

# **UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

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

## **Analyse der Qualität der Übergabegespräche von Studierenden im Praktischen Jahr im Rahmen eines simulierten ersten Arbeitstages im Krankenhaus**

### **Dissertation**

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

vorgelegt von

André Ingenabel  
aus Düsseldorf

Hamburg 2021

**Angenommen von der Medizinischen Fakultät der Universität Hamburg am:  
21.06.2021**

**Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität  
Hamburg.**

**Prüfungsausschuss, der Vorsitzende: Prof. Dr. Alexander Schwoerer**

**Prüfungsausschuss, zweite Gutachterin: Prof. Dr. Sigrid Harendza**

# Inhaltsverzeichnis

1	Einleitung .....	1
1.1	Bedeutung des Übergabegespräches im klinischen Alltag.....	1
1.2	Übergabegespräche im Medizinstudium.....	3
1.3	Bewertungszugänge der Qualität einer Übergabesituation .....	5
1.4	ÄKHOM (Ärztliche Kompetenzen: Hamburg-Oldenburg-München)-Projekt .....	7
2	Arbeitshypothese und Fragestellung .....	9
3	Material und Methoden .....	10
3.1	Ablauf und Strukturierung einer 360°-Kompetenzprüfung .....	10
3.2	Übergabegespräche (Phase 4) einer 360°-Kompetenzprüfung.....	11
3.3	Simulierte Fallszenarien .....	11
3.3.1	Fallkonstellation A .....	11
3.3.2	Fallkonstellation B .....	12
3.4	Prüfungsteilnehmende .....	12
3.5	Entwicklung des HOPERA-Bewertungsbogens inklusive Ankerbeispielen zur Analyse der Übergabegespräche .....	13
3.6	Pilottestung und Optimierung des HOPERA-Bewertungsbogens .....	14
3.7	Analyse der Übergabegespräche mittels HOPERA .....	15
3.8	Statistische Analyse .....	15
4	Ergebnisse .....	17
4.1	Ermittlung der Bewertenden-Übereinstimmung .....	17
4.2	Bewertung der Übergabegespräche insgesamt und auf Itemebene .....	18
4.2.1	Operationalisierte Bewertungen (HOPERA) auf Itemebene.....	18
4.2.2	Peerbewertungen (HOPA) auf Itemebene .....	19
4.2.3	Auswertung der Gruppenunterschiede der Gesamtleistung.....	20
4.3	Vergleich der Peerbewertungen mit operationalisierten Fremdbewertungen .....	22
4.4	Faktorenanalyse basierend auf den operationalisierten Bewertungen .....	24
5	Diskussion .....	27
6	Zusammenfassung .....	35
7	Summary .....	36

8	Abkürzungsverzeichnis .....	37
9	Abbildungsverzeichnis .....	38
10	Tabellenverzeichnis.....	39
11	Literaturverzeichnis.....	40
12	Anhang .....	46
12.1	Anhang 1 – HOPA.....	46
12.2	Anhang 2 – HOPERA.....	47
12.3	Anhang 3 – HOPERA operationalisierte Bewertung .....	48
13	Danksagung.....	51
14	Lebenslauf .....	52
15	Eidesstattliche Erklärung.....	53

# 1 Einleitung

## 1.1 Bedeutung des Übergabegespräches im klinischen Alltag

Die Übergabesituation im ärztlichen Schichtdienst stellt eine wichtige Grundlage für die Übermittlung klinischer Informationen zu Patientinnen und Patienten im Sinne des korrekten ärztlichen Arbeitens dar. Sie ist somit ein elementarer Bestandteil des medizinischen Alltags und tritt bedingt durch Regulierungsbestrebungen der Arbeitszeit mit einhergehender Umstrukturierung der Schichtdienste zunehmend gehäuft auf (Horwitz et al. 2006). Laut Jeffcott et al. (2009) lässt sich die Patientenübergabe wie folgt definieren: „Patientenübergabe ist der Transfer von Patienteninformation, professioneller Zuständigkeit und Verantwortlichkeit zwischen einzelnen Personen und Teams innerhalb des gesamten Gesundheitssystems“.

Die klinische Übergabe von Patientendaten findet in diversen Situationen statt: anlässlich des Schichtwechsels oder bei Patientenaufnahme, -verlegung, -verweisung und -entlassung. Der Begriff der Patientenübergabe wird daher in der Literatur nicht konsistent für eine bestimmte Situation im klinischen Kontext genutzt, sondern beschreibt Übergabeinteraktionen von jedweder routinemäßigen Kommunikation bis hin zu solcher in spezifischen Bereichen, wie beispielsweise dem perioperativen Prozess.

Die Übergabe ist in Hinblick auf die Behandlungsstrategie sowie die Sicherheit der Patientinnen und Patienten von entscheidender Bedeutung. Dabei kann das Übergabegespräch als Zusammenspiel vieler komplexer dynamischer Prozesse verstanden werden, die neben dem reinen Informationstransfer auch Elemente der Diskussion sowie Supervision beinhalten (Yee et al. 2006). Die Übergabesituation gestaltet sich äußerst variabel, meist unstrukturiert und dadurch besonders fehleranfällig (Manser und Foster 2011). Denson et al. (2015) demonstrierten, dass mit zunehmender Anzahl der erforderlichen Übergaben die Mortalität eines Patienten während eines Krankenhausaufenthaltes zunimmt. Als einer der Hauptfaktoren, die die Patientensicherheit gefährden, lassen sich Kommunikationsfehler ausmachen. Dies zeigte eine Analyse unerwünschter Ereignisse, bei der Kommunikationsfehler neben Ausbildungsdefiziten als entscheidend beitragender Faktor zur Reduktion der Patientensicherheit beobachtet werden konnten (Pronovost et al. 2006). In Betrachtung aller schweren unerwünschten Ereignisse in der Medizin, tragen laut der US-amerikanischen Akkreditierungsorganisation Joint Commission Kommunikationsfehler in 65% der Fälle zu diesen bei (Merkel et al. 2017). Dies äußert sich beispielsweise in Verzögerungen bei Diagnostik und Therapie, Behandlungsfehlern, erhöhten Kosten und verlängerten Krankenhausaufenthalten (Wong et al. 2008, World Health Organization 2020). Der Übergabesituation sind hohe kommunikative Anforderungen inhärent. Oftmals ist die Qualität der Übergabe aufgrund des Mangels an übermittelten Informationen nicht ausreichend und

führt so zu unerwünschten Ereignissen (Horwitz et al. 2009a, Horwitz et al. 2009b). Bei diesen unerwünschten Ereignissen handelt es sich um vermeidbare Zwischenfälle oder Situationen, in denen weitere Informationen hilfreich gewesen wären (Borowitz et al. 2008). Im Zuge einer Untersuchung von 334 berichteten Zwischenfällen aufgrund von problematischen Übergabegesprächen konnte neben einer unvollständigen Übergabe (45,2%) ein komplettes Ausbleiben der Übergabe (29,3%) als zweithäufigste Ursache beobachtet werden (Pezzolesi et al. 2010). Eine weitere Hauptfehlerquelle innerhalb der Kommunikation während eines Übergabegesprächs war neben einem unzureichenden Informationsaustausch die Wahl eines fehleranfälligen Kommunikationsprozesses, wie beispielsweise der Austausch über das Telefon oder in schriftlicher Form mit unleserlichen Anmerkungen (Arora et al. 2005).

Die World Health Organisation (WHO) forderte erstmals im Jahr 2005 ein strukturiertes Übergabekonzept für das Gesundheitssystem, das alle relevanten Informationen in einer definierten Reihenfolge erfasst (Merkel et al. 2017). Im Zuge dessen wurden im medizinischen Kontext standardisierte Übergaben mittels Checklisten etabliert, die unter anderem durch Vermeidung von Kommunikationsfehlern zur Verbesserung der Patientensicherheit beitragen (Müller et al. 2018). Ein systematischer Review von Riesenberg et al. (2009) identifizierte 24 unterschiedliche standardisierte Übergabekonzepte. Am häufigsten wird dabei die Verwendung der SBAR-Mnemotechnik zur Strukturierung von Kommunikationsabläufen erwähnt. SBAR ist ein englisches Akronym und beinhaltet die vier essentiellen Punkte: Situation, Background (engl. Hintergrund), Assessment (engl. Einschätzung) und Recommendation (engl. Empfehlung). Seit 2016 wird der Gebrauch des SBAR-Schemas durch die Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin (DGAI) für diesen Bereich empfohlen (von Dossow und Zwissler 2016). Das SBAR-Konzept findet insbesondere in Hochrisikosituationen Verwendung und ist dem Militärbereich entlehnt. Informationen sollen dort schnell, effektiv und konsistent übergeben werden. So wird es im medizinischen Kontext in erster Linie im Zuge des perioperativen Managements eingesetzt. Aufgrund multipler Übergabesituationen in kürzester Zeit handelt es sich hierbei um einen besonders anfälligen Prozess in Hinsicht auf Kommunikationsfehler. Bei Verwendung des SBAR-Konzeptes konnte ein signifikanter Rückgang der Anzahl unerwarteter Todesfälle beobachtet werden (De Meester et al. 2013). Diese Erkenntnis steht in einer Linie mit einer weltweit durchgeführten Studie, die einen positiven Effekt auf Morbidität und Mortalität chirurgischer Patientinnen und Patienten durch verbesserte Teamkommunikation nach Einführung eines standardisierten präoperativen Briefings zeigte (Haynes et al. 2009). Über die alleinige Verwendung des SBAR-Konzeptes hinaus konnte im Rahmen einer multimodalen Intervention mit zusätzlichem didaktischem Webinar und Simulationstraining die Qualität der Übergabe nachhaltig verbessert werden (Weinger et al. 2015). Einen entscheidenden Punkt standardisierter

Übergaben stellt der Gebrauch von automatisch generierten Übergabeprotokollen aus der elektronischen Patientenakte dar. Lediglich mündlich übergebene Informationen weisen eine Diskrepanz zum Umfang jener in der Patientenakte auf, beziehungsweise auch verglichen mit ersichtlichen Informationen eines vorgegebenen Übergabeprotokolls (Catchpole et al. 2007, Bhabra et al. 2007). Der Einsatz automatisch generierter Übergabeprotokolle auf Basis der elektronischen Patientenakte funktioniert zuverlässig und führt zu zunehmender Vollständigkeit der relevanten Patienteninformationen (Pickering et al. 2009, Flanagan et al. 2009). Ein systematischer Review von Arora et al. (2009) analysierte Interventionen zur Verbesserung der Patientenübergabe und konstatierte, dass die Verwendung technologiebasierter Unterstützungssysteme mit einer reduzierten Anzahl vermeidbarer unerwünschter Ereignisse einhergeht. Darüber hinaus ließ sich eine gesteigerte Zufriedenheit der Teilnehmenden mit der Qualität der Übergabe beobachten, wenn technologiebasierte Unterstützungssysteme eingesetzt wurden (Ram und Block 1992). Vor dem Hintergrund dieser empirischen Belege zeigt sich die Notwendigkeit standardisierte Vorgehensweisen zu etablieren. Dafür ist es von entscheidender Bedeutung zu untersuchen, welche Aspekte eine gut gelungene Übergabe definieren und deren Qualität bestimmen. Riesenberg et al. (2009) identifizierten im Zuge ihres systematischen Reviews 24 standardisierte Übergabetechniken, von denen experimentelle Untersuchungen selten sind. In einer experimentellen Untersuchung durch Bhabra et al. (2007) sah die Aufgabe vor, die im Rahmen einer Übergabe erhaltenen Patienteninformationen einer anderen Person zu übergeben. Hier wurde experimentell die Art und Weise der Übergabe dahingehend manipuliert, dass diese entweder rein verbal oder notizgestützt oder mit Hilfe eines computergenerierten Übergabeprotokolls erfolgte. Allerdings fehlt in dieser Studie der Bezug zum klinischen Arbeiten, da die an der Übergabe beteiligten Personen nie in irgendeinem klinischen Bezug zu den entsprechenden Patientinnen und Patienten standen und die Übergabe keine selbst generierten neuen Informationen enthielt (Bhabra et al. 2007).

## **1.2 Übergabegespräche im Medizinstudium**

Im Rahmen des Medizinstudiums spielen Schulungsmaßnahmen zu Übergabetechniken an den meisten Universitäten eine untergeordnete Rolle. Es werden wenige bis keine Lehrveranstaltungen angeboten, die eine relevante Vorgehensweise hinsichtlich der strukturellen, inhaltlichen sowie kollegialen Gestaltung solcher Übergabegespräche vermitteln. Eine Umfrage von Stojan et al. (2015) zeigte, dass nur 35% der in den USA befragten ärztlichen Berufsanfängerinnen und Berufsanfänger während des Medizinstudiums im Umgang mit Übergaben geschult wurden. Weitere 51% der Befragten gaben an, wenigstens Feedback zu ihren Übergabefähigkeiten erhalten zu haben (Stojan et al. 2015). Eine US-amerikanische Studie von Horwitz et al. (2006) konnte bei 202 erfassten Programmen für

ärztliche Berufsanfängerinnen und Berufsanfänger im Bereich der Inneren Medizin nur in 60% Schulungsmaßnahmen zu Übergabetechniken identifizieren. Eine weitere Untersuchung von Sinha et al. (2007) ergab, dass in 185 explorierten US-amerikanischen Notaufnahmen nur in 25,6% didaktische Inhalte mit Fokus auf die Übergabe vermittelt wurden. Die Mehrheit der Ausbildenden (71,6%) gab an, dass die Patientenversorgung durch entsprechende Handlungsleitlinien im Rahmen der Übergabe verbessert werden kann. 72,3% der Ausbildenden äußerten, dass eine standardisierte Übergabe die Kommunikation verbessere und das Auftreten von konsekutiven Fehlern reduziere (Sinha et al. 2007). Ohne entsprechende Inhalte zum Umgang mit Übergabesituationen sind Berufsanfängerinnen und Berufsanfänger schlecht vorbereitet (Cleland et al. 2009, Monrouxe et al. 2017). Ein Review von Masterson et al. (2013) zum Thema Ressourcen für Schulungsmaßnahmen rund um die Übergabe bestätigt, dass eine Schulung zu Übergabetechniken Berufsanfängerinnen und Berufsanfänger selbstsicherer mit entsprechenden Situationen umgehen lässt. Eine weitere Studie konnte zeigen, dass sich im Rahmen simulierter Übergaben über das Telefon durch Medizinstudierende und junge Ärztinnen und Ärzte die Kommunikation verbesserte, wenn das standardisierte Übergabekonzept SBAR Verwendung fand (Cunningham et al. 2012). Es existiert wenig wissenschaftliche Literatur zum Thema Schulungsmaßnahmen für Übergabegespräche. Laut eines Reviews von Gordon und Findley (2011) konnten zum damaligen Zeitpunkt zehn Interventionsstudien ausgemacht werden, die einen Lernzuwachs von Studierenden der Medizin, Pflegekräften sowie Ärzten und Ärztinnen im Umgang mit Übergabesituationen nachweisen konnten. Rollenspiele und Simulationsübungen wurden durch Teilnehmende der Interventionen gegenüber rein didaktischen Seminaren als besser geeignet wahrgenommen (Desmedt et al. 2021). Die vermittelten Inhalte wurden nachhaltig internalisiert und konnten auf spätere Situationen am Arbeitsplatz übertragen werden (Gakhar und Spencer 2010). Ein aktuelles deutsches Pilotprojekt an der Rheinisch-Westfälischen Technischen Hochschule (RWTH) Aachen, das ein Training von Übergabesituationen in das reformierte Curriculum implementierte, führte sowohl zu einem gesteigerten Wissen und selbstsichererem Umgang mit standardisierten Übergabetools als auch zu einer verbesserten Qualität der Übergabegespräche (Thaeter et al. 2018). In diesem Projekt wurden an der medizinischen Fakultät der RWTH Aachen drei didaktische Module zur Übergabesituation in einen vierwöchigen Kurs des Themenblocks Allgemeinanästhesie und Intensivmedizin integriert. Weitere Projekte dieser Art fehlen an anderen deutschen medizinischen Fakultäten. Dies steht nicht in einer Linie mit den Forderungen der WHO, welche die Implementierung von Unterrichtsinhalten in das Curriculum fordert, die die Patientensicherheit stärken (Walton et al. 2011). Die Übergabesituation leistet einen entscheidenden Beitrag zur Patientensicherheit (Manser und Foster 2011). Studierenden der Medizin ist es wichtig, Ausbildung unter dem Aspekt der Patientensicherheit zu erfahren (Walton et al. 2011). Die Umsetzung von weiteren

Projekten wie jenes von Thaeter et al. (2018) sind daher zukünftig im Interesse der Studierenden der Medizin wünschenswert. Die laufende Reformierung des Studiums der Humanmedizin an vielen deutschen Fakultäten bietet derzeit die Möglichkeit, frühzeitig klinisch relevante Themen in der Lehre zu adressieren. Studierende eines Modellstudienganges, wie an der RWTH Aachen, sollten voraussichtlich eher mit der Relevanz der Übergabe konfrontiert worden sein. Inwieweit sich Studierende eines Regelstudienganges von denjenigen eines reformierten Curriculums in Hinblick auf die Qualität der Übergabe unterscheiden, ist jedoch unbekannt.

### **1.3 Bewertungszugänge der Qualität einer Übergabesituation**

Je nachdem unter welchem primären Gesichtspunkt die Übergabe bewertet wird, kann die wahrgenommene Qualität variieren. Um die Komplexität der Übergabe adäquat abzubilden, muss die Prüfung der Qualität einer Übergabe in drei Säulen untergliedert werden, die mittels jeweiliger Messung (1) den Inhalt, (2) den Prozess oder (3) die Ergebnisse evaluieren (Cheung et al. 2010).

