pm 1,18$	$1,56 \pm 0,75$	0,97	$2,42 \pm 1,22$	$1,81 \pm 0,92$	0,56
Daten	$3,06 \pm 0,62$	$3,15 \pm 0,53$	$2,96 \pm 0,72$	0,30	$3,11 \pm 0,56$	$3,08 \pm 0,67$	0,05
Gesamtscore	$28,03 \pm 3,65$	$29,08 \pm 2,96$	$26,72 \pm 4,12$	0,66	$29,24 \pm 3,28$	$27,19 \pm 3,20$	0,63

Tabelle 7. Vergleich der Studierenden anhand des Patientengesprächs 2

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d, SD = Standardabweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München, PJ =Praktisches Jahr

4.4.3 Patientengespräch 3

Beim dritten Patientengespräch („58-jährige Frau mit anhaltenden Bauchschmerzen“) wurde das Kriterium „Kontrolle über das Gespräch übernehmen“ insgesamt am höchsten ($4,87 \pm 0,38$) und das Kriterium „Zusammenfassen“ insgesamt am niedrigsten ($2,37 \pm 1,28$)

bewertet. Der Gesamtscore aller Teilnehmenden lag bei $30,13 \pm 3,65$. Auch hier wurden die PJ-Studierenden wieder höher ($30,68 \pm 3,08$) bewertet als die Studierenden aus dem 10. Semester ($29,84 \pm 4,03$). Hierfür ergab sich ein kleiner Effekt ($d = 0,24$). Im Vergleich der Universitäten wurden die Hamburger Studierenden leicht höher ($30,68 \pm 3,3$) bewertet als die Studierenden aus München ($30,27 \pm 3,28$). Auch hier war dieser Unterschied signifikant ($p = 0.005$). Nach dem Cohens d konnte dafür jedoch kein Effekt beobachtet werden ($d = 0,12$) (Tabelle 8).

Items	Gesamt (n=68) MW \pm SD	PJ (n=40) MW \pm SD	Semester 10 (n=25) MW \pm SD	d	UHH (n=37) MW \pm SD	TUM (n=26) MW \pm SD	d
Kontrolle	4,87 \pm 0,38	4,88 \pm 0,40	4,92 \pm 0,27	-0,13	4,84 \pm 0,44	4,88 \pm 0,32	-0,12
Informationen	3,66 \pm 0,50	3,63 \pm 0,48	3,84 \pm 0,37	-0,51	3,65 \pm 0,48	3,69 \pm 0,54	-0,09
Konkretisieren	4,25 \pm 0,81	4,23 \pm 0,85	4,32 \pm 0,79	-0,12	4,22 \pm 0,70	4,42 \pm 0,84	-0,27
Pathophysiologie	4,06 \pm 1,01	4,15 \pm 0,76	4,16 \pm 1,05	-0,01	4,00 \pm 0,93	4,38 \pm 0,88	-0,42
Reihenfolge	4,21 \pm 0,70	4,23 \pm 0,76	4,16 \pm 0,61	0,09	4,30 \pm 0,65	4,15 \pm 0,77	0,20
Rückversichern	3,56 \pm 1,08	3,78 \pm 1,08	3,20 \pm 0,98	0,56	3,89 \pm 0,86	3,38 \pm 1,11	0,51
Zusammenfassen	2,37 \pm 1,28	2,60 \pm 1,28	2,08 \pm 1,26	0,41	2,59 \pm 1,26	2,15 \pm 1,17	0,36
Daten	3,16 \pm 0,56	3,20 \pm 0,51	3,16 \pm 0,61	0,07	3,19 \pm 0,56	3,19 \pm 0,56	-0,01
Gesamtscore	30,13 \pm 3,65	30,68 \pm 3,08	29,84 \pm 4,03	0,24	30,68 \pm 3,30	30,27 \pm 3,28	0,12

Tabelle 8. Vergleich der Studierenden anhand des Patientengesprächs 3

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d , SD = Standardabweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München, PJ = Praktisches Jahr

4.4.4 Patientengespräch 4

Das insgesamt am höchsten bewertete Kriterium beim vierten Patientengespräch („54-jähriger Mann mit anhaltenden Flankenschmerzen“) war erneut „Kontrolle über das Gespräch übernehmen“ ($4,51 \pm 0,61$) und das insgesamt am niedrigsten bewertete Kriterium „Zusammenfassen“ ($2,07 \pm 1,15$). Der Gesamtscore lag bei $28,37 \pm 3,54$. Auch hier wurden die PJ-Studierenden ($28,8 \pm 3,11$) ein wenig besser beurteilt als die Studierenden des 10. Semesters ($28,16 \pm 3,86$), ohne dass mit Hilfe des Cohens d ein Effekt beschrieben werden konnte ($d = 0,18$). Beim Vergleich der Hamburger ($29,65 \pm 3,41$) und der Münchener

Studierenden ($26,81 \pm 3,06$) konnte hingegen ein mittlerer Effekt ($d = 0,88$) in Richtung einer höheren Punktzahl der Hamburger Teilnehmenden beobachtet werden. Auch dieser Unterschied stellte sich als signifikant heraus ($p = 0,005$) (Tabelle 9).