Die Beurteilung der Qualität des Inhalts einer Übergabe umfasst dabei den Anteil der übermittelten relevanten Schlüsselinformationen. Bhabra et al. (2007) legten beispielsweise in ihrem experimentellen Vergleich unterschiedlicher Übergabemethoden einen Umfang von je 20 Schlüsselinformationen für alle Patientinnen und Patienten fest. Diese sollten repräsentativ für Informationen stehen, die typischerweise im Rahmen der Übergabe vermittelt werden. Zudem erfolgte eine Gewichtung der 20 Schlüsselinformationen in essentielle und weniger bedeutsame Informationen zu der jeweiligen Patientin beziehungsweise dem jeweiligen Patienten (Bhabra et al. 2007). Bisherige Untersuchungen, die die Qualität einer Übergabe erfassen, haben sich auf den Aspekt der Genauigkeit und Vollständigkeit der übermittelten Informationen konzentriert (Bhabra et al. 2007, Wayne et al. 2008). Die Qualität einer Übergabe bemisst sich in Studien dieser Art hauptsächlich an der Adhärenz zum Übergabeprotokoll, also einer vollständigen Vermittlung der zuvor festgelegten Schlüsselinformationen. Unter dem Aspekt der Vollständigkeit wird auch die korrekt wiedergegebene Anzahl der zu übergebenden Patientinnen und Patienten verstanden (Van Eaton et al. 2005). Darüber hinaus gilt es, weitere Aspekte entsprechend der von Cheung et al. (2010) beschriebenen Säulen der Qualitätsmessung einer Übergabe zu berücksichtigen.

Messungen der Prozessqualität der Übergabe fokussieren sich insbesondere auf die Verhaltensebene. Auf Verhaltensebene spielt die Einbeziehung des Gegenübers im Sinne des gemeinsamen Planens sowie des gemeinsamen Entscheidungsfindens eine entscheidende Rolle und bildet die Teamwork-Fähigkeit in Übergabesituationen ab. Die

Übergabeempfängerin beziehungsweise der Übergabeempfänger gestaltet die Qualität des Prozesses der Übergabe mit, indem sie oder er die bestehende Dokumentation kritisch hinterfragt und die erhaltenen Informationen kommentiert und abschließend zusammenfasst. Einen weiteren Unterpunkt der Prozessqualität stellen die Umfeldbedingungen der Übergabesituation dar und dies beinhaltet unter anderem das Auftreten von Unterbrechungen einer Übergabe, die dem laufenden klinischen Betrieb geschuldet sind (Cheung et al. 2010). Die Messung der Prozessqualität hat bisher nur wenig Einzug in Studien gehalten und wird beispielsweise von Manser et al. (2010) in ihrem Bewertungszugang der Qualität der Übergabe berücksichtigt. Manser et al. (2010) zeigten unter anderem auf, dass die Arbeitsatmosphäre, im Sinne von Spannungen zwischen den Beteiligten, einen entscheidenden Beitrag zur Qualität der Übergabe leistet.

Das Ergebnis einer Übergabe lässt sich zum einen anhand der Zufriedenheit der Teilnehmenden messen, zum anderen an den relevanten Konsequenzen hinsichtlich der Patientensicherheit. Da die Übergabesituation immer zwei Perspektiven beinhaltet, ist es wichtig, dass sowohl diejenige des Übergabenden, als auch jene des Empfangenden in die Bewertung der Übergabe einfließen. Cheung et al. (2010) haben einen Vorschlag erarbeitet, welche Aspekte der Übergabe sich auf die Patientensicherheit auswirken und wie sich das Ergebnis einer Übergabe darauf basierend quantifizieren lässt. Unter anderem berücksichtigt ihr Bewertungszugang die Ebene eines „shared mental model“, welches beinhaltet, ob während der Übergabe geplante Aufgaben nachfolgend im klinischen Betrieb umgesetzt werden oder zum Beispiel zukünftige Patientenergebnisse antizipiert werden können (Cheung et al. 2010).

Laut einer Analyse der multidimensionalen Struktur der Übergabesituation durch Manser et al. (2010) lässt sich die Qualität der Übergabe insbesondere an drei Faktoren festmachen: (1) Informationsaustausch, (2) einheitliches Verständnis und (3) Arbeitsatmosphäre. Diese Faktoren finden sich auch in der von Cheung et al. (2010) beschriebenen Dreiteilung der Evaluation einer Übergabe wieder. Der Faktor des Informationsaustausches geht in dem von Cheung et al. (2010) beschriebenen Modell in der Säule der Messung des Inhalts auf. Unter das einheitliche Verständnis fällt in der Analyse von Manser et al. (2010) die Sicherstellung der Vollständigkeit der Übergabe, das aktive Erörtern von Fragen und Unklarheiten sowie das Antizipieren von möglichen Risiken und Komplikationen im Sinne eines geschlossenen Kommunikationszyklus. Der Faktor des einheitlichen Verständnisses wird durch die Messung des Ergebnisses einer Übergabe abgebildet, welche das Etablieren eines „shared mental model“ berücksichtigt. Die Beschreibung des „shared mental model“ durch Cheung et al. (2010) bezieht sich dabei auf ein gemeinsames Verständnis des Krankheitsbildes des

Patienten, des jüngsten klinischen Verlaufs, der verabreichten Therapien, der Gründe für ausstehende diagnostische Tests und Therapien sowie der wahrscheinlichen Ursache der Erkrankung. Der dritte von Manser et al. (2010) herausgearbeitete Faktor der Arbeitsatmosphäre wird durch die von Cheung et al. (2010) beschriebene Messung der Prozessqualität berücksichtigt. Die Messung der Prozessqualität nach Cheung et al. (2010) greift dabei auch auf Aspekte des Faktors des einheitlichen Verständnisses nach Manser et al. (2010) zurück, indem sie die kritisch hinterfragende, kommentierende sowie resümierende Mitwirkung der Übergabeempfängerin beziehungsweise des Übergabeempfängers als Qualität des Prozesses versteht. Messungen der Prozessqualität und der Ergebnisqualität sind also nicht gänzlich als distinkt zu betrachten, da beide Aspekte der Evaluation des einheitlichen Verständnisses beinhalten.

Alle Beobachtungen von Cheung et al. (2010) und Manser et al. (2010) zusammengenommen, scheint es bei der Evaluation der Qualität einer Übergabe naheliegend, dass ihre komplexe Struktur durch die Berücksichtigung dreier Faktoren adäquat abgebildet werden kann. Diese setzen sich im Wesentlichen mit den folgenden Punkten auseinander: (1) Inhalt, (2) Ergebnisse im Sinne eines einheitlichen Verständnisses oder der Zufriedenheit der Teilnehmenden und (3) Prozesseigenschaften, wie zum Beispiel die Arbeitsatmosphäre.

## **1.4 ÄKHOM (Ärztliche Kompetenzen: Hamburg-Oldenburg-München)-Projekt**

Das durch das Bundesministerium für Bildung und Forschung (BMBF) geförderte Verbundprojekt Ärztliche Kompetenzen Hamburg, Oldenburg, TU München (ÄKHOM) ermöglichte erstmals in einer validierten Simulation eines ersten Arbeitstages im Krankenhaus den direkten Vergleich ärztlicher Kompetenzen von Medizinstudierenden dreier verschiedener Hochschulen mit unterschiedlichen Curricula (Prediger et al. 2020). Das ÄKHOM-Projekt ging aus einem bereits validierten internationalen Projekt der kooperierenden Universitäten Hamburg und Utrecht zur Erprobung eines neuen kompetenzbasierten Prüfungsmodells für Medizinstudierende im letzten Studienjahr (Praktisches Jahr, PJ) hervor (Wijnen-Meijer et al. 2013b). Der erste Durchlauf im Jahr 2017 demonstrierte die Umsetzbarkeit des Prüfungsmodells mit hoher Zufriedenheit der Teilnehmenden (Harendza et al. 2017a). Inhalt des kompetenzbasierten Prüfungsmodells ist eine Simulation des ersten Arbeitstages einer Ärztin oder eines Arztes in der Weiterbildung im Krankenhaus. Die Auswahl der zu evaluierenden Kompetenzen basiert auf einer Ranking-Studie durch berufserfahrene Medizinerinnen und Mediziner (Fürstenberg et al. 2017). Bei dem Prüfungsmodell handelt es sich um eine 360°-Kompetenzprüfung, da die Kompetenzen der PJ-Studierenden sowohl anhand eines 360°-Feedbacks, als auch mittels einer Selbsteinschätzung evaluiert werden

(Harendza et al. 2017a). Das 360°-Feedback beinhaltet eine Bewertung der PJ-Studierenden durch die teilnehmenden Schauspielpatientinnen und -patienten, Oberärztinnen und Oberärzte, Peers und Pflegekräfte. Essentielle Bestandteile der Simulation sind eine Sprechstunde sowie das anschließende Patientenmanagement und eine abschließende Übergabesituation (Harendza et al. 2017a). Das ÄKHOM-Projekt bietet Medizinstudierenden in ihrem letzten Studienjahr daher die Möglichkeit, den Umgang mit Übergabesituationen in größerem Umfang zu üben.

Im Rahmen dieser Arbeit fand zudem ein neues Messinstrument zur Evaluation der Qualität von Patientenübergaben Verwendung, welches auf Teilaspekten des im Rahmen des ÄHKOM-Projektes validierten „Facets of Competence“-Modells (FOC) basiert (Prediger et al. 2019). Auf Grundlage dieses Kompetenzmodells kam bisher mit dem „Handover Partner Assessment“-Fragebogen (HOPA) innerhalb des ÄKHOM-Projektes ein Messinstrument zur Evaluation der Qualität der Übergabe durch die Übergabeempfängerin beziehungsweise den Übergabeempfänger, sogenannte Peer-Evaluation, zum Einsatz (Prediger et al. 2020). Um einen weiteren Bewertungszugang der Qualität einer Übergabe zu ermöglichen, welcher frei von möglichen Verzerrungen der Peer-Evaluation ist, wurde der HOPA-Fragebogen in der vorliegenden Arbeit dahingehend zu dem neuen Messinstrument „Handover Performance External Rater Assessment“ (HOPERA) modifiziert und operationalisiert, sodass eine Bewertung der Qualität der Übergabe durch externe Beobachter realisiert werden konnte. Im Zuge dessen war es im Rahmen des ÄKHOM-Projektes erstmals möglich, die Qualität der Übergabegespräche von PJ-Studierenden unterschiedlicher Hochschulen mit divergierender Ausrichtung des Curriculums unabhängig von der Peerevaluation zu ermitteln und zu vergleichen.

## 2 Arbeitshypothese und Fragestellung

Übergabesituationen sind essenzieller Bestandteil der medizinisch-klinischen Tätigkeit und ein kritisches Element im Hinblick auf die Patientensicherheit. Während des Medizinstudiums werden in der Regel keine Unterrichtsveranstaltungen zur strukturellen, inhaltlichen sowie kollegialen Gestaltung solcher Übergabegespräche angeboten. Einzelne Aspekte der Übergabe werden im Rahmen von klinisch-praktischen Veranstaltungen am Krankenbett oder im Praktischen Jahr zumeist indirekt gelernt, bei denen im Nachgang einer verantwortlichen Ärztin oder einem verantwortlichen Arzt oder einer Kommilitonin oder einem Kommilitonen über entsprechende Patientenkontakte zu berichten ist. In diesen Situationen bleibt jedoch mehrheitlich ein Feedback aus. Es ist unklar, wie die Qualität der Übergabe unter Studierenden des Praktischen Jahres im deutschsprachigen Raum ausfällt. Eine 360°-Kompetenzprüfung eines ersten simulierten Arbeitstages im Krankenhaus für PJ-Studierende in der Ärztinnen-beziehungsweise Arztrolle bot die Möglichkeit, die Qualität von Übergabegesprächen mit einem neu entwickelten Instrument systematisch zu erfassen und im Hinblick auf verschiedene qualitative Kriterien zu analysieren.

Die vorliegende Arbeit untersucht daher mit Hilfe der auf Video aufgezeichneten Übergabegespräche des kompetenzbasierten Simulationsmodells folgende Fragestellungen:

1. Welche Ausprägung haben spezifische Qualitätskriterien in Übergabegesprächen von PJ-Studierenden in der Ärztinnen- beziehungsweise Arztrolle?
2. Finden sich Unterschiede in der Übergabequalität zwischen Absolventinnen und Absolventen eines Modellstudienganges oder eines Regelstudienganges?
3. Gibt es Unterschiede in der qualitativen Bewertung der Übergabe durch die subjektive Perspektive der Übergabeempfängerin beziehungsweise des Übergabeempfängers verglichen mit einer operationalisierten Fremdbewertung?
4. Lassen sich grundlegende Faktoren identifizieren, die für die Übergabesituation relevant sind?

## **3 Material und Methoden**

### **3.1 Ablauf und Strukturierung einer 360°-Kompetenzprüfung**

Im Rahmen des BMBF-Verbundprojektes ÄKHOM (Ärztliche Kompetenzen: Hamburg-Oldenburg-München) wurde eine 360°-Kompetenzprüfung mit fünf Phasen als Simulation eines ersten ärztlichen Arbeitstages im Krankenhaus konzipiert und validiert (Prediger et al. 2020). Zuerst erfolgte ein 20-minütiges Briefing (Phase 1), gefolgt von einem 5-minütigen Gespräch mit der für die jeweiligen Teilnehmenden zuständigen Oberärztin beziehungsweise dem zuständigen Oberarzt. Diesem schloss sich eine 30-minütige Sprechstunde an (Phase 2), in der jeweils drei Teilnehmende drei unterschiedliche Simulationspatientinnen und -patienten antrafen. Insgesamt gab es sechs unterschiedliche simulierte Fallszenarien, die fest in zwei Fallkonstellationen eingeteilt wurden. In einem Durchlauf befanden sich also sechs Teilnehmende, von denen sich jeweils drei mit den gleichen simulierten Fallszenarien konfrontiert sahen. Die Anamnesegespräche wurden auf Video aufgezeichnet. Hierauf folgte eine 2,5-stündige Patientenmanagement-Phase (Phase 3). Die Teilnehmenden konnten in dieser Phase medizinische Testungen von einem virtuellen Labor durchführen lassen sowie radiologische Befunde anfordern. Der Zugang zu diesen medizinischen Diensten erfolgte per Laptop oder telefonisch. Darüber hinaus stand den Teilnehmenden jederzeit uneingeschränkter Internetzugang zur Verfügung, sodass weitere frei zugängliche Informationen eingeholt werden konnten. Die Aufgabe in dieser Phase bestand darin, auf Basis der gesammelten Informationen eine Behandlungsstrategie für die jeweiligen drei zu betreuenden Patientinnen und Patienten zu entwerfen. Während dieser Phase wurden die Teilnehmenden auch mit weiteren Aufgaben konfrontiert, die üblicherweise während des Klinikalltages auftreten. Beispielsweise kam es bei Interaktionen mit den Pflegekräften zu außerplanmäßigen Anforderungen und eingehende Telefonanrufe simulierter Kolleginnen oder Kollegen mussten beantwortet werden. Zu jedem Zeitpunkt hatten die Teilnehmenden die Möglichkeit ihre Oberärztin oder ihren Oberarzt telefonisch zu kontaktieren. Diese standen nach Ablauf der ersten Hälfte dieser Phase für fünf Minuten persönlich für Nachfragen und Diskussion des weiteren Vorgehens zur Verfügung. Nach Ablauf der Patientenmanagement-Phase kam es zu einem maximal 15-minütigen Übergabegespräch (Phase 4). In dieser Situation berichtete jeweils eine teilnehmende Person (übergabender Part) einer anderen teilnehmenden Person (übergabeempfangender Part) über ihre drei Patientinnen und Patienten. Dem übergabeempfangenden Part waren dabei die Fallszenarien des übergabenden Parts nicht bekannt. Nach jedem Übergabegespräch rotierte eine der teilnehmenden Personen in eine neue Dyade und erfüllte die jeweils andere Rolle (von übergabeempfangendem Part zu übergabendem Part oder vice versa). Die betreuenden Oberärztinnen beziehungsweise Oberärzte waren im Raum anwesend, interagierten aber nicht mit den Teilnehmenden. Diese Phase wurde ebenfalls auf Video aufgezeichnet. Ein 30-

minütiges Debriefing (Phase 5) markierte das Ende der Kompetenzprüfung. Teilnehmende hatten hier die Möglichkeit, ihre Erfahrungen im Rahmen des simulierten ersten Arbeitstages zu evaluieren. Diese Phase wurde als Audiospur aufgenommen.

### **3.2 Übergabegespräche (Phase 4) einer 360°-Kompetenzprüfung**

Der Schwerpunkt dieser Arbeit liegt auf der Analyse von Daten, die während der vierten Phase im Rahmen der Übergabesituation gesammelt wurden. Im Anschluss an das Übergabegespräch (Phase 4) bewertete die Übergabeempfängerin beziehungsweise der Übergabeempfänger die jeweilige übergebende Gesprächspartnerin oder den Gesprächspartner mittels des "Handover Partner Assessment"- Fragebogens (HOPA) (Anhang 1). Die Items des HOPA-Fragebogens stammen aus Ankerbeispielen des Facets of Competence-Fragebogens (FOC), der bereits im Rahmen eines ersten Durchlaufes einer 360°-Kompetenzprüfung ebenfalls als Messinstrument zum Einsatz kam (Harendza et al. 2017). Das Konzept der Facets of Competence erfasst dabei folgende sieben Dimensionen: (1) Verantwortung, (2) Teamwork und Kollegialität, (3) Kennen von persönlichen Grenzen/Möglichkeiten, (4) Struktur, Arbeitsplanung und Dringlichkeit, (5) Umgang mit Fehlern, (6) wissenschaftlich begründete Arbeitsmethoden und (7) verbale Kommunikation mit Kolleginnen oder Kollegen sowie Supervisorinnen oder Supervisoren (Prediger et al. 2019). Die Analysen der FOC-Bewertungen legten nahe, dass Ankerbeispiele für eine objektivere Bewertung entwickelt werden sollten. Einzelne, neu konzipierte Ankerbeispiele für den FOC-Fragebogen wurden auf Basis inhaltlicher Überlegungen zur Konstruktion des HOPA herangezogen. Der HOPA-Fragebogen ist bisher nicht validiert worden. Die mittels des HOPA-Fragebogens gewonnenen Bewertungen der Übergebenden durch die jeweils zuhörenden Teilnehmenden stellen sogenannte Peerbewertungen dar, welche im Rahmen dieser Arbeit analysiert werden.

### **3.3 Simulierte Fallszenarien**

#### **3.3.1 Fallkonstellation A**

Ein 46-jähriger Patient stellt sich mit epigastrischen Schmerzen vor. Die Schmerzsymptomatik besteht seit vier Monaten und wurde bei Verdacht auf Gastritis mit einem unbekanntem Medikament behandelt, dessen Einnahme zu keiner klinischen Verbesserung führte. Hilfreich zur Diagnosefindung sind eine Ösophago-Duodenoskopie, die einen Soorbefall zeigt, sowie eine Überlegung zu den Gründen für das Auftreten von Soorösophagitis (Immunsuppression; in diesem Fall eine bisher nicht bekannte HIV-Infektion). Eine vertiefte Anamnese hätte einen Anhalt für einen Primäraffekt ergeben sowie für ein riskantes Sexualverhalten.