Items	Gesamt (n=68) MW \pm SD	PJ (n=40) MW \pm SD	Semester 10 (n=25) MW \pm SD	d	UHH (n=37) MW \pm SD	TUM (n=26) MW \pm SD	d
Kontrolle	4,51 \pm 0,61	4,63 \pm 0,48	4,48 \pm 0,64	0,26	4,62 \pm 0,59	4,35 \pm 0,62	0,46
Informationen	3,31 \pm 0,65	3,23 \pm 0,65	3,48 \pm 0,64	-0,39	3,30 \pm 0,61	3,35 \pm 0,68	-0,08
Konkretisieren	3,60 \pm 0,75	3,70 \pm 0,64	3,48 \pm 0,90	0,29	3,70 \pm 0,69	3,42 \pm 0,74	0,39
Pathophysiologie	3,65 \pm 0,98	3,70 \pm 0,95	3,64 \pm 1,02	0,06	3,92 \pm 0,91	3,46 \pm 0,97	0,49
Reihenfolge	4,04 \pm 0,65	4,18 \pm 0,67	3,92 \pm 0,56	0,42	4,16 \pm 0,72	3,88 \pm 0,58	0,43
Rückversichern	3,41 \pm 0,97	3,48 \pm 0,95	3,40 \pm 0,94	0,08	3,68 \pm 0,90	3,12 \pm 1,01	0,59
Zusammenfassen	2,07 \pm 1,15	2,13 \pm 1,19	2,00 \pm 1,13	0,11	2,30 \pm 1,14	1,69 \pm 1,07	0,55
Daten	3,76 \pm 0,62	3,78 \pm 0,65	3,76 \pm 0,59	0,02	3,97 \pm 0,64	3,54 \pm 0,50	0,77
Gesamtscore	28,37 \pm 3,54	28,80 \pm 3,11	28,16 \pm 3,86	0,18	29,65 \pm 3,41	26,81 \pm 3,06	0,88

Tabelle 9. Vergleich der Studierenden anhand des Patientengesprächs 4

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d , SD = Standardabweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München, PJ = Praktisches Jahr

4.4.5 Patientengespräch 5

Insgesamt am höchsten bewertet von den Kriterien wurde auch beim fünften Patientengespräch („36-jährige Frau mit Schmerzen in der linken Brust- und Schulterregion“) „Kontrolle über das Gespräch übernehmen“ ($4,62 \pm 0,64$) und am niedrigsten „Zusammenfassen“ ($2,79 \pm 1,31$). Der Gesamtscore lag bei ($30,46 \pm 3,78$). Die PJ-Studierenden ($31,2 \pm 3,74$) wurden insgesamt höher bewertet als die Studierenden aus dem 10. Semester ($3,64 \pm 0,62$). Hier konnte ein kleiner Effekt beobachtet werden ($d = 0,44$). Beim Vergleich der Universitäten haben die Hamburger Studierenden eine höhere Gesamtpunktzahl ($31 \pm 4,15$) erhalten als die Studierenden aus München ($30,31 \pm 3,12$). Ein Effekt nach Cohens d konnte jedoch nicht beobachtet werden (Tabelle 10).

Items	Gesamt (n=68) MW ± SD	PJ (n=40) MW ± SD	Semester 10 (n=25) MW ± SD	d	UHH (n=37) MW ± SD	TUM (n=26) MW ± SD	d
Kontrolle	4,62 ± 0,64	4,60 ± 0,70	4,64 ± 0,56	-0,06	4,65 ± 0,71	4,54 ± 0,57	0,17
Informationen	3,53 ± 0,65	3,53 ± 0,67	3,60 ± 0,57	-0,12	3,38 ± 0,75	3,73 ± 0,44	-0,59
Konkretisieren	4,35 ± 0,78	4,45 ± 0,74	4,28 ± 0,83	0,22	4,32 ± 0,81	4,58 ± 0,57	-0,37
Pathophysiologie	3,79 ± 0,83	3,90 ± 0,83	3,76 ± 0,76	0,18	3,76 ± 0,79	3,88 ± 0,93	-0,15
Reihenfolge	4,07 ± 0,73	4,20 ± 0,71	3,88 ± 0,71	0,45	4,19 ± 0,73	4,04 ± 0,71	0,21
Rückversichern	3,53 ± 1,14	3,65 ± 1,17	3,36 ± 1,13	0,25	3,68 ± 1,14	3,42 ± 1,15	0,22
Zusammenfassen	2,79 ± 1,31	3,05 ± 1,32	2,40 ± 1,23	0,51	3,24 ± 1,22	2,32 ± 1,14	0,80
Daten	3,76 ± 0,57	3,83 ± 0,54	3,64 ± 0,62	0,32	3,78 ± 0,62	3,81 ± 0,48	-0,04
Gesamtscore	30,46 ± 3,78	31,2 ± 3,74	29,56 ± 3,72	0,44	31,00 ± 4,15	30,31 ± 3,12	0,19

Tabelle 10. Vergleich der Studierenden anhand des Patientengesprächs 5

Die Bewertungskriterien sind hier in gekürzter Form definiert, die vollständigen Bezeichnungen sind bei „Material und Methoden“ aufgeführt. d = Cohens d, SD = Standard

abweichung, MW = Mittelwert, UHH = Universität Hamburg, TUM = Technische Universität München, PJ = Praktisches Jahr