Ein 44-jähriger Patient imponiert klinisch auffällig mit starken abdominellen Schmerzen. Die Schmerzen sind seit acht Jahren bekannt und treten regelmäßig in anfallsweisen Episoden

auf. Sämtliche Diagnostik erfolgte bereits im Vorfeld ohne pathologischen Befund. Schmerztherapeutisch wurde der Patient zuvor mit diversen Opiaten in Kombination behandelt und zeigt Anzeichen für eine Abhängigkeit. Die Erhebung einer positiven Familienanamnese für Migräne sowie Erfragen auslösender Faktoren sind wegweisend für die Differentialdiagnose einer abdominellen Migräne.

Eine 46-jährige Patientin wird vom Hausarzt aufgrund eines postoperativen Kreatininanstiegs nach Cholezystektomie ohne Symptome zur weiteren Abklärung vorgestellt. Ursache für das akute Nierenversagen ist eine Hanta-Virus Infektion bei beruflich bedingtem Kontakt mit Stallungen und Tierexkrementen. Die Erhebung der Berufsanamnese und das Erstellen eines Blutbildes mit Thrombopenie, sowie der Nachweis des präoperativ normalen Kreatinins unter Ausschluss weiterer Ursachen für das akute Nierenversagen sind für die Diagnosefindung entscheidend.

### **3.3.2 Fallkonstellation B**

Ein 52-jähriger Patient leidet an postprandial zunehmenden Bauchschmerzen gepaart mit Übelkeit und Brechreiz. Die Schmerzsymptomatik besteht seit einer Woche und war bereits ein Jahr zuvor schon einmal aufgetreten. Klinisch wegweisend für eine chronische Cholezystitis sind das positive Murphy-Zeichen sowie ein auffälliger sonographischer Befund der Gallenblase bei leichten Entzündungszeichen.

Eine 35-jährige Frau verspürt plötzlich starke Unterleibschmerzen bei bestehendem Kinderwunsch und Aussetzen der Kontrazeption. Die Schmerzen sind zuvor mit geringerer Intensität über Wochen periodisch aufgetreten. Ursächlich ist eine stielgedrehte Ovarialzyste. Wegweisend sind der negative Schwangerschaftstest und eine Sonographie des Abdomens sowie eine anschließende gynäkologische intravaginale Sonographie.

Eine 48-jährige Patientin berichtet über abnehmende physische und kognitive Leistungsfähigkeit. Weiterhin gibt es einen Anhalt für Störungen des Affekts nach einem bariatrischen Eingriff in Form einer Sleeve-Gastrektomie bei Adipositas vor einem halben Jahr. Ursächlich ist ein Vitamin B6 Mangel bei ausgebliebender postoperativer Nachsorge. Zielführend ist die neurologische Untersuchung sowie Bestimmung des Vitamin-B-Status im Labor und ein typischer Aspekt im cMRT.

## **3.4 Prüfungsteilnehmende**

An dem Durchlauf der 360°-Kompetenzprüfung im Juli 2018 nahmen ausschließlich PJ-Studierende von drei verschiedenen Universitäten (Universität Hamburg, Universität

Oldenburg, TU München) mit jeweils unterschiedlichen Curricula teil: Universität Hamburg (vertikal-integrierter Modellstudiengang), Universität Oldenburg (vertikal-integriertes Curriculum in Kooperation mit der Universität Groningen), TU München (normaler Regelstudiengang). Insgesamt konnten 89 PJ-Studierende für die 360°-Kompetenzprüfung gewonnen werden. Von den 89 Teilnehmenden studierten 40 an der Universität Hamburg, 44 an der TU München und fünf an der Universität Oldenburg. 60 Teilnehmende waren weiblich und 29 männlich. Das mittlere Alter lag bei  $26,87 \pm 3,59$  Jahren. Frauen waren im Mittel  $26,47 \pm 3,7$  Jahre alt und Männer  $27,69 \pm 3,24$  Jahre. Als Anreiz zur Mitwirkung erhielten alle Teilnehmenden einen Buchgutschein im Wert von 25 Euro.

### **3.5 Entwicklung des HOPERA-Bewertungsbogens inklusive Ankerbeispielen zur Analyse der Übergabegespräche**

Aus dem HOPA-Fragebogen (Anhang 1), welcher der Peerbewertung dient, wurde ein weiterer analoger Bewertungsbogen, der „Handover Performance External Rater Assessment“ - HOPERA-Fragebogen, zur Fremdeinschätzung durch Dritte entwickelt (Anhang 2). Die Entwicklung umfasste eine Umformulierung der Items, sodass der HOPERA die Perspektive der Fremdbeobachtung ermöglichte; von „Meine Kollegin/ mein Kollege...“ zu „Die Studentin/ der Student...“. Weiterhin wurde basierend auf qualitativ-strukturellen Überlegungen die Reihenfolge der Items variiert. Es ergab sich folgende Clusterung der Items: vier Items zu medizinisch-inhaltlichen Aspekten (HOPERA Item 1,2,3,4), sowie vier Items zur medizinisch-inhaltlichen Einbindung der Kollegin beziehungsweise des Kollegen (HOPERA Item 5,6,8,9) und drei Items zum kollegialen Verhalten insgesamt (HOPERA Item 7,10,11). Ob sich diese Dimensionen auch durch statistische Verfahren extrahieren lassen, ist Gegenstand dieser Arbeit.

Der HOPERA besteht aus elf Items (Anhang 2) in folgender Reihenfolge: Item 1 beinhaltet, ob in einer schlüssigen Weise über die Patientinnen und Patienten berichtet wurde, so dass die Kollegin oder der Kollege kaum Nachfragen stellen musste. Item 2 evaluiert, ob Diagnostik- und Behandlungsschritte gut nachvollziehbar begründet wurden. Item 3 bewertet, ob für die Patientinnen und Patienten eine Strategie berichtet wurde, mit der die Kollegin oder der Kollege weiterarbeiten sollte. Item 4 erschließt, inwiefern relevante Ergebnisse der Patientenfälle in adäquater Sprache fokussiert berichtet wurden. Item 5 wertet, ob kommuniziert wurde, manche Dinge nicht erledigt beziehungsweise nicht geschafft zu haben. Item 6 beurteilt, ob auf verschiedene Art und Weise geäußert wurde, dass man nicht weitergekommen war. Item 7 hält fest, ob nicht Andere für Fehler verantwortlich gemacht wurden. Item 8 bildet ab, ob Ansichten, Wissen und Expertise der Kollegin oder des Kollegen im Gespräch berücksichtigt wurden. Item 9 evaluiert, ob nach der Meinung der Kollegin oder

des Kollegen zu den Patientinnen oder Patienten gefragt wurde. Item 10 erschließt, inwiefern sich gegenüber der Kollegin oder dem Kollegen respektvoll verhalten wurde. Item 11 beschäftigt sich abschließend damit, ob man sich kritikfähig gezeigt hatte.

Für jedes Item des HOPERA stand analog zum HOPA eine fünfstufige Likert-Skala (1: „trifft voll und ganz zu“ bis 5: „trifft überhaupt nicht zu“) zur Verfügung. Entsprechend jeder der fünf Bewertungsstufen der einzelnen elf Items wurden insgesamt 55 Ankerbeispiele entwickelt (Anhang 3). Diese Ankerbeispiele sollten die Objektivität der Fremdbeobachtung durch Dritte besser gewährleisten. Auf Basis des HOPERA-Fragebogens mitsamt seinen Ankerbeispielen wird im Rahmen dieser Arbeit die Qualität der Übergabe der Teilnehmenden anhand der Videoaufnahmen bewertet.

### **3.6 Pilottestung und Optimierung des HOPERA-Bewertungsbogens**

Im Zuge eines ersten Durchlaufs wurden die HOPERA-Bewertungsbögen mitsamt ihrer Ankerbeispiele von einer Ärztin (Betreuerin dieses Projektes) und einem Medizinstudenten (Promovierender) evaluiert. Es erfolgte dabei eine unabhängige Bewertung von zehn Videos, die jeweils fünfmal die Fallkonstellationen A und B beinhalteten. Die Videos wurden weiterhin nach dem Kriterium ausgewählt, dass keine teilnehmende Person sowohl in ihrer aktiven, als auch passiven Rolle zu sehen war. So sollte Verzerrungen der Bewertungen eines Teilnehmenden hinsichtlich eines weiteren Eindrucks über mehr als ein Video vorgebeugt werden. Jedes Video wurde zweimal hintereinander geschaut, um sich bei der Bewertung auf einzelne distinkte Aspekte konzentrieren zu können. So lag der Fokus beim ersten Mal gezielt auf den medizinisch-inhaltlichen Aspekten (Item 1 bis 4) und beim zweiten Mal auf den verbliebenen Unterpunkten. Zugunsten der Objektivität der operationalisierten Fremdbeobachtung wurden die Bewertungen im Anschluss hinsichtlich ihrer Interrater-Reliabilität ausgewertet, um aufzuzeigen inwieweit die Ergebnisse als vom Beobachter unabhängig anzusehen sind. Jene Bewertungen, die nur von einem Rater als „nicht beurteilbar“ eingestuft wurden, wurden nicht in die Analyse des Übereinstimmungsmaßes einbezogen. Wurde eine Bewertung von beiden als „nicht beurteilbar“ eingestuft, so ist diese als übereinstimmend in die Analyse eingegangen.

Nach Diskussion der Bewertungsunterschiede wurden die Ankerbeispiele auf Formulierungsebene sowie ihre Zuordnung zu den jeweiligen Bewertungsstufen überarbeitet und angepasst. In einem zweiten Durchlauf wurden erneut zehn weitere Videos nach selbigem Schema ausgewählt und gesichtet. Analog zum ersten Durchgang erfolgten eine erneute Auswertung hinsichtlich der Interrater-Reliabilität sowie eine Diskussion verbliebender

Bewertungsunterschiede. Die Ankerbeispiele wurden abschließend nochmals in ihrer Formulierung und ihrer Zuordnung zu den jeweiligen Bewertungsstufen leicht angepasst.

### **3.7 Analyse der Übergabegespräche mittels HOPERA**

Insgesamt 89 Videos mit einer Länge zwischen 5:24 min und 15:06 min wurden vom Promovierenden anhand des Bewertungsbogens HOPERA ausgewertet. Die Reihenfolge erfolgte zufällig. Jedes Video wurde zweimal gesichtet, um weiterhin analog zur Pilottesting bei jedem Durchlauf den Fokus in Bezug auf bestimmte Aspekte zu legen. Dem Promovierenden waren während der Auswertung keine weiteren Informationen zu den Teilnehmenden hinsichtlich ihres Studienortes oder anderer personenbezogenen Daten bekannt.

### **3.8 Statistische Analyse**

Insgesamt gingen die Daten von 82 Teilnehmenden in die Analyse ein. Die fünf Teilnehmenden aus Oldenburg wurden von der Analyse ausgeschlossen, da die Gruppe insgesamt zu klein war für einen Studienstandortvergleich. In einem Fall war eine teilnehmende Person aus Hamburg dem Promovierenden bekannt, sodass beide Übergabesituationen, in denen diese Person beteiligt war, nicht in der Analyse berücksichtigt wurden.

Nach erfolgreicher Sichtung des gesamten Videomaterials und abgeschlossener Bewertung wurden die Daten mit dem Statistikprogramm IBM® SPSS® Statistics 24.0 ausgewertet. Sowohl die durch die Analyse des Videomaterials gewonnenen Bewertungen über den HOPERA, als auch die Peerbewertungen über den HOPA wurden im Rahmen der statistischen Auswertung aufgrund der zugrundeliegenden fünfstufigen Likert-Skala als quasimetrisch behandelt. Die Interrater-Reliabilität wurde mittels prozentualer Übereinstimmung der operationalisierten Bewertungen (HOPERA) ermittelt. Zur Bestimmung der Qualität der Übergabegespräche wurde für jeden Teilnehmenden ein Prozentscore zur Erfassung seiner Gesamtleistung ermittelt (Gesamtscore in %). Insgesamt wurden dabei in Abhängigkeit der Raterperspektive zwei getrennte Gesamtscores berechnet; einer für die Peerbewertung (HOPA) und einer für die operationalisierte Fremdbewertung (HOPERA). Zur Konstruktion des Gesamtscores wurde der Mittelwert über alle beurteilten Items der operationalisierten Fremdbeziehungsweise Peerbewertung gebildet und prozentual an der möglichen Maximalleistung gemessen (zur besseren Verdeutlichung: Sind beliebig viele Items von möglichen elf Items als beurteilbar eingestuft worden und wurden im Mittel mit vier Punkten bewertet, ergibt dies gemessen an der Maximalleistung von fünf Punkten pro Item einen Gesamtscore von 80%). Es erfolgte eine deskriptive Auswertung der einzelnen Itembewertungen sowie des Gesamtscores getrennt für die jeweilige Raterperspektive. Zur Analyse der

Gruppenunterschiede zwischen den Hochschulstandorten auf Itemebene wurde über alle elf Items hinweg ein t-Test gerechnet. Bei kleiner Stichprobengröße und nicht erfüllter Grundannahme der Normalverteilung wurde auf eine verteilungsfreie Alternative wie dem Mann-Whitney-U-Test zurückgegriffen. Der Gesamtscore auf Basis der operationalisierten Fremdbewertung wurde zur besseren Annäherung an die Normalverteilung mittels der monomolecular model-Transformation  $\ln(101/(101-\text{Gesamtscore}))$  unter Zuhilfenahme einer Konstante  $[k=1$  (1% des Maximalscores mit 100 Punkten)] transformiert. Dies war nötig, da es Teilnehmende mit 100 Prozentpunkten gab, für deren Score aufgrund einer Division durch Null im entsprechenden Term sonst keine Transformation möglich gewesen wäre. Im ersten Schritt wurden für den transformierten Gesamtscore univariate Varianzanalysen mit dem Studienort als einzige unabhängige Variable gerechnet (einfaktorielle ANOVA). In anschließenden univariaten Varianzanalysen wurden weitere, sich möglicherweise auf den Gesamtscore auswirkende Kovariaten wie Geschlecht und Alter sowie studienspezifische Parameter (PJ-Tertial, Wahlfach bereits absolviert, Note im Zweiten Abschnitt der Ärztlichen Prüfung) in das Modell eingeschlossen (einfaktorielle ANCOVA). Levene-Tests zur Abklärung der Varianzhomogenität wurden bei jeder Varianzanalyse und t-Testung durchgeführt. Im Falle von inhomogenen Varianzen wurde ergänzend ein Welch-Test gerechnet.

Für eine vergleichende Auswertung der operationalisierten Fremd- zur Peerbewertung wurden t-Tests für zwei unabhängige Stichproben durchgeführt. Dabei wurden die beiden Bewertungsperspektiven sowohl getrennt für beide Hochschulstandorte, als auch in Summe einander gegenübergestellt. Hier wurde ebenfalls im Falle einer kleinen Stichprobengröße und nicht erfüllter Grundannahme der Normalverteilung auf eine verteilungsfreie Alternative wie dem Mann-Whitney-U-Test zurückgegriffen. In die Analyse der Unterschiede der Peerbewertungen von den operationalisierten Fremdbewertungen gingen analog zu den vorherigen Auswertungen die Daten von 82 Teilnehmenden ein.

Anschließend erfolgten explorative Faktorenanalysen zur Ermittlung von übergeordneten Kategorien innerhalb der Itemstruktur des HOPERA. Es wurden über alle Itembewertungen rotierte Faktorenanalysen (Hauptkomponente & Varimax) durchgeführt, um mögliche Unterskalen des HOPERA zu identifizieren. Die Eignung der entsprechenden Variablen für die Durchführung einer Faktorenanalyse wurde mittels bivariater Korrelationen, inverser Korrelationsmatrix und KMO-Wert sowie Bartlett-Test überprüft.

## 4 Ergebnisse

### 4.1 Ermittlung der Bewertenden-Übereinstimmung

Im ersten Probedurchlauf konnte insgesamt eine Übereinstimmung von 74,47% erreicht werden. Es sind 94 aus 110 Itembewertungen in die Analyse eingegangen (Tabelle 1).

Itemnummer	Summe der übereinstimmenden Bewertungen	Anzahl der berücksichtigten Bewertungen	Anzahl der übereinstimmenden Bewertungen in Prozent
Item 1	6	10	60,00
Item 2	5	10	50,00
Item 3	9	10	90,00
Item 4	6	10	60,00
Item 5	5	6	83,33
Item 6	3	4	75,00
Item 7	8	10	80,00
Item 8	5	9	55,56
Item 9	5	5	100,00
Item 10	8	10	80,00
Item 11	10	10	100,00
<b>Gesamt</b>	<b>70</b>	<b>94</b>	<b>74,47</b>

Tabelle 1: Prozentuale Übereinstimmung der Rater-Bewertungen beim ersten Probedurchlauf

Itemnummer	Summe der übereinstimmenden Bewertungen	Anzahl der berücksichtigten Bewertungen	Anzahl der übereinstimmenden Bewertungen in Prozent
Item 1	7	10	70,00
Item 2	7	10	70,00
Item 3	9	10	90,00
Item 4	7	10	70,00
Item 5	7	10	70,00
Item 6	7	10	70,00
Item 7	9	9	100,00
Item 8	8	9	88,89
Item 9	7	9	77,78
Item 10	9	10	90,00
Item 11	10	10	100,00
<b>Gesamt</b>	<b>87</b>	<b>107</b>	<b>81,31</b>

Tabelle 2: Prozentuale Übereinstimmung der Rater-Bewertungen beim zweiten Probedurchlauf

In Rahmen des zweiten Probedurchlaufs konnte über alle elf Items (Anhang 2) eine prozentuale Übereinstimmung von 81,31% erreicht werden. Hierbei sind 107 von 110 möglichen Itembewertungen in die Analyse eingegangen (Tabelle 2).