5. Diskussion

Die Entwicklung des Bewertungsbogens in dieser Arbeit diente dem Zweck, „Clinical Reasoning“-Fähigkeiten an Hand eines Anamnesegespräches beurteilen zu können. Bei der Bewertung von „Clinical Reasoning“-Fähigkeiten in Bezug auf eine Interaktion zwischen einer Ärztin oder einem Arzt mit Patientinnen oder Patienten ist es wichtig, über entsprechende solide und praxisnahe Instrumente zu verfügen (Kogan et al. 2009). Mehrere Faktoren können dabei die Beurteilung beeinflussen. So können einzelne Bewertende unterschiedliche Referenzen (z.B. sich selbst) haben, mit denen sie das Gesehene vergleichen. Auch die Herangehensweise Beobachtungen auf eine numerische Skala zu übertragen kann sich unterscheiden (Kogan et al. 2009). Von den verschiedenen, bisher verwendeten Bewertungsinstrumenten für „Clinical Reasoning“-Fähigkeiten haben sich zwei als am besten geeignet herausgestellt. Die sogenannten „extended matching items“ (EMIs) basieren auf „Multiple Choice“ Fragen. Bei den sogenannten „script concordance tests“ (SCTs) handelt es sich um klinische Szenarien, die bearbeitet werden (Nazim et al. 2019). Beide Testformen sind valide und verlässliche Instrumente, die jedoch zur Bewertung von einer Arzt-Patienten Interaktion oder spezifischer eines Anamnesegespräches nicht geeignet sind (Nazim et al. 2019). Der für diese Arbeit entwickelte Bewertungsbogen basierte auf einer qualitativen Erhebung von beobachtbaren Aspekten, an denen sich „Clinical Reasoning“ in Anamnesegesprächen erkennen lässt (Haring et al. 2017). Es zeigte sich durch die parallele Pilottestung zweier Bewertender von ausgewählten Videos, dass die Qualität des Bogens insgesamt zufriedenstellend und für die weitere Auswertung durch nur eine Person geeignet ist.

Insgesamt wurden die Studierenden in der Kategorie „Kontrolle über das Gespräch übernehmen“ am besten bewertet, gefolgt von „Stellen der Fragen in einer logischen Reihenfolge“. Die Studierenden waren offenbar gut in der Lage, abstrakte klinische Zeichen oder Symptome in ein bereits durch Wissen geformtes und bestehendes System einzuordnen und so Zusammenhänge zu erkennen, die für eine fokussierte Anamnese im Sinne des „Clinical Reasoning“ erforderlich sind (Bordage & Lemieux 1986). Das Beziehen neuer Informationen auf bereits bestehendes Wissen und Gedankenstrukturen zeichnet den Ablauf der Diagnosefindung bei erfahrenen Ärztinnen und Ärzten aus (Bordage & Lemieux 1986). Zudem kann das gute Abschneiden der Teilnehmenden in den beiden genannten Kategorien ein Hinweis darauf sein, dass die Studierenden zumindest in Teilen ihre Anamnese an Leitsymptomen ausrichten konnten und sich nicht nur starr an ein gelerntes Anamneseschema hielten (Nierenberg 2017). Das wiederum zeigt, dass die Studierenden relevante von weniger bedeutsamen Informationen unterscheiden konnten, was ebenfalls

eine wichtige Qualität des „Clinical Reasoning“ ist, die erfahrene Medizinerinnen und Mediziner auszeichnet (Nierenberg 2017).

Die „Clinical Reasoning“-Indikatoren „Konkretisieren von Symptomen“ und „Zeichen für pathophysiologisches Denken“ wurden insgesamt im Mittel mit der dritt bzw. viert höchsten Punktzahl bewertet. Beide sind Zeichen dafür, dass die Teilnehmenden in einem gewissen Umfang, während der Gespräche schnell und intuitiv, eine Liste von Differenzialdiagnosen aufgestellt haben und diese dann in einem zweiten Schritt in Form des „analytischen Reasoning“ verifiziert haben (Brush et al. 2017). Besonders der erste intuitive Schritt ist hier wichtig und erhöht die Wahrscheinlichkeit einer richtigen Diagnose (Brush et al. 2017). Intuitiv ablaufende Gedanken sind jedoch auch anfällig dafür, einem sogenannten Bestätigungsfehler („Confirmation Bias“) zu unterliegen (Hergovich et al. 2010). Dieser beschreibt die Tendenz dazu, neue Informationen zu suchen, die unseren Erwartungen entsprechen („Confirming Evidence“) (Mahoney 1977), was prinzipiell ein normales Vorgehen im Clinical Reasoning Prozess ist. Im konkreten Beispiel aus dieser Arbeit könnten Studierende bestimmte Informationen der Patientinnen oder Patienten als bedeutsam einordnen, weil sie sich bereits eine Meinung über das vorliegende Krankheitsbild gebildet haben, auch wenn sie mit ihrer Einschätzung in diesem Fall falsch lagen. Andererseits könnte etwa der Bewertende oder die Bewertende eine Geste oder ein Kopfnicken eines oder einer Teilnehmenden fälschlicherweise als Zeichen deuten, dass der oder die Studierende eine bestimmte Aussage eines Patienten oder einer Patientin als wichtig einordnet. Ein Grund für eine solche Überinterpretation könnte das bereits vorhandene Wissen des Bewertenden über den Ablauf des jeweiligen Patientenfalles sein.

Weiterhin ist das „Zusammenfassen“ ein bedeutender Teil eines Anamnesegespräches. Es gibt der Ärztin bzw. dem Arzt die Möglichkeit, eventuelle Missverständnisse auszuräumen und für sich selber die erfolgte Anamnese auf Vollständigkeit zu prüfen (Lichstein 1990). In der Bewertung der Studierenden war diese Kategorie in der Gesamtpunktzahl auffällig niedriger als die restlichen und lag als einzige unter dem Durchschnitt. Ob dem schlechteren Abschneiden möglicherweise zu Grunde liegt, dass das „Zusammenfassen“ nicht ausreichend gelehrt wurde oder es daran lag, dass die Studierenden wegen der vorgegebenen Zeit von zehn Minuten pro Gespräch einen Zeitdruck verspürten, lässt sich nicht abschließend klären. Zudem berichten Studierende, dass sie das „Zusammenfassen“ als besonders schwierig zu erlernen empfinden, da es in ihren alltäglichen Gesprächen so nicht angewandt wird (Boyle et al. 2005). An der Medical School of Colorado beispielsweise ist das „Zusammenfassen“ explizit ein Baustein eines patientenorientierten Kommunikationstrainings. Es wird dort als Teil der sogenannten „ILS“-Technik („Invite,