## 4.2 Bewertung der Übergabegespräche insgesamt und auf Itemebene

### 4.2.1 Operationalisierte Bewertungen (HOPERA) auf Itemebene

Die deskriptive Statistik auf Itemebene (Tabelle 3) zeigte, dass über alle Teilnehmenden hinweg die Leistung in Bezug auf Item 7 mit  $4,50 \pm 0,80$  Punkten am besten und Item 11 mit  $3,14 \pm 0,90$  Punkten am schlechtesten bewertet wurde (Tabelle 3). Es zeigte sich weiterhin, dass insbesondere bei Item 7 ( $n$  gesamt = 12) 70 Bewertungen und bei Item 11 ( $n$  gesamt = 7) 75 Bewertungen als „nicht beurteilbar“ eingestuft wurden. Somit lag über alle Teilnehmende hinweg im Falle von Item 7 in 14,63% und im Falle von Item 11 in 8,54% eine Bewertung vor. In Bezug auf alle anderen Items lagen über alle Teilnehmenden Bewertungen in mindestens 90% der Fälle vor. Unter diesen Items wurden über alle PJ-Studierende hinweg Item 10 mit  $4,48 \pm 0,88$  Punkten am besten und Item 9 mit  $3,91 \pm 1,16$  Punkten am schlechtesten evaluiert.

Itemnummer	Gesamt			Hamburg			München		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Item 1	82	4,07	1,05	38	3,84	1,05	44	4,27	1,02
Item 2	82	4,13	1,02	38	3,97	0,94	44	4,27	1,06
Item 3	82	4,46	0,80	38	4,34	0,85	44	4,57	0,76
Item 4	82	4,33	0,75	38	4,21	0,70	44	4,43	0,79
Item 5	82	4,38	0,86	38	4,34	0,81	44	4,41	0,90
Item 6	80	4,18	1,09	38	4,16	1,13	42	4,19	1,06
Item 7	12	4,50	0,80	5	4,00	1,00	7	4,86	0,38
Item 8	74	4,39	0,81	34	4,32	0,88	40	4,45	0,75
Item 9	74	3,91	1,16	34	3,53	1,16	40	4,23*	1,07
Item 10	82	4,48	0,88	38	4,47	0,92	44	4,48	0,85
Item 11	7	3,14	0,90	3	2,67	1,15	4	3,50	0,58
<b>Gesamtscore %</b>	82	85,23	13,33	38	82,98	13,43	44	87,18	13,08

**Tabelle 3:** Itemstatistik und Gesamtscore (%) der operationalisierten Fremdbewertung ( $n$  = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \* =  $p < ,05$ )

Die Betrachtung der jeweiligen Untergruppen der PJ-Studierenden ergab, dass Teilnehmende der Universität Hamburg im Mittel mit einer Leistung von  $2,67 \pm 1,15$  (Item 11) bis  $4,47 \pm 0,92$  Punkten (Item 10) bewertet wurden, wobei Item 9 mit  $3,53 \pm 1,16$  Punkten unter den Items, bei denen Bewertungen in mehr als 90% der Fälle vorlagen, am schlechtesten bewertet wurde.

Bei Teilnehmenden der TU München lag die Bewertung im Mittel bei  $3,50 \pm 0,58$  (Item 11) bis  $4,86 \pm 0,38$  Punkten (Item 7) beziehungsweise bei  $4,19 \pm 1,06$  (Item 6) bis  $4,57 \pm 0,76$  (Item 3) unter Berücksichtigung der Items, die Bewertungen in mindestens 90% der Fälle vorwiesen.

Eine Testung möglicher Gruppenunterschiede mittels t-Tests auf Itemebene konnte bei Item 9 (...hat nach der Meinung der Kollegin oder des Kollegen zu Patienten oder Patientinnen gefragt) einen signifikanten Unterschied hinsichtlich des Studienorts aufzeigen ( $t(72) = -2,675$ ,  $p = ,009$ ). Münchener-Teilnehmende schnitten dabei hinsichtlich Item 9 ( $4,23 \pm 1,07$ ;  $n = 40$ ) signifikant ( $p < ,05$ ) besser ab als Hamburger Teilnehmende bei Item 9 ( $3,53 \pm 1,16$ ;  $n = 34$ ). In Bezug auf andere Items und auf den Gesamtscore konnten keine weiteren signifikanten Unterschiede in Abhängigkeit des Hochschulstandorts ermittelt werden.

#### 4.2.2 Peerbewertungen (HOPA) auf Itemebene

Die deskriptive Statistik auf Itemebene (Tabelle 4) zeigte, dass über alle Teilnehmende hinweg die Leistung bei Item 10 mit  $4,95 \pm 0,22$  Punkten am besten und bei Item 9 mit  $3,06 \pm 1,27$  Punkten am schlechtesten bewertet wurde. Die Itemstatistik zeigte, dass insbesondere bei Item 11 ( $n$  gesamt = 58) 24 Bewertungen als „nicht beurteilbar“ eingestuft wurden. Somit lag über alle Teilnehmende hinweg im Falle von Item 11 in 70,73% eine Bewertung vor. In Bezug auf alle anderen Items lagen über alle Teilnehmenden Bewertungen in mindestens 90% der Fälle vor.

Itemnummer	Gesamt			Hamburg			München		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Item 1	82	4,10	0,73	38	3,97	0,75	44	4,20	0,70
Item 2	82	4,18	0,74	38	4,08	0,78	44	4,27	0,69
Item 3	82	4,24	0,84	38	4,18	0,80	44	4,30	0,88
Item 4	82	4,35	0,74	38	4,13	0,84	44	4,55*	0,59
Item 5	81	4,54	0,76	38	4,47	0,89	43	4,60	0,62
Item 6	77	3,68	1,20	37	3,62	1,28	40	3,73	1,13
Item 7	75	4,61	1,00	36	4,39	1,23	39	4,82	0,68
Item 8	80	4,49	0,62	37	4,46	0,69	43	4,51	0,55
Item 9	77	3,06	1,27	35	3,03	1,29	42	3,10	1,27
Item 10	82	4,95	0,22	38	4,95	0,23	44	4,95	0,21
Item 11	58	4,38	0,77	27	4,37	0,88	31	4,39	0,67
<b>Gesamtscore %</b>	82	84,82	9,00	38	82,95	9,92	44	86,44	7,89

**Tabelle 4:** Itemstatistik und Gesamtscore (%) der Peerbewertung, Itemreihenfolge an HOPERA angepasst ( $n$  = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \* =  $p < ,05$ )

Die Betrachtung der jeweiligen Untergruppen der PJ-Studierenden ergab, dass Teilnehmende der Universität Hamburg im Mittel mit einer Leistung von  $3,03 \pm 1,29$  (Item 9) bis  $4,95 \pm 0,23$  Punkten (Item 10) bewertet wurden. Bei Teilnehmenden der TU München lag die Bewertung im Mittel bei  $3,10 \pm 1,27$  (Item 9) bis  $4,95 \pm 0,21$  Punkten (Item 10).

Eine Testung möglicher Gruppenunterschiede mittels t-Tests auf Itemebene konnte bei Item 4 (...hat relevante Ergebnisse der Patientenfälle in adäquater Sprache fokussiert berichtet) einen signifikanten Unterschied hinsichtlich des Studienorts aufzeigen ( $t(80) = -2.603$ ,  $p = ,011$ ). Münchener-Teilnehmende schnitten dabei hinsichtlich Item 4 ( $4,55 \pm 0,59$ ;  $n = 44$ ) signifikant ( $p < ,05$ ) besser ab als Hamburger Teilnehmende bei Item 4 ( $4,13 \pm 0,84$ ;  $n = 38$ ). In Bezug auf andere Items und auf den Gesamtscore konnten keine weiteren signifikanten Unterschiede in Anhängigkeit des Hochschulstandorts ermittelt werden.

#### 4.2.3 Auswertung der Gruppenunterschiede der Gesamtleistung

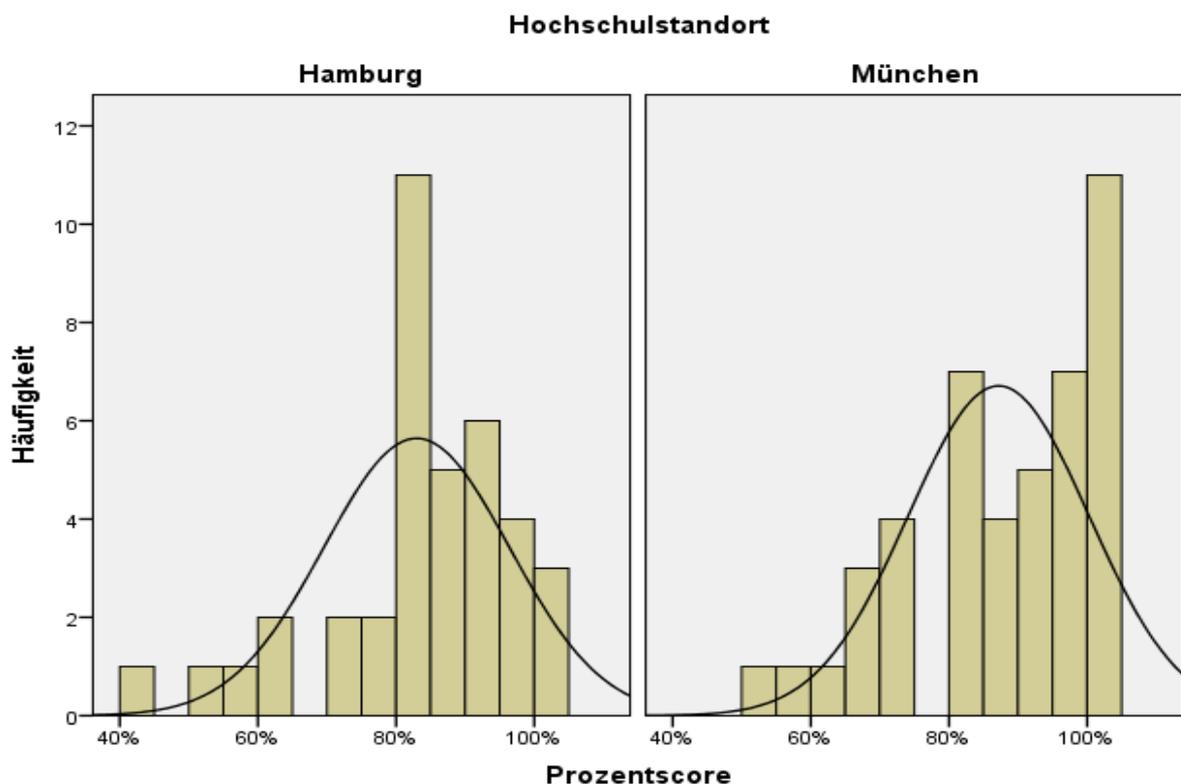
Gesamtscore	Gesamt			Hamburg			München		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
<b>Gesamtscore Peer %</b>	82	84,82	9,00	38	82,95	9,92	44	86,44	7,89
<b>Gesamtscore operationalisierte Bewertung %</b>	82	85,23	13,33	38	82,98	13,43	44	87,18	13,08
<b>Transformierter Gesamtscore der operationalisierten Bewertung</b>	82	2,38	1,22	38	2,07	1,00	44	2,64*	1,34

**Tabelle 5:** Untransformierter Gesamtscore (%) der Peer- und operationalisierten Bewertung sowie transformierter Gesamtscore der operationalisierten Bewertung ( $n$  = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \* =  $p < ,05$ )

Auf deskriptiver Ebene zeigte sich in Bezug auf den Gesamtscore (%), welcher auf den Peerbewertungen basiert, eine Mittelwertdiskrepanz von 3,49 Prozentpunkten. Hierbei erzielten PJ-Studierende der TU München ( $86,44 \pm 7,89$ ;  $n = 44$ ) bessere Ergebnisse als PJ-Studierende der Universität Hamburg ( $82,95 \pm 9,92$ ;  $n = 38$ ) (Tabelle 5). Die Auswertung der Gruppenunterschiede zwischen den Hochschulstandorten der auf den Peerbewertungen basierenden Gesamtscores (%) ergab keinen signifikanten Unterschied ( $F(1,80) = 3,142$ ,  $p = ,080$ , partielles  $\eta^2 = ,038$ ,  $n = 82$ ).

Auf deskriptiver Ebene zeigte sich in Bezug auf den Gesamtscore (%), welcher auf den operationalisierten Bewertungen basiert, eine Mittelwertdiskrepanz von 4,20 Prozentpunkten.

Dabei erzielten PJ-Studierende der TU München ( $87,18 \pm 13,08$ ;  $n = 44$ ) bessere Ergebnisse als PJ-Studierende der Universität Hamburg ( $82,98 \pm 13,43$ ;  $n = 38$ ) (Tabelle 5). Die graphische Analyse der Verteilung der auf den operationalisierten Bewertungen basierenden Gesamtscores (%) über die Studierenden der beiden Hochschulstandorte zeigte, dass sich die Gesamtscores (%) eher linksschief darstellen, insbesondere im Fall der TU München (Abbildung 1). Aus diesem Grunde wurden die auf den operationalisierten Bewertungen basierenden Gesamtscores (%) über die zuvor beschriebene monomolekulare Model-Transformation besser an die Normalverteilung angenähert.



**Abbildung 1:** Histogramm über den Gesamtscore (%) getrennt für die Hochschulstandorte Hamburg und München mit Referenzkurve einer Normalverteilung

Unter Einbezug des transformierten Gesamtscores in eine ANOVA stellte sich ein signifikanter Unterschied in Abhängigkeit des Studienortes dar ( $F(1,80) = 4,654$ ,  $p = ,034$ , partielles  $\eta^2 = ,055$ ,  $n = 82$ ). Gemäß Levene-Test lagen inhomogene Varianzen vor ( $F(1,80) = 7,592$ ,  $p = ,007$ ). Ein ergänzend durchgeführter Welch-Test war ebenfalls signifikant ( $p = ,030$ ). In Bezug auf den transformierten Gesamtscore erreichten Münchener Teilnehmende ( $2,64 \pm 1,34$ ;  $n = 44$ ) signifikant höhere Werte als Hamburger Teilnehmende ( $2,07 \pm 1,00$ ;  $n = 38$ ) (Tabelle 5). Der bereits in der deskriptiven Analyse des untransformierten Gesamtscores (%) beobachtete Mittelwertunterschied von 4,20 Prozentpunkten zeigte sich im transformierten Modell also als signifikant (Tabelle 5). Die Analyse der geschätzten Randmittel des transformierten Modells ergab nach Rücktransformation einen geschätzten Mittelwert von 88,3 Prozentpunkten für Hamburg (95%KI 82,3-92,4) und von 93,8 für München (95%KI 90,7-96).

Eine Analyse der einfaktoriellen ANCOVA unter Einbezug der Kovariaten Alter und Geschlecht zeigte, dass für den transformierten Gesamtscore weiterhin ein signifikanter Unterschied zwischen den beiden Hochschulstandorten ( $F(1,78) = 4,242$ ,  $p = ,043$ , partielles  $\eta^2 = ,052$ ,  $n = 82$ ) ausgemacht werden konnte. Eine weitere einfaktorielle ANCOVA unter Berücksichtigung der studienspezifischen Kovariate „Wahlfach bereits absolviert“ zeigte, dass in diesem Modell weiterhin ein signifikanter Unterschied zwischen den beiden Hochschulstandorten hinsichtlich des transformierten Gesamtscores beobachtet werden konnte ( $F(1,79) = 6,419$ ,  $p = ,034$ , partielles  $\eta^2 = ,056$ ,  $n = 82$ ).

Im Unterscheid dazu zeigte sich in zwei weiteren einfaktoriellen ANCOVAs mit Anzahl der PJ-Semester oder der Note im zweiten Staatsexamen als Kovariate, dass der Unterschied des transformierten Gesamtscores über den Hochschulstandort nicht mehr signifikant war (bei PJ-Semester als Kovariate:  $F(1,70) = 3,457$ ,  $p = ,067$ , partielles  $\eta^2 = ,047$ ,  $n = 73$ ; bei Note im zweiten Staatsexamen als Kovariate  $F(1,75) = 3,879$ ,  $p = ,053$ , partielles  $\eta^2 = ,049$ ,  $n = 78$ ). Eine univariate Varianzanalyse des Einflusses des PJ-Semesters allein auf den transformierten Gesamtscore brachte dahingegen keinen signifikanten Unterschied hervor ( $F(1,76) = 1,235$ ,  $p = ,270$ , partielles  $\eta^2 = ,016$ ,  $n = 78$ ). Gleiches galt bei Eingang der Note im zweiten Staatsexamen in die Analyse ( $F(1,78) = 2,276$ ,  $p = ,200$ , partielles  $\eta^2 = ,057$ ,  $n = 82$ ).

### 4.3 Vergleich der Peerbewertungen mit operationalisierten Fremdbewertungen

Itemnummer	Fremdbewertung			Peerbewertung		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Item 1	82	4,07	1,05	82	4,10	0,73
Item 2	82	4,13	1,02	82	4,18	0,74
Item 3	82	4,46	0,80	82	4,24	0,84
Item 4	82	4,33	0,75	82	4,35	0,74
Item 5	82	4,38	0,86	81	4,54	0,76
Item 6	80	4,17*	1,09	77	3,68	1,20
Item 7	12	4,50	0,80	75	4,61	1,00
Item 8	74	4,39	0,81	80	4,49	0,62
Item 9	74	3,91**	1,16	77	3,06	1,27
Item 10	82	4,48	0,88	82	4,95**	0,22
Item 11	7	3,14	0,90	58	4,38**	0,77

**Tabelle 6:** Itemstatistik über beide Raterperspektiven der operationalisierten Fremd- und Peerbewertung (n = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \* =  $p < ,05$ , \*\* =  $p < ,005$ )

In der Zusammenschau über beide Hochschulstandorte zeigten sich signifikante Mittelwertunterschiede bei gleichen Varianzen für Item 6 (...hat auf verschiedene Art und Weise geäußert, dass er oder sie nicht weitergekommen ist) ( $t(155) = 2,739$ ,  $p = ,007$ ), Item 9 (...hat nach der Meinung der Kollegin oder des Kollegen zu Patienten oder Patientinnen gefragt) ( $t(149) = 4,239$ ,  $p < ,001$ ) sowie bei ungleichen Varianzen für Item 10 (...hat sich der Kollegin oder dem Kollegen gegenüber respektvoll verhalten) ( $t(90,833) = -4,762$ ,  $p < ,001$ ). Ein ergänzend durchgeführter Welch-Test für Item 10 war ebenfalls signifikant ( $p < ,001$ ). Ein weiterer signifikanter Gruppenunterschied zwischen den beiden Raterperspektiven konnte bei Item 11 (...hat sich kritikfähig gezeigt) nachgewiesen werden ( $U = 62$ ,  $Z = -3,225$ ,  $p = ,001$ ). Peerbewertungen von Item 10 ( $4,95 \pm 0,22$ ;  $n = 82$ ) und Item 11 ( $4,38 \pm 0,77$ ;  $n = 58$ ) zeigten dabei, dass diese im Mittel signifikant höher ausfielen als die operationalisierten Fremdbewertungen von Item 10 ( $4,48 \pm 0,88$ ;  $n = 82$ ) und Item 11 ( $3,14 \pm 0,90$ ;  $n = 7$ ). Item 6 ( $4,17 \pm 1,09$ ;  $n = 80$ ) und Item 9 ( $3,91 \pm 1,16$ ;  $n = 74$ ) wurden im Mittel signifikant besser durch die operationalisierten Fremdbewertung bewertet, im Vergleich zur Peerbewertung von Item 6 ( $3,68 \pm 1,20$ ;  $n = 77$ ) und Item 9 ( $3,06 \pm 1,27$ ;  $n = 77$ ) (Tabelle 6).