Listen, Summarize“) gelehrt (Boyle et al. 2005). Die Hamburger Studierenden aus dem vertikal-integrierten Studiengang iMED wurden sowohl in der Gesamtbewertung, als auch bei sieben von acht „Clinical Reasoning“-Indikatoren höher bewertet als die Studierenden des Regelstudiengangs von der TU München. Als signifikant erwies sich der Unterschied bei den Kategorien „Rückversichern bei den Patienten“, „Stellen der Fragen in einer logischen Reihenfolge“ und „Zeichen für pathophysiologisches Denken“. Diese Unterschiede waren zunächst überraschend, da das Training von Anamnesefähigkeiten seit über 30 Jahren nahezu unverändert fest in der medizinischen Ausbildung implementiert ist (Kefenheim et al. 2015). Die Basisfertigkeiten der Kommunikation können sich jedoch im Laufe des Studiums verschlechtern, wenn diese nicht adäquat und gezielt repetitiv geschult werden (Kefenheim et al. 2015). Studierende in vertikal-integrierten Curricula scheinen häufiger die Möglichkeit zu haben, in „problem-orientierten“ Lerneinheiten fokussierte Anamnesen zu üben, die sich am „Clinical Reasoning“ orientieren und nicht an festgelegten Anamneseschemata (Epstein 2004). Zudem haben Studierende vertikal-integrierter Studiengänge vermehrt Unterrichtseinheiten, die eine höhere Interaktivität in kleinen Gruppen (Deveugele et al. 2005) und die Vernetzung vorklinischer mit klinischen Inhalten ermöglichen (Bandiera et al. 2013). Ein fallbasierter „Clinical Reasoning“-Kurs (Keemink et al. 2018) als Teil des vertikal-integrierten Curriculums an der medizinischen Hochschule Utrecht (Wijnen-Meijer et al. 2010), führte beispielsweise zu einer Stärkung von „Illness Scripts“ bei den Studierenden und zur Verbesserung der diagnostischen Fähigkeiten (Keemink et al. 2018).

Biomedizinisches Grundwissen wie Kenntnisse über Anatomie, Physiologie, Biochemie usw., helfen dabei, die zugrundeliegenden Mechanismen von Krankheitsbildern zu verstehen und eine Verbindung zwischen diesen und den klinischen Symptomen von Patienten und Patientinnen herzustellen. Es fördert zudem die Anwendung von „Clinical Reasoning“ und hilft eine korrekte Diagnose zu stellen (Woods 2007). Wenn Studierende jedoch mit der Zeit dieses Grundwissen wieder verlieren, wird das Herstellen solcher Verbindungen schwieriger und die Studierenden verlassen sich immer mehr ausschließlich auf das Erkennen von Mustern, hauptsächlich an Hand von klinischen Aspekten. Prüfungen mit Multiple Choice Klausuren verstärken diesen Effekt (Freiwald et al. 2014). Der Aufbau von traditionellen Curricula medizinischer Hochschulen, in denen zunächst das Grundlagenwissen und anschließend klinische Aspekte getrennt gelehrt werden, scheint weniger zu den beschriebenen Denkvorgängen und damit weniger zur Anregung von „Clinical Reasoning“ zu führen (Woods 2007). Ein integrierter Ansatz, bei dem die Zusammenhänge von biomedizinischen und klinischen Aspekten deutlicher werden, könnte es den Studierenden erleichtern solche Verbindungen auch nach längerer Zeit noch zu erkennen (Woods 2007). Dies könnte ebenfalls ein Grund für die zum Teil signifikant öfter erkannten Anzeichen von

„Clinical Reasoning“ bei Studierenden aus dem vertikal-integriertem Studiengang sein. Das Zusammenbringen von erlernten Inhalten und dem aktiven Durchführen von Anamnesen führt insgesamt zum Ausbau der Anamnese- und damit auch der „Clinical Reasoning“-Fähigkeiten (Kurtz et al. 2003). Dies könnte das bessere Abschneiden von Studierenden aus dem PJ im Vergleich zu jenen aus dem 10. Semester erklären. Dieser Unterschied erwies sich zwar nicht als signifikant, war jedoch in sieben von acht Kategorien zu beobachten. Während ihres PJs haben die Studierenden nicht nur häufig die Gelegenheit praktisch Anamnesen zu erheben, sondern auch problemorientiert zu lernen (Epstein 2004). Die Anamnesefähigkeiten der Studierenden aus dem 10. Semester beruhen wiederum zumeist auf Lehrveranstaltungen, welche noch immer oft auf das Anwenden bestimmter Anamneseschemata fokussiert sind und weniger auf die Art und Weise vorbereiten wie wichtige Informationen erlangt werden können (Maguire et Rutter 1976). Dieses Vorgehen beschreiben Maguire und Rutter bereits 1976 als nicht ausreichend. Zudem könnten die Lehrveranstaltungen zur Anamneseerhebung von den Studierenden aus dem 10. Semester eventuell bereits einige Semester früher im Studium durchgeführt worden sein. Durch die in der Zwischenzeit zusätzlich gelernten Inhalte könnte so das Gelernte über die Anamneseerhebung in den Hintergrund geraten sein (Pfeiffer et al. 1998).

Insgesamt wurden bei den Patientenfällen 3 („58-jährige Frau mit seit drei Tagen anhaltenden Bauchschmerzen“) und 5 („36-jährige Frau mit Fieber unter Immunsuppression“) die höchsten Punktwerte erzielt. Es lässt sich vermuten, dass diese Leitsymptome in den beiden unterschiedlichen Curricula, der an der Studie beteiligten Hochschulen, intensiver behandelt wurden als andere und so die Teilnehmenden von ähnlichen, ihnen bereits bekannten Fällen profitiert haben könnten (Young et al. 2007). Unterschiedliche Aufgaben regen allerdings auch die Anwendung von „Clinical Reasoning“ in der anteiligen Beanspruchung des schnellen Systems 1 oder des langsamen Systems 2 der „Dual Processing“-Theorie verschieden stark an. Bei den beiden genannten Fällen könnten die Studierenden durch die beschriebene Symptomatik der Patienten leichter Muster erkannt und auf bereits entwickelte „Illness Scripts“ zurückgegriffen haben (Norman et al. 2013). Insgesamt haben die Teilnehmenden bei allen Patientenfällen, mit denen sie konfrontiert waren, überdurchschnittliche Zeichen für das Anwenden von „Clinical Reasoning“ gezeigt. Ein Grund hierfür könnte sein, dass die Leitsymptome der Patientenfälle, also „Schwindel und Palpitationen“, „Bluthusten“, „Bauchschmerzen“, „Flankenschmerzen“ und „Fieber unter Immunsuppression“, als typische Beratungsanlässe im Unterricht vorkommen. Das Wissen über klinische Aspekte von bestimmten Krankheitsbildern und deren biomedizinische Hintergründe ist eine wesentliche Voraussetzung für das Anwenden von „Clinical Reasoning“ (Woods 2007).

Der in dieser Arbeit anhand der von Haring et al. (2017) beschriebenen „Clinical Reasoning“-Indikatoren empirisch entwickelte Bewertungsbogen, wurde nicht validiert. Dies ist eine Schwäche der Arbeit. Die Validierung des neuen Bewertungsinstrument ist entscheidend, da von ihr abhängt inwiefern die beobachtbaren Ergebnisse angemessen und interpretierbar sind (Tor et al. 2016). Insofern können die Ergebnisse zunächst nur im Sinne einer Pilotstudie betrachtet werden. Zum Einsatz in einem größeren Rahmen als Bewertungsinstrument sollte eine Validierungsstudie erfolgen. Der Bewertungsbogen zeigte insgesamt mit 95,8% auf Itemebene und mit 86,1% auf Gesprächsebene eine zufriedenstellende Bewertungsübereinstimmung. Es hätte diese Untersuchung allerdings noch gestärkt, wenn eine Pilottestung des Bogens mit mehreren, im besten Falle weiteren erfahrenen Medizinerinnen oder Medizinerinnen stattgefunden hätte. Außerdem hätte, wenn alle Filme (und nicht nur zehn) von den zwei Bewertenden angesehen worden wären, eine Intraklassen-Korrelation berechnet werden können. Diese hätte die Qualität der Studie noch weiter gestärkt. Eine weitere Schwäche dieser Arbeit liegt darin, dass zur Beurteilung der Studierenden nur fünf verschiedene klinische Szenarien verwendet wurden. Hierdurch könnte die Möglichkeit, die „Clinical Reasoning“-Fähigkeiten einzuschätzen, in einem gewissen Ausmaß eingeschränkt gewesen sein. Andererseits wurde bei der Auswahl der Patientenfälle auf eine gewisse Bandbreite des nötigen klinischen Wissens und ausgeglichene Schwierigkeitsgrade geachtet. Zu den Stärken dieser Arbeit gehört die große Anzahl an Anamnesevideos (n=340), die in die Auswertung mit einbezogen wurden. Zudem kamen die Teilnehmenden von unterschiedlichen Universitäten mit verschiedenen Curricula. Dies erhöht die Wahrscheinlichkeit, mit der man davon ausgehen kann, dass die erzielten Ergebnisse allgemein übertragbar sind. Die Studierenden der Universität Oldenburg konnten aufgrund ihrer geringen Anzahl beim Vergleich der Universitäten nicht berücksichtigt werden. Im Vergleich der 10. Semester und der PJ-Studierenden sind sie allerdings mit eingeschlossen.

Im Gegensatz zu anderen, oben beschriebenen, bereits etablierten Instrumenten zur Messung von „Clinical Reasoning“ (EMI und SCT) ermöglicht es der neu entwickelte Bogen Anamnesegespräche zu untersuchen. Um die Unterschiede zwischen den Bewertenden in Zukunft kleiner zu gestalten, sollten zukünftige Bewertende gezielt in der Verwendung des Bogens geschult werden. Die Schulung von Bewertenden sollte ebenfalls eine Strategie enthalten, wie die Beobachtungen möglichst objektiv und in einer operationalisierten Weise auf eine numerische Skala übertragen werden können, um eine gute Konstanz bei den Ergebnissen zu erreichen. Außerdem könnte die Bewertungsskala um Anker ergänzt werden, wie sie in einem kürzlich publizierten Bogen eingeführt wurden (Haring et al. 2020).

Dann wäre das neu entwickelte Instrument mit den oben genannten Einschränkungen auch in einem größeren Rahmen verwendbar.