Item-nummer	Hamburg						München					
	Fremdbewertung			Peerbewertung			Fremdbewertung			Peerbewertung		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Item 1	38	3,84	1,05	38	3,97	0,75	44	4,27	1,02	44	4,20	0,70
Item 2	38	3,97	0,94	38	4,08	0,78	44	4,27	1,07	44	4,27	0,69
Item 3	38	4,34	0,85	38	4,18	0,80	44	4,57	0,76	44	4,30	0,88
Item 4	38	4,21	0,70	38	4,13	0,84	44	4,43	0,79	44	4,55	0,59
Item 5	38	4,34	0,82	38	4,47	0,89	44	4,41	0,90	43	4,60	0,62
Item 6	38	4,16	1,13	37	3,62	1,28	42	4,19	1,07	40	3,73	1,13
Item 7	5	4,00	1,00	36	4,39	1,23	7	4,86	0,38	39	4,82	0,68
Item 8	34	4,32	0,88	37	4,46	0,69	40	4,45	0,75	43	4,51	0,55
Item 9	34	3,53	1,16	35	3,03	1,29	40	4,23**	1,07	42	3,10	1,27
Item 10	38	4,47	0,92	38	4,95**	0,23	44	4,48	0,85	44	4,95**	0,21
Item 11	3	2,67	1,16	27	4,37**	0,88	4	3,50	0,58	31	4,39**	0,67

**Tabelle 7:** Vergleich der Peerbewertung mit der operationalisierten Fremdbewertung ( $n$  = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \*\* =  $p < ,005$ )

In der Untergruppe der Hamburger PJ-Studierenden bot sich bei ungleichen Varianzen ein signifikanter Mittelwertunterschied bei Item 10 (...hat sich der Kollegin oder dem Kollegen gegenüber respektvoll verhalten) ( $t(41,439) = -3,075$ ,  $p = ,004$ ). Ein ergänzend durchgeführter Welch-Test war ebenfalls signifikant ( $p = ,004$ ). Ein weiterer signifikanter Gruppenunterschied zwischen den beiden Raterperspektiven konnte bei Item 11 (...hat sich kritikfähig gezeigt) nachgewiesen werden ( $U = 9$ ,  $Z = -2,386$ ,  $p = ,017$ ). Dabei fielen die Peerbewertungen von

Item 10 ( $4,95 \pm 0,23$ ;  $n = 38$ ) und Item 11 ( $4,37 \pm 0,88$ ;  $n = 27$ ) jeweils höher aus als die operationalisierten Fremdbewertungen von Item 10 ( $4,47 \pm 0,92$ ;  $n = 38$ ) und Item 11 ( $2,67 \pm 1,16$ ;  $n = 3$ ) (Tabelle 7). In der Untergruppe der Münchener Teilnehmenden ergaben sich signifikante Mittelwertunterschiede bei gleichen Varianzen für Item 9 (...hat nach der Meinung der Kollegin oder des Kollegen zu Patienten oder Patientinnen gefragt) ( $t(80) = 4,349$ ,  $p < ,001$ ) sowie bei ungleichen Varianzen für Item 10 (...hat sich der Kollegin oder dem Kollegen gegenüber respektvoll verhalten) ( $t(48,280) = -3,620$ ,  $p = ,001$ ). Ein ergänzend durchgeführter Welch-Test war ebenfalls signifikant ( $p = ,001$ ). Ein weiterer signifikanter Gruppenunterschied zwischen den beiden Raterperspektiven konnte bei Item 11 (...hat sich kritikfähig gezeigt) nachgewiesen werden ( $U = 22$ ,  $Z = -2,262$ ,  $p = ,024$ ). Die Peerbewertung fiel bezüglich Item 10 ( $4,95 \pm 0,21$ ;  $n = 44$ ) und Item 11 ( $4,39 \pm 0,67$ ;  $n = 31$ ) besser aus als die entsprechenden operationalisierten Fremdbewertungen von Item 10 ( $4,48 \pm 0,85$ ;  $n = 44$ ) und Item 11 ( $3,50 \pm 0,58$ ;  $n = 4$ ). Hinsichtlich Item 9 lag eine bessere operationalisierte Fremdbewertung ( $4,23 \pm 1,07$ ;  $n = 40$ ) im Vergleich zur Peerbewertung ( $3,10 \pm 1,27$ ;  $n = 42$ ) vor (Tabelle 7).

#### **4.4 Faktorenanalyse basierend auf den operationalisierten Bewertungen**

Um eine Aussage treffen zu können, worin sich Hamburger und Münchener Teilnehmende unterscheiden, wurden alle operationalisierten Fremdbewertungen der Items einer explorativen Faktorenanalyse unterzogen. Da Item 7 (...hat nicht Andere für Fehler verantwortlich gemacht) und Item 11 (...hat sich kritikfähig gezeigt) vorrangig als „nicht beurteilbar“ eingestuft wurden, sind sie aus der Faktorenanalyse ausgeschlossen worden, da sonst zu wenige Fälle für eine aussagekräftige Faktorenanalyse zur Verfügung gestanden hätten. Insgesamt sind so 72 gültige Fälle mit erfolgter Itembewertung über alle 9 verbliebenen Items in die Faktorenanalyse eingegangen. Die Voraussetzungen für eine Faktorenanalyse waren nach Kaiser-Meyer-Olkin Measure of Sampling Adequacy ( $KMO = .829$ ) und Bartlett-Test ( $\text{Chi-Quadrat}(36) = 454.609$ ,  $p < ,001$ ) erfüllt. Es wurde eine Hauptkomponentenanalyse mit Varimax-Rotation durchgeführt. Die rotierte Faktorenlösung deckt unter Anwendung des Kaiser-Kriteriums, sowie der Analyse des Screeplots zwei Faktoren auf, die zusammen 69,05% der Varianz erklären. Anhand der zwei Faktoren ist es möglich, die zuletzt 9 Items umfassende Skala zu unterteilen (Tabelle 8). Die erste Subskala beinhaltet dabei sicher Item 2 (...hat seine Diagnostik- und Behandlungsschritte gut nachvollziehbar begründet), Item 1 (...hat in einer schlüssigen Weise über die Patienten oder Patientinnen berichtet, so dass seine Kollegin oder sein Kollege kaum Nachfragen stellen musste), Item 4 (...hat relevante Ergebnisse der Patientenfälle in adäquater Sprache fokussiert berichtet), Item 3 (...hat für die Patienten oder Patientinnen eine Strategie berichtet, wie seine Kollegin oder sein Kollege mit diesen weiter arbeiten sollte) und Item 9 (...hat nach der Meinung der Kollegin oder des

Kollegen zu Patienten oder Patientinnen gefragt). Diese Items weisen Faktorladungen von  $\geq ,742$  bei vernachlässigbaren Querladungen von  $\leq ,319$  und Differenzen  $\geq ,434$  auf. Die Zuordnung von Item 6 (...hat auf verschiedene Art und Weise geäußert, dass er oder sie nicht weitergekommen ist) zu dieser ersten Skala erfolgt bei vorliegenden Querladungen sowie geringer Differenz zwischen den Ladungen von ,344 auf die beiden Faktoren (auf Faktor 1 = ,540 und auf Faktor 2 = ,196) nicht eindeutig. Die zweite Subskala wird durch Item 10 (...hat sich der Kollegin oder dem Kollegen gegenüber respektvoll verhalten), Item 8 (...hat Ansichten, Wissen und Expertise der Kollegin oder des Kollegen im Gespräch berücksichtigt) und Item 5 (...hat kommuniziert, manche Dinge nicht erledigt oder geschafft zu haben) gebildet. Diese Items weisen Faktorladungen von  $\geq ,676$  bei vernachlässigbaren Querladungen von  $\leq ,306$  und Differenzen  $\geq ,391$  auf.

Itemnummer	Faktor 1	Faktor 2
Item 2	,936	,141
Item 1	,908	,195
Item 4	,845	,152
Item 3	,753	,319
Item 9	,742	,308
Item 6	,540	,196
Item 10	,076	,859
Item 8	,306	,836
Item 5	,285	,676

**Tabelle 8:** Faktorladungen der rotierten Faktorenmatrix auf Basis der operationalisierten Bewertungen

Auf deskriptiver Ebene zeigte sich in Bezug auf den Gesamtscore von Faktor 1 (%) eine Mittelwertdiskrepanz von 6,88 Prozentpunkten. Hierbei erzielten PJ-Studierende der TU München ( $86,75 \pm 16,85$ ;  $n = 44$ ) bessere Ergebnisse als PJ-Studierende der Universität Hamburg ( $79,87 \pm 15,68$ ;  $n = 38$ ) (Tabelle 9). Die Auswertung der Gruppenunterschiede zwischen den Hochschulstandorten der Gesamtscores (%) ergab keinen signifikanten Unterschied ( $F(1,80) = 3,627$ ,  $p = ,060$ , partielles  $\eta^2 = ,043$ ,  $n = 82$ ).

Hinsichtlich des Gesamtscores von Faktor 2 (%) lag eine Mittelwertdiskrepanz von 0,43 Prozentpunkten vor. Dabei erzielten PJ-Studierende der TU München ( $88,41 \pm 14,19$ ;  $n = 44$ ) bessere Ergebnisse als PJ-Studierende der Universität Hamburg ( $87,98 \pm 14,36$ ;  $n = 38$ ) (Tabelle 9).

Faktor	Gesamt			Hamburg			München		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
<b>Gesamtscore Faktor 1 %</b>	82	83,56	16,58	38	79,87	15,68	44	86,75	16,85
<b>Transformierter Gesamtscore Faktor 1</b>	82	2,42	1,38	38	2,04	1,19	44	2,75*	1,46
<b>Gesamtscore Faktor 2 %</b>	82	88,21	14,18	38	87,98	14,36	44	88,41	14,19
<b>Transformierter Gesamtscore Faktor 2</b>	82	2,93	1,50	38	2,89	1,49	44	2,97	1,52

**Tabelle 9:** Untransformierter Gesamtscore (%) der operationalisierten Bewertung sowie transformierter Gesamtscore der operationalisierten Bewertung in Hinblick auf die ermittelten Faktoren (n = Anzahl der Bewertungen,  $\bar{x}$  = Mittelwert, SD = Standardabweichung, \* =  $p < ,05$ )

Analog zu der Transformation der Gesamtskala über alle Items wurde auch jeweils der Gesamtscore von Faktor 1 und Faktor 2 mittels der monomolekularen Model-Transformation besser an die Normalverteilung angenähert. Unter Einbezug des transformierten Gesamtscores von Faktor 1 in eine ANOVA stellte sich ein signifikanter Unterschied in Abhängigkeit des Studienortes dar ( $F(1,80) = 7,445$ ,  $p = ,008$ , partielles  $\eta^2 = ,085$ ,  $n = 82$ ). Gemäß Levene-Test lagen inhomogene Varianzen vor ( $F(1,80) = 8,169$ ,  $p = ,005$ ). Ein ergänzend durchgeführter Welch-Test war ebenfalls signifikant ( $p = ,007$ ). In Bezug auf den transformierten Gesamtscore von Faktor 1 erreichten Münchener Teilnehmende ( $2,75 \pm 1,46$ ;  $n = 44$ ) signifikant höhere Werte als Hamburger Teilnehmende ( $2,04 \pm 1,19$ ;  $n = 38$ ) (Tabelle 9). Der bereits in der deskriptiven Analyse des untransformierten Gesamtscores von Faktor 1 (%) beobachtete Mittelwertunterschied von 6,88 Prozentpunkten zeigte sich im transformierten Modell also als signifikant (Tabelle 9). Die Analyse der geschätzten Randmittel des transformierten Modells ergab nach Rücktransformation einen geschätzten Mittelwert von 88,5 Prozentpunkten für Hamburg (95%KI 80,8-93,3) und von 95,9 für München (95%KI 93,0-97,7). Hinsichtlich des transformierten Gesamtscores von Faktor 2 konnte mittels ANOVA kein signifikanter Unterschied in Abhängigkeit des Studienortes ermittelt werden ( $F(1,80) = 0,59$ ,  $p = ,809$ , partielles  $\eta^2 = ,001$ ,  $n = 82$ ).

## 5 Diskussion

Die Übergabesituation ist ein wichtiger und essentieller Bestandteil des klinischen Alltages. Sie tritt täglich wiederkehrend in verschiedenen Kontexten ein. Die Übergabe stellt dabei ein kritisches Element dar, weil sie mit Diskontinuität der Patientenbetreuung durch wechselndes medizinisches Personal einhergeht und zur Zunahme vermeidbarer unerwünschter Ereignisse in der Patientenversorgung führt (Petersen et al. 1994). Insbesondere Kommunikationsfehler wirken sich negativ auf den Erfolg einer Übergabe aus, indem sie unter anderem Unsicherheit in der Entscheidungsfindung hinsichtlich der Patientenversorgung zur Folge haben (Arora et al. 2005). Eine ungenügende Qualität der Übergabe wirkt sich negativ auf die Patientensicherheit aus (Manser und Foster 2011, Desmedt et al. 2021). Im Zuge dessen ist es relevant, Kenntnis über den Stand der Qualität von Übergaben zu erlangen, um kritisch ein mögliches Defizit der Patientensicherheit zu ermitteln und entsprechenden Interventionsbedarf abzuleiten. Folglich wurden weitere Untersuchungen mit systematischem Ansatz zur Etablierung valider Messmethoden der Qualität der Übergabe gefordert (Manser und Foster 2011). Die Etablierung des kompetenzbasierten Simulationsmodells eines klinischen Arbeitstages ermöglicht dabei unter anderem erstmals eine vergleichende Analyse der Qualität der Übergabesituation von PJ-Studierenden unterschiedlicher deutscher Hochschulstandorte mit divergierender Ausrichtung des Curriculums (Harendza et al. 2017a, Prediger et al. 2020). Dabei kam ein neuer Bewertungszugang zur Ermittlung der Qualität der Übergabe zum Einsatz, welcher auf den wichtigsten zehn Kompetenzen für Berufsanfängerinnen und Berufsanfänger in der Medizin aufbaut (Fürstenberg et al. 2017) und bereits im Rahmen des kompetenzbasierten Simulationsmodells etabliert wurde (Prediger et al. 2020). Absolventinnen und Absolventen der teilnehmenden Hochschulstandorte unterschieden sich dabei nicht in ihrer Einstellung hinsichtlich der wichtigsten Kompetenzen für Berufsanfängerinnen und Berufsanfänger (Fürstenberg et al. 2017). Die Bewertung der Übergabegespräche erfolgte durch zwei unterschiedliche Perspektiven: (1) subjektive Peerbewertung mittels „Handover Partner Assessment“ (HOPA), (2) operationalisierte Fremdbewertung mittels „Handover Performance External Rater Assessment“ (HOPERA). Die Entwicklung der operationalisierten Fremdbewertung in dieser Arbeit sollte dem möglichen Bias der Reziprozität von Peerbewertungen entgegenreten und variierenden Bewertungsstrengen anhand von Ankerbeispielen zu den einzelnen Bewertungsstufen vorbeugen (Magin 2001). Untersuchungen mittels Fokusgruppen konnten zeigen, dass Peerbewertungen das Risiko bergen, unabhängig von der eigentlich beobachteten Leistung auszufallen, um soziale Spannungen in der Gruppe zu minimieren (Biesma et al. 2019). Unter Berücksichtigung beider Perspektiven fiel die Qualität von Übergabegesprächen deutschsprachiger PJ-Studierender in der Arzt- oder Ärztinnenrolle im Rahmen eines simulierten Arbeitstages insgesamt gut aus. Die teilnehmenden PJ-Studierenden der

Hochschulstandorte Universität Hamburg und TU München unterschieden sich dabei allerdings signifikant bezüglich der Qualität ihrer Übergabe, wenn diese mittels der neuentwickelten operationalisierten Bewertungsgrundlage (HOPERA) ermittelt wurde.