6. Zusammenfassung

Beim „Clinical Reasoning“ handelt es sich um einen komplexen kognitiven Prozess, der bei Ärztinnen und Ärzten beispielsweise während eines Patientenkontaktes abläuft und einen großen Anteil an der medizinischen Entscheidungsfindung hat. Aufgrund seiner Komplexität und der vielen unterschiedlichen Voraussetzungen, darunter insbesondere Erfahrung und biomedizinisches Wissen, ist dieser Prozess schwer zu messen oder abzubilden. Das Ziel dieser Arbeit war es, einen Bewertungsbogen aus qualitativ identifizierten „Clinical-Reasoning“-Indikatoren zu entwickeln und die „Clinical Reasoning“-Fähigkeiten von Medizinstudierenden im Zuge von simulierten Anamnesegesprächen zu beurteilen. Der Bewertungsbogen wurde anhand von „Clinical Reasoning“-Indikatoren aus einer publizierten Studie erstellt und umfasst acht Kriterien, die auf einer Likert-Skala von einem bis fünf Punkten bewertet werden können. Die Bewertung erfolgte mit Hilfe von Anamnesegesprächen, die während einer 360-Grad Kompetenzprüfung aufgezeichnet worden waren. Das Studierendenkollektiv bestand aus insgesamt 69 fortgeschrittenen Medizinstudierenden, größtenteils aus dem 10. Semester oder dem PJ. Die Studierenden stammten von drei verschiedenen Universitäten mit unterschiedlichen Curricula. Insgesamt wurden 340 Videos ausgewertet und die Ergebnisse verglichen. Die Studierenden aus dem vertikal-integrierten Studiengang (Hamburg) erzielten einen signifikant höheren Gesamtscore als die Studierenden aus dem nicht vertikal-integrierten Studiengang (TU München), $148,65 \pm 14,26$ versus $140,19 \pm 13,16$ ($p < 0,01$). Ebenfalls signifikant besser bewertet wurden die Studierenden aus Hamburg gegenüber den Studierenden der TU München in den Einzelkategorien: „Stellen der Fragen in einer logischen Reihenfolge“, $4,10 \pm 0,44$ versus $3,33 \pm 0,49$ ($p < 0,01$), „Zeichen für pathophysiologisches Denken“, $3,92 \pm 0,49$ versus $3,85 \pm 0,59$ ($p < 0,05$) und „Rückversichern bei den Patienten“, $3,58 \pm 0,44$ versus $3,02 \pm 0,67$ ($p < 0,01$). Auch bei den Gesamtscores in Bezug auf die jeweiligen Patientenfälle wurden die Studierenden aus Hamburg in vier von fünf Fällen signifikant besser bewertet. Im Vergleich der Studierenden aus dem 10. Semester und dem PJ zeigten sich keine signifikanten Unterschiede. Der neu entwickelte Bewertungsbogen zeigte eine zufriedenstellende Qualität zur quantitativen Bewertung von „Clinical Reasoning“ im Zuge von Anamnesegesprächen, sollte jedoch für die weitere Nutzung noch validiert werden. Ob die Unterschiede in den Ergebnissen der Studierenden aus Hamburg und der TU München durch die verschiedenen Curricula zu erklären sind, lässt sich zum jetzigen Zeitpunkt nicht sicher feststellen. Um dies weiter zu untersuchen wären Studien erforderlich, die Einblicke in den genauen Curriculumsverlauf der Universitäten und das Abschneiden der Studierenden in den jeweiligen Prüfungen gewähren.

7. Summary

“Clinical reasoning” is a complex cognitive process that occurs, for example, during patient contact and plays a major role in medical decision-making. Due to its complexity and the many different prerequisites, including in particular experience and biomedical knowledge, this process is difficult to measure or map. The aim of this thesis was to develop an assessment form consisting of qualitatively identified "clinical reasoning" indicators and to assess the "clinical reasoning" skills of medical students during simulated history taking. The assessment form was developed using clinical reasoning indicators from a published study and includes eight criteria that can be scored on a Likert scale of one to five points. The evaluation was based on recordings of simulated patient interviews during a 360-degree competence assessment. The student collective consisted of a total of 69 advanced medical students, mostly from the 10th semester or final year (PJ). The students came from three different universities with different undergraduate curricula. A total of 340 videos were evaluated and the results compared. Students from the vertically integrated program (Hamburg) achieved a significantly higher overall score than students from the non-vertically integrated program (TU Munich), 148.65 ± 14.26 versus 140.19 ± 13.16 ($p < 0.01$). Students from Hamburg also scored significantly better compared with students from the TU Munich in the individual categories: "Putting questions in a logical order", 4.10 ± 0.44 versus 3.33 ± 0.49 ($p < 0.01$), "Asking specific questions that point to pathophysiologic thinking", 3.92 ± 0.49 versus 3.85 ± 0.59 ($p < 0.05$) and "Checking with patient", 3.58 ± 0.44 versus 3.02 ± 0.67 ($p < 0.01$). The students from Hamburg also scored significantly better in four out of five patient cases in the overall scores. In the comparison of the students from the 10th semester and the PJ, no significant differences were found. The newly developed assessment form showed a satisfactory quality for the quantitative evaluation of "Clinical Reasoning" in the course of history taking, but should still be validated for further use. Whether the differences in the results of the students from Hamburg and the TU Munich can be explained by the different curricula cannot be determined with certainty at this time. In order to investigate this further, studies would be required that provide insights into the exact curriculum of the universities and the students' performance in the respective examinations.

8. Abkürzungsverzeichnis

CCD	<i>Clinical Case Discussions</i>
d	<i>Cohens d</i>
EMI	<i>extended matching items</i>
ILS	<i>Invite, Listen, Summarize</i>
iMED	<i>integrierter Modellstudiengang Medizin Hamburg</i>
MW	<i>Mittelwert</i>
OSCE	<i>Objective Structured Clinical Examinations</i>
PJ	<i>Praktisches Jahr</i>
SCT	<i>script concordance test</i>
SD	<i>Standardabweichung</i>
TU	<i>Technische Universität</i>
TUM	<i>Technische Universität München</i>
UHH	<i>Universität Hamburg</i>

9. Tabellenverzeichnis

Tabelle 1. Anzahl zu erlangender Daten für bestimmte Punktzahlen pro Patientenfall.....	15
Tabelle 2. Vergleich der Bewertungen auf Itemebene	16
Tabelle 3. Vergleich der Gesamtpunktzahlen der Bewertenden auf Gesprächsebene	17
Tabelle 4. Vergleich der Studierenden nach Studienzeit auf Itemebene für alle Patientenfälle	18
Tabelle 5. Vergleich der Studierenden nach Universität auf Itemebene für alle Patientenfälle	19
Tabelle 6. Vergleich der Studierenden anhand des Patientengesprächs 1	20
Tabelle 7. Vergleich der Studierenden anhand des Patientengesprächs 2	21
Tabelle 8. Vergleich der Studierenden anhand des Patientengesprächs 3	22
Tabelle 9. Vergleich der Studierenden anhand des Patientengesprächs 4	23
Tabelle 10. Vergleich der Studierenden anhand des Patientengesprächs 5	24

10. Literaturverzeichnis

Bandiera G, Boucher A, Neville A, Kuper A, Hodges B. Integration and timing of basic and clinical sciences education. *Med Teach*. 2013;35(5):381-387.

Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med*. 1991;66:70-72.

Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355(21):2217-2225.

Boyle D, Dwinnell B, Platt F. Invite, listen, and summarize: a patient-centered communication technique. *Acad Med*. 2005;80(1):29-32.

Brush JE Jr, Sherbino J, Norman GR. How expert clinicians intuitively recognize a medical diagnosis. *Am J Med*. 2017;130(6):629-634.

Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ*. 2007;41(12):1178-1184.

Custers EJ. Thirty years of illness scripts: theoretical origins and practical applications. *Med Teach*. 2015;37(5):457-462.

Deveugele M, Derese A, De Maesschalck S, Willems S, Van Driel M, De Maeseneer J. Teaching communication skills to medical students, a challenge in the curriculum? *Patient Educ Couns*. 2005;58(3):265-270.

Durning S, Artino AR Jr, Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.

Durning SJ, Ratcliffe T, Artino AR Jr, van der Vleuten C, Beckman TJ, Holmboe E, Lipner RS, Schuwirth L. How is clinical reasoning developed, maintained, and objectively assessed? Views from expert internists and internal medicine interns. *J Contin Educ Health Prof*. 2013;33(4):215-223.

Epstein RJ. Learning from the problems of problem-based learning. *BMC Med Educ.* 2004;4:1.

Freiwald T, Salimi M, Khaljani E, Harendza S. Pattern recognition as a concept for multiple-choice questions in a national licensing exam. *BMC Med Educ.* 2014;14:232.

Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *Br Med J.* 1975;2(5969):486-489.

Harendza S, Berberat PO, Kadmon M. Assessing competences in medical students with a newly designed 360-degree examination of a simulated first day of residency: a feasibility study. *J Community Med Health Educ.* 2017a;7:4.

Harendza S, Krenz I, Klinge A, Wendt U, Janneck M. Implementation of a clinical reasoning course in the internal medicine trimester of the final year of undergraduate medical training and its effect on students' case presentation and differential diagnostic skills. *GMS J Med Educ.* 2017b;34(5):Doc66.

Haring CM, Cools BM, van Gorp PJM, van der Meer JWM, Postma CT. Observable phenomena that reveal medical students' clinical reasoning ability during expert assessment of their history taking: a qualitative study. *BMC Med Educ.* 2017;17(1):147.

Haring CM, Klaarwater CCR, Bouwmans GA, Cools BM, van Gorp PJM, van der Meer JWM, Postma CT. Validity, reliability and feasibility of a new observation rating tool and a post encounter rating tool for the assessment of clinical reasoning skills of medical students during their internal medicine clerkship: a pilot study. *BMC Med Educ.* 2020;20(1):198.

Hergovich, Andreas & Schott, Reinhard & Burger, Christoph. (2010). Biased evaluation of abstracts depending on topic and conclusion: further evidence of a confirmation bias within scientific psychology. *Current Psychology.* 2010;29:188-209.

Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Acad Med.* 2004;79(3):276-280.

Kahneman, D. *Thinking, fast and slow.* New York: Farrar, Straus and Giroux, 2011.

Kassirer JP. Diagnostic reasoning. *Ann Intern Med.* 1989;110(11):893-900.

Keemink Y, Custers EJFM, van Dijk S, Ten Cate O. Illness script development in pre-clinical education through case-based clinical reasoning training. *Int J Med Educ.* 2018;9:35-41.

Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, Zipfel S, Herrmann-Werner A. Teaching history taking to medical students: a systematic review. *BMC Med Educ.* 2015;15:159.

Khin-Htun S, Kushairi A. Twelve tips for developing clinical reasoning skills in the pre-clinical and clinical stages of medical school. *Med Teach.* 2019;41(9):1007-1011.

Klein M, Otto B, Fischer MR, Stark R. Fostering medical students' clinical reasoning by learning from errors in clinical case vignettes: effects and conditions of additional prompting procedures to foster self-explanations. *Adv Health Sci Educ Theory Pract.* 2019;24(2):331-351.

Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA.* 2009;302(12):1316-1326.

Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. *Med Educ.* 2011;45(10):1048-1060.

Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. *Acad Med.* 2003;78(8):802-809.

Lichstein PR. The Medical Interview. In: Walker HK, Hall WD, Hurst JW, editors. *Clinical methods: the history, physical, and laboratory examinations.* 3rd ed. Boston: Butterworths; 1990: Chapter 3.

Maguire GP, Rutter DR. History-taking for medical students. I-Deficiencies in performance. *Lancet.* 1976;2(7985):556-558.

Mahoney, MJ. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cogn Ther Res.* 1977; 1: 161-175.

Nazim SM, Talati JJ, Pinjani S, Biyabani SR, Ather MH, Norcini JJ. Assessing clinical reasoning skills using script concordance test (SCTsd) and extended matching questions (EMQs): A pilot for urology trainees. *J Adv Med Educ Prof.* 2019;7(1):7-13.

Nierenberg R. The chief complaint driven medical history: implications for medical education. *Int J Med Educ.* 2017;8:205-206.

Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach.* 2007;29(9):855-871.

Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39(4):418-427.

Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract.* 2009;14 Suppl 1:37-49.