Münchener PJ-Studierende überzeugten insgesamt durch eine höhere Qualität ihrer Übergabegespräche im Vergleich zu jener der Hamburger PJ-Studierenden. Die beiden Hochschulstandorte bieten zwei verschiedene Modelle des Studiums der Humanmedizin an. Die TU München strukturiert die Lehre in Form eines Regelstudienganges, während sich die Universität Hamburg seit dem Wintersemester 2012/2013 der Reformierung der Struktur im Sinne eines vertikal integrierten Modellstudienganges verpflichtet hat (Wissenschaftsrat 2014). Die vertikale Integration bei Medizinstudiengängen beinhaltet vier wesentliche Elemente: (1) frühe klinische Erfahrung, (2) Integration biomedizinischer Wissenschaften und klinischer Fälle, (3) lange klinische Praxiszeiten während der letzten Studienjahre und (4) Förderung der klinischen Verantwortungsbereitschaft (Wijnen-Meijer et al. 2013a). So sollen praktische Inhalte über die gesamte Dauer des Studiums besser verteilt und integriert werden und somit die strikte Trennung von vorklinischen und klinischen Studiumsinhalten entzerrt werden (Wissenschaftsrat 2014). Mit Einführung der vertikalen Integration soll der Übergang zum Berufseinstieg erleichtert werden (Wijnen-Meijer et al. 2010). Studien konnten nachweisen, dass sich Absolventinnen und Absolventen eines reformierten Studienganges besser vorbereitet fühlen und sich selbst auch kompetenter einschätzen (Wijnen-Meijer et al. 2010, Schmidt et al. 2006). Übergaben sind ein elementarer Bestandteil klinischer Tätigkeiten. Studierende, die im Rahmen eines vertikal integrierten Studienganges frühzeitig klinische Erfahrungen sammeln und intensiver dem klinischen Alltag ausgesetzt sind, sollten daher voraussichtlich auch häufiger mit Übergabesituationen konfrontiert sein. Sheng et al. (2020) demonstrierten, dass vertikal integrierte Curricula die konkrete Möglichkeit zur Schulung im Umgang mit der Übergabesituation bieten und einen Lernzuwachs zur Folge haben. Im Rahmen der vorliegenden Arbeit konnte jedoch kein positiver Effekt des vertikal integrierten Studienganges der Universität Hamburg in Hinblick auf die Qualität der Übergabe beobachtet werden. Im Gegenteil zeigte sich, dass PJ-Studierende des Regelstudienganges an der TU München besser abschnitten. Vorausgesetzt, Studierende eines vertikal integrierten Modellstudienganges sammeln tatsächlich früher und umfangreicher Erfahrungen im klinischen Kontext, scheint sich dies folglich nicht in einem relevanten Schulungsprozess zum Umgang mit Übergabesituation niederschlagen. Dies mag daran liegen, dass nur indirekt entsprechende Fertigkeiten vermittelt werden und meistens keine definierten Unterrichtseinheiten zum Thema Übergabe in das Curriculum integriert sind (Stojan et al. 2015). Zukünftig könnte der Modellstudiengang an der Universität Hamburg daher in der klinisch orientierten Lehre stärker die Übergabesituation adressieren. Ein vielversprechender

Effekt diesbezüglich konnte beispielsweise durch die von Thaeter et al. (2018) erprobte Implementierung eines Trainings von Übergabesituationen in das Curriculum an einer deutschen Hochschule beschrieben werden. Absolventinnen und Absolventen dieses Trainings überzeugten unter anderem durch einen Lernzuwachs und verbesserte Leistung im Rahmen der Übergabesituation. Inwiefern Studierende zum Beispiel von der Implementierung standardisierter Übergabekonzepte und Einführung technologiebasierter Übergabeprotokolle profitieren, ließe sich im Anschluss in einem erneuten Durchlauf des kompetenzbasierten Simulationsmodells ermitteln. Dabei wäre in Zukunft zudem eine experimentelle Untersuchung der in der Literatur beschriebenen Übergabetechniken im Setting des kompetenzbasierten Simulationsmodells denkbar (Riesenberg et al. 2009). Es könnte beispielsweise der Parameter der Verfügbarkeit technologiebasierter Übergabeprotokolle manipuliert werden (Bhabra et al. 2007). Das kompetenzbasierte Simulationsmodell bietet derweil den Vorteil, dass es aufgrund seiner Komplexität und Realitätsnähe engen Bezug zur klinischen Tätigkeit aufweist und gewonnene Erkenntnisse zur Übergabequalität auf die Praxis übertragen werden können. Darüber hinaus könnten auch im Anschluss an die Teilnahme ein umfangreiches Feedback und Schulungsmaßnahmen zur Übergabe stattfinden und der Effekt der Intervention bei einer erneuten Teilnahme ausgewertet werden. In jedem Falle muss bei erneuten Durchläufen des kompetenzbasierten Simulationsmodells eine ausreichende Anzahl von Teilnehmenden weiterer Hochschulstandorte sichergestellt werden. Hier muss kritisch der Ausschluss des Studienstandortes Oldenburg in dieser Arbeit erwähnt werden, welcher aufgrund einer zu geringen Stichprobengröße erfolgte. Der resultierende Vergleich von zwei Hochschulstandorten mit jeweils unterschiedlicher Ausrichtung des Curriculums kann nicht abschließend repräsentativ interpretiert werden, sondern liefert Ergebnisse, die es in weiteren Umsetzungen des kompetenzbasierten Simulationsmodells zu untersuchen gilt.

Um eine genauere Aussage treffen zu können, inwiefern sich Münchener PJ-Studierende von Hamburger PJ-Studierenden hinsichtlich der Qualität ihrer Übergabe unterscheiden, hilft es, das Ergebnis der faktoranalytischen Untersuchungen heranzuziehen. Anhand von Faktorenanalysen konnten in der Struktur des verwendeten Messinstruments zur Beurteilung der Qualität der Übergabegespräche (HOPERA) statistisch zwei Komponenten extrahiert werden: (1) eine medizinisch-inhaltliche Komponente und (2) eine global-kommunikative Komponente. Lediglich die erstgenannte Komponente konnte inhaltlich sinnvoll interpretiert werden. In der Literatur sind nur wenige vergleichbare Analysen der inhärenten Struktur von Messinstrumenten zur Qualität einer Übergabe beschrieben. Manser et al. (2010) extrahierten aus ihrem 19 Items umfassenden Bewertungsbogen zur Ermittlung der Qualität einer Übergabe drei Faktoren. Diese benannten sie (1) Informationsaustausch, (2) einheitliches Verständnis und (3) Arbeitsatmosphäre (Manser et al. 2010). Diese Dreiteilung lässt sich gut

mit den von Cheung et al. (2010) herausgearbeiteten drei Ebenen der Messung der Qualität einer Übergabe übereinbringen: (1) Inhalt, (2) Prozess, (3) Ergebnisse. Die Ebene des Inhalts wird durch den Faktor des Informationsaustausches abgebildet, wohingegen Eigenschaften der Übergabe, wie beispielsweise die Arbeitsatmosphäre, den Prozess beschreiben und Ergebnisse der Übergabe anhand der Etablierung eines einheitlichen Verständnisses oder der Zufriedenheit der Teilnehmenden evaluiert werden. Weitere Unterteilungen der Qualitätsmessung der Übergabe in drei Dimensionen finden sich bei Horwitz et al. (2009b): (1) Inhalt, (2) Deutlichkeit der Sprache und (3) Umgebung sowie bei Jeffcott et al. (2009): (1) Transfer von Informationen, (2) Übertragung von Verantwortung und/ oder Rechenschaftspflicht sowie (3) Systemelemente, wie insbesondere Teamwork und Arbeitsumgebung. Allen Bewertungszugängen ist dabei die Komponente des Inhalts beziehungsweise der vermittelten Information während der Übergabe gemein. Die erste der beiden extrahierten Komponenten des in dieser Arbeit verwendeten Messinstruments zur Beurteilung der Qualität der Übergabegespräche (HOPERA) entspricht eben dieser in der Literatur beschriebenen Komponente des Inhalts beziehungsweise der vermittelten Information während der Übergabe. Die tiefgehende Analyse der Leistungsunterschiede der PJ-Studierenden unterschiedlicher Studiengangsmodele ergab, dass Münchener PJ-Studierende sich insbesondere auf der Ebene medizinisch-inhaltlicher Aspekte der Übergabe leistungstechnisch abheben. Die Bewertungen der zugehörigen Items dieser Komponente fielen dabei unabhängig von der Perspektive der Bewertung (Peer- vs. Fremdbewertung) gleich aus, mit Ausnahme eines Items. Die abweichende Bewertung dieses Items (Item 9) erklärt sich durch eine Modifikation der Bewertungsgrundlage, welche in der Konzeption der operationalisierten Fremdbewertung Einzug fand. Die Peerbewertung beinhaltete lediglich, ob die Übergabeempfängerin oder der Übergabeempfänger nach der eigenen Meinung zum übergebenen Fall gefragt wurde. Die operationalisierte Fremdbewertung berücksichtigte bei der Wertung, ob für die übergebende Person generell eine Notwendigkeit bestand, ihr Gegenüber nach der Meinung zu fragen. Waren im Rahmen der Übergabesituation die medizinisch-inhaltlichen Aspekte weitestgehend geklärt und die beziehungsweise der Teilnehmende sah dementsprechend keine Notwendigkeit nach der Meinung zu fragen, so wurde dieses Item schlechtesten Falls auf der Bewertungsstufe drei von fünf gewertet. Dies sollte einer repräsentativeren Ermittlung einer Gesamtleistung dienen, führt aber dazu, dass sich die Peerbewertung nicht sinnvoll mit der operationalisierten Fremdbewertung vergleichen lässt. Die medizinisch inhaltliche Komponente der Übergabe sollte daher in Zukunft unter Ausschluss dieses Items bewertet werden, um konsistent zwischen den Bewertungsperspektiven auszufallen.

In Bezug auf die zweite faktoranalytisch ermittelte Komponente wurde kein Qualitätsunterschied der Übergabe zwischen den teilnehmenden Hochschulen gefunden. Allerdings ließ sich diese Komponente auch weder eindeutig zu einer der im Vorfeld theoretisch angenommenen Komponenten, noch zu einer in der Literatur beschriebenen Komponente zuordnen.

Aufgrund dieser Ergebnisse lässt sich vermuten, dass PJ-Studierende der TU München eventuell auf ein breiteres theoretisches medizinisch-inhaltliches Wissen zurückgreifen, welches ihnen ermöglicht, Übergaben zu medizinisch-inhaltlichen Aspekten im Rahmen des Transfers von Patienteninformationen detaillierter und vollständiger zu gestalten. Die Ergebnisse lieferten einen Hinweis darauf, dass medizinisch-inhaltliches Wissen die Unterschiede in der Qualität der Übergabe zwischen den Hochschulstandorten erklärt. Die Anzahl der bereits absolvierten PJ-Semester und die Note im zweiten Staatsexamen der Teilnehmenden hatten einen nicht zu vernachlässigenden Einfluss auf die Qualität der Übergabe. Bei Kontrolle des Effekts des Hochschulstandorts auf die Qualität der Übergabe durch die Anzahl der PJ-Semester oder die Note im zweiten Staatsexamen zeigte sich dieser nicht mehr signifikant. Dies erscheint insbesondere vor dem Hintergrund der ausschlaggebenden Komponente der medizinisch-inhaltlichen Aspekte bei der Beurteilung schlüssig. Eine bessere Note im zweiten Staatsexamen kann als Indikator für ein umfangreicheres medizinisch-inhaltliches Wissen gesehen werden. Darüber hinaus kann ein fortgeschrittenes PJ-Semester ebenfalls mit Zuwachs medizinisch-inhaltlichen Wissens in Verbindung gebracht werden (Raupach et al. 2013) und sich so auf die Qualität der Übergabe auswirken, auch wenn PJ-Studierende und Studierende niedrigerer Semester sich gleichermaßen selbstsicher im Umgang mit der Übergabesituation im Rahmen des simulierten Arbeitstages einschätzten (Harendza et al. 2017a). Unterschiede in der Qualität der Übergabe konnten dabei weder durch die Anzahl der PJ-Semester, noch anhand der Note im zweiten Staatsexamen primär erklärt werden. Der direkte deskriptive Vergleich der durchschnittlichen Ergebnisse des zweiten Staatsexamens anhand der vom Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) veröffentlichten Statistiken zeigt, dass Studierende im Regelstudiengang der TU München im Mittel besser abschnitten als Studierende der Universität Hamburg im Modellstudiengang (IMPP 2019). Es ist anzunehmen, dass sich dies folglich indirekt in der Qualität der Übergabe auf der Ebene der medizinisch-inhaltlichen Komponente niederschlägt. Diesbezüglich lässt sich weitergehend ein Unterschied in der Fähigkeit des Clinical Reasonings diskutieren. Clinical Reasoning stellt einen kognitiven Prozess dar, dessen sich Ärztinnen und Ärzte bedienen, um auf Basis gesammelter Informationen problemorientiert Entscheidungen zu treffen (Harendza et al. 2017b). Die Qualität von Fallvorstellungen, welche als ein Indikator der Fähigkeit des Clinical Reasonings

gilt, konnte im Kontext des kompetenzbasierten Simulationsmodells anhand des erhobenen medizinischen Wissensstands der Studierenden vorhergesagt werden (Fürstenberg et al. 2019). Es ist anzunehmen, dass die Qualität von Fallvorstellungen in direktem Bezug zur Qualität der medizinisch-inhaltlichen Komponente der Übergabe steht. Eventuell erwerben Studierende im Regelstudiengang der TU München nicht nur ein umfangreicheres Faktenwissen, sondern überzeugen auch durch ihre Clinical Reasoning Kompetenz, bei der es unter anderem darauf ankommt, Fakten zu begründen. So schneiden sie bei der Ermittlung der Qualität ihrer Übergabe besser ab, weil die Übergabe logischer erfasst werden kann. In einer vorangegangenen Untersuchung des kompetenzbasierten Simulationsmodells zeigte sich allerdings kein Effekt der Art des Studienganges auf die Fähigkeit des Clinical Reasonings (Fürstenberg et al. 2019). Es bleibt daher fraglich, ob die Fähigkeit des Clinical Reasonings elementaren Einfluss auf die Qualität der Übergabe nimmt, oder die Komponente medizinisch-inhaltlichen Wissens dafür ausschlaggebender ist und der Fähigkeit des Clinical Reasonings nur eine moderierende Rolle des Effekts zugeschrieben werden kann.

Inwieweit sich PJ-Studierende unterschiedlicher Hochschulen mit divergierender Ausrichtung des Curriculums noch in anderen relevanten Facetten in Bezug auf die Qualität der Übergabe unterscheiden, ist mittels des verwendeten Messverfahrens HOPERA nicht abschließend zu beurteilen. Die Limitation in der Anwendung des HOPERA steht in einer Reihe mit den Ergebnissen des systematischen Reviews von Desmedt et al. (2021), welcher konstatierte, dass bisher kein Instrument existiere, das alleine zur bestmöglichen Evaluation der Übergabesituation genüge. Die Ergebnisse der faktoranalytischen Untersuchungen des HOPERA stellten sich mit Ausnahme der medizinisch-inhaltlichen Komponente nicht kongruent mit den weiteren in der Literatur beschriebenen Komponenten dar (Manser et al. 2010, Jeffcott et al. 2009, Horwitz et al. 2009b). Dies ist insbesondere kritisch zu werten, als dass die theoretischen Überlegungen zu der Struktur des neuentwickelten Messinstruments zur Beurteilung der Qualität der Übergabegespräche (HOPERA) ähnliche Komponenten beinhalteten. Zum Beispiel ging die Auffassung der Übergabe von Manser et al. (2010) über den reinen Informationstransfer hinaus und beschrieb diese als Teamwork-Aufgabe mit Etablierung eines einheitlichen Verständnisses unter Berücksichtigung der Arbeitsatmosphäre. Der HOPERA sollte neben einer Komponente medizinisch-inhaltlicher Aspekte zwei weitere Komponenten abbilden, welche kollegiales Verhalten und die medizinisch-inhaltliche Einbindung der Kollegin oder des Kollegen erfassen sollten. Die grundlegende Struktur des Bewertungsbogens von Manser et al. (2010) spiegelte folglich annähernd die intendierte Ausrichtung des operationalisierten Bewertungsbogens (HOPERA) wider, insofern die medizinisch-inhaltliche Einbindung der Kollegin oder des Kollegen als Äquivalent zur Komponente des einheitlichen Verständnisses und das kollegiale Verhalten

mindestens als Teilaspekt der Arbeitsatmosphäre gedeutet wird. Dass sich diese Struktur nicht durch die Faktorenanalyse extrahieren ließ, ist dadurch zu erklären, dass entsprechende Items des Messinstruments aufgrund einer kleinen Anzahl an Bewertungen aus der Analyse ausgeschlossen wurden. Die geringe Anzahl an Bewertungen kam dadurch zustande, dass diese Items nur unter Erfüllung einer Bedingung bewertet werden konnten. Im Einzelnen musste dazu ein unterlaufener Fehler erkannt werden (Item 7) und es musste Kritik durch die Übergabeempfängerin oder den Übergabeempfänger erfolgt sein (Item 11). Nach Ausschluss dieser Items verblieben zu wenige Items, die in die Faktorenanalyse eingingen, um feinere Abstufungen zu ermöglichen. Mit dem Anspruch, die Übergabequalität umfassend zu evaluieren, sollten daher entsprechende Items, wie die von Manser et al. (2010) erarbeiteten, in den Aufbau des HOPERA implementiert werden, die durchgehend eine bedingungsfreie Bewertung erlauben. Items, deren Bewertung an die Erfüllung einer Bedingung gebunden sind, eignen sich nicht, um vergleichende Aussagen der Qualität der Übergabe zu treffen. Eine umfangreichere Berücksichtigung des Aspekts der Etablierung eines einheitlichen Verständnisses als Teil der ergebnis- und prozessorientierten Messung könnte bei weiteren Anwendungen des HOPERA zur Abgrenzung der Komponenten beitragen, welche durch die medizinisch-inhaltliche Einbindung der Kollegin oder des Kollegen und durch das kollegiale Verhalten beschrieben werden. Zusätzlich sollte die Prüfung der Zufriedenheit der Teilnehmenden mit der erfolgten Übergabe erfasst werden, um so eine ergebnisorientierte Messung der Qualität der Übergabe sicherzustellen (Cheung et al. 2010). Zukünftig ließe sich so bei weiteren Analysen des ergänzten HOPERA wahrscheinlich auch die theoretisch angenommene und in der Literatur beschriebene Grundstruktur über drei Komponenten bestätigen (Manser et al. 2010, Jeffcott et al. 2009, Horwitz et al. 2009b).

In der Zusammenschau der Ergebnisse lässt sich festhalten, dass die Qualität der Übergabe PJ-Studierender unterschiedlicher Hochschulen im deutschsprachigen Raum insgesamt gut ausfiel. Das kompetenzbasierte Simulationsmodell eignet sich gut, um Erkenntnisse über die Qualität der Übergabe zu gewinnen. Die medizinisch-inhaltliche Komponente der Qualität der Übergabe kann sowohl mittels operationalisierter Fremdbewertung durch den HOPERA, als auch durch Peerbewertung mittels HOPA evaluiert werden. Der HOPERA hat das Potential, die Übergabe frei von Bewertungsfehlern umfassend zu evaluieren und Aussagen über die medizinisch-inhaltliche Komponente hinaus zu treffen, wenn die Itemstruktur überarbeitet und ergänzt wird. Um weitere valide Erkenntnisse zu erlangen, sollte die Untersuchung der Übergabesituation als Teil des kompetenzbasierten Simulationsmodells mit einer größeren Kohorte wiederholt werden. Insbesondere sollten weitere Hochschulstandorte mit unterschiedlichen Ausrichtungen des Curriculums partizipieren, um den Effekt des Studiengangmodells auf die Qualität der Übergabe weiterführend zu untersuchen. Zukünftig

können dann auch gezielte Interventionen entsprechend der einzelnen Komponenten der Übergabe entworfen und in das Curriculum integriert werden. Nur durch hohe Standards der Übergabequalität kann die Patientensicherheit gewährleistet werden.

## 6 Zusammenfassung

Die Übergabe stellt eine kritische Situation im klinischen Alltag dar, welche für die Patientensicherheit von zentraler Bedeutung ist. Es ist daher elementar, die Qualität von Übergaben zu ermitteln um kritisch einen möglichen Interventionsbedarf zu diskutieren. Das kompetenzbasierte Simulationsmodell eines ersten Arbeitstages für PJ-Studierende in der Ärztinnen- oder Arztrolle bot erstmals die Möglichkeit, die Qualität von Übergaben PJ-Studierender unterschiedlicher deutscher Hochschulen mit divergierender Ausrichtung des Curriculums zu untersuchen. Insgesamt nahmen 89 PJ-Studierende der Universität Hamburg, der TU München und der Universität Oldenburg teil. Teilnehmende übergaben Patienteninformationen, welche sie zuvor selbstständig während einer Sprechstunde mit drei simulierten Patientinnen und Patienten und in einer nachfolgenden Patientenmanagementphase gesammelt hatten, an andere Teilnehmende (Peers), die nicht mit den Patientinnen und Patienten vertraut waren. Die Bewertung der Qualität der Übergabe erfolgte durch die Peers mittels „Handover Partner Assessment“ (HOPA) und durch eine in dieser Arbeit etablierte und operationalisierte Fremdbewertung „Handover Performance External Rater Assessment“ (HOPERA). Die Struktur des HOPERA wurde mittels Faktorenanalysen untersucht, um mögliche grundlegende Komponenten der Qualität der Übergabe herauszuarbeiten. Die Qualität von Übergabegesprächen von PJ-Studierenden in der Ärztinnen- oder Arztrolle fiel insgesamt gut aus. Dabei konnten hochschulabhängige Unterschiede festgemacht werden, wenn die Qualität der Übergabe mittels des neuentwickelten operationalisierten Fremdbewertungsbogens HOPERA evaluiert wurde. Absolventinnen und Absolventen eines Regelstudienganges der TU München schnitten im Vergleich zu jenen eines Modellstudienganges der Universität Hamburg besser ab. Dieser Unterscheid zeigte sich insbesondere auf der Ebene des identifizierten medizinisch-inhaltlichen Faktors einer Übergabe. Bei der Bewertung der medizinisch-inhaltlichen Aspekte machte es dabei allerdings keinen Unterschied, ob diese anhand einer operationalisierten Bewertung oder mittels Peerbewertung erfolgte. Der Einfluss des Hochschulstandorts mit entsprechendem Studiengangmodell schwand mit fortgeschrittenem PJ-Semester und konnte ebenfalls durch besseres Abschneiden im zweiten Staatsexamen kompensiert werden. Der neuentwickelte operationalisierte Fremdbewertungsbogen HOPERA zeigte sich hinsichtlich der Beurteilung weiterer Faktoren einer Übergabe nicht gut geeignet. Mit dem Anspruch, die Komplexität der Übergabe besser abzubilden, bedarf es daher in Zukunft einer weiteren Überarbeitung und einer erneuten Validierung im Setting des kompetenzbasierten Simulationsmodells.

## 7 Summary

The handover is a critical situation in everyday clinical practice, which is of central importance for patient safety. Therefore, it is essential to assess the quality of handovers in order to highlight a possible need for intervention. The competence-based simulation model of a first working day for final year medical students in the physician's role offered the opportunity to examine the quality of handovers. Thus, handover quality of final year medical students from different German universities with different undergraduate curricula could be compared for the first time. A total of 89 final year students from the University of Hamburg, the Technical University of Munich, and the University of Oldenburg participated. Handover was performed after patient information was previously gathered independently during a consultation hour with three simulated patients and a following patient management phase. The patient information was handed over to other participants (peers), who were not familiar with the patients. The quality of the handover was assessed by the peers using the "Handover Partner Assessment" (HOPA) and by an operationalized external assessment "Handover Performance External Rater Assessment" (HOPERA), the latter being established in this project. The structure of the HOPERA was examined by factor analyzes in order to identify possible essential components of the quality of the handover. Overall, the quality of final year students' handovers in the doctor's role was good. Differences between the universities could be identified when the quality of the handover was assessed using the newly developed operationalized external evaluation tool HOPERA. Graduates of a regular undergraduate curriculum at the Technical University of Munich exceeded those of a model curriculum at the University of Hamburg. This difference was particularly evident at the level of the identified medical-content factor of a handover. However, when evaluating the medical content-related aspects, it made no difference whether this was done using the operationalized external evaluation or the peer evaluation. The influence of the university with the corresponding model of curriculum faded with advanced final year semesters and could also be compensated by better results in the second state examination. The newly developed operationalized external assessment HOPERA showed deficits in terms of the assessment of other factors involved in a handover. In order to better feature the complexity of the handover, a further revision and renewed validation in the setting of the competence-based simulation model is required in the future.

## 8 Abkürzungsverzeichnis

ÄKHOM	Ärztliche Kompetenzen Hamburg, Oldenburg, TU München
ANCOVA	Kovarianzanalyse
ANOVA	Varianzanalyse
BMBF	Bundesministerium für Bildung und Forschung
cMRT	kraniale Magnetresonanztomographie
DGAI	Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin
F	Wert der F-Verteilung aus Varianzanalyse
FOC	Facets of Competence
HIV	Humanes Immundefizienz-Virus
HOPA	Handover Partner Assessment
HOPERA	Handover Performance External Rater Assessment
IMPP	Institut für medizinische und pharmazeutische Prüfungsfragen
k	Konstante
KI	Konfidenzintervall
KMO	Kaiser-Meyer-Olkin
ln	natürlicher Logarithmus
n	Stichprobengröße, Anzahl der Bewertungen
p	Signifikanzniveau
PJ	Praktisches Jahr
RWTH	Rheinisch-Westfälische Technische Hochschule
SBAR	Situation, Background, Assessment, Recommendation
SD	Standardabweichung
t	T-Wert aus T-Test
TU	Technische Universität
U	U-Wert des Mann-Whitney-U-Tests
US	United States
USA	United States of America
WHO	World Health Organization
$\bar{x}$	Mittelwert
Z	Z-Wert des Mann-Whitney-U-Tests
$\eta^2$	Effektstärke: deskriptives Maß für den erklärten Varianzanteil in der Varianzanalyse

## 9 **Abbildungsverzeichnis**

<b>Abbildung 1:</b> Histogramm über den Gesamtscore (%) getrennt für die Hochschulstandorte Hamburg und München mit Referenzkurve einer Normalverteilung.....	21
---	----

## 10 Tabellenverzeichnis

<b>Tabelle 1:</b>	Prozentuale Übereinstimmung der Rater-Bewertungen beim ersten Probedurchlauf.....	17
<b>Tabelle 2:</b>	Prozentuale Übereinstimmung der Rater-Bewertungen beim zweiten Probedurchlauf.....	17
<b>Tabelle 3:</b>	Itemstatistik und Gesamtscore (%) der operationalisierten Fremdbewertung (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, * = p < ,05).....	18
<b>Tabelle 4:</b>	Itemstatistik und Gesamtscore (%) der Peerbewertung, Itemreihenfolge an HOPERA angepasst (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, * = p < ,05).....	19
<b>Tabelle 5:</b>	Untransformierter Gesamtscore (%) der Peer- und operationalisierten Bewertung sowie transformierter Gesamtscore der operationalisierten Bewertung (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, * = p < ,05).....	20
<b>Tabelle 6:</b>	Itemstatistik über beide Raterperspektiven der operationalisierten Fremd- und Peerbewertung (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, * = p < ,05, ** = p < ,005).....	22
<b>Tabelle 7:</b>	Vergleich der Peerbewertung mit der operationalisierten Fremdbewertung (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, ** = p < ,005) .....	23
<b>Tabelle 8:</b>	Faktorladungen der rotierten Faktorenmatrix auf Basis der operationalisierten Bewertungen.....	25
<b>Tabelle 9:</b>	Untransformierter Gesamtscore (%) der operationalisierten Bewertung sowie transformierter Gesamtscore der operationalisierten Bewertung in Hinblick auf die ermittelten Faktoren (n = Anzahl der Bewertungen, $\bar{x}$ = Mittelwert, SD = Standardabweichung, * = p < ,05).....	26

## 11 Literaturverzeichnis

- ARORA, V., JOHNSON, J., LOVINGER, D., HUMPHREY, H. J. & MELTZER, D. O. 2005. Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis. *Qual Saf Health Care*, 14, 401-7.
- ARORA, V. M., MANJARREZ, E., DRESSLER, D. D., BASAVIAH, P., HALASYAMANI, L. & KRIPALANI, S. 2009. Hospitalist handoffs: a systematic review and task force recommendations. *J Hosp Med*, 4, 433-40.
- BHABRA, G., MACKEITH, S., MONTEIRO, P. & POTHIER, D. D. 2007. An experimental comparison of handover methods. *Ann R Coll Surg Engl*, 89, 298-300.
- BIESMA, R., KENNEDY, M. C., PAWLIKOWSKA, T., BRUGHA, R., CONROY, R. & DOYLE, F. 2019. Peer assessment to improve medical student's contributions to team-based projects: randomised controlled trial and qualitative follow-up. *BMC Med Educ*, 19, 371.
- BOROWITZ, S. M., WAGGONER-FOUNTAIN, L. A., BASS, E. J. & SLEDD, R. M. 2008. Adequacy of information transferred at resident sign-out (in-hospital handover of care): a prospective survey. *Qual Saf Health Care*, 17, 6-10.
- CATCHPOLE, K. R., DE LEVAL, M. R., MCEWAN, A., PIGOTT, N., ELLIOTT, M. J., MCQUILLAN, A., MACDONALD, C. & GOLDMAN, A. J. 2007. Patient handover from surgery to intensive care: using Formula 1 pit-stop and aviation models to improve safety and quality. *Paediatr Anaesth*, 17, 470-8.
- CHEUNG, D. S., KELLY, J. J., BEACH, C., BERKELEY, R. P., BITTERMAN, R. A., BROIDA, R. I., DALSEY, W. C., FARLEY, H. L., FULLER, D. C., GARVEY, D. J., KLAUER, K. M., MCCULLOUGH, L. B., PATTERSON, E. S., PHAM, J. C., PHELAN, M. P., PINES, J. M., SCHENKEL, S. M., TOMOLO, A., TURBIAK, T. W., VOZENILEK, J. A., WEARS, R. L., WHITE, M. L., SECTION OF QUALITY, I. & PATIENT SAFETY, A. C. O. E. P. 2010. Improving handoffs in the emergency department. *Ann Emerg Med*, 55, 171-80.
- CLELAND, J. A., ROSS, S., MILLER, S. C. & PATEY, R. 2009. "There is a chain of Chinese whispers ...": empirical data support the call to formally teach handover to prequalification doctors. *Qual Saf Health Care*, 18, 267-71.
- CUNNINGHAM, N. J., WEILAND, T. J., VAN DIJK, J., PADDLE, P., SHILKOFSKI, N. & CUNNINGHAM, N. Y. 2012. Telephone referrals by junior doctors: a randomised controlled trial assessing the impact of SBAR in a simulated setting. *Postgrad Med J*, 88, 619-26.
- DE MEESTER, K., VERSPUY, M., MONSIEURS, K. G. & VAN BOGAERT, P. 2013. SBAR improves nurse-physician communication and reduces unexpected death: a pre and post intervention study. *Resuscitation*, 84, 1192-6.

- DENSON, J. L., MCCARTY, M., FANG, Y., UPPAL, A. & EVANS, L. 2015. Increased mortality rates during resident handoff periods and the effect of ACGME duty hour regulations. *Am J Med*, 128, 994-1000.
- DESMEDT, M., ULENAERS, D., GROSEMANS, J., HELLINGS, J. & BERGS, J. 2021. Clinical handover and handoff in healthcare: a systematic review of systematic reviews. *Int J Qual Health Care*, 33.
- FLANAGAN, M. E., PATTERSON, E. S., FRANKEL, R. M. & DOEBBELING, B. N. 2009. Evaluation of a physician informatics tool to improve patient handoffs. *J Am Med Inform Assoc*, 16, 509-15.
- FÜRSTENBERG, S., OUBAID, V., BERBERAT, P. O., KADMON, M. & HARENDZA, S. 2019. Medical knowledge and teamwork predict the quality of case summary statements as an indicator of clinical reasoning in undergraduate medical students. *GMS J Med Educ*, 36, Doc83.
- FÜRSTENBERG, S., SCHICK, K., DEPPERMAN, J., PREDIGER, S., BERBERAT, P. O., KADMON, M. & HARENDZA, S. 2017. Competencies for first year residents - physicians' views from medical schools with different undergraduate curricula. *BMC Med Educ*, 17, 154.
- GAKHAR, B. & SPENCER, A. L. 2010. Using direct observation, formal evaluation, and an interactive curriculum to improve the sign-out practices of internal medicine interns. *Acad Med*, 85, 1182-8.
- GORDON, M. & FINDLEY, R. 2011. Educational interventions to improve handover in health care: a systematic review. *Med Educ*, 45, 1081-9.
- HARENDZA, S., BERBERAT, P. & KADMON, M. 2017a. 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*, 07, 550.
- HARENDZA, S., KRENZ, I., KLINGE, A., WENDT, U. & JANNECK, M. 2017b. 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*, 34, Doc66.
- HAYNES, A. B., WEISER, T. G., BERRY, W. R., LIPSITZ, S. R., BREIZAT, A. H., DELLINGER, E. P., HERBOSA, T., JOSEPH, S., KIBATALA, P. L., LAPITAN, M. C., MERRY, A. F., MOORTHY, K., REZNICK, R. K., TAYLOR, B., GAWANDE, A. A. & SAFE SURGERY SAVES LIVES STUDY, G. 2009. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*, 360, 491-9.
- HORWITZ, L. I., KRUMHOLZ, H. M., GREEN, M. L. & HUOT, S. J. 2006. Transfers of patient care between house staff on internal medicine wards: a national survey. *Arch Intern Med*, 166, 1173-7.

- HORWITZ, L. I., MEREDITH, T., SCHUUR, J. D., SHAH, N. R., KULKARNI, R. G. & JENQ, G. Y. 2009a. Dropping the baton: a qualitative analysis of failures during the transition from emergency department to inpatient care. *Ann Emerg Med*, 53, 701-10 e4.
- HORWITZ, L. I., MOIN, T., KRUMHOLZ, H. M., WANG, L. & BRADLEY, E. H. 2009b. What are covering doctors told about their patients? Analysis of sign-out among internal medicine house staff. *Qual Saf Health Care*, 18, 248-55.
- IMPP. 2019. Archiv Ergebnisse Medizin. Mainz [Online im Internet]. URL: <https://www.impp.de/pruefungen/medizin/archiv-medizin.html> [Stand: 14 November 2019, 17:33].
- JEFFCOTT, S. A., EVANS, S. M., CAMERON, P. A., CHIN, G. S. & IBRAHIM, J. E. 2009. Improving measurement in clinical handover. *Qual Saf Health Care*, 18, 272-7.
- MAGIN, D. 2001. Reciprocity as a source of bias in multiple peer assessment of group work. *Stud High Educ*, 26, 53-63.
- MANSER, T. & FOSTER, S. 2011. Effective handover communication: an overview of research and improvement efforts. *Best Pract Res Clin Anaesthesiol*, 25, 181-91.
- MANSER, T., FOSTER, S., GISIN, S., JAECKEL, D. & UMMENHOFER, W. 2010. Assessing the quality of patient handoffs at care transitions. *Qual Saf Health Care*, 19, e44.
- MASTERSON, M. F., GILL, R. S., TURNER, S. R., SHRICHAND, P. & GIULIANI, M. 2013. A systematic review of educational resources for teaching patient handover skills to resident physicians and other healthcare professionals. *Can Med Educ J*, 4, e96-e110.
- MERKEL, M. J., VON DOSSOW, V. & ZWISSLER, B. 2017. Structured patient handovers in perioperative medicine: rationale and implementation in clinical practice. *Anaesthesist*, 66, 396-403.
- MONROUXE, L. V., GRUNDY, L., MANN, M., JOHN, Z., PANAGOULAS, E., BULLOCK, A. & MATTICK, K. 2017. How prepared are UK medical graduates for practice? A rapid review of the literature 2009-2014. *BMJ Open*, 7, e013656.
- MÜLLER, M., JÜRGENS, J., REDAËLLI, M., KLINGBERG, K., HAUTZ, W. E. & STOCK, S. 2018. Impact of the communication and patient hand-off tool SBAR on patient safety: a systematic review. *BMJ Open*, 8, e022202.
- PETERSEN, L. A., BRENNAN, T. A., O'NEIL, A. C., COOK, E. F. & LEE, T. H. 1994. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med*, 121, 866-72.
- PEZZOLESI, C., SCHIFANO, F., PICKLES, J., RANDELL, W., HUSSAIN, Z., MUIR, H. & DHILLON, S. 2010. Clinical handover incident reporting in one UK general hospital. *Int J Qual Health Care*, 22, 396-401.

- PICKERING, B. W., HURLEY, K. & MARSH, B. 2009. Identification of patient information corruption in the intensive care unit: using a scoring tool to direct quality improvements in handover. *Crit Care Med*, 37, 2905-12.
- PREDIGER, S., FÜRSTENBERG, S., BERBERAT, P. O., KADMON, M. & HARENDZA, S. 2019. Interprofessional assessment of medical students' competences with an instrument suitable for physicians and nurses. *BMC Med Educ*, 19, 46.
- PREDIGER, S., SCHICK, K., FINCKE, F., FURSTENBERG, S., OUBAID, V., KADMON, M., BERBERAT, P. O. & HARENDZA, S. 2020. Validation of a competence-based assessment of medical students' performance in the physician's role. *BMC Med Educ*, 20, 6.
- PRONOVOST, P. J., THOMPSON, D. A., HOLZMUELLER, C. G., LUBOMSKI, L. H., DORMAN, T., DICKMAN, F., FAHEY, M., STEINWACHS, D. M., ENGINEER, L., SEXTON, J. B., WU, A. W. & MORLOCK, L. L. 2006. Toward learning from patient safety reporting systems. *J Crit Care*, 21, 305-15.
- RAM, R. & BLOCK, B. 1992. Signing out patients for off-hours coverage: comparison of manual and computer-aided methods. *Proc Annu Symp Comput Appl Med Care*, 114-8.
- RAUPACH, T., VOGEL, D., SCHIEKIRKA, S., KEIJSERS, C., TEN CATE, O. & HARENDZA, S. 2013. Increase in medical knowledge during the final year of undergraduate medical education in Germany. *GMS Z Med Ausbild*, 30, Doc33.
- RIESENBERG, L. A., LEITZSCH, J. & LITTLE, B. W. 2009. Systematic review of handoff mnemonics literature. *Am J Med Qual*, 24, 196-204.
- SCHMIDT, H. G., VERMEULEN, L. & VAN DER MOLEN, H. T. 2006. Longterm effects of problem-based learning: a comparison of competencies acquired by graduates of a problem-based and a conventional medical school. *Med Educ*, 40, 562-7.
- SHENG, J., MANJUNATH, S., MICHAEL, M., GAJERA, P., WANG, E., SENDELBACH, D., GREILICH, P. & AMBARDEKAR, A. 2020. Integrating handover curricula in medical school. *Clin Teach*, 17, 661-8.
- SINHA, M., SHRIKI, J., SALNESS, R. & BLACKBURN, P. A. 2007. Need for standardized sign-out in the emergency department: a survey of emergency medicine residency and pediatric emergency medicine fellowship program directors. *Acad Emerg Med*, 14, 192-6.
- STOJAN, J. N., SCHILLER, J. H., MULLAN, P., FITZGERALD, J. T., CHRISTNER, J., ROSS, P. T., MIDDLEMAS, S., HAFTEL, H., STANSFIELD, R. B. & LYPSON, M. L. 2015. Medical school handoff education improves postgraduate trainee performance and confidence. *Med Teach*, 37, 281-8.

- THAETER, L., SCHRODER, H., HENZE, L., BUTTE, J., HENN, P., ROSSAINT, R. & SOPKA, S. 2018. Handover training for medical students: a controlled educational trial of a pilot curriculum in Germany. *BMJ Open*, 8, e021202.
- VAN EATON, E. G., HORVATH, K. D., LOBER, W. B., ROSSINI, A. J. & PELLEGRINI, C. A. 2005. A randomized, controlled trial evaluating the impact of a computerized rounding and sign-out system on continuity of care and resident work hours. *J Am Coll Surg*, 200, 538-45.
- VON DOSSOW, V. & ZWISSLER, B. 2016. Recommendations of the german association of anesthesiology and intensive care medicine (DGAI) on structured patient handover in the perioperative setting : The SBAR concept. *Anaesthetist*, 65, 1-4.
- WALTON, M., WOODWARD, H., VAN STAALDUINEN, S., LEMER, C., GREAVES, F., NOBLE, D., ELLIS, B., DONALDSON, L., BARRACLOUGH, B. & EXPERT GROUP CONVENED BY THE WORLD ALLIANCE OF PATIENT SAFETY, A. E. L. F. T. S.-P. 2011. Republished paper: The WHO patient safety curriculum guide for medical schools. *Postgrad Med J*, 87, 317-21.
- WAYNE, J. D., TYAGI, R., REINHARDT, G., ROONEY, D., MAKOUL, G., CHOPRA, S. & DAROSA, D. A. 2008. Simple standardized patient handoff system that increases accuracy and completeness. *J Surg Educ*, 65, 476-85.
- WEINGER, M. B., SLAGLE, J. M., KUNTZ, A. H., SCHILDCROUT, J. S., BANERJEE, A., MERCALDO, N. D., BILLS, J. L., WALLSTON, K. A., SPEROFF, T., PATTERSON, E. S. & FRANCE, D. J. 2015. A multimodal intervention improves postanesthesia care unit handovers. *Anesth Analg*, 121, 957-71.
- WIJNEN-MEIJER, M., TEN CATE, O., VAN DER SCHAAF, M. & HARENDZA, S. 2013a. Graduates from vertically integrated curricula. *Clin Teach*, 10, 155-9.
- WIJNEN-MEIJER, M., TEN CATE, O. T., VAN DER SCHAAF, M. & BORLEFFS, J. C. 2010. Vertical integration in medical school: effect on the transition to postgraduate training. *Med Educ*, 44, 272-9.
- WIJNEN-MEIJER, M., VAN DER SCHAAF, M., BOOIJ, E., HARENDZA, S., BOSCARDIN, C., VAN WIJNGAARDEN, J. & TEN CATE, T. J. 2013b. An argument-based approach to the validation of UHTRUST: can we measure how recent graduates can be trusted with unfamiliar tasks? *Adv Health Sci Educ Theory Pract*, 18, 1009-27.
- WISSENSCHAFTSRAT. 2014. Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge. Dresden [Online im Internet]. URL: [https://www.wissenschaftsrat.de/download/archiv/4017-14.pdf?\\_\\_blob=publicationFile&v=1](https://www.wissenschaftsrat.de/download/archiv/4017-14.pdf?__blob=publicationFile&v=1) [Stand: 25 Januar 2021, 12:07].

- WONG, M. C., YEE, K. C. & TURNER, P. 2008. A structured evidence-based literature review regarding the effectiveness of improvement interventions in clinical handover. University of Tasmania, Australia [Online im Internet]. URL: <https://www.safetyandquality.gov.au/sites/default/files/migrated/Clinical-Handover-Literature-Review-for-release.pdf> [Stand: 18 November 2020, 19:56].
- WORLD HEALTH ORGANIZATION. 2020. Daten und Statistiken zum Thema Patientensicherheit. Kopenhagen [Online im Internet]. URL: <https://www.euro.who.int/en/health-topics/Health-systems/patient-safety/data-and-statistics> [Stand: 18 November 2020, 16:45].
- YEE, K. C., WONG, M. C. & TURNER, P. 2006. Medical error management and the role of information technology - a new approach to investigating medical handover in acute care settings. *Stud Health Technol Inform*, 124, 679-84.

## 12 Anhang

### 12.1 Anhang 1 – HOPA

Bitte bewerten Sie die Kollegin/ den Kollegen, die/ der Ihnen ihre/ seine Patientenfälle übergeben hat!

Meine Kollegin/ mein Kollege ...		trifft ... zu					nicht beurteilbar
		überhaupt nicht				voll und ganz	
1.	hat in einer schlüssigen Weise über die Patient*innen berichtet, so dass ich kaum Nachfragen stellen musste.	<input type="radio"/>					
2.	hat seine Diagnostik- und Behandlungsschritte für mich gut nachvollziehbar begründet.	<input type="radio"/>					
3.	hat mir für die Patient*innen eine Strategie berichtet, wie ich mit diesen weiter arbeiten soll.	<input type="radio"/>					
4.	hat relevante Ergebnisse der Patientenfälle in adäquater Sprache fokussiert berichtet.	<input type="radio"/>					
5.	hat meine Ansichten, mein Wissen und meine Expertise im Gespräch berücksichtigt.	<input type="radio"/>					
6.	hat sich mir gegenüber respektvoll verhalten.	<input type="radio"/>					
7.	hat auf verschiedene Art und Weise geäußert, dass sie/ er <u>nicht</u> weitergekommen ist.	<input type="radio"/>					
8.	hat mich nach meiner Meinung zu Patient*innen gefragt, bei denen noch nicht alles geklärt war.	<input type="radio"/>					
9.	hat kommuniziert, manche Dinge <u>nicht</u> erledigt/ geschafft zu haben.	<input type="radio"/>					
10.	hat sich kritikfähig gezeigt.	<input type="radio"/>					
11.	hat <u>nicht</u> Andere (z.B. die Mitarbeiter*innen des Labors/ der Röntgenabteilung) für Fehler verantwortlich gemacht.	<input type="radio"/>					

## 12.2 Anhang 2 – HOPERA

Bitte bewerten Sie die Übergabe der Studentin/ des Studenten

Die Studentin/ der Student ...		trifft ... zu					nicht beurteilbar
		überhaupt nicht				voll und ganz	
1.	hat in einer schlüssigen Weise über die Patient*innen berichtet, so dass seine Kollegin/ sein Kollege kaum Nachfragen stellen musste.	<input type="radio"/>					
2.	hat seine Diagnostik- und Behandlungsschritte gut nachvollziehbar begründet.	<input type="radio"/>					
3.	hat für die Patient*innen eine Strategie berichtet, wie seine Kollegin/ sein Kollege mit diesen weiter arbeiten sollte.	<input type="radio"/>					
4.	hat relevante Ergebnisse der Patientenfälle in adäquater Sprache fokussiert berichtet.	<input type="radio"/>					
5.	hat kommuniziert, manche Dinge <u>nicht</u> erledigt/ geschafft zu haben.	<input type="radio"/>					
6.	hat auf verschiedene Art und Weise geäußert, dass sie/ er <u>nicht</u> weitergekommen ist.	<input type="radio"/>					
7.	hat <u>nicht</u> Andere (z.B. die Mitarbeiter*innen des Labors/ der Röntgenabteilung) für Fehler verantwortlich gemacht.	<input type="radio"/>					
8.	hat Ansichten, Wissen und Expertise der Kollegin/ des Kollegen im Gespräch berücksichtigt.	<input type="radio"/>					
9.	hat nach der Meinung der Kollegin/ des Kollegen zu Patient*innen gefragt	<input type="radio"/>					
10.	hat sich der Kollegin/ dem Kollegen gegenüber respektvoll verhalten.	<input type="radio"/>					
11.	hat sich kritikfähig gezeigt.	<input type="radio"/>					

## 12.3 Anhang 3 – HOPERA operationalisierte Bewertung

Item 1	<p>1 =berichtet unschlüssig; Konzepte wirken nicht durchdacht; Nachfragen werden ebenfalls unschlüssig beantwortet</p> <p>2 =berichtet eher unschlüssig; Konzepte sind zum kleineren Teil erkennbar; Nachfragen werden zum Teil nicht schlüssig beantwortet</p> <p>3 =berichtet zum Teil unschlüssig; Nachfragen werden eher schlüssig beantwortet</p> <p>4 =berichtet überwiegend schlüssig; vermehrt Nachfragen, die eher schlüssig beantwortet werden können</p> <p>5 =berichtet durchgängig schlüssig; Nachfragen werden immer schlüssig beantwortet</p>
Item 2	<p>1 =keine Begründung</p> <p>2 =versucht zu begründen, aber kaum nachvollziehbar</p> <p>3 =versucht zu begründen, teilweise nachvollziehbar</p> <p>4 =begründet überwiegend nachvollziehbar</p> <p>5 =begründet durchgängig nachvollziehbar</p>
Item 3	<p>1 =berichtet keine Strategien für die Patient*innen und hat auch auf Rückfragen keine Strategievorschläge</p> <p>2 =berichtet zum Teil keine Strategien, die ggf. auch nicht weiter diskutiert werden; hat auch auf Rückfragen keine Strategievorschläge</p> <p>3 =berichtet teilweise Strategien; hat auf Rückfrage weitere Ansätze einer Strategie berichtet; diskutiert ggf. weitere Ansätze</p> <p>4 =berichtet überwiegend Strategien; hat auf Rückfragen überwiegend weitere Strategievorschläge</p> <p>5 =berichtet durchgängig Strategien; hat auf Rückfragen weitere Vorschläge oder nimmt Vorschläge auf</p>
Item 4	<p>1 =keine Fachsprache oder fundamental inkorrekte Fachsprache; unfokussiert</p> <p>2 =wenig Fachsprache oder häufig inkorrekte Fachsprache; überwiegend unfokussiert</p> <p>3 =partiell Fachsprache und wenig inkorrekte Fachsprache; teilweise fokussiert</p> <p>4 =mehrheitlich Fachsprache und ausschließlich korrekte Fachsprache; überwiegend fokussiert</p> <p>5 =intensiver Gebrauch korrekter Fachsprache; fokussiert</p>
Item 5	<p>1 =keine Nachfrage; keine Kommunikation</p> <p>2 =auf Nachfrage nicht adäquat eingegangen; fälschlicherweise behauptet alles erledigt zu haben</p> <p>3 =ausschließlich auf Nachfrage</p> <p>4 =eher auf Nachfrage; seltener von sich aus</p> <p>5 =mehr als einmal von sich aus</p>
Item 6	<p>1 =ist offensichtlich stellenweise nicht weitergekommen, äußert dies aber nicht.</p> <p>2 =ist offensichtlich stellenweise nicht weitergekommen; äußert dies jedoch nur implizit (z.B. über Körpersprache oder quengelt)</p> <p>3 =gibt auf Nachfrage an, recherchiert haben zu müssen oder Hilfe telefonisch kontaktiert zu haben, da sie/ er nicht weitergekommen ist</p> <p>4 =gibt teilweise von sich aus zu, recherchiert haben zu müssen oder Hilfe kontaktiert zu haben, da er/ sie nicht weitergekommen ist; steht dazu, dass sie/ er nicht weitergekommen ist.</p> <p>5 =formuliert klar, dass sie/ er nicht weitergekommen ist; hat explizit um Hilfe gebeten; hat eigene Ideen entwickelt und diskutiert diese bei der Übergabe</p>

Item 7	<p>1 =hat Andere für Fehler verantwortlich gemacht und despektierliche Begriffe verwendet und/ oder Beteiligte diffamiert</p> <p>2 =hat Andere für Fehler verantwortlich gemacht und beschwert sich moderat, hat dafür kein Verständnis</p> <p>3 =hat einen Fehler bemerkt und äußert sich verständnisvoll „kann ja mal passieren“; sieht sich nicht in der Verantwortung, beschuldigt aber auch nicht klar Andere</p> <p>4 =hat einen Fehler bemerkt aber nicht korrigiert; übernimmt Verantwortung für die falsche Interpretation oder das Weiterarbeiten mit einem falschen Befund</p> <p>5 =hat einen Fehler bemerkt und gesagt, wie sie/ er ihn gelöst hat, ohne jemanden zu beschuldigen; weist Kollegin/ Kollegen im Sinne eines nachhaltigen Fehlermanagements darauf hin</p> <p>nicht beurteilbar = spricht nicht an, dass ihr/ ihm ein fehlerhafter Befund vorgelegt worden ist; fällt nicht auf, dass ihr/ ihm ein fehlerhafter Befund vorgelegt worden ist, hat ggf. den falschen Befund auch nicht erhalten</p>
Item 8	<p>1 =negiert Input seiner Gesprächspartnerin/ seines Gesprächspartners; widerspricht</p> <p>2 =hat Input seiner Gesprächspartnerin/ seines Gesprächspartners mehrfach ignoriert</p> <p>3 =geht auf Input seiner Gesprächspartnerin/ seines Gesprächspartners meist nur oberflächlich ein; nimmt Input zur Kenntnis ohne ihn weiter zu diskutieren; hält an eigener Idee fest</p> <p>4 =eher offen gegenüber Anregungen, verhaltene Diskussionsbereitschaft, hält eher an eigener Idee fest</p> <p>5 =kann sich zurücknehmen, diskutiert Input, bereit andere Ideen weiter zu verfolgen</p>
Item 9	<p>1 =Dinge ungeklärt und hat nie nach Meinung gefragt und geht auf spontane Meinungsäußerungen* der Gesprächspartnerin/ des Gesprächspartners nicht ein; tut Meinungen oberflächlich ab</p> <p>2 =Dinge ungeklärt und hat nicht nach Meinung gefragt, geht auf spontane Meinungsäußerungen* nur oberflächlich ein</p> <p>3 =Dinge ungeklärt und hat nicht nach Meinung gefragt, geht aber auf spontane Meinungsäußerungen* der Gesprächspartnerin/ des Gesprächspartners ein, kommt darüber in einen Dialog und Meinungs austausch</p> <p>4 =Dinge ungeklärt und hat einmalig zur Meinung gefragt und ist dann auf Meinung eingegangen ODER: bei Patienten war weitestgehend alles geklärt, <b>keine</b> Meinung eingeholt, sofern spontane Meinungsäußerungen* vorhanden, wird auf diese allerdings <b>nicht</b> eingegangen</p> <p>5 =Dinge ungeklärt und hat mehr als einmal nach Meinung gefragt und ist dann auf diese Meinungen eingegangen ODER: bei Patienten war weitestgehend alles geklärt, <b>selten oder nie</b> Meinung eingeholt, sofern spontane Meinungsäußerungen* vorhanden, wird auf diese eingegangen</p> <p>*oder Rückfragen die Vorschläge implizieren</p>

Item 10	<p>1 =verhält sich überwiegend unkollegial; beleidigt Kollegin/ Kollegen</p> <p>2 =verhält sich eher unkollegial; behandelt Kollegin/ Kollegen von oben herab</p> <p>3 =verhält sich überwiegend kollegial; neigt jedoch dazu persönliche Grenzen zu überschreiten; oder grenzt sich zu stark ab, beschäftigt sich mehr mit sich; nimmt Kollegin/ Kollegen nicht immer ernst</p> <p>4 =verhält sich kollegial; nimmt Kollegin/ Kollegen überwiegend ernst; zeigt sich aufmerksam</p> <p>5 =verhält sich kollegial; nimmt Kollegin/ Kollegen immer ernst; lobt Vorschläge der Kollegin/ des Kollegen</p>
Item 11	<p>1 =kann Kritik nicht annehmen; schlägt zurück</p> <p>2 =kann Kritik kaum annehmen; fühlt sich angegriffen; nimmt Kritik persönlich; versucht sich mit Ausreden zu retten; keine Bereitschaft zur Veränderung</p> <p>3 =kann Kritik nur teilweise annehmen; empfindet Kritik als ungerecht; rechtfertigt sich; wenig Bereitschaft zur Veränderung</p> <p>4 =kann Kritik eher annehmen; eingeschränkte Bereitschaft zur Veränderung; zum Teil konstruktiver Umgang</p> <p>5 =kann Kritik voll annehmen; schätzt Kritik wert; kann Kritik produktiv verarbeiten; lernt aus Kritik; hohe Bereitschaft zur Veränderung</p>

## **13 Danksagung**

An dieser Stelle möchte ich allen beteiligten Personen danken, die in das Projekt ÄKHOM involviert sind. Insbesondere möchte ich mich bei meiner Doktormutter Frau Prof. Dr. med. Sigrid Harendza für die ausgezeichnete Betreuung bei der Umsetzung dieser Arbeit bedanken. Meiner Familie danke ich für ihre Unterstützung in allen Lebenslagen.

## **14 Lebenslauf**

Der Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

## 15 Eidesstattliche Erklärung

Ich versichere ausdrücklich, dass ich die Arbeit selbstständig 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: .....