Norman G, Monteiro S, Sherbino J. Is clinical cognition binary or continuous? *Acad Med.* 2013;88(8):1058-1060.

Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med.* 2017;92(1):23-30.

Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *West J Med.* 1992;156(2):163-165.

Pfeiffer C, Madray H, Ardolino A, Willms J. The rise and fall of students' skill in obtaining a medical history. *Med Educ.* 1998;32(3):283-288.

Pulito AR, Donnelly MB, Plymale M, Mentzer RM Jr. What do faculty observe of medical students' clinical performance? *Teach Learn Med.* 2006;18(2):99-104.

Schmitt BP, Kushner MS, Wiener SL. The diagnostic usefulness of the history of the patient with dyspnea. *J Gen Intern Med.* 1986;1(6):386-393.

Thampy H, Willert E, Ramani S. Assessing clinical reasoning: targeting the higher levels of the pyramid. *J Gen Intern Med.* 2019;34(8):1631-1636.

Tor E, Steketee C, Mak D. Clinical audit project in undergraduate medical education curriculum: an assessment validation study. *Int J Med Educ.* 2016;7:309-319.

Weidenbusch M, Lenzer B, Sailer M, Strobel C, Kunisch R, Kiesewetter J, Fischer MR, Zottmann JM. Can clinical case discussions foster clinical reasoning skills in undergraduate medical education? A randomised controlled trial. *BMJ Open.* 2019;9(9):e025973.

Wijnen-Meijer M, ten Cate OT, van der Schaaf M, Borleffs JC. Vertical integration in medical school: effect on the transition to postgraduate training. *Med Educ.* 2010;44(3):272-279.

Wijnen-Meijer M, Ten Cate O, van der Schaaf M, Burgers C, Borleffs J, Harendza S. Vertically integrated medical education and the readiness for practice of graduates. *BMC Med Educ.* 2015;15:229.

Woods NN. Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ.* 2007;41(12):1173-1177.

Yazdani S, Hosseinzadeh M, Hosseini F. Models of clinical reasoning with a focus on general practice: a critical review. *J Adv Med Educ Prof.* 2017;5(4):177-184.

Young M, Brooks L, Norman G. Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Med Educ.* 2007;41(12):1146-1151.

Zayyan M. Objective structured clinical examination: the assessment of choice. *Oman Med J.* 2011;26(4):219-222.

11. Anhang

1 Bewertungsbogen der „Clinical Reasoning“-Indikatoren im Anamnesegespräch 41



"Clinical Reasoning"-Indikatoren im Anamnesegespräch

Indikatoren	trifft nicht zu					trifft voll zu				
Kontrolle über das Gespräch übernehmen Die Studierenden übernehmen die Kontrolle über die Gesprächsführung, um so an erforderliche Informationen zu gelangen.	0	0	0	0	0	0	0	0	0	0
Erkennen von und reagieren auf relevante Informationen Die Studierenden zeigen, dass sie relevante Informationen erkennen, indem sie z.B. mit offensichtlichem Interesse auf diese reagieren.	0	0	0	0	0	0	0	0	0	0
Konkretisieren von Symptomen Die Studierenden stellen gezielte Nachfragen, um die Symptome, die sie als wichtig erachten, detaillierter zu erfassen.	0	0	0	0	0	0	0	0	0	0
Zeichen für pathophysiologisches Denken An den Fragen der Studierenden wird ersichtlich, dass sie sich gedanklich mit spezifischen Ursachen bestimmter Symptome auseinandersetzen.	0	0	0	0	0	0	0	0	0	0
Stellen der Fragen in einer logischen Reihenfolge Die Studierenden stellen die Fragen in einer logischen Reihenfolge und nicht nach einem Anamneseschema.	0	0	0	0	0	0	0	0	0	0
Rückversichern bei den Patienten Die Studierenden versichern sich durch Rückfragen bei den Patienten, dass ihre klinischen Gedankengänge auf richtigen Informationen basieren.	0	0	0	0	0	0	0	0	0	0
Zusammenfassen Die Studierenden fassen ihre gesammelten Informationen laut zusammen, sobald sie ein sinnvolles Maß erlangt haben.	0	0	0	0	0	0	0	0	0	0
Gesammelte Daten und Effektivität Die Studierenden sammeln ausreichend, qualitativ hochwertige Daten in angemessener Geschwindigkeit.	0	0	0	0	0	0	0	0	0	0

(1) (2) (3) (4) (5)
 0 0 0 0 0
 trifft nicht zu trifft eher nicht zu trifft teilweise zu trifft eher zu trifft voll zu

12. Lebenslauf

Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt

13. Danksagung

An dieser Stelle möchte ich mich bei allen denjenigen bedanken, die mich während meiner Anfertigung dieser Arbeit unterstützt haben:

Zunächst einmal geht mein besonderer Dank an Frau Prof. Dr. Sigrid Harendza für die hervorragende Betreuung dieser Arbeit. Die stets freundliche Unterstützung und guten Ratschläge waren für mich immer wieder entscheidend, um den Faden nicht zu verlieren und haben mich stets ermutigt weiter voranzugehen.

Zudem Danke ich Frau Sophie Fürstenberg für die gute Kooperation und Unterstützung bei Fragen zur statistischen Analyse.

Ein großer Dank geht auch an alle Organisatoren und Studierenden, die im Zuge des ÄKHOM-Projektes die Grundlagen für diese Arbeit geschaffen haben.

Zuletzt möchte ich meiner Familie danken für die liebevolle Unterstützung und Ermutigung während der Anfertigung dieser Arbeit und während des gesamten Studiums. Besonders danke ich meiner Mutter und meinem Bruder Jonathan für den klaren Blick von außen, der so manchen sprachlichen oder gedanklichen Knoten lösen konnte.

14. Eidesstattliche Erklärung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: