

Institut für Didaktik und Ausbildungsforschung in der Medizin

Institut der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Martin Fischer

Eine qualitative und quantitative Analyse der mündlichen Facharztprüfung Innere Medizin

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Marc Aurelio Beckers

aus Geldern

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

1. Berichterstatter: Prof. Dr. Ralf Schmidmaier

2. Berichterstatter: Prof. Dr. Jörg Schelling

Mitberichtersteller: PD Dr. Stefanie Heinze

Mitbetreuung durch die
promovierten Mitarbeiter: Dr. phil. Nurith Epstein

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 10.03.2022

Danksagung:

An erster Stelle gilt mein Dank Herrn Professor Dr. Ralf Schmidmaier für die Vermittlung der Doktorarbeit und die über den gesamten Verlauf der Dissertation schnelle und kompetente Beratung.

Ebenfalls möchte ich Herrn Professor Dr. Martin Fischer für die Unterstützung und wichtige Vorschläge bei dem Aufbau des Ergebnisteils danken.

Ich möchte auch Frau Dr. Nurith Epstein für die Betreuung und Mithilfe bei der Konzeption der Arbeit, ganz besonders zu Beginn der Planung und Durchführung, danken.

Danken möchte ich außerdem Leah Braun für die wichtige und konstruktive Kritik an der Struktur und der inhaltlichen Darstellung dieser Monographie.

Auch Katharina Borrmann möchte ich für die Mithilfe bei der Durchführung der Interkoder-Reliabilitätsprüfung danken.

Zuletzt möchte ich meinen Eltern danken, die mich während dieser Arbeit zu jeder Zeit bedingungslos unterstützt haben.

Ich widme diese Doktorarbeit meiner Familie

INHALTSVERZEICHNIS

ZUSAMMENFASSUNG.....	VII
SUMMARY.....	VIII
1 ZIELSETZUNG DER ARBEIT.....	1
2 EINLEITUNG.....	1
2.1 KOMPETENZEN IN DER MEDIZIN	1
2.2 PRÜFUNGSFORMEN IN DER MEDIZINISCHEN LEHRE	3
2.2.1 <i>Allgemeines</i>	3
2.2.2 <i>Verschiedene Prüfungsformen</i>	4
2.3 HINTERGRUND DER DISSERTATION	6
2.4 RELEVANZ	8
3 FRAGESTELLUNG.....	8
3.1 DIE FORSCHUNGSFRAGE.....	9
3.2 HYPOTHESEN	9
4 MATERIAL UND METHODEN.....	10
4.1 ZUGRUNDE LIEGENDES MATERIAL	10
4.2 DATENMANAGEMENT UND AUSWERTUNG	10
4.2.1 <i>Prinzipien der strukturierten Inhaltsanalyse nach Mayring</i>	10
4.2.2 <i>Anforderungen an das Kategoriensystem</i>	11
4.2.3 <i>Das ULME-Kompetenzstufenmodell</i>	12
4.2.4 <i>Der Kodierprozess</i>	15
4.2.5 <i>Qualitätssicherung des Kodierprozesses</i>	15
4.3 DIE ANALYSESOFTWARE	16
5 ERGEBNISSE	17
5.1 STICHPROBE.....	17
5.1.1 <i>Aufteilung der Prüfungen/Protokolle nach Bundesland</i>	17
5.1.2 <i>Facharztbezeichnung</i>	17
5.1.3 <i>Neue/Alte Weiterbildungsordnung</i>	18
5.1.4 <i>Spezialisierung der Prüfer</i>	18
5.2 SPEZIFISCHE VARIANZ UNTERSCHIEDLICHER THEMENGEBIETE.....	19
5.2.1 <i>Die Anzahl der gestellten Fragen pro Prüfung</i>	19
5.2.2 <i>Die Anzahl unterschiedlicher Themengebiete pro Prüfung</i>	19
5.3 GLOBALE AUFTEILUNG DER THEMENGEBIETE	20
5.3.1 <i>Aufteilung der Fallvorstellungen nach Themengebiet</i>	21
5.3.2 <i>Aufteilung der Einzelfragen nach Themengebiet</i>	21
5.3.3 <i>Überschneidungen von Kodierungen und Themengebiet</i>	22
5.3.4 <i>Aufteilung der inhaltlichen Themengebiete im Vergleich zum „Blueprint“ des American Board of Internal Medicine und Guidance on the Swiss Board Examination in General Internal Medicine</i>	25
5.4 AUFTEILUNG UNTERSCHIEDLICHER UNTERSUCHUNGS-UND BEHANDLUNGSMETHODEN.....	27
5.4.1 <i>Aufteilung der Untersuchungs- und Behandlungsmethoden</i>	27
5.5 AUFTEILUNG DES KOMPETENZNIVEAUS NACH DER ULME-KLASSIFIKATIONSMATRIX	28
5.5.1 <i>Aufteilung der Leistungskategorien</i>	28
5.5.2 <i>Aufteilung der Wissensdimensionen</i>	28
5.5.3 <i>Die Verknüpfung von Leistungskategorie und Wissensdimension</i>	29
6 DISKUSSION.....	31
6.1 ZUSAMMENFASSUNG DER ARBEIT	31
6.2 DISKUSSION DER METHODEN.....	31
6.2.1 <i>Stärken und Schwächen</i>	31
6.2.2 <i>Wissenschaftlicher Vergleich</i>	32
6.2.3 <i>Ausblick</i>	33
6.3 DISKUSSION DER ERGEBNISSE.....	34
6.3.1 <i>Diskussion zu Hypothese 1</i>	34

6.3.2	<i>Diskussion zu Hypothese 2</i>	35
6.3.3	<i>Diskussion zu Hypothese 3</i>	36
6.3.4	<i>Diskussion zu Hypothese 4</i>	37
6.4	AUSBLICK.....	37
7	LITERATURVERZEICHNIS	39
8	ABBILDUNGSVERZEICHNIS	41
9	TABELLENVERZEICHNIS	41
10	ABKÜRZUNGSVERZEICHNIS	42
11	ANHANG	43
11.1	ZUSÄTZLICHE ABBILDUNGEN	43
11.2	EIDESSTATTLICHE ERKLÄRUNG	45

Zusammenfassung

Die Facharztprüfung ist die letzte Prüfung in der ärztlichen Aus- und Weiterbildung und deswegen von besonderer Bedeutung für die Qualität der ärztlichen Versorgung. Die Prüfung soll die Inhalte der Weiterbildungsordnung abdecken und findet in mündlicher Form statt. Es ist allerdings bisher wenig über die tatsächlich abgefragten Inhalte und die Qualität der Prüfung bekannt. In dieser Studie sollte die folgende Frage beantwortet werden:

Welche inhaltlichen und strukturellen Charakteristika haben Facharztprüfungen in Deutschland am Beispiel der Inneren Medizin?

100 Protokolle von Facharztprüfungen aus den Jahren 2013-2016, die von den Prüflingen verfasst wurden, wurden ausgewertet. Es erfolgte eine qualitative Analyse der Daten nach Mayring. Die Daten wurden im Hinblick auf geprüfte Schwerpunktbereiche und abgefragte Kompetenzen analysiert.

Die Fragetypen wurden in die drei Kategorien: „Fallvorstellungen“, „Kurzfragen“ und „Medien“ eingeteilt. Zusammengenommen wurden in den einzelnen Protokollen im Mittel 27 (SD=10) Fragen gestellt. Die Themengebiete Gastroenterologie und Kardiologie sind hierbei mit jeweils 17% am häufigsten abgefragt worden. Die Varianz unterschiedlicher Themengebiete in den einzelnen Prüfungen war hoch. Es erfolgte zudem eine Einteilung der „Kurzfragen“ in jeweils eine Wissensdimension (Fakten, Konzepte, Prozeduren) und eine Leistungskategorie (Reproduzieren, Verstehen und Anwenden, Kritisierten und Reflektieren). Das Reproduzieren (49,01%) von Fakten (43,04%) machte dabei den Großteil der abgefragten Kompetenzen aus. Im Vergleich mit dem Blueprint des „American Board of Internal Medicine“ stimmten 5 der 6 am häufigsten abgefragten Schwerpunktbereiche mit der Aufteilung der Protokolle überein.

In der mündlichen Facharztprüfung Innere Medizin werden viele Themengebiete und Wissensstichproben abgeprüft, die Varianz zwischen einzelnen Prüfungen ist dabei aber hoch.

Inhaltlich kam der Abfrage von Verständnis und kritischem Reflektieren klinischer Inhalte zudem nur eine untergeordnete Rolle zu. Eine bessere Standardisierung der Prüfung im Sinne der Entwicklung eines Blueprints und die Überlegung, andere Prüfungsformate zu der mündlichen Prüfung hinzuzufügen, könnten die Bedingungen für alle Kandidaten angleichen und im Sinne der Testgütekriterien verbessern.

Summary

The medical specialist exam is the final assessment in medical education and therefore, it inherits a great importance regarding the quality of medical care. The exam should cover all contents of the education guidelines for physicians (WBO) and is conducted in the fashion of an (unstructured) oral exam. However, little is known about the actual queried contents and quality of this exam. In this survey, the following question is ought to be answered: *Which contentual and structural characteristics do medical specialist exams have in Germany on the example of the Internal Medicine exam?*

100 protocols of medical specialist exams composed by candidates from the years 2013-2016 have been analysed. A qualitative analysis of the gathered data according to Mayring followed. The data was analysed in regards to queried thematic areas and competences.

The types of questions then were categorized into three classes: “case presentations”, “short questions” and “medial questions”. Collectively for the three classes, 27 (SD=10) questions were asked per individual protocol. With both 17% the thematic areas Gastroenterology and Cardiology were queried most frequently. Furthermore, a classification of “short questions” into knowledge dimensions (facts, concepts, procedures) and performance categories (reproducing, understanding and applying, criticising and reflecting on) was conducted. Comparing the Blueprint of the American Board of Internal Medicine, 5 out of 6 most frequently queried thematic areas of the American Blueprint corresponded with the distribution of the protocols.

On average the quantity of queried thematic areas and knowledge samples was sufficient, however the variety between individual exams was high. Content related comprehension and critical reflection of clinical content played a subordinate role. Amongst other things, the determination of obligate contents for query referring to the conceptualisation of a Blueprint and consideration of additional assessment formats could lead to a higher conformity in conditions for all examinees and an improvement of exam quality in regards to assessment quality criteria.

1 Zielsetzung der Arbeit

Die von den Facharztanwärtern geforderten Kompetenzen, die zum Erreichen des Facharztstitels notwendig sind, können aus der Weiterbildungsordnung (WBO) des jeweiligen Bundeslandes entnommen werden.

Diese Arbeit hat das Ziel, näher zu beleuchten, wie diese einzelnen Anforderungen in der internistischen Facharztprüfung umgesetzt werden. Dazu erfolgte eine Einteilung der geprüften Fragen in unterschiedliche Schwerpunkte. Diese Einteilung gibt Aufschluss auf die quantitative Verteilung der gestellten Fragen und zugehörigen Schwerpunkte.

Während Länder wie die Schweiz oder die USA einen „Blueprint“ für die Facharztprüfung bereitstellen, der eine prozentuale Aufgliederung der Schwerpunktbereiche vorgibt, ist dies in Deutschland nicht der Fall. Mit den Ergebnissen dieser Arbeit soll eine prozentuale Aufteilung der Fachgebiete untereinander erstellt werden (1).

Des Weiteren sollten Informationen über das Kompetenzniveau offengelegt werden. Dazu wurden die Fragen nach vorher definierten Kompetenzebenen analysiert und kategorisiert. Damit soll ermöglicht werden, dass zu jeder gestellten Frage eine Aussage über den Schwerpunktbereich und das geforderte Kompetenzlevel getroffen werden kann.

Diese Arbeit soll einen Beitrag zur Qualitätsanalyse und zur Bewertung der Standardisierung der internistischen Facharztprüfung leisten. Die Arbeit richtet sich außerdem der Fragestellung, ob ein Bedarf an Änderung der inhaltlichen und methodischen Konzeption der Facharztprüfungen in Deutschland besteht.

2 Einleitung

2.1 Kompetenzen in der Medizin

Die Qualitätssicherung der medizinischen Versorgung liegt in der Verantwortlichkeit ärztlicher Standesorganisationen. Diese sollen die kompetente und zuverlässige Berufsausübung ihrer Mitglieder gewährleisten. Durch eine konsistent hohe professionelle ärztliche Kompetenz kann die adäquate Versorgung der Patienten sichergestellt werden (1).

Zum Erreichen des Ziels der anerkannten Ausübung ärztlicher Tätigkeiten sind in Deutschland zwei Ausbildungsschritte notwendig: Der Abschluss des Studiums der Humanmedizin und der nachfolgenden Approbationserteilung definiert hierbei den

ersten Abschnitt der ärztlichen Ausbildung. Der NKLM (Nationale Kompetenzbasierte Lernzielkataloge Medizin) gibt Lernziele und geforderte Kompetenzen vor. Nach Erfüllen der gesetzlichen Vorgaben der ÄAppO (Ärztliche Approbationsordnung), ist der Absolvent dazu berechtigt, die ärztliche Weiterbildung zu beginnen.

Die ärztliche Weiterbildung wird durch die Weiterbildungsordnung der zuständigen Landesärztekammern geregelt. Durch Erfüllung der in der WBO geforderten Dokumentation vordefinierter Tätigkeiten und Bestehen einer abschließenden Facharztprüfung ist der Absolvent dazu berechtigt, den Facharzttitel als Nachweis für erworbene Kompetenz zu tragen (2). Um zu gewährleisten, dass die geforderten Kompetenzstandards eingehalten werden, ist es notwendig, die Kompetenzen der Absolventen der medizinischen Aus- und Weiterbildung zu prüfen (3).

Für die Messung von Kompetenz ist zunächst eine Definition erforderlich. Verschiedene Ansätze haben sich hierzu etabliert:

Die Musterweiterbildungsordnung der Bundesärztekammer bezeichnet Kompetenz als die Teilmenge der Inhalte eines Gebietes, die zum Erwerb von Kenntnissen, Erfahrungen und Fertigkeiten in einer Facharztweiterbildung notwendig sind und durch eine Prüfung nachgewiesen werden (2).

Der NKLM definiert Kompetenz noch ausführlicher als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (4)

Kompetenz wird dabei weiter in verschiedene Kompetenzebenen unterteilt, welche die Lerntiefe kennzeichnen (5). Die Definition des NKLM entstand im Kontext bereits etablierter Standardtaxonomien von Kompetenzebenen. Hier sind besonders die Miller-Pyramide und die Taxonomie des Schweizer Lernzielkatalogs von Relevanz (siehe Abb. 1) (6) (7). Insgesamt drei Kompetenzebenen wurden dabei für die NKLM-Klassifikation definiert:

1. Das Faktenwissen steht dabei für das Nennen und Beschreiben von deskriptiven Wissen.
2. Das Handlungs- und Begründungswissen steht für das Erklären von Sachverhalten und Zusammenhängen, die in einen klinisch-wissenschaftlichen Kontext eingeordnet und datenbasiert bewertet werden.
3. Die Handlungskompetenz ist noch einmal in zwei Untergruppen unterteilt: Zum Einen in die Durchführung und Demonstration einer Handlung unter Anleitung und zum

Anderen in die selbstständige, situationsadäquate Durchführung einer Handlung in Kenntnis der Konsequenzen dieser (5).

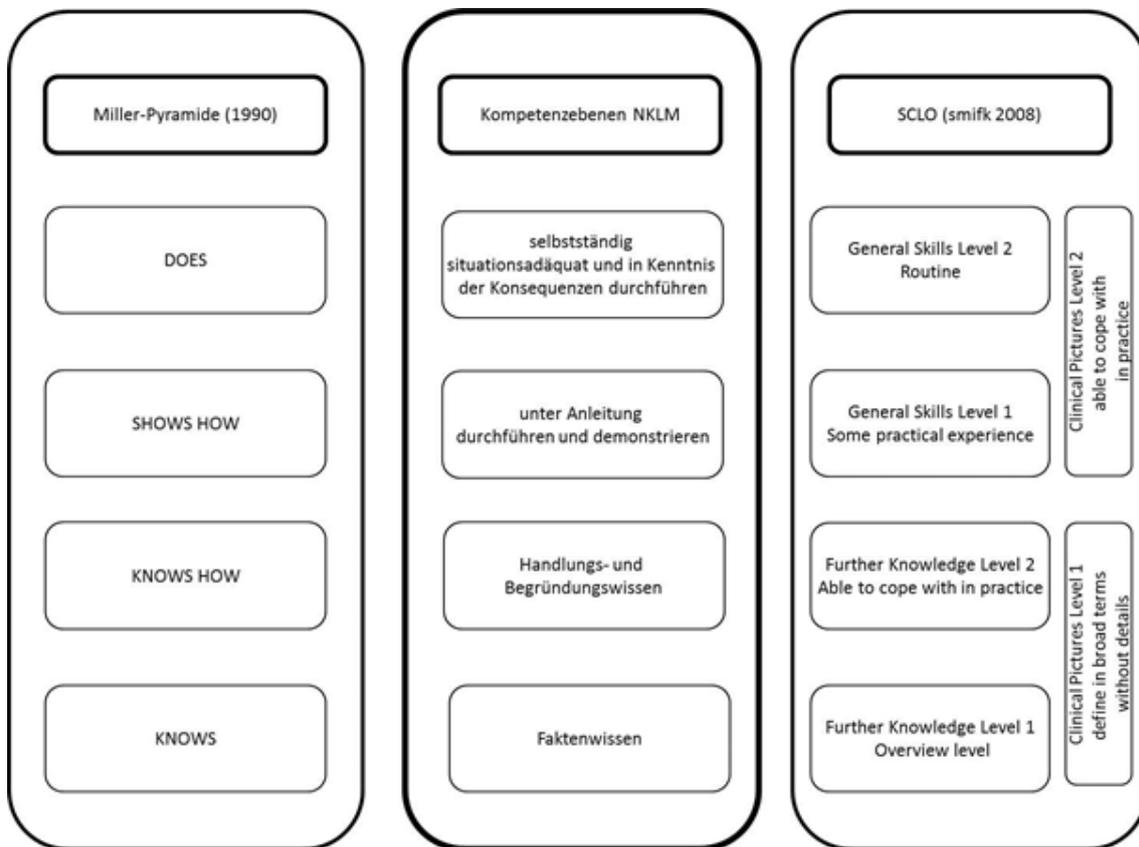


Abbildung 1: Kompetenzebenen von NKLM im Vergleich zu Miller-Pyramide und Swiss catalogue of learning Objectives for Undergraduate Medical Training (SCLO) (5)

2.2 Prüfungsformen in der medizinischen Lehre

2.2.1 Allgemeines

Grundsätzlich soll mithilfe einer Prüfung die Leistungsfähigkeit eines Prüflings gemessen werden (8). Bezogen auf die Facharztweiterbildung, soll eine abschließende Prüfung die sichere und hochwertige Patientenversorgung durch die Absolventen der Weiterbildung sicherstellen (9) (10).

Bevor allerdings die Leistung eines Prüflings bewertet werden kann, müssen Qualitätskriterien an die Prüfung selbst gestellt werden. Um die Qualität einer Prüfung zu messen, werden in der medizinischen Lehre häufig die allgemeinen Testgütekriterien der psychologischen Testtheorie herangezogen. Von besonderer Relevanz sind dabei die Hauptgütekriterien der Objektivität, Reliabilität und Validität (11) (12).

Verschiedene Prüfungsformen sind notwendig, da einige besser spezifische Kompetenzen prüfen können als andere (1).

Welche Vor- und Nachteile einzelne Prüfungsformen besitzen, ist für die bestmögliche Objektivierung von kompetenzbasierten Leistungsmessungen von Bedeutung.

Die Objektivität gibt an, inwieweit eine (Leistungs)-Messung vom Prüfenden und den externen Prüfungsbedingungen unabhängig ist. In Multiple-Choice-Prüfungen, bei denen für Prüflinge ähnliche Prüfungsbedingungen herrschen, ist die Objektivität beispielsweise leichter zu gewährleisten, als bei Prüfungen in denen qualitative Informationen wie Kommunikationskompetenz oder Professionalität bewertet werden sollen (9) (1).

Die Reliabilität ist ein Maß dafür, wie zuverlässig eine Prüfung (Leistung/Kompetenzen) misst. Ist die Reliabilität einer Prüfung hoch, so werden bei einer Wiederholung der Messung die Resultate gleich oder sehr ähnlich sein (11). Theoretisch kann die Reliabilität Werte zwischen 0 und 1 annehmen. Eine Reliabilität von 0 würde dabei implizieren, dass die Prüfungsergebnisse vollkommen zufällig sind. Eine Reliabilität von 1 würde bedeuten, dass ein unterschiedliches Resultat gänzlich auf die unterschiedliche Kompetenz (Leistung) zurückzuführen ist (1). Für die Berechnung der Reliabilität wird häufig das Cronbachs α verwendet. Bei bestehens- und notenrelevanten Prüfungen, sollte eine Reliabilität von mindestens $\alpha \geq 0,8$ erreicht werden (8).

Die Validität gibt den Grad der Genauigkeit an, mit dem die Prüfung misst, was bei der Erstellung der Prüfung beabsichtigt war zu messen (13) (12).

Für bestimmte Kompetenzen bieten sich verschiedene Prüfungsmethoden mehr als andere an. Um ein hohes Maß an Validität zu gewährleisten, sollte die Wahl der Prüfungsmethode sich hauptsächlich nach den zu prüfenden Kompetenzen richten (1).

2.2.2 *Verschiedene Prüfungsformen*

2.2.2.1 Die Multiple-Choice-Prüfung

Bei der als Multiple-Choice- oder Wahlantwortverfahren bezeichneten Prüfungsmethode werden verschiedene standardisierte schriftliche Aufgaben („Items“) von den Kandidaten bearbeitet. Hierbei müssen aus mehreren, vorgegebenen Antworten eine oder mehrere ausgewählt werden (1) (14).

Da alle Kandidaten die gleiche Aufgabenstellung erhalten, kann eine hohe Durchführungsobjektivität gewährleistet werden. Generell minimiert der hohe

Standardisierungsgrad subjektive Einflüsse (9). Zudem ermöglicht die hohe Fragenzahl pro Zeiteinheit eine hohe Quantität von Inhaltsstichproben. Dies reduziert den Einfluss von Störfaktoren, wie Raten oder spezifischen fragebezogenen Problematiken (1). Prüfungsökonomisch ist durch die computergestützte Auswertung eine Korrektur vieler Prüfungen in wenig Zeit und unter relativ geringen Kosten möglich (9).

Nachteile der Prüfungsmethodik sind unter anderem der eingeschränkte Evaluationsbereich und die Verleitung, einfaches Faktenwissen zu prüfen. Fraglich ist ebenfalls, ob sich das Verwenden von MC-Prüfungen negativ auf das Lernverhalten der Kandidaten auswirkt, da diese sich der Prüfung eher mit einem oberflächlichen Lernansatz nähern könnten (15) (16).

Dennoch hat die MC-Prüfung bei der Überprüfung von Fakten- und (teilweise) Anwendungswissen eine feste Rolle in der medizinischen Lehre.

2.2.2.2 Das OSCE

Bei der Prüfungsmethodik des OSCE (objektives strukturiertes klinisches Examen) werden anhand eines Blueprints (Blaupause) verschiedene Stationen entwickelt. Diese Stationen sollen jeweils unterschiedliche Kompetenzkomponenten, vor allem im Bereich der klinisch-praktischen Kompetenzen, prüfen (14). Hierbei kommen unter anderem auch Schauspielpatienten zum Einsatz und ermöglichen das Abfragen von Prüfungsinhalten wie Anamnesegesprächen oder der Demonstration von verschiedenen klinischen Untersuchungstechniken. Die Bewertung der Leistung erfolgt mithilfe vorher entwickelter Checklisten (9) (17).

Durch das Prüfen von Fertigkeiten und Handlungskompetenzen existiert mit dem OSCE eine wichtige Ergänzung zu reinen Wissensprüfungen. Außerdem erlaubt das OSCE den Prüfern nicht nur Einblicke in die Handlungsergebnisse der Kandidaten, sondern auch in deren Ablauf. Zudem kann das OSCE durch den praktisch orientierten und von Prüfern supervisierten Stationsaufbau, eine hohe lernsteuernde Wirkung entwickeln (9) (18).

Nachteile des OSCE sind die zeit- und personalaufwändige Entwicklung und die hohen dabei entstehenden Kosten. Da die Stationen kreisförmig aufgebaut sind, kann zudem ein einzelner Störfaktor den Arbeitsfluss aller Stationen beeinträchtigen (1)

Wegen seines starken klinisch-praktischen Bezugs wird das OSCE trotz hohen Personalaufwands und Kosten in der medizinischen Ausbildung in Deutschland sowie international, beispielsweise im Rahmen einiger Facharztprüfungen in der Schweiz, angewandt (19).

2.2.2.3 Die mündliche Prüfung

Das mündliche Prüfungsformat wird in der medizinischen Ausbildung sowohl während des Studiums als auch in Facharztprüfungen verwendet (9). Die Prüfung kann in unterschiedlicher Ausprägung in Form einer Befragung, Diskussion, einer Fallvorstellung oder einer Mischung dieser stattfinden (1). Dieses Format findet vorwiegend Anwendung, um zu prüfen, wie Kandidaten besonders ihr Verständnis- und Begründungswissen anwenden (9).

Vorteile dieser Prüfungsform sind der breite Evaluationsbereich und die hohe Flexibilität im Prüfungsfluss. Ebenso kann die sprachliche Ausdrucks- und Kommunikationsfähigkeit des Kandidaten miterfasst werden (1).

Nachteile der Prüfungsmethode sind die zeitliche Begrenzung der Prüfung, die mit einer geringen Stichprobe des geforderten Wissensspektrums einhergeht (9). Die Auswirkungen persönlicher Gefühle wie Sympathie oder Antipathie auf Seiten der Prüfer und beispielsweise Prüfungsangst auf Seiten der Absolventen, stellen ebenfalls einen Störfaktor dar, der größer als in anderen standardisierten Prüfungsmethoden zum Tragen kommt (1).

Die unter Verwendung von Protokollformularen und Checklisten durchgeführte strukturierte mündliche Prüfung (SMP) ist im Vergleich zur unstrukturierten mündlichen Prüfung vorzuziehen, da diese Maßnahmen die Prüfungsqualität im Sinne der Testgütekriterien verbessern (1) (20).

2.3 Hintergrund der Dissertation

Der Erwerb von Kompetenzen in der Medizin und die daraus folgende Herausforderung, diese in Form von Prüfungen bestmöglich messbar zu machen, ist ein fundamentaler Bestandteil der medizinischen Lehre (9).

Facharztprüfungen, die im besonderen Maße zum einen für den Absolventen, zum anderen für den Berufsstand des Arztes und der damit einhergehenden Rolle in der Gesellschaft (im Sinne der ärztlichen Standesorganisationen) relevant sind, sollten dabei von besonders hoher Testqualität sein.

Die Facharztanerkennung in Deutschland gibt in der Weiterbildungsordnung einen Teil abzuleistender Tätigkeiten im klinischen Arbeitsumfeld und nach erfolgreichem Ableisten dieser, eine Prüfung vor (2). Erst nach bestehen aller Erfordernisse gilt die Facharztkompetenz als bewiesen und berechtigt den Absolventen zum Tragen des

Facharztstitels. Die erforderliche Prüfung wird durch zwei Prüfer und einen protokollierenden Vorsitzenden geleitet. Methodisch wird die Prüfung in einem Einzelgespräch in Form einer mündlichen Prüfung durchgeführt. Hierbei können alle vorgeschriebenen Weiterbildungsinhalte geprüft werden (2). Der Prüfling erfährt am Ende der Prüfung über Bestehen oder Nichtbestehen. Eine Skalierung der erbrachten Leistung in Form einer Note wird nicht angegeben. Die schriftliche oder praktische Prüfung der geforderten Kompetenzen findet nicht statt.

In anderen Ländern wird die Überprüfung von erforderlichen Kompetenzen zum Erreichen des Facharztstitels unterschiedlich gehandhabt. In der Schweiz besteht die Prüfung beispielsweise aus einer schriftlichen Prüfung mit Multiple Choice (MC), Kurzantwort-Fragen (KAF), und/oder weiteren Fragetypen (21). Hierbei gelten für die verschiedenen Facharzttrichtungen jeweils andere Prüfungsmodalitäten mit abweichenden Prüfungsmethoden (19).

In den USA setzt die Verleihung des Facharztstitels bzw. die Zertifizierung durch das Medical Board eines Fachgebiets eine staatliche Lizenz zur freien Berufsausübung voraus (22). In Form einer computerunterstützten Multiple-Choice-Prüfung müssen in vier Blöcken jeweils 60 „single answer“ Fragen beantwortet werden. Für die internistische Facharztprüfung gibt das American Board of Internal Medicine einen Blueprint vor, der die Aufteilung der Fragen nach Fachgebiet unter Vorbehalt geringer Änderungen vorgibt (23). Ein weiterer Unterschied zur deutschen Facharztprüfung ist die „maintenance of certification“ (MOC). Dabei werden klar definierte Kompetenzen und Fähigkeitsgebiete geprüft. Diese Prüfung müssen die bereits zertifizierten Fachärzte in einem zeitlich festgelegten Rahmen von 10 Jahren kontinuierlich wiederholen (24).

Eine Neuerung in Bezug auf Facharztprüfungen ist die Etablierung länderübergreifender Fachärzte wie beispielsweise dem europäischen Facharzt für Notfallmedizin (25). Hier wird die Facharztprüfung mittels einer MC-Prüfung und zusätzlichen OSCE-Stationen durchgeführt. Die Durchfallquoten sind dabei mit bis zu 64% im mündlich-praktischen Teil vergleichsweise hoch (26).

In Bezug auf die deutsche Facharztprüfung werden beispielsweise in Bayern die Durchfallquoten regelmäßig veröffentlicht (27). Von den in den Jahren 2014/2015 abgelegten Facharztprüfungen im Bereich der Inneren Medizin bestanden dabei nur 3 Prüflinge (1,3%) nicht (28).

Zusammenfassend ist in den letzten Jahrzehnten eine Entwicklung medizinischer Prüfungen weg von vorherrschenden schriftlichen Aufsätzen und mündlichen Prüfungen, hin zu standardisierten Tests (OSCE, MC-Prüfung) zu beobachten (29). Diese Bemühungen dienen vor allem dem Ziel der Verbesserung der Testgütekriterien (30).

Zudem erhält die kompetenzorientierte Ausbildung in der medizinischen Lehre einen immer größeren Stellenwert (31) .

2.4 Relevanz

Eine tiefgreifende Untersuchung der internistischen Facharztprüfung ist aus mehreren Gründen sehr relevant:

Vor dem Hintergrund der Bedeutung der Facharztprüfung für den Berufsstand des Arztes, der Gesellschaft und des individuellen Absolventen, sind in Deutschland nur sehr wenige Informationen über die Prüfung zugänglich

Die Weiterbildungsordnungen der Bundesärztekammern geben zwar an, welche Kompetenzen in den jeweiligen Fachbereichen erworben werden sollen, stellen aber keine Auskunft über quantitative Verhältnisse der Fragen und Fallvorstellungen oder das Kompetenzniveau beispielsweise im Rahmen eines Blueprints vor (2).

Da in Deutschland die medizinische Aus- und Weiterbildung auf dem Prinzip Kompetenzbasierter Curricula aufgebaut ist, liegt der Fokus nicht auf der Abarbeitung von Lernzielen, sondern auf dem Outcome, also dem Erreichen oder nicht Erreichen von Kompetenz (2). Diese Arbeit ist von Bedeutung, weil sie den Fokus der Kompetenzbasierten Curricula, dem Assessment der erreichten Kompetenz im Sinne einer Prüfung, untersuchen soll.

Bisher gibt es keine offiziellen, konkreten Vorschriften über den Inhalt und die didaktische Methodik der Prüfungen. Zudem ist der Bedarf an Strukturvorgaben und/oder der Prüferschulung nicht bekannt, da keine Informationen darüber zugänglich sind, was in den Prüfungen wirklich stattfindet.

Um diese Wissenslücke zu füllen, ist diese Arbeit von Relevanz. Absolventen können so weitergehende Informationen über den Prüfungshergang erhalten und Prüfern wird die Möglichkeit gegeben, die Methodik ihrer eigenen Prüfungen zu reflektieren.

3 Fragestellung

3.1 Die Forschungsfrage

Vor dem beschriebenen theoretischen Hintergrund ergibt sich die folgende Forschungsfrage:

Welche inhaltlichen und strukturellen Charakteristika haben Facharztprüfungen in Deutschland am Beispiel der Inneren Medizin?

3.2 Hypothesen

1. Die Weiterbildungsinhalte werden durch Abfrage ausreichend vieler Wissensstichproben aus unterschiedlichen Schwerpunktfächern abgedeckt.
2. Es wird vor allem Verständnis und kritisches Reflektieren klinischer Zusammenhänge abgefragt.
3. Das Abgefragte Spektrum inhaltlicher Themengebiete ist in seiner Verteilung dem amerikanischen „Blueprint“ ähnlich.
4. Von der WBO geforderte Untersuchungs- und Behandlungsverfahren werden in Form verschiedener Medien geprüft.

4 Material und Methoden

4.1 Zugrunde liegendes Material

Basis der Studie waren die von den Prüflingen angefertigten Gedächtnisprotokolle aus der Sammlung des „Intensivkurs Innere Medizin – Vorbereitung auf die Facharztprüfung und Refreshkurs“ (32). Der Kurs wird seit 1994 unter der Schirmherrschaft und Qualitätskontrolle der deutschen Gesellschaft für Innere Medizin (DEGIM) abgehalten. Ebenfalls wird der Kurs in Zusammenarbeit mit der Bayrischen Landesärztekammer (BLÄK) durchgeführt. Herausgeber der Sammlung ist das Klinikum der Ludwig-Maximilians-Universität München (KUM) und die Münchner Akademie für Ärztliche Fortbildung e.V. (MAÄF e.V.). Die digitale Version beinhaltete insgesamt 196 Prüfungsprotokolle aus den Jahren 2013 bis 2017. Ab einer Stichprobengröße von 50-60 kann von einem Sättigungseffekt ausgegangen werden (33).

Aus den 196 vorhandenen Protokollen wurden daher für die Jahre 2013 bis 2016 jeweils 25 Protokolle (insgesamt 100) zufällig ausgewählt und analysiert.

Die Prüflinge wurden angehalten, Rahmeninformationen wie das Datum der Prüfung, den Wohnort, den Ort der Prüfung, die Spezialisierung der Prüfer und Prüfung nach alter oder neuer Weiterbildungsordnung, anzugeben. Diese Informationen befanden sich immer in der Kopfzeile des jeweiligen Protokolls.

Den Inhalt der Prüfung sollten die Prüflinge so ausführlich wie möglich und auch unter nennen von eigenen Falschantworten niederschreiben. Die Teilnehmer wurden von den Kursleitern durch eine interne Preisausschreibung für vollständige Protokolle dazu motiviert, qualitativ hochwertige Niederschriften abzuliefern,

4.2 Datenmanagement und Auswertung

4.2.1 Prinzipien der strukturierten Inhaltsanalyse nach Mayring

Um für die Beantwortung einer Fragestellung relevante Informationen aus einem Datensatz zu gewinnen, ist es notwendig, die enthaltenen Daten zu reduzieren und zu ordnen (34). Mit der strukturierten Inhaltsanalyse nach Mayring wird es durch die Verwendung eines angepassten Kategoriensystems möglich, auch größere Datenmengen zu bearbeiten (35). Durch diese Methode können einzelne Textpassagen verschiedenen Übergruppen zugeordnet werden. Dies ermöglicht ein tiefergreifendes

Verständnis über die in den vorliegenden Daten befindlichen relevanten Informationen (35).

Die Entwicklung des Kategoriensystems kann induktiv, vom Datenmaterial ausgehend oder deduktiv auf Grundlage bereits vorhandener Informationen vorgenommen werden. Wurden einzelne Kategorien definiert, können passende Textpassagen diesen zugeordnet werden. Im Austausch mit anderen Kodierenden und unter kritischer Betrachtung bereits kodierter Segmente, findet eine Reevaluation des vorläufigen Kategoriensystems statt. Der Prozess der Kategorienbildung kann als abgeschlossen betrachtet werden, wenn alle für die Fragestellung relevanten Textpassagen einem oder mehreren Codes im Kategoriensystem zugeordnet werden konnten.

Nach einer Modifikation oder abschließenden Reevaluation des Kodier Systems unter sich wiederholenden Qualitätskontrollen, kann schließlich die Hauptkodierung erfolgen.

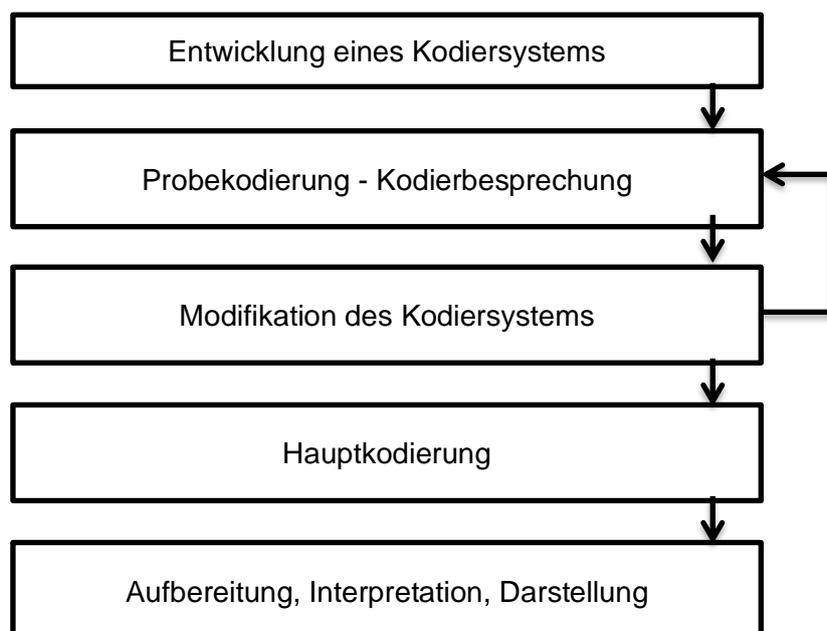


Abbildung 2: Arbeitsschritte der Analyse nach Mayring

4.2.2 Anforderungen an das Kategoriensystem

Das Kategoriensystem, welches in dieser Arbeit Anwendung findet, sollte im Wesentlichen in der Lage sein, die Aufteilung der inhaltlichen Themengebiete im Sinne der vorher definierten Fragestellung (Hypothese 1, Hypothese 3), in die einzelnen Fachgebiete einzuordnen. Außerdem sollte eine quantitative Einteilung der abgefragten Untersuchungs- und Behandlungsverfahren wie beispielsweise der Elektrokardiographie

oder Laborbefunden, ermöglicht werden (weitere Hypothese 4). Zur Kategorisierung dieser Informationen bedurfte es einer Einteilung der jeweiligen Textpassage in eine Gruppe (z.B. Kardiologie, Gastroenterologie etc.). Diese Gruppen konnten weiter zu einer Einteilung in Kode-Familien zusammengefasst werden (Fallvorstellung, Einzelfrage, Medien).

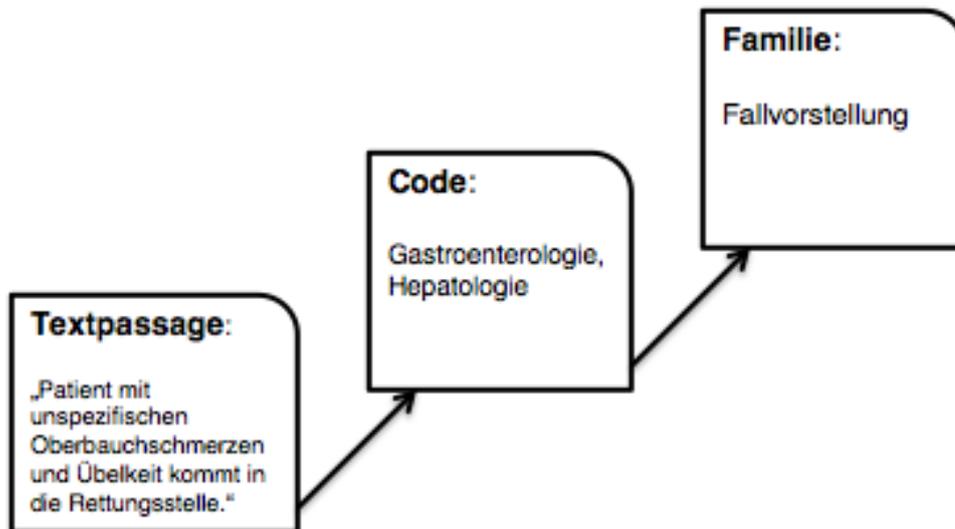


Abbildung 3: Beispiel der Kodierschematik

Die Fragestellung, welches Kompetenzniveau jede Einzelfrage aus dem jeweiligen Fachgebiet (im Sinne der Hypothese 2) besitzt, konnte auf diese Weise nicht mit einem einzigen Kode und einer übergeordneten Kode-Familie kodiert werden.

Zur eindeutigen Beantwortung dieser Fragestellung wurde bei der Literaturrecherche die Anwendung des ULME-Kompetenzstufenmodells als Möglichkeit gesehen, Einzelfragen in jeweils eine vom Modell klar definierte Leistungskategorie und eine Wissensdimension einzuordnen (36).

4.2.3 Das ULME-Kompetenzstufenmodell

Das ULME-Kompetenzstufenmodell ist eine Klassifikationsmatrix, die der Analyse des Anspruchsniveaus von Testaufgaben dient. Sie beruht überwiegend auf der von Anderson und Kathwohl (2001) revidierten und erweiterten Bloom'schen Taxonomie (37).

Zur Sicherung der curricularen Validität müssen in Prüfungen die vorher definierten Wissensbereiche und das angeforderte Leistungsniveau bei der Erstellung von Fragen und Testaufgaben berücksichtigt werden. Ursprünglich dient die Matrix bei der Erstellung von Prüfungsfragen als Hilfsmittel, um das in den Aufgaben erwünschte Leistungsniveau bezogen auf die entsprechenden Inhaltsbereiche festzustellen.

Tabelle 1: Die Kompetenzmatrix nach dem ULME-Modell

Leistungskategorie	Reproduzieren	Anwenden/Verstehe n	Kritisieren/Reflektiere n
Wissensdimension			
Fakten	<i>R+F</i>	<i>A/V+F</i>	<i>K/R+F</i>
Konzepte	<i>R+K</i>	<i>A/V+K</i>	<i>K/R+K</i>
Prozeduren	<i>R+P</i>	<i>A/V+P</i>	<i>K/R+P</i>

In dieser Arbeit soll die Klassifikationsmatrix ebenfalls zur Feststellung von Wissensdimensionen und Leistungskategorien einzelner Fragen verwendet werden. Allerdings soll nicht von den Kategorien, welche die Matrix bietet, eine Prüfungsfrage erstellt werden, sondern eine bereits bestehende Kurzfrage aus dem vorliegenden Material (aus den Gedächtnisprotokollen der mündlichen Facharztprüfung) auf die entsprechende Verknüpfung von Wissensdimension und Leistungskategorie rückbezogen werden.

Die einzelnen Untergruppen der Leistungskategorien und Wissensdimensionen werden nach dem ULME-Kompetenzstufenmodell in einem festen Rahmen definiert (36). Bei der Kodierung passender Textpassagen aus den Gedächtnisprotokollen und der subsequenten Einteilung dieser in die Matrix, wurden die Einteilungskriterien nach der Definition des ULME-Kompetenzstufenmodells übernommen.

Tabelle 2: Charakterisierung der Wissenskategorie nach dem ULME-Modell (36)

Wissenskategorie	Charakterisierung
Fakten	-Aussagen über spezifische Bezeichnungen -Definitionen im Sinne von Worterklärungen

	-Aussagen über Normen, sowie über die Existenz, Quantität, Qualität und Zusammensetzung von Phänomenen
Konzepte	-Sacherklärungen (Bildung von Oberbegriffen) -Begriffshierarchien ermöglichen Kategorien übergreifende Schlussfolgerungen -Aussagen zu Prozessen und Zusammenhängen zwischen empirischen Phänomenen -Aussagen zu Ursache-Wirkung-Beziehungen
Prozeduren	-Zur Bewältigung berufstypischer Handlungssituationen -Lösungswege und Handlungsergebnisse können bekannt sein oder müssen erst gefunden werden

Tabelle 3: Charakterisierung der Wissenskategorie nach dem ULME-Modell (36)

Leistungskategorie	Charakterisierung
Reproduzieren	-Wiedergeben/Erkennen von erlerntem Wissen -keine Situative Einbindung erforderlich -typisch: nennen, wiedergeben, beschreiben
Verstehen und Anwenden	-erlerntes Schema auf neue Situation anwenden -erlerntes Schema auf strukturell gleiche aber sprachlich anders gefasste Aufgabe beziehen -typisch: erklären, ermitteln, begründen, beurteilen, vergleichen
Kritisieren und Reflektieren	-Grenzen von Schemata bewusst reflektieren -Anforderungen einer Veränderung der Struktur sind zu identifizieren -typisch: Lösung beurteilen, Probleme aufzeigen, Leistungsfähigkeit abwägen

4.2.4 Der Kodierprozess

Nachdem die grundlegenden Anforderungen an das Kategoriensystem definiert wurden, fand eine erste Bearbeitung des Materials und eine initiale Kategorienbildung statt. Eine Differenzierung der aus dem Datenmaterial zu extrahierenden Informationen wurde entwickelt.

Folgende „Fragescenarien“ können unterschieden werden: Fallvorstellungen, Kurzfragen und Medienfragen.

Hierbei besitzen Fallvorstellungen die Eigenschaft, eine klinische Situation zu beschreiben, enthalten selbst aber noch keine konkrete Fragestellung. Sie werden ausschließlich in inhaltliche Kategorien eingeteilt.

Kurzfragen können nach einer konkreten Fallvorstellung oder ungeordnet aufeinander folgen. Die Kurzfragen der Prüfer wurden zuerst in inhaltliche Kategorien eingeteilt, werden mit Hilfe der ULME-Kompetenzmatrix aber auch nach Leistungskategorie und Wissensdimension kategorisiert. So besitzt eine Einzelfrage immer mindestens eine Kodierung für die inhaltliche Kategorie und obligat jeweils eine Kodierung für die Wissensdimension und Leistungskategorie.

Wenn Fallvorstellungen und Kurzfragen zum Teil Komponenten verschiedener inhaltlicher Kategorien besaßen (z.B. Kardiologie *und* Notfallmedizin), wurden diese für beide Kategorien kodiert.

Untersuchungs- und Behandlungsfragen, die mit Hilfe von Fotos, EKG-Streifen, Projektoren und ähnlichem gestellt wurden, sind unter der Familie „Medien“ weiter in einzelne Untersuchungstechniken kategorisiert worden. Hierbei wurde jede voneinander unabhängige Darstellung (auch der gleichen Untersuchungstechnik) individuell kodiert.

Ebenfalls wurden Rahmeninformationen wie der Ort der Facharztprüfung oder die Spezialisierung der Prüfer in eigene Kategorien unterteilt.

4.2.5 Qualitätssicherung des Kodierprozesses

Nach mehrmaliger Revision und Modifikation des Kategoriensystems erfolgte eine Intercoder Reliabilitäts-Prüfung. Hierbei wurden jeweils 3 randomisierte Fälle von 2013 bis 2016 (insgesamt 12) ausgewählt und von einem „externen Coder“ kodiert.

Bei 100 untersuchten Protokollen entspricht dies einer doppelten Codierung von 12 Prozent. Nach Krippendorff's alpha berechnet, betrug die die Intercoder-Reliabilität $\kappa = 0,89$.

4.3 Die Analysesoftware

Für die Auswertung des Materials wurde die Analysesoftware MAXQDA (Maxqda Standard 12, Version 12.3.2, Software für qualitative Datenanalyse, 1989 – 2018, VERBI Software. Consult. Sozialforschung GmbH, Berlin, Deutschland) verwendet (38).

Mit dem Programm wurde eine Dokumentengruppe eröffnet, die alle 100 Fälle aus den Jahren 2013 bis 2016 enthielt. Relevante Textpassagen konnten direkt im Dokument kodiert werden und wurden in das Kategoriensystem eingeordnet. Insgesamt wurden 10978 Kodierungen vorgenommen, die sich in die oben genannten definierten Kode-Gruppen und Familien aufteilten. Das Programm ermöglichte zudem die Erhebung der Interkoder-Reliabilität (Anhand der Berechnung durch Krippendorff's alpha).

Die deskriptive Statistik wurde mit SPSS Version 24 und Excel durchgeführt.

5 Ergebnisse

Insgesamt wurden 100 Protokolle analysiert. Diese unterteilen sich in jeweils 25 Protokolle für die Jahre 2013 bis 2016. Die Protokolle waren im Durchschnitt 720,35 Worte lang. Von den 100 zufällig ausgewählten Protokollen, musste keines von der Analyse ausgeschlossen werden.

5.1 Stichprobe

5.1.1 Aufteilung der Prüfungen/Protokolle nach Bundesland

Die Prüfungsprotokolle stammten mit 54 Prozent zum größten Teil aus Bayern. Drei Protokollanten gaben den Ort ihrer Prüfung nicht an. Insgesamt konnten Protokolle aus mindestens 11 verschiedenen Bundesländern ausgewertet werden.

Tabelle 4: Aufteilung des Prüfungsortes nach Häufigkeit

Bundesland	Häufigkeit	Prozent
Bayern	54	54%
NRW	16	16%
Baden-Württemberg	6	6%
Schleswig-Holstein	5	5%
Hamburg	3	3%
Berlin	3	3%
Rheinland-Pfalz	3	3%
Hessen	3	3%
Ort nicht angegeben	3	3%
Sachsen	2	2%
Thüringen	1	1%
Niedersachsen	1	1%
Gesamt	100	100%

5.1.2 Facharztbezeichnung

In den Prüfungen wurde in 79 Prozent der Fälle die Berufsbezeichnung: *“Facharzt für Innere Medizin“* ohne weitere Spezialisierung angegeben. Vier Prozent der Protokollanten gaben an, eine Facharzt-Spezialisierung angestrebt zu haben. In 17% der Protokolle wurde keine nähere Facharztbezeichnung angegeben.

Tabelle 5: Aufteilung der Facharzt-Spezialisierungen

Facharzt und Facharzt-Spezialisierung	Prozent
Facharzt für Innere Medizin	79%
Facharzt für Innere Medizin und Kardiologie	2%
Facharzt für Innere Medizin und Hämatookologie	1%
Facharzt für Innere Medizin und Pneumologie	1%
Keine nähere Bezeichnung angegeben	17%

5.1.3 Neue/Alte Weiterbildungsordnung

In den 100 Protokollen wurde von den Verfassern insgesamt 31 mal eine Weiterbildung nach neuer WBO angegeben. 3 Kandidaten gaben an, nach alter WBO geprüft worden zu sein. Die übrigen 66 Protokollanten gaben nicht an, ob sie nach alter oder neuer WBO geprüft wurden.

5.1.4 Spezialisierung der Prüfer

Die Facharzt-Spezialisierung der Prüfer wurde insgesamt 124 mal kodiert. Dabei waren die meisten Prüfer Gastroenterologen (32,3%) und Kardiologen (16,9%). In einigen Protokollen wurde die Spezialisierung der Prüfer nicht näher angegeben. Dies spiegelt sich in 20,9% der Kodierungen mit dem Namen „*Internist ohne Bezeichnung*“ wieder.

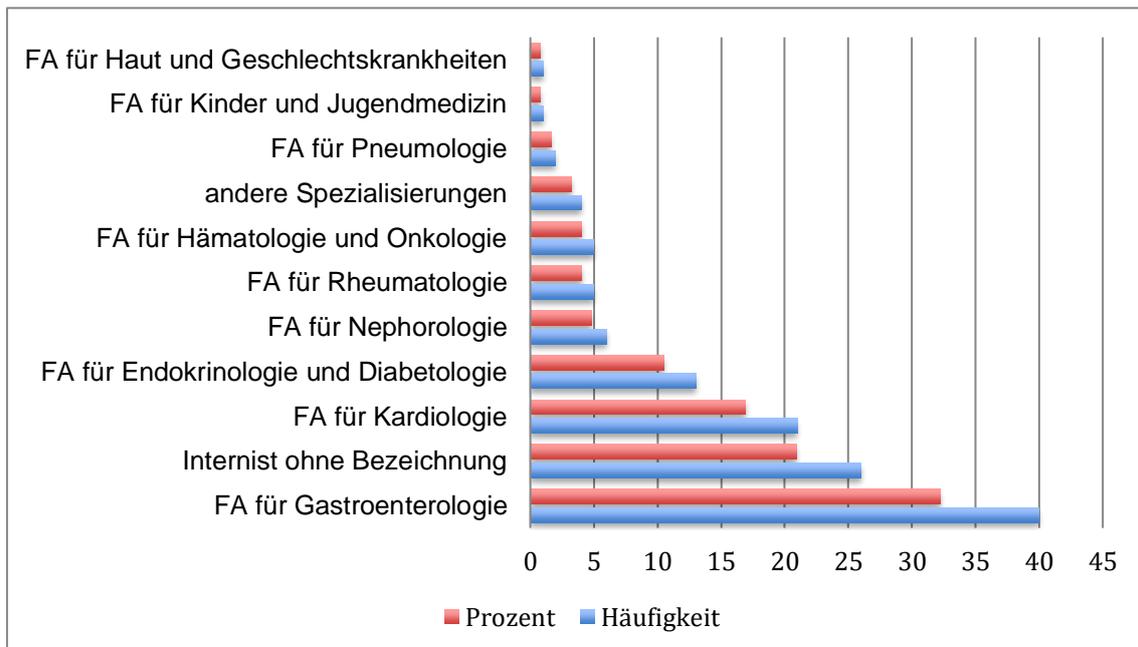


Abbildung 4: Spezialisierungen der Prüfer

5.2 Spezifische Varianz unterschiedlicher Themengebiete

In diesem Abschnitt sollen hauptsächlich die Ergebnisse der Haupthypothese 1 dargestellt werden. Zum einen wurde überprüft, wie viele Wissensstichproben im Sinne von Fragen pro Prüfung durchschnittlich abgefragt wurden. Zum anderen wurde analysiert, wie viele unterschiedliche Themengebiete von den Prüfungen durchschnittlich abgedeckt wurden. Die Anzahl der Fragen und Varianz der Themengebiete wurde so für jede Prüfung bestimmt und jeweils individuell in die Kategorien „Fallvorstellungen“, „Kurzfragen“ und „Medien“ eingeteilt.

5.2.1 Die Anzahl der gestellten Fragen pro Prüfung

Insgesamt wurden im Schnitt 27,20 (SD = 10,23) Fragen pro Prüfung gestellt. Die Minimalanzahl an Fragen pro Prüfling liegt bei 9, während das Maximum bei 56 ist. Von den drei verschiedenen Fragetypen wurden am häufigsten Einzelfragen gestellt (M = 19,73 (SD = 9,10)). Fälle wurden häufig als Einstieg in die Prüfung genutzt (M = 2,37 (SD = 1,20)). Es wurden durchschnittlich 5,10 (SD = 4,03) Medienfragen gestellt.

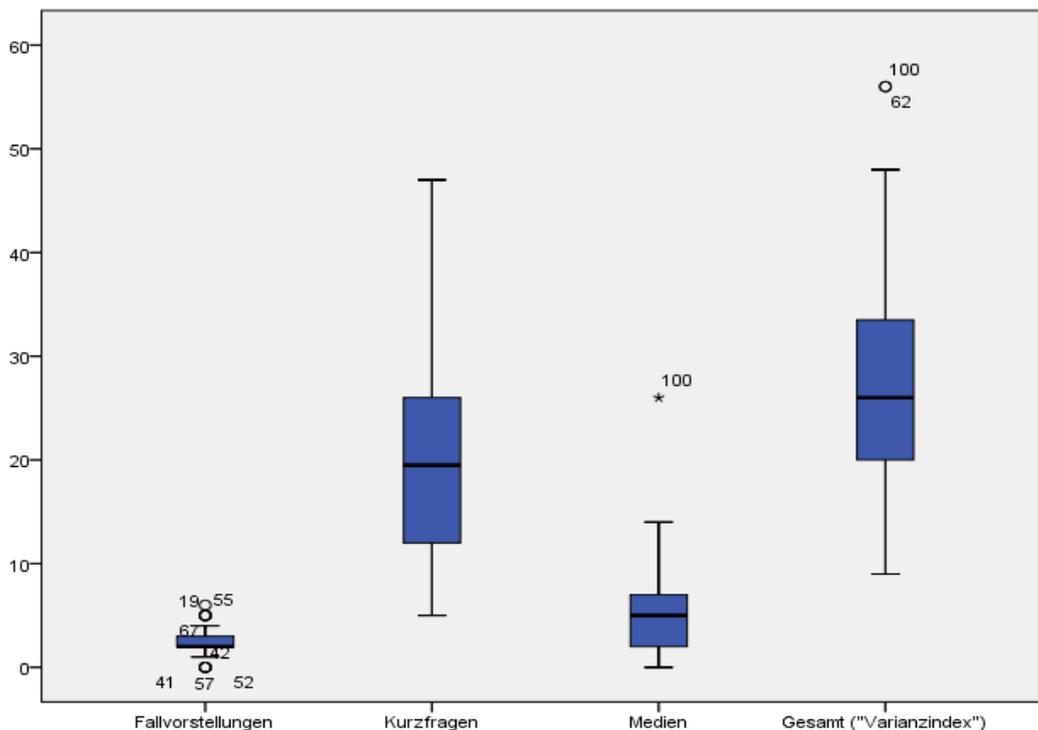


Abbildung 5: Häufigkeit der unterschiedlichen Themengebiete

5.2.2 Die Anzahl unterschiedlicher Themengebiete pro Prüfung

Die Anzahl unterschiedlicher Themengebiete pro Prüfung wurde für jede der drei Hauptkategorien „Fallvorstellungen“, „Kurzfragen“ und „Medien“ untersucht. Die Anzahl

der unterschiedlichen Themengebiete bei den „Fallvorstellungen“ lag hier bei einem Mittelwert von 2,76 bei einer Standardabweichung von 1,46. In der Kategorie „Kurzfragen“ lag der Mittelwert der unterschiedlichen Themengebiete bei einem Mittelwert von 5,80 bei einer Standardabweichung von 1,93. In der Kategorie „Medien“ wurde ein Mittelwert von 2,47 mit einer Standardabweichung von 1,40 gemessen.

Die Anzahl der unterschiedlichen Themengebiete bezogen auf die Gesamtheit der Hauptkategorien betrug im Mittel 11,03 mit einer Standardabweichung von 3,05.

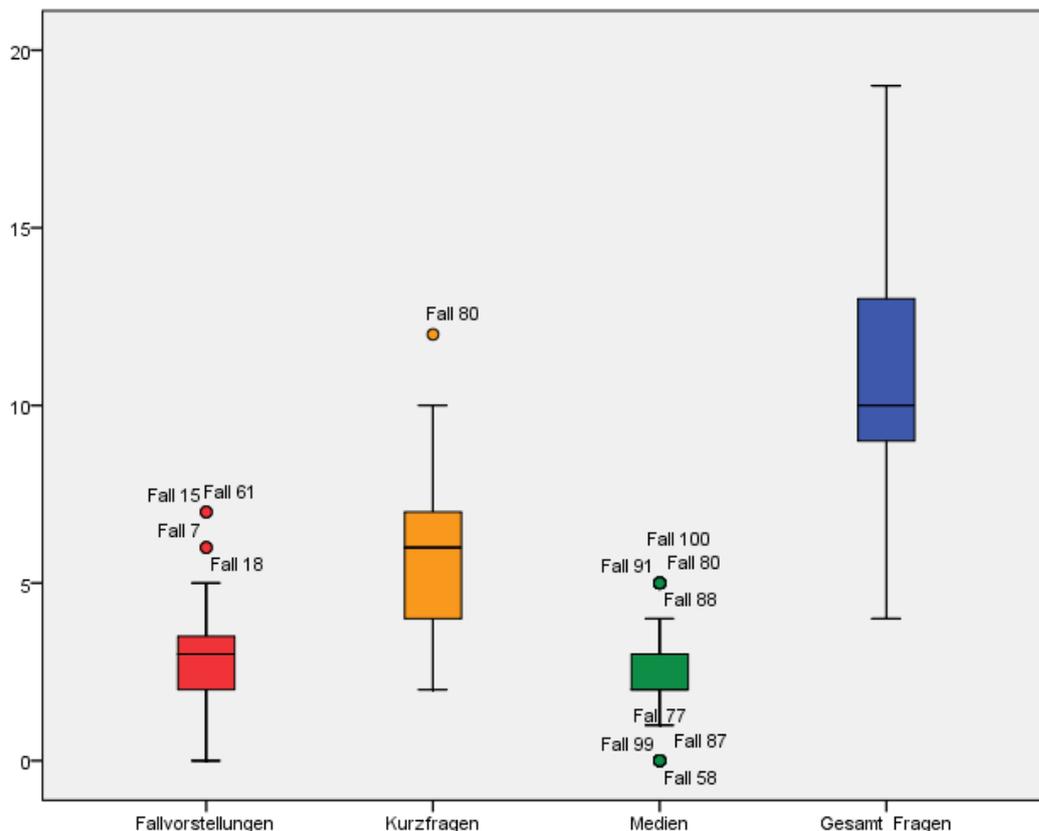


Abbildung 6: Die Varianz der Protokolle bezogen auf die unterschiedlichen Themengebiete in den Hauptkategorien: "Fallvorstellungen", "Kurzfragen" und "Medien"

5.3 Globale Aufteilung der Themengebiete

In diesem Abschnitt wurden die Ergebnisse in Bezug auf die Hypothese 1 dargestellt. Es erfolgte eine Einteilung der „Fallvorstellungen“ und „Kurzfragen“ in die unterschiedlichen Themengebiete.

Zudem wurden Fragen, bei denen sich Themengebiete bei der Kodierung überschneiden haben, dargestellt.

Die Ergebnisse der Aufteilung der Themengebiete in „Fallvorstellungen“ und „Kurzfragen“ wurden abschließend mit dem Blueprint des American Board of Internal Medicine sowie dem schweizerischen Blueprint verglichen.

5.3.1 Aufteilung der Fallvorstellungen nach Themengebiet

Insgesamt wurde in 320 Fällen eine Kodierung der Kategorie „Fallvorstellungen“ vorgenommen. Die Unterkategorien „Gastroenterologie, Hepatologie“ (17,8%) und „Kardiologie“ (17,2%) wurden hierbei am häufigsten kodiert. Die Unterkategorien „Intoxikationen, Nebenerscheinungen, Alkoholismus“, „Rheumatologie, Knochenkrankheiten“, „Neurologie, Psychiatrie, Psychosomatik“, „Geriatric“, „Intensivmedizin“ und „Ernährung“ bilden mit jeweils unter 2% der Kodierungen in dieser Kategorie die am wenigsten repräsentierten Kodierungen.

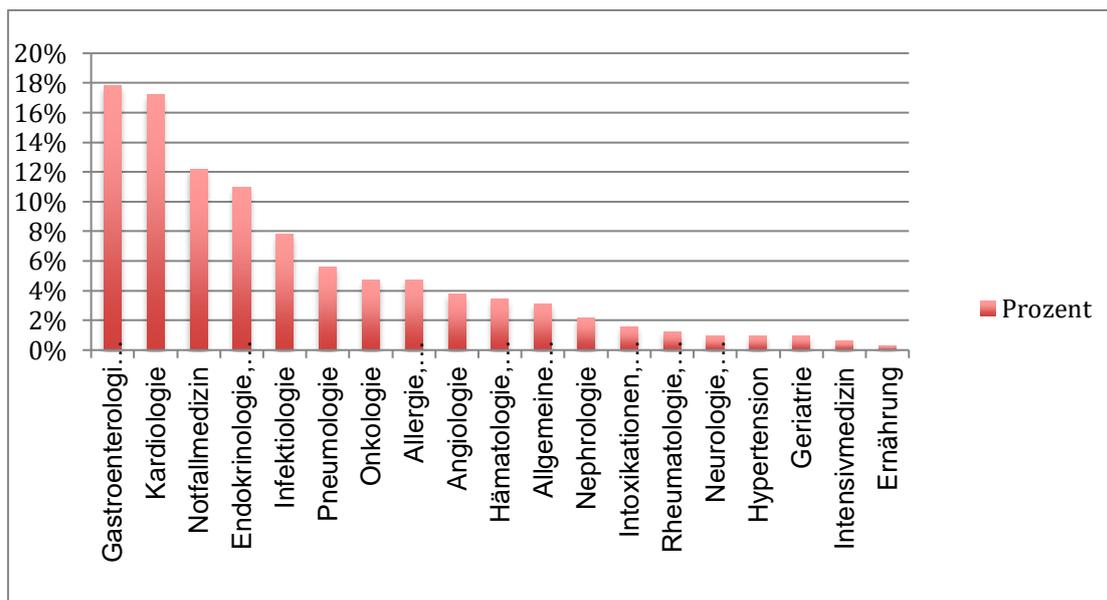


Abbildung 7: Aufteilung der Fallvorstellungen nach Themengebiet

5.3.2 Aufteilung der Einzelfragen nach Themengebiet

Insgesamt wurden 2659 Textpassagen in die einzelnen Unterkategorien der Kategorie „Einzelfragen“ eingeordnet. So wie in der Kategorie „Fallvorstellungen“, wurden die Unterkategorien „Gastroenterologie, Hepatologie“ (17,75%) und „Kardiologie“ (17,34%) auch hier am häufigsten kodiert. Während bei den „Fallvorstellungen“ das Themengebiet

„Notfallmedizin“ mit 12,18% am dritthäufigsten kodiert wurde, macht dieses Fach bei der Aufteilung der „Kurzfragen“ 4,89% aus.

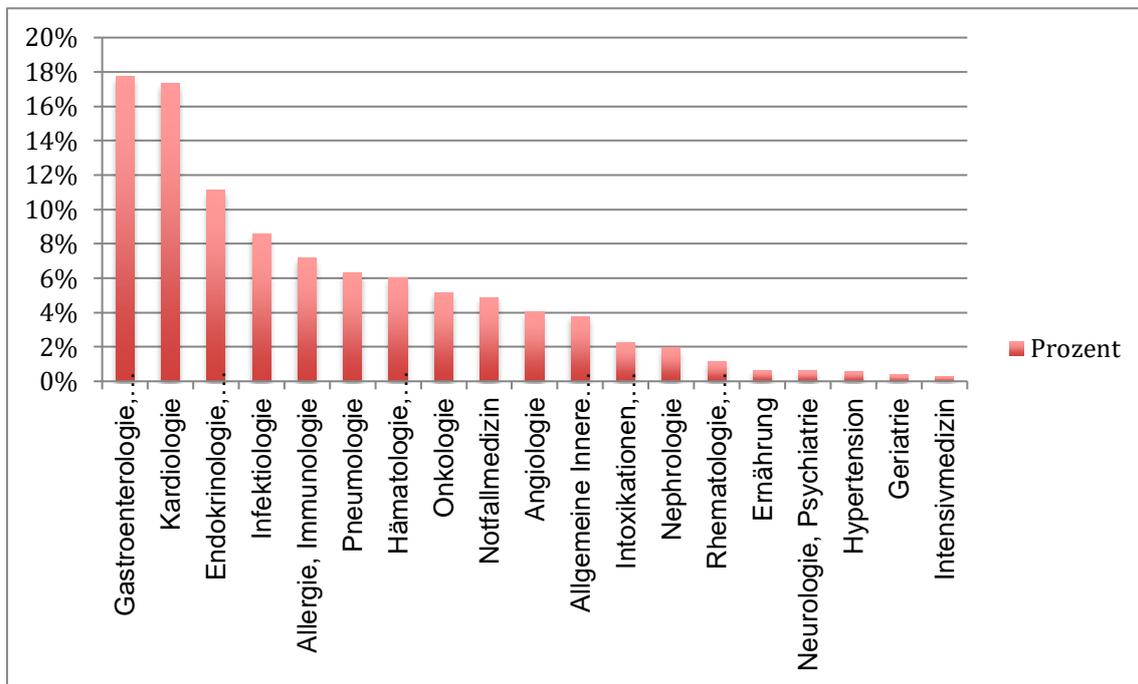


Abbildung 8: Aufteilung der Einzelfragen nach Themengebiet

5.3.3 Überschneidungen von Kodierungen und Themengebiet

Aus den 320 Textpassagen die mit dem Code „Fallvorstellungen“ kodiert wurden, waren in 184 Fällen Überschneidungen festzustellen. Hierbei sind die häufigsten Überschneidungen in den Unterkategorien „Notfallmedizin“ (30 Überschneidungen), „Kardiologie“ (29 Überschneidungen) und „Infektiologie“ (20) aufgetreten.

In der Abbildung 7 sind die häufigsten Überschneidungen dargestellt, ein vollständiges Diagramm befindet sich im Anhang dieser Arbeit.

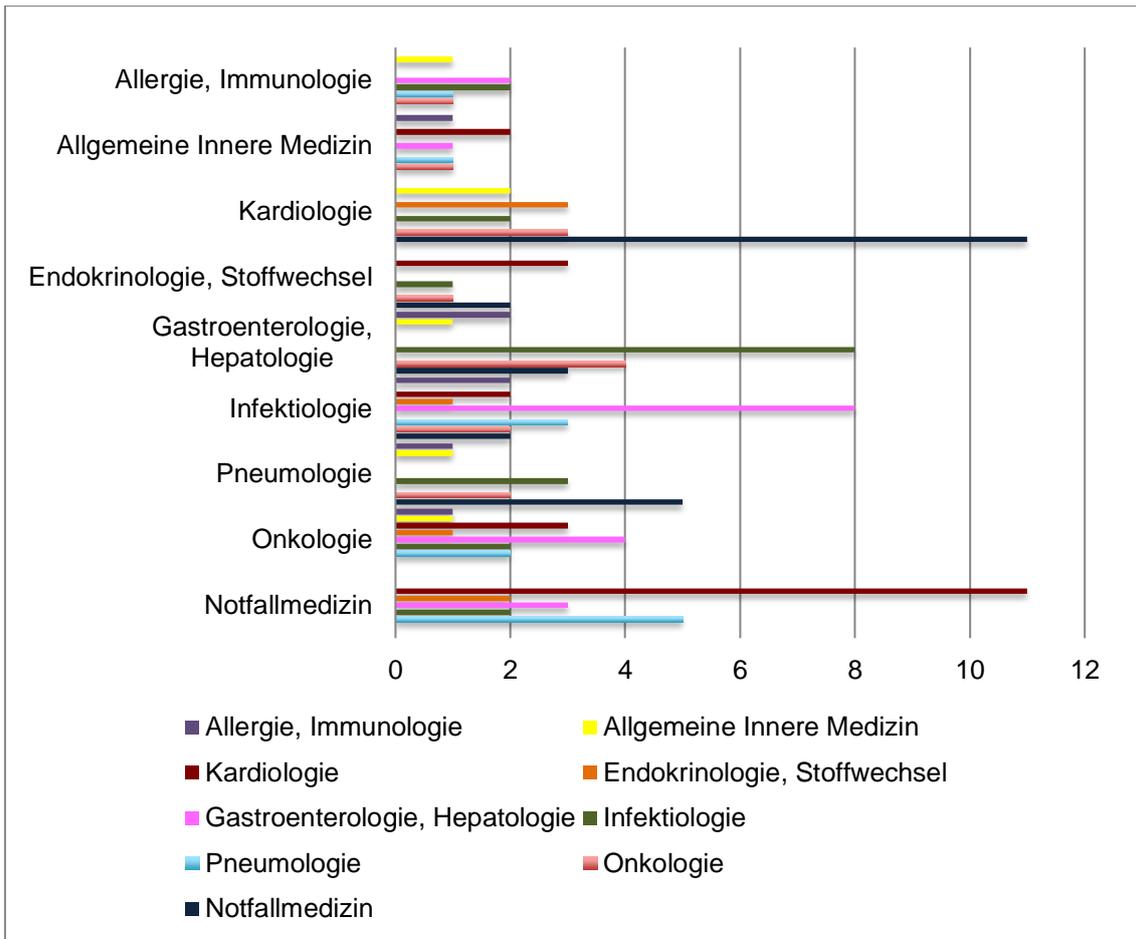


Abbildung 9: Häufige Überschneidungen der Kodierung bei Fallvorstellungen

Aus den 2659 Textpassagen die mit dem Code „Kurzfragen“ kodiert wurden, waren in 1263 Fällen Überschneidungen festzustellen. Hierbei sind die häufigsten Überschneidungen in den Unterkategorien „Infektiologie“ (167), „Gastroenterologie, Hepatologie“ (149) und „Allergie, Immunologie“ (145) aufgetreten.

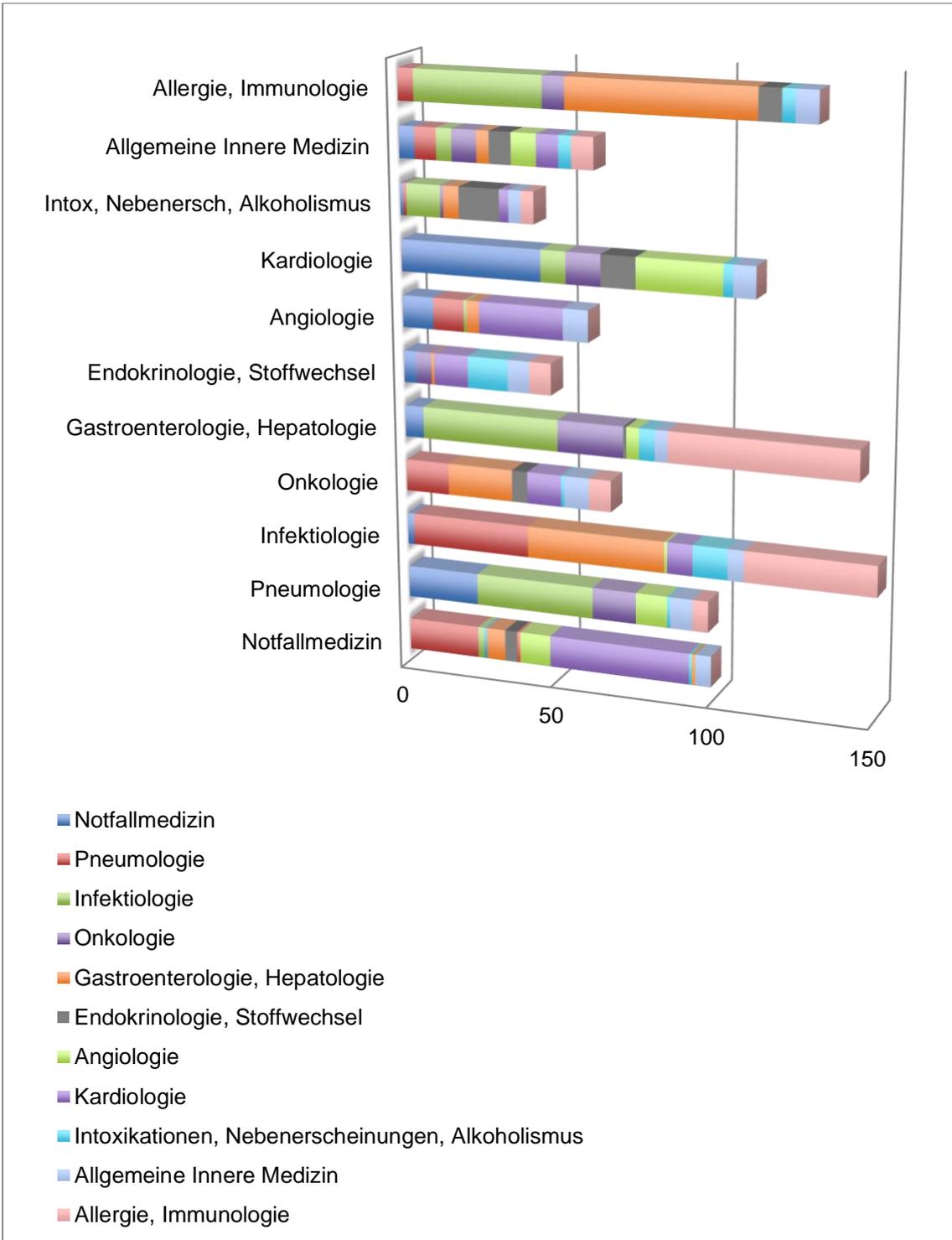


Abbildung 10: Häufige Überschneidungen der Kodierung von Kurzfragen

5.3.4 *Aufteilung der inhaltlichen Themengebiete im Vergleich zum „Blueprint“ des American Board of Internal Medicine und Guidance on the Swiss Board Examination in General Internal Medicine*

Die Definitionen der unterschiedlichen Themengebiete des Blueprints des American Board of Internal Medicine und der Einteilung, die in dieser Arbeit verwendet wurde, weisen Unterschiede auf. Ein Vergleich des „Rankings“ vieler Themengebiete ist aber trotzdem möglich. 5 der 6 Themengebiete, die im Blueprint des ABOIM den höchsten prozentualen Anteil der Themengebiete untereinander besitzen, finden sich ebenfalls in den 6 Themengebieten der Gedächtnisprotokolle, die den größten Anteil untereinander besitzen. Diese Themengebiete beinhalten: „Kardiologie“, „Gastroenterologie, Hepatologie“, „Endokrinologie, Stoffwechsel“, „Infektiologie“ und „Pneumologie“. So sind auch 7 von den anteilig größten 9 Themengebieten des ABOIM in den 9 anteilig größten Themengebieten der Protokolle vorhanden.

Ein Unterschied in den 6 häufigsten ABOIM Blueprint Themengebieten und den Themengebieten der Gedächtnisprotokolle, ist die prozentuale Verteilung der Kategorie „Rheumatology and Orthopedics“ mit 9% im Blueprint und 1,17% und 1,25% (respektiv „Kurzfragen“ und „Fallvorstellungen“) in den Gedächtnisprotokollen. In der prozentualen Verteilung der Themengebiete aus den Gedächtnisprotokollen befindet sich unter den häufigsten 6 Themengebieten mit 4,69% in der Kategorie „Fallvorstellungen“ und 7,18% in der Kategorie „Kurzfragen“, „Allergie, Immunologie“. Im ABOIM Blueprint wird das Themengebiet „Allergy and Immunology“ mit 2% angegeben. Die schweizerische Gesellschaft für allgemeine Innere Medizin hat ebenfalls einen Blueprint herausgegeben (7). Mit 14% ist das Themengebiet „General considerations incl. prevention, palliative medicine, rehabilitation, health insurance and economics, ethics, statistics and questions not allocated to a subspecialty“ am größten. Es folgt „Kardiologie“ mit 11% und „Rheumatologie“ mit 9%. Danach sind die Fächer „Endokrinologie“, „Nephrologie“, „Gastroenterologie“, „Hämatologie“, „Infektiologie“, und „Pulmonologie“ mit jeweils 8% vertreten. „Neurologie“ ist mit 7%, „Intensivmedizin“ und „Dermatologie“ jeweils mit 4% Anteil an der Prüfung angegeben. Im Vergleich zu den Ergebnissen der Verteilung der Fächer in den ausgewerteten Gedächtnisprotokollen wird dort erheblich weniger häufig nach generellen Überlegungen in Prävention, Gesundheitswesen, Ökonomie etc. (siehe oben) gefragt. Ebenfalls sind die Fächer „Kardiologie“ und „Gastroenterologie“ in den Protokollen mit 17% im Vergleich häufiger vertreten als im Schweizer Blueprint.

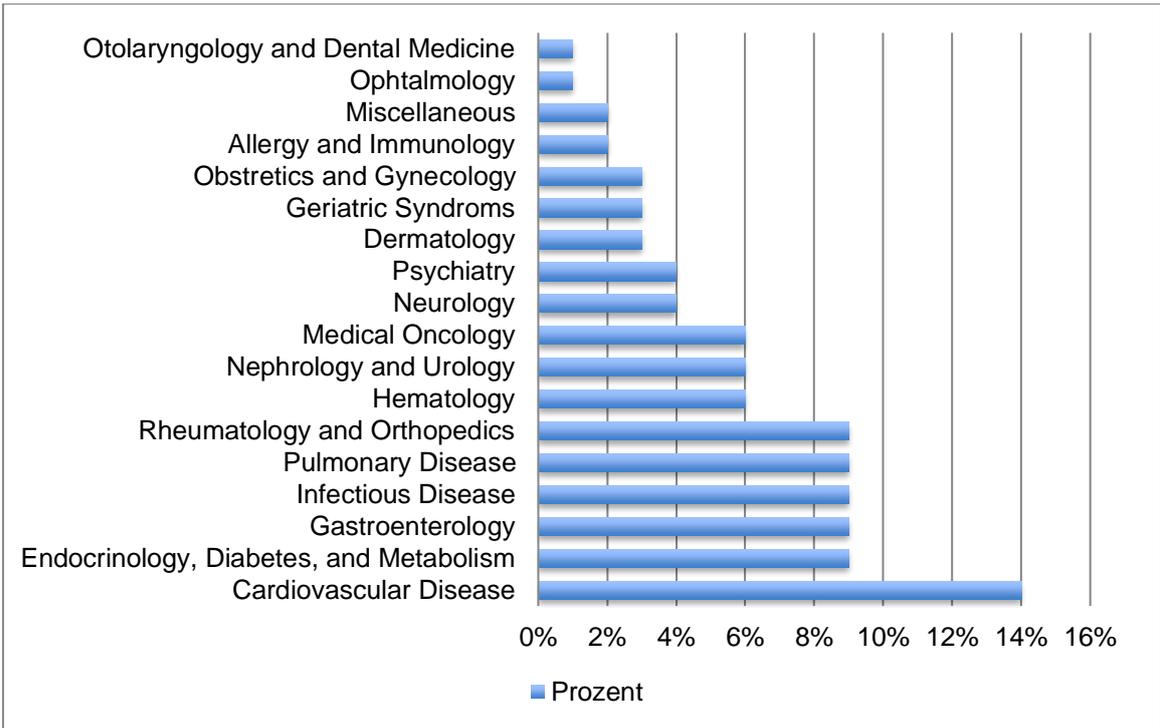


Abbildung 11: Blueprint des American Board of Internal Medicine

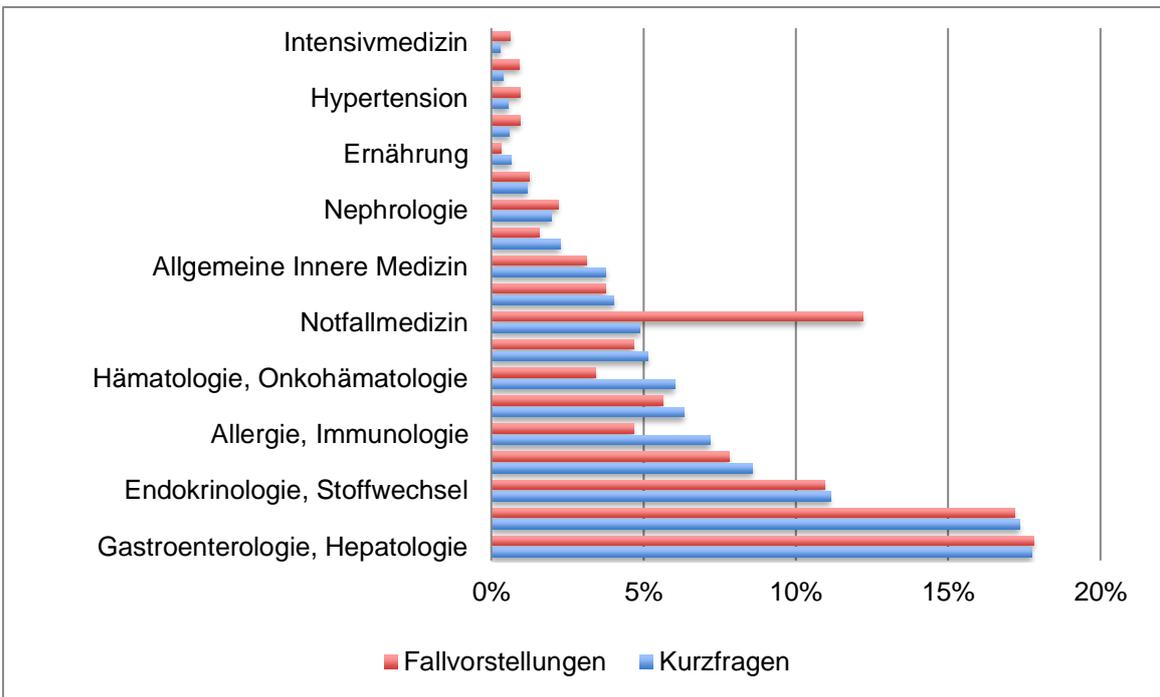


Abbildung 12: "Blueprint-Rekonstruktion" der Gedächtnisprotokolle

5.4 Aufteilung unterschiedlicher Untersuchungs- und Behandlungsmethoden

In diesem Abschnitt wurden die Ergebnisse in Bezug auf die Nebenhypothese 2 dargestellt. Es wurde überprüft, wie viele unterschiedliche Untersuchungs- und Behandlungsmethoden abgefragt werden und welche relative Aufteilung diese zueinander besitzen.

5.4.1 Aufteilung der Untersuchungs- und Behandlungsmethoden

Unter der Hauptkodierung „Medien“ wurden insgesamt 508 Kodierungen vorgenommen. Die Unterkategorie „Sonographie“ wurde mit 50,20% am häufigsten kodiert. Mit 14,96% als zweithäufigste Unterkategorie, wurde das „EKG“ kodiert.

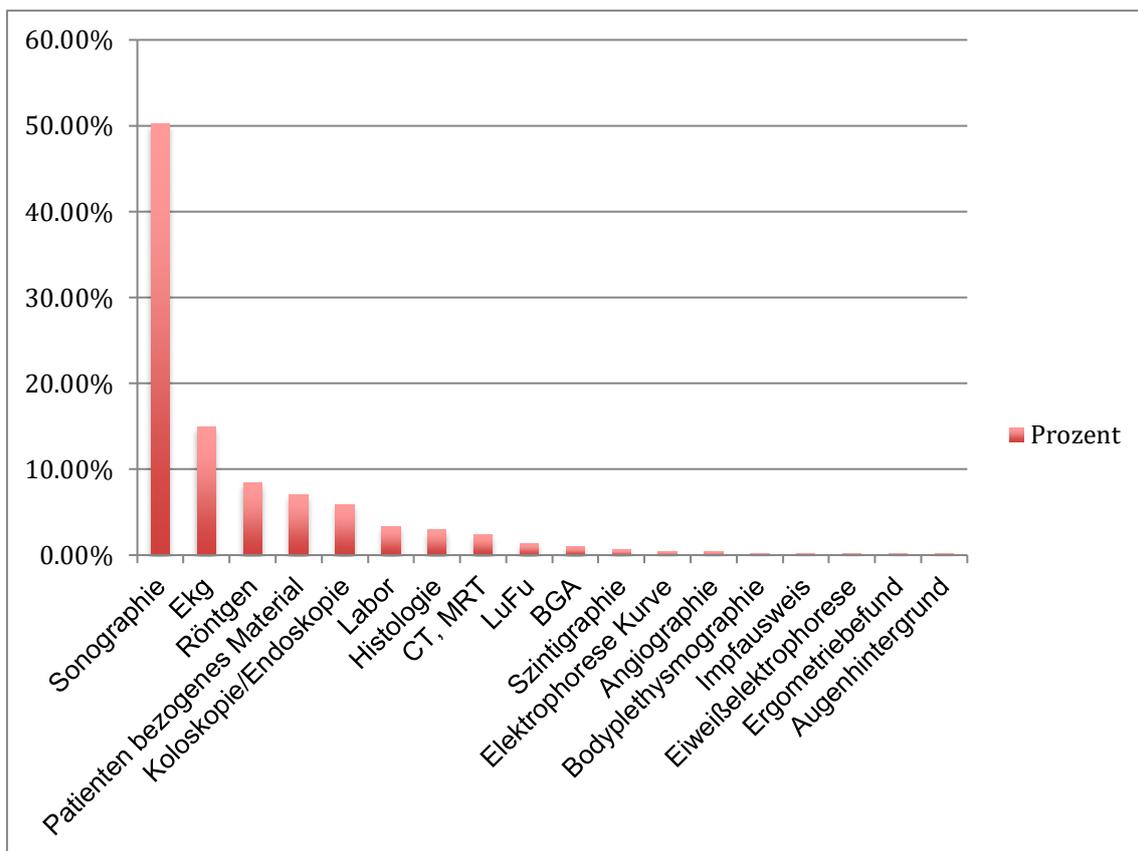


Abbildung 13: Aufteilung der Untersuchungs- und Behandlungsmethoden

Legende: EKG = Elektrokardiogramm, Lufu = Lungenfunktionsprüfung, BGA = Blutgasanalyse, CT = Computertomographie, MRT = Magnetresonanztomographie

5.5 Aufteilung des Kompetenzniveaus nach der ULME-Klassifikationsmatrix

In diesem Abschnitt sollen die Ergebnisse in Bezug auf die Haupthypothese 2 dargestellt werden. Um festzustellen, ob vor allem das Verständnis und kritisches Reflektieren klinischer Zusammenhänge abgeprüft werden, wurden alle „Kurzfragen“ im Sinne des ULME-Kompetenzstufenmodells in die jeweils drei Leistungs- und Wissensdimensionen eingeteilt. Abschließend wurden die Ergebnisse dieser Einteilung in die Kompetenzmatrix eingefügt und dargestellt.

5.5.1 Aufteilung der Leistungskategorien

Insgesamt 2053 Textpassagen wurden mit dem Code „Leistungskategorie“ kodiert. Hierbei wurden 1029 Passagen der Untergruppe „Verstehen und Anwenden“, 1011 der Untergruppe „Reproduzieren“ und 13 der Untergruppe „Kritisieren und Reflektieren“ zugeordnet.

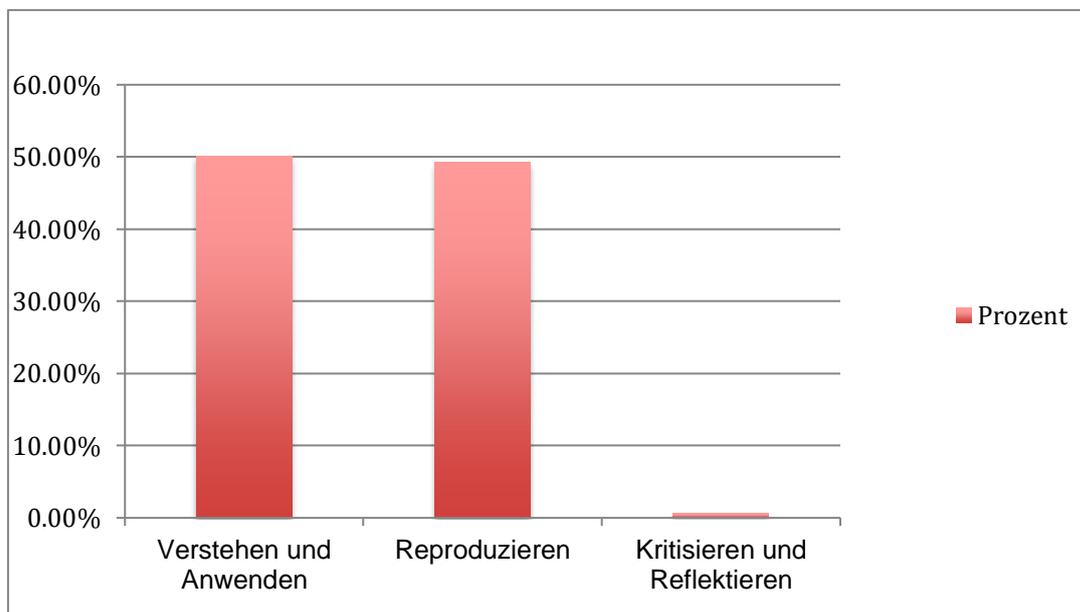


Abbildung 14: Aufteilung der Leistungskategorien

5.5.2 Aufteilung der Wissensdimensionen

Die insgesamt 2050 Kodierungen der Kategorie „Wissensdimension“ teilten sich in 886 Kodierungen der Untergruppe „Fakten“, 742 Kodierungen der Gruppe „Prozeduren“ und 422 Kodierungen der Untergruppe „Konzepte“.

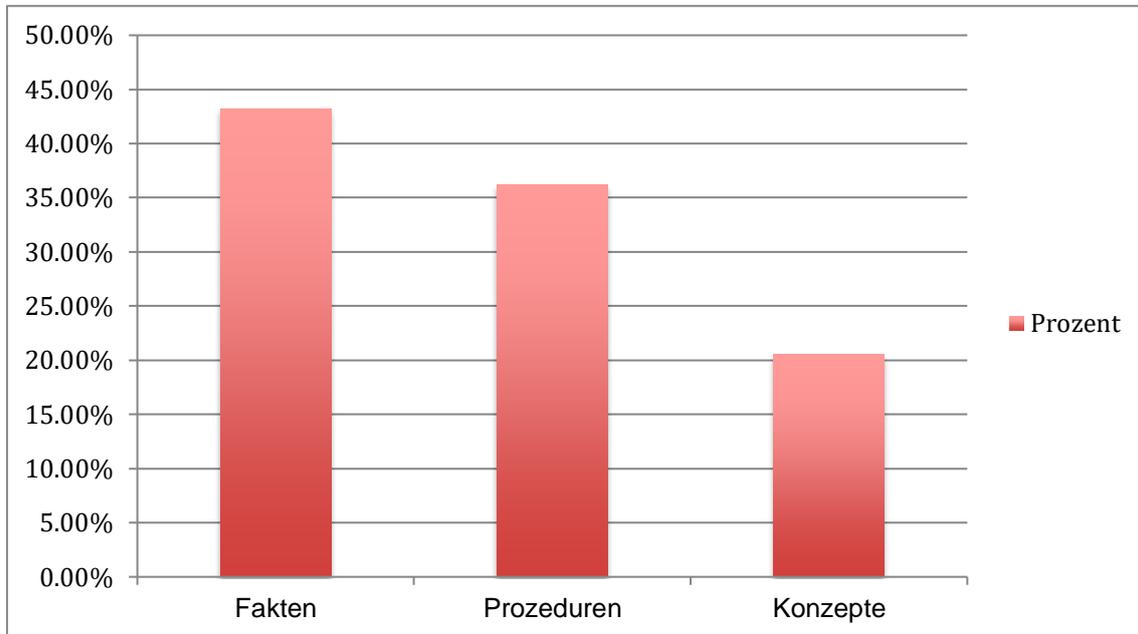


Abbildung 15: Aufteilung der Wissensdimensionen

5.5.3 Die Verknüpfung von Leistungskategorie und Wissensdimension

Die Leistungskategorie „Reproduzieren“ wurde 1010 mal kodiert und steht in Verbindung zu den drei Wissensdimensionen: „Fakten“, „Konzepte“ und „Prozeduren“.

Von den 1010 Kodierungen wurden 842 in Verbindung mit „Fakten“ kodiert, 2 in Verbindung mit „Konzepte(n)“ und 166 in Verbindung mit „Prozeduren“.

Die Leistungskategorie „Verstehen und Anwenden“ wurde 1030 mal kodiert. Von den 1030 Kodierungen wurden 38 in Verbindung mit „Fakten“ kodiert, 420 in Verbindung mit „Konzepte(n)“ und 572 in Verbindung mit „Prozeduren“.

Die Leistungskategorie Kritisieren und Reflektieren wurde 13 mal kodiert. Von den 13 Kodierungen wurde 9 mal mit „Fakten“ kodiert, kein mal mit „Konzepte(n)“ und 4 mal mit „Prozeduren“.

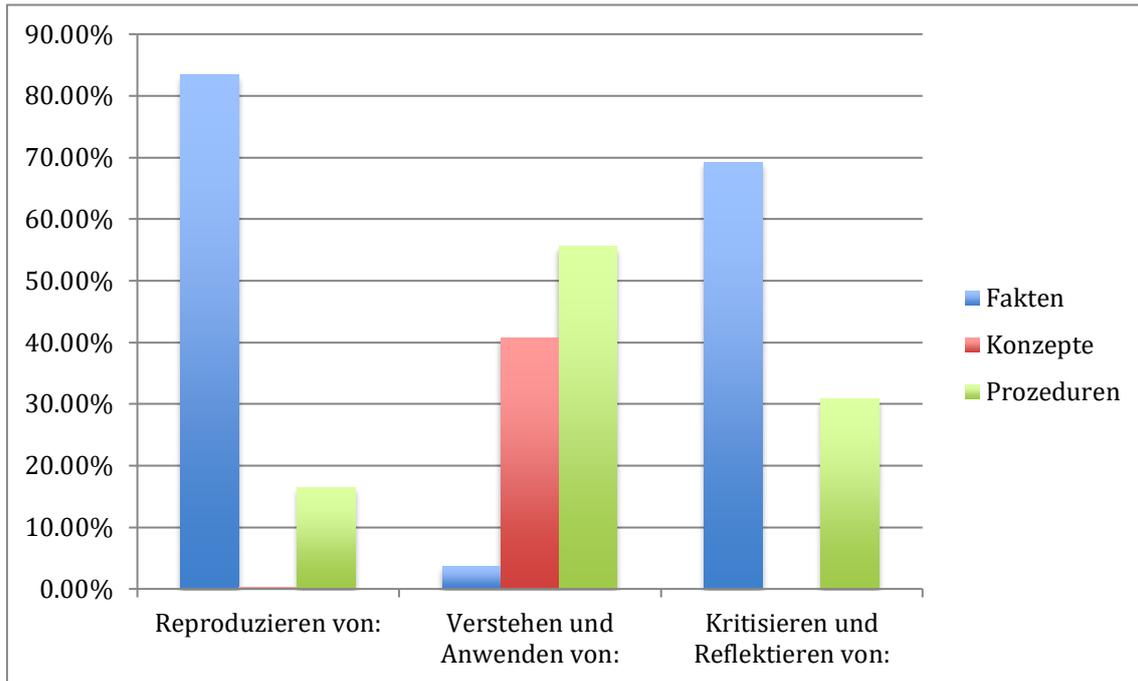


Abbildung 16: Verknüpfung von Leistungskategorie und Wissensdimension

Tabelle 6: Verknüpfung von Leistungskategorie und Wissensdimension anhand der ULME-Klassifikationsmatrix in absoluten Zahlen

Leistungskategorie	Reproduzieren	Anwenden/Verstehe	Kritisieren/Reflektiere
Wissensdimension		n	n
Fakten	842	38	9
Konzepte	2	420	0
Prozeduren	166	572	4

6 Diskussion

6.1 Zusammenfassung der Arbeit

Die Facharztprüfung ist von besonderer Bedeutung für die Qualität der ärztlichen Versorgung, da sie die letzte Prüfung der ärztlichen Weiterbildung vor dem Erwerb des Facharztstitels, ist. Das Bestehen der in mündlicher Form abgehaltenen Prüfung, soll dem Kandidaten das Erreichen der „Facharztkompetenz“ bescheinigen. Hierbei sollen die Inhalte der Weiterbildungsordnung abgedeckt werden. Die in der Prüfung abgefragten Inhalte und die Qualität der Prüfung wurden bislang nicht tiefgehend untersucht. In dieser Studie soll durch eine strukturierte Inhaltsanalyse nach Mayring und mittels der Einteilung von Fragen in das ULME-Kompetenzstufenmodell, die folgende Frage beantwortet werden: Welche inhaltlichen und strukturellen Charakteristika haben Facharztprüfungen in Deutschland am Beispiel der Inneren Medizin?

6.2 Diskussion der Methoden

Das dieser Arbeit zugrunde liegende Material lag in Form von schriftlich verfassten Gedächtnisprotokollen der Teilnehmer des DEGIM Kurses „Intensivkurs Innere Medizin – Vorbereitung auf die Facharztprüfung und Refreshkurs“ vor. Durch eine strukturierte Inhaltsanalyse nach Mayring konnten große Datenmengen bearbeitet und auf relevante Informationen reduziert werden, welche in spezifische Kategorien eingeordnet wurden. Um Informationen auf das Kompetenzlevel der Fragen zu erhalten, wurde das ULME-Kompetenzstufenmodell angewandt, welches auf der nach Anderson und Kathwohl modifizierten Bloom'schen Taxonomie aufbaut. Dadurch war es möglich, Fragen in Leistungskategorien und Wissensdimensionen einzuordnen und eine Matrix mit verschiedenen Kompetenzleveln zu bilden.

6.2.1 Stärken und Schwächen

Stärke der angewandten Vorgehensweise und Nutzung des genannten Materials ist unter anderem die relativ hohe Anzahl von bearbeiteten Protokollen (100). Außerdem konnte durch die interne Preisausschreibung mit Gewinnen für hochwertige Gedächtnisprotokolle eine hohe Motivation bei den Teilnehmern generiert werden. Die Teilnehmer waren zudem durch den unter der Schirmherrschaft der DEGIM veranstalteten Refresher-Kurs sicher zu identifizieren und dem ärztlichen Berufsstand zuzuordnen. Es wurden umfangreiche Daten erfasst, die sowohl Informationen über

abgefragte Schwerpunktbereiche als auch auf das Kompetenzlevel von Fragen ermöglichen. So konnte mit dem mehrdimensionalen Kodierschema eine tiefgreifende Auswertung des Kompetenzlevels erfolgen, die weitreichendere Informationen bietet, als beispielsweise ein eindimensionales Kodier Schema (1).

Schwächen bezogen auf das verwendete Material könnte man auf die nicht standardisierte Form der Gedächtnisprotokolle beziehen. Aus diesem Umstand ergeben sich teilweise nicht vollständige Protokolle und eine höhere Varianz in Bezug auf die Textlänge der individuellen Protokolle. Ebenfalls war das Bundesland Bayern in dieser Arbeit anteilmäßig überrepräsentiert. Die gesammelten Informationen geben zudem nur die Prüfung aus Sicht der Prüflinge wieder. Obwohl diese dazu aufgefordert wurden objektiv und ausführlich zu berichten, fehlt hier die Sicht der Prüfer auf die jeweilige Facharztprüfung.

6.2.2 Wissenschaftlicher Vergleich

In dieser Arbeit wurden Informationen aus niedergeschriebenen Gedächtnisprotokollen gewonnen. Die strukturierte Inhaltsanalyse hätte auch mit anderen medialen Formaten, wie beispielsweise der Analyse von Sprachaufnahmen (Sprach-Memos) durchgeführt werden können (35) (39). Diese hätten den Vorteil einer ganzheitlichen Darstellung der in der Prüfung Abgefragten Inhalte ermöglicht. Tatsächlich werden in der Facharztprüfung regelmäßig Sprachaufnahmen erstellt. Diese dienen bei Nichtbestehen der Prüfung zur Revision durch ein Prüfungskomitee (2). Die Aufnahmen stehen jedoch unter Verschluss und standen der wissenschaftlichen Analyse nicht zur Verfügung. Auf medizinischen Web-Seiten und von (wenigen) Verlagen publizierten Manuskripten finden sich ausschließlich Erfahrungsberichte und Gedächtnisprotokolle in schriftlicher Form. Aus diesen Gründen wurde auch bei dieser Arbeit die Analyse auf Gedächtnisprotokolle bezogen.

Bei der Analyse der Protokolle wurde die strukturelle Inhaltsanalyse nach Mayring gewählt. Sie gehört zu den Techniken der qualitativen Analyse, die zu einer Standardmethode der Textanalyse geworden ist (40). Trotz der nicht standardisierten Form der Protokolle konnte so ein ganzheitlicher Ansatz der Analyse verfolgt werden, in dem neue Textelemente direkt adaptiert und in eine Änderung des Kategoriensystems übertragen werden können (41) (40). Das System ermöglicht eine genaue Anpassung der Kategorienbildung auf die Kernhypothesen, dies ist besonders gut geeignet für Arbeiten wie diese, da durch ihren explorativen Charakter keine paradigmatischen Kategorien und Analysesysteme für mündliche FA-Prüfungen existieren.

Vor dem Hintergrund der bereits etablierten Bloom'schen „Taxonomy of educational objectives“ muss die Verwendung weniger bekannter Klassifikationsmodelle für die Einteilung von Kompetenzen argumentiert werden. Das in dieser Arbeit verwendete ULME-Kompetenzstufenmodell baut in fundamentalen Bestandteilen auf der Bloom'schen Taxonomie und der von Anderson und Krathwohl modifizierten Bloom'schen Taxonomie auf (42) (37). Die von Anderson und Krathwohl revidierte und erweiterte Bloomsche Taxonomie zeichnet dabei besonders aus, dass sie im Vergleich zu Bloom die Wissensdimension von der Leistungsdimension trennt. So entstand bei Anderson und Krathwohl eine Matrix mit vier Wissensdimensionen und sechs Leistungsdimensionen (37). Das ULME-Kompetenzstufenmodell fasst diese Matrix noch einmal zusammen und bildet drei Wissensdimensionen und drei Leistungskategorien (36) (siehe S. 14). Bei den in den Gedächtnisprotokollen dargestellten Fragen war es oftmals nicht möglich, eine exakte Kategorisierung in die ausdifferenzierten Leistungskategorien und Wissensdimensionen des Modells von Anderson und Krathwohl zu tätigen. Daher bot sich das ULME-Modell an, da hier eine geringfügig breitere Definition für die unterschiedlichen Dimensionen existiert, die zur Bearbeitung im Sinne der aufgestellten Hypothesen aber vollständig ausreichte.

Tabelle 7: Vergleich der Klassifikationsmatrix von Anderson und Krathwohl zur Kompetenzmatrix des ULME-Modells

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Meta-cognitive Knowledge						

Leistung		Reproduzieren	Anwenden/ Verstehen	Kritisieren/ Reflektieren
Wissen				
Fakten	A B			
Konzepte	A B			
Prozeduren	A B			

6.2.3 Ausblick

Ein transparenter, institutionell geplant und durchgeführter Fragebogen, der auf Seiten von Absolventen und Prüfern prüfungsbezogene Daten erhebt, könnte in Bezug auf die Qualitätssicherung der Prüfung und für die wissenschaftlich (noch bessere) Analyse der Prüfung interessant sein.

6.3 Diskussion der Ergebnisse

Die Ergebnisse dieser Arbeit können die aufgestellten Hypothesen in weiten Teilen nicht unterstützen. Die abgefragten Wissensstichproben wiesen trotz hohen Mittelwerten bei der Verwendung unterschiedlicher Themengebiete eine hohe prüfungsabhängige Varianz auf. Das praktische Handlungswissen wurde im Vergleich zum Reproduzieren von Fakten weniger häufig geprüft. Der Blueprint des ABOIM und die Aufteilung der unterschiedlichen Themengebiete in den Protokollen wiesen einige Gemeinsamkeiten auf, im Gegensatz zu der Verteilung der Themengebiete in den Protokollen gilt diese Verteilung im ABOIM Blueprint jedoch für alle Prüflinge gleichermaßen. Die Untersuchungs- und Behandlungsverfahren wurden regelhaft geprüft, allerdings war die Verwendung unterschiedlicher Verfahren nicht gleichmäßig verteilt, sondern stark auf die Unterkategorie „Sonographie“ verschoben.

6.3.1 Diskussion zu Hypothese 1

Hypothese 1: *„Die Weiterbildungsinhalte werden durch Abfrage ausreichend vieler Wissensstichproben aus unterschiedlichen Schwerpunktfächern abgedeckt.“*

Bei durchschnittlich 27 Fragen pro Prüfung kann davon ausgegangen werden, dass ausreichend viele Wissensproben aus den unterschiedlichen Richtungen abgefragt wurden. Die hohe Varianz in der Anzahl abgeprüfter Fragen könnte hierbei darauf hinweisen, dass aufgrund der geringen Standardisierung eine hohe interindividuelle Variabilität besteht.

Die Auswertung der Anzahl unterschiedlicher Themengebiete bezogen auf die Gesamtheit der Hauptkategorien betrug im Mittel 11,03 mit einer Standardabweichung von 3,05. Die maximale Anzahl unterschiedlicher Themengebiete betrug neunzehn, die minimale Anzahl lag bei vier. Es wurden durchschnittlich also relativ viele unterschiedliche Themengebiete abgefragt, die Varianz in den einzelnen Prüfungen war jedoch hoch. Die Ergebnisse dieser Untersuchung stimmen mit der in verschiedenen Studien belegten niedrigen Reliabilität und hohen Divergenz in der Durchführung mündlicher Prüfungen überein. Konkret kommen A.Yeung et al und H. van den Busche et al in ihren Studien zu dem Ergebnis, dass eine hohe Divergenz und geringe Reliabilität zwischen den Ergebnissen mündlicher und schriftlicher Prüfungen besteht (43) (44). Bezogen auf JC Yang et al kam es zu einer deutlichen Verbesserung der Reliabilität, wenn anstatt einer unstrukturierten mündlichen Prüfung, ein strukturiertes

Evaluationsinstrument zum Einsatz kam (20). Zusammenfassend konnte festgestellt werden, dass nicht sicher ist, ob in jeder Prüfung ausreichend viele unterschiedliche Themengebiete abgefragt werden. Im Hinblick auf die Ergebnisse dieser Fragestellung muss darauf hingewiesen werden, dass sich einige Protokollanten möglicherweise besser an die Prüfungsfragen erinnern konnten als andere. Dies könnte zu Abweichungen der Resultate geführt haben.

6.3.2 Diskussion zu Hypothese 2

Hypothese 2: „Es wird vor allem das Verständnis und kritisches Reflektieren klinischer Zusammenhänge abgefragt“

Die Bearbeitung dieser Fragestellung kam zu dem Ergebnis, dass fast die Hälfte (49,01%) aller geforderten Leistungen auf das Reproduzieren von verschiedenen Wissensdimensionen bezogen war. Von den verschiedenen Wissensdimensionen wurde das Faktenwissen am Häufigsten abgefragt (43,04%). In Bezug auf die der Methode immanenten Vorteile der mündlichen Prüfung ist aufgrund verschiedener Literaturquellen zu erwarten, dass besonders praktisches Handlungswissen abgefragt wird (9) (1). In einer Studie von Öchner W et al wurde ebenfalls festgestellt, dass bei der mündlichen Prüfung des zweiten Staatsexamens überwiegend Faktenwissen geprüft wurde, diese Problematik betrifft also nicht nur die mündliche Facharztprüfung (45). Aus testökonomischer Sicht ist die Abfrage von reinem Faktenwissen in mündlichen Prüfungen nicht wirtschaftlich und eher eine Domäne der Wahlantwortverfahren/MCQ-Prüfungen (9). Vor dem Hintergrund anderer Prüfungsmodelle, wie beispielsweise der europäischen Notarztprüfung, in der OSCE-Stationen und MCQ-Fragen eingesetzt werden, kann argumentiert werden, dass die deutsche Facharztprüfung der Inneren Medizin mit ihrer Prüfung nicht die für diese Methode optimalen inhaltlichen Kompetenzen prüft (25). Als mildernder Umstand sollte allerdings betrachtet werden, dass im ersten Schritt der fachärztlichen Weiterbildung besonders praktische Fähigkeiten und Handlungskompetenzen erworben werden, die der Weiterzubildende dokumentiert und von seinem Unterweisungsbefugten bestätigen lassen muss. Erst bei Nachweis der Ableistung dieser Fähigkeiten und Handlungen kann der Weiterzubildende die mündliche Facharztprüfung antreten (2).

6.3.3 Diskussion zu Hypothese 3

Hypothese 3: „Das abgefragte Spektrum inhaltlicher Themengebiete ist in seiner Verteilung dem amerikanischen „Blueprint“ ähnlich.“

Im Vergleich mit dem amerikanischen Blueprint des American Board of Internal Medicine und den Ergebnissen der Aufteilung inhaltlicher Themengebiete der Gedächtnisprotokolle sind viele Ähnlichkeiten festzustellen. Hierbei waren besonders viele Gemeinsamkeiten in der prozentualen Aufteilung der Themengebiete mit dem größten relativen Anteil vorhanden. So sind beispielsweise 5 der 6 im ABOIM Blueprint am häufigsten vertretenen Themengebiete auch in den häufigsten 6 Themengebieten der Gedächtnisprotokolle vertreten. Bei dem Vergleich beider „Blueprints“ darf jedoch ein Unterschied von großer Bedeutung nicht unterschlagen werden. Der amerikanische Blueprint gibt eine Aufteilung der Themengebiete vor, die für jeden Prüfling exakt gleich ist. Die Verteilung der Themengebiete der Gedächtnisprotokolle ist zwar sehr ähnlich, stellt aber die Aufteilung der Themengebiete nicht für jede einzelne Prüfung, sondern summativ für die gesamten 100 Prüfungsprotokolle dar. Es können also in spezifischen Prüfungen erhebliche Abweichungen bestehen. Daher ist die Entwicklung von prädefinierten Fragen und strukturierten Checklisten, im Sinne eines Blueprints auch für die deutsche Facharztprüfung sinnvoll. Die Empfehlung der Verwendung von Checklisten und Blueprints findet sich in zahlreichen Publikationen (46) (47) (1). Im Vergleich zu Facharztprüfungen in Kanada, USA, der Schweiz und Österreich ist die Methodik der unstrukturierten mündlichen Prüfung ohne Blueprint zudem ein Einzelfall (48). Zudem könnte nach Cilliers et al die Entwicklung eines Blueprints eine positive lernsteuernde Wirkung auf die Absolventen haben (16). So könnten bezogen auf Van der Vleuten et al der Aufwand, die verwendeten Materialien und Lernformen die von Prüflingen genutzt werden, von den Entwicklern des Blueprints an die gestellten Anforderungen angepasst werden (49) (9). Auch der „Guidance on the Swiss Board Examination in General Internal Medicine“ Blueprint gibt genaue Verteilungsangaben für die Prüfungsthemen an. Hier war im Vergleich zu den Protokollen besonders auffällig, dass der Schweizer Blueprint generelle Überlegungen zu Prävention, Palliativmedizin, Rehabilitation, Gesundheitswesen und Ökonomie fest in seine Themenwahl implementiert, während dieser Ansatz in den deutschen Protokollen praktisch nicht zum Tragen kommt (7).

6.3.4 Diskussion zu Hypothese 4

Hypothese 4: „Von der WBO geforderte Untersuchungs- und Behandlungsverfahren werden in Form verschiedener Medien geprüft.“

Die Bearbeitung der Fragestellung ergab, dass insgesamt 18 unterschiedliche Untersuchungs- und Behandlungsverfahren geprüft wurden. Mit sehr wenigen Ausnahmen wurden diese Verfahren in jeder Prüfung abgefragt. Auffällig war besonders die hohe Anzahl der Untersuchungsmethode „Sonographie“ (50,49%). Im Vergleich der Facharztprüfungen von Ländern wie beispielsweise der USA, bei denen klinische Informationen und Behandlungsverfahren in die computergestützte MCQ-Prüfung eingepflegt und somit für alle Prüflinge gleich zu beantworten sind, kann auch hier die geringe Standardisierung der deutschen Facharztprüfung kritisiert werden (23). Ebenfalls bietet die Abfrage der Untersuchungs- und Behandlungsverfahren in der deutschen Facharztprüfung kein ganzheitliches Assessment der klinischen Fähigkeiten der Absolventen. Der bei der Anwendung eines OSCEs beobachtbare praktische Handlungsaspekt, wie beispielsweise die korrekte Durchführung einer Sonographie um initial zu der klinisch pathologischen Bilddarstellung zu gelangen, fehlt hier (9).

6.4 Ausblick

Die Ergebnisse dieser Arbeit konnten die aufgestellten Hypothesen in weiten Teilen nicht unterstützen. In Zusammenschau der Ergebnisse dieser Arbeit ergeben sich folgende Empfehlungen für die zukünftige Entwicklung der ärztlichen Weiterbildung in Deutschland bezogen auf die Facharztprüfung:

Der Einsatz verschiedener Prüfungsmethoden (Triangulation) sollte genutzt werden, um alle Bereiche geforderter Kompetenzen mit adäquaten Prüfungsmethoden optimal prüfen zu können. Dabei könnte auch eine höhere Prüfungsvalidität erreicht werden (50).

Die Festlegung von zu prüfenden Inhalten im Rahmen der Konzeptionierung eines Blueprints könnte zu einer Verbesserung der Gewichtung relevanter Lernziele dienen und die Prüfung im Hinblick auf die Qualitätsmerkmale der Testgütekriterien optimieren (46).

Klar strukturierte Checklisten und die Entwicklung eines Erwartungshorizontes könnten die Interpretations- und Auswertungsobjektivität verbessern.

Um eine Mindestreliabilität von 0,8 zu erreichen, könnte die Anzahl der gestellten Fragen erhöht werden (8) (9).

Sollte die mündliche Prüfung als Prüfungsform beibehalten werden, wäre eine vermehrte Fokussierung auf das Abfragen praktischer Handlungskompetenzen im Vergleich zu dem Reproduzieren von Faktenwissen wünschenswert. Zudem sollte die Leistungskategorie „Kritisieren und Reflektieren“, die sich bei den Ergebnissen dieser Arbeit als stark unterrepräsentiert zeigte, häufiger abgefragt werden.

Um eine gleichmäßig hohe Kompetenz der Prüfer zu gewährleisten, würden sich Lehrgänge anbieten, welche die Kommunikationsfähigkeiten der Prüfer und die Prüfungsstruktur in den jeweiligen Prüfungen stärken könnten (51).

Vor dem Hintergrund der Novellierung der (Muster-)Weiterbildungsordnung, welche die Bundesärztekammer 2018 verabschieden möchte und in der eine stärkere Kompetenzorientierung der ärztlichen Weiterbildung angestrebt wird, wäre es sinnvoll, das Momentum der Veränderungsprozesse zu nutzen und auch die mündliche Facharztprüfung zu reformieren (52). Bei den Reformationen könnte man den Dialog zu ausländischen Institutionen wie dem ABOIM oder des deutschsprachigen Auslandes (DACH-Region) suchen, die bereits die oben beschriebenen Empfehlungen in weiten Teilen in ihre Weiterbildungsordnung implementiert haben.

7 Literaturverzeichnis

1. Bloch R, Hofer D, Krebs R, Schläppi P, Weis S, Westkämper R. Kompetent prüfen. Handbuch zur Planung, Durchführung und Auswertung von Facharztprüfungen Medizinische Fakultät Universität Bern, Bern/Wien: Institut für Aus-, Weiter-und Fortbildung. 1999.
2. Bundesärztekammer. (Muster-) Weiterbildungsordnung und (Muster-) Richtlinie. Stand: 23.10.2015.
3. Ratnapalan S, Hilliard RI. Needs assessment in postgraduate medical education: a review. *Medical Education Online*. 2002;7(1):4542.
4. Weinert F. S. 27f; Vergleichende Leistungsmessung in Schulen—eine umstrittene Selbstverständlichkeit. Weinert, Leistungsmessung in Schulen; Weinheim/Basel. 2002;11.
5. ed. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM). In: e.V MMFdB, editor. 2015
6. Ge M. The assessment of clinical skills/competence/performance. *Academic Medicine*. 1990;65(09).
7. al. BHe. Swiss catalogue of learning objectives for undergraduate medical training. 2008 ff. Stand: 04.02.2018.
8. Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild*. 2006;23(3):2006-23.
9. Bauer D, Görlitz A, Huber J, Fischer MR, Holzer M. Theorie und Praxis: Prüfungen in der Medizin. In: Brockmann J, Pilniok A, editors. *Prüfen in der Rechtswissenschaft: Probleme, Praxis und Perspektiven*. 1 ed. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG; 2013. p. 75-94.
10. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical teacher*. 2011;33(3):206-14.
11. Döring N, Bortz J. *Forschungsmethoden und Evaluation*: Springer; 2016.
12. Fabry G. *Medizindidaktik: ein Handbuch für die Praxis*: Huber; 2008.
13. Lienert GA, Raatz U. *Testaufbau und testanalyse*: Beltz; 1998.
14. Harden RM, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical education*. 1979;13(1):39-54.
15. Scouller K. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*. 1998;35(4):453-72.
16. Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, Van der Vleuten CP. The mechanism of impact of summative assessment on medical students' learning. *Advances in health sciences education*. 2010;15(5):695-715.
17. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *ACADEMIC MEDICINE-PHILADELPHIA-*. 1993;68:443-.
18. Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse education today*. 2007;27(5):481-90.
19. Montagne S, Meng-Hentschel J, Huwendiek S. Prüfungen: Schweizer Perspektiven (Vorträge). Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA).121.
20. Yang JC, Laube DW. Improvement of reliability of an oral examination by a structured evaluation instrument. *Journal of medical education*. 1983;58(11):864-72.
21. Fortbildung SSIfÄW-u. Facharzt für Allgemeine Innere Medizin - sgaim. 01.01.2011, letzte Revision 05.03.2015.
22. Specialties ABoM. ABMS Board Certification Report 2015-2016 December 2016 [Available from: www.abms.org].

-
23. Medicine ABol. Internal Medicine, Certification Examination Blueprint 2018 [updated January 2018. Available from: www.abim.org.
 24. Specialties ABoM. Standards for the ABMS Program for Maintenance of Certification 2015 [updated 2014. Available from: www.abms.org.
 25. Dißmann P. Europäische Facharztprüfung für Notfallmedizin 2014. 697-700 p.
 26. Dißmann P. Europäische Facharztprüfung für Notfallmedizin. Notfall+ Rettungsmedizin. 2014;17(8):697-700.
 27. Landesärztekammer B. Tätigkeitsbericht der Bayerischen Landesärztekammer 2014/15 dem 74. Bayerischen Ärztetag vorgelegt. . Bayerische Landesärztekammer; 2015 2015.
 28. Landesärztekammer B. Tätigkeitsbericht der bayrischen Landesärztekammer 2016/2017. Bayerische Landesärztekammer; 2017.
 29. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. Medical education. 2004;38(2):199-203.
 30. Vleuten CVD, Verwijnen G, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. Medical Teacher. 1996;18(2):103-9.
 31. Kadmon M, Ganschow P, Gillen S, Hofmann H, Braune N, Johannink J, et al. Der kompetente Chirurg. Der Chirurg. 2013;84(10):859-68.
 32. Vorbereitung auf die Facharztprüfung und Refresherkurs Available from: http://www.klinikum.uni-muenchen.de/Medizinische-Klinik-und-Poliklinik-IV/de/Ueberuns/download/Intensiv_Kurse_Innere_Medizin/IIM_2019/index.html.
 33. Mason M, editor Sample size and saturation in PhD studies using qualitative interviews. Forum qualitative Sozialforschung/Forum: qualitative social research; 2010.
 34. Lamnek S. Gruppendiskussion. Theorie und Praxis. 2. überarbeitete und erweiterte Auflage 2005 [1998]. Weinheim ua: Beltz; 1998.
 35. Flick U. Qualitative Sozialforschung–Eine Einführung, vollständig überarbeitete und erweiterte Neuauflage. Reinbek bei Hamburg: Rowohlt. 2007.
 36. Hofmeister W. Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. Prüfungen und Standards in der beruflichen Bildung In bwp@-Berufs-und Wirtschaftspädagogik online, Ausgabe Nr. 2005;8.
 37. Anderson LW, Krathwohl DR, Airasian P, Cruikshank K, Mayer R, Pintrich P, et al. A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy. New York Longman Publishing Artz, AF, & Armour-Thomas, E(1992) Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups Cognition and Instruction. 2001;9(2):137-75.
 38. MAXQDA, Software für qualitative Datenanalyse 2018 [cited 2018 30.03.2018]. Available from: <https://www.maxqda.de/faq/wie-zitiert-man-maxqda>.
 39. Kvale S. Doing interviews: Sage; 2008.
 40. Mayring P. Qualitative Inhaltsanalyse. Weinheim [u.a.]: Beltz; 2015. 1 Online-Ressource (152 S.) : p.
 41. Mayring P. Einführung in die qualitative Sozialforschung: Beltz Weinheim; 2016.
 42. Bloom BS, College Co, Examiners U. Taxonomy of educational objectives: Longmans, Green New York; 1964.
 43. Yeung A, Booth T, Larkin T, McCoubrie P, McKnight L. The FRCR 2B oral examination: Is it reliable? Clinical radiology. 2013;68(5):466-71.
 44. van den Bussche H, Wegscheider K, Zimmermann T. Der Ausbildungserfolg im Vergleich (III). Dtsch Arztebl. 2006;103:B2762-B5.
 45. Öchsner W PA, Huber-Lang M. Auf dem Kompetenzprüfstand: Die mündliche-praktische Prüfung des Zweiten Abschnitts der Ärztlichen Prüfung. Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ), Leipzig. 2015.
 46. Hays R. Assessment in medical education: roles for clinical teachers. The Clinical Teacher. 2008;5(1):23-7.

47. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Advances in health sciences education*. 2010;15(2):277-89.
48. van der Heiden N, Gesundheitsberufe C-LSW. Der Weg zur fachlich eigenverantwortlichen ärztlichen Tätigkeit in der Schweiz und in ausgewählten Ländern.
49. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education*. 1996;1(1):41-67.
50. Flick U. *Triangulation*. Wiesbaden: VS, Verl. für Sozialwiss.; 2011. 127 S. p.
51. Streitlein-Böhme I, Niebling W, Fabry G, Böhme K. Three years of experience with a workshop for medical specialty examiners in South Baden: A project report and initial evaluation results. *GMS journal for medical education*. 2017;34(5).
52. Novellierung der (Muster-)Weiterbildungsordnung [Available from: <http://www.bundesaerztekammer.de/aerzte/aus-weiterbildung/weiterbildung/novellierung/>].

8 Abbildungsverzeichnis

<i>Abbildung 1: Kompetenzebenen von NKLM im Vergleich zu Miller-Pyramide und Swiss catalogue of learning Objectives for Undergraduate Medical Training (SCLO) (5)</i>	3
<i>Abbildung 2: Arbeitsschritte der Analyse nach Mayring</i>	11
<i>Abbildung 3: Beispiel der Kodierschematik</i>	12
<i>Abbildung 4: Spezialisierungen der Prüfer</i>	18
<i>Abbildung 5: Häufigkeit der unterschiedlichen Themengebiete</i>	19
<i>Abbildung 6: Die Varianz der Protokolle bezogen auf die unterschiedlichen Themengebiete in den Hauptkategorien: "Fallvorstellungen", "Kurzfragen" und "Medien"</i>	20
<i>Abbildung 7: Aufteilung der Fallvorstellungen nach Themengebiet</i>	21
<i>Abbildung 8: Aufteilung der Einzelfragen nach Themengebiet</i>	22
<i>Abbildung 9: Häufige Überschneidungen der Kodierung bei Fallvorstellungen</i>	23
<i>Abbildung 10: Häufige Überschneidungen der Kodierung von Kurzfragen</i>	24
<i>Abbildung 11: Blueprint des American Board of Internal Medicine</i>	26
<i>Abbildung 12: "Blueprint-Rekonstruktion" der Gedächtnisprotokolle</i>	26
<i>Abbildung 13: Aufteilung der Untersuchungs- und Behandlungsmethoden</i>	27
<i>Abbildung 14: Aufteilung der Leistungskategorien</i>	28
<i>Abbildung 15: Aufteilung der Wissensdimensionen</i>	29
<i>Abbildung 16: Verknüpfung von Leistungskategorie und Wissensdimension</i>	30
<i>Abbildung 17: Vollständige Abbildung der Überschneidungen von "Fallvorstellungen"</i>	43
<i>Abbildung 18: Vollständige Abbildung der Überschneidungen der Themengebiete in der Kategorie "Kurzfragen"</i>	44

9 Tabellenverzeichnis

<i>Tabelle 1: Die Kompetenzmatrix nach dem ULME-Modell</i>	13
<i>Tabelle 2: Charakterisierung der Wissenskategorie nach dem ULME-Modell (36)</i>	13
<i>Tabelle 3: Charakterisierung der Wissenskategorie nach dem ULME-Modell (36)</i>	14
<i>Tabelle 4: Aufteilung des Prüfungsortes nach Häufigkeit</i>	17
<i>Tabelle 5: Aufteilung der Facharzt-Spezialisierungen</i>	17
<i>Tabelle 6: Verknüpfung von Leistungskategorie und Wissensdimension anhand der ULME-Klassifikationsmatrix in absoluten Zahlen</i>	30
<i>Tabelle 7: Vergleich der Klassifikationsmatrix von Anderson und Krathwohl zur Kompetenzmatrix des ULME-Modells</i>	33

10 Abkürzungsverzeichnis

NKLM	<i>Nationale Kompetenzbasierte Lernzielkataloge Medizin</i>
ÄAppO	<i>Approbationsordnung für Ärzte</i>
WBO	<i>Weiterbildungsordnung</i>
MC-Prüfung	<i>Multiple Choice Prüfung</i>
MCQ-Prüfung	<i>Multiple Choice Question Prüfung</i>
OSCE	<i>Objective structured clinical examination</i>
SMP	<i>Strukturierte mündliche Prüfung</i>
KAF	<i>Kurzantwortfragen</i>
ABOIM	<i>American Board of Internal Medicine</i>
MOC	<i>Maintenance of certification</i>
DEGIM	<i>Deutsche Gesellschaft für Innere Medizin</i>
BLÄK	<i>Bayrische Bundesärztekammer</i>
KUM	<i>Klinikum der Ludwigs-Maximilians-Universität München</i>
MAÄF e.V.	<i>Münchner Akademie für Ärztliche Fortbildung</i>
DACH-Region	<i>Akronym für Deutschland, Österreich und Schweiz</i>

11 Anhang

11.1 Zusätzliche Abbildungen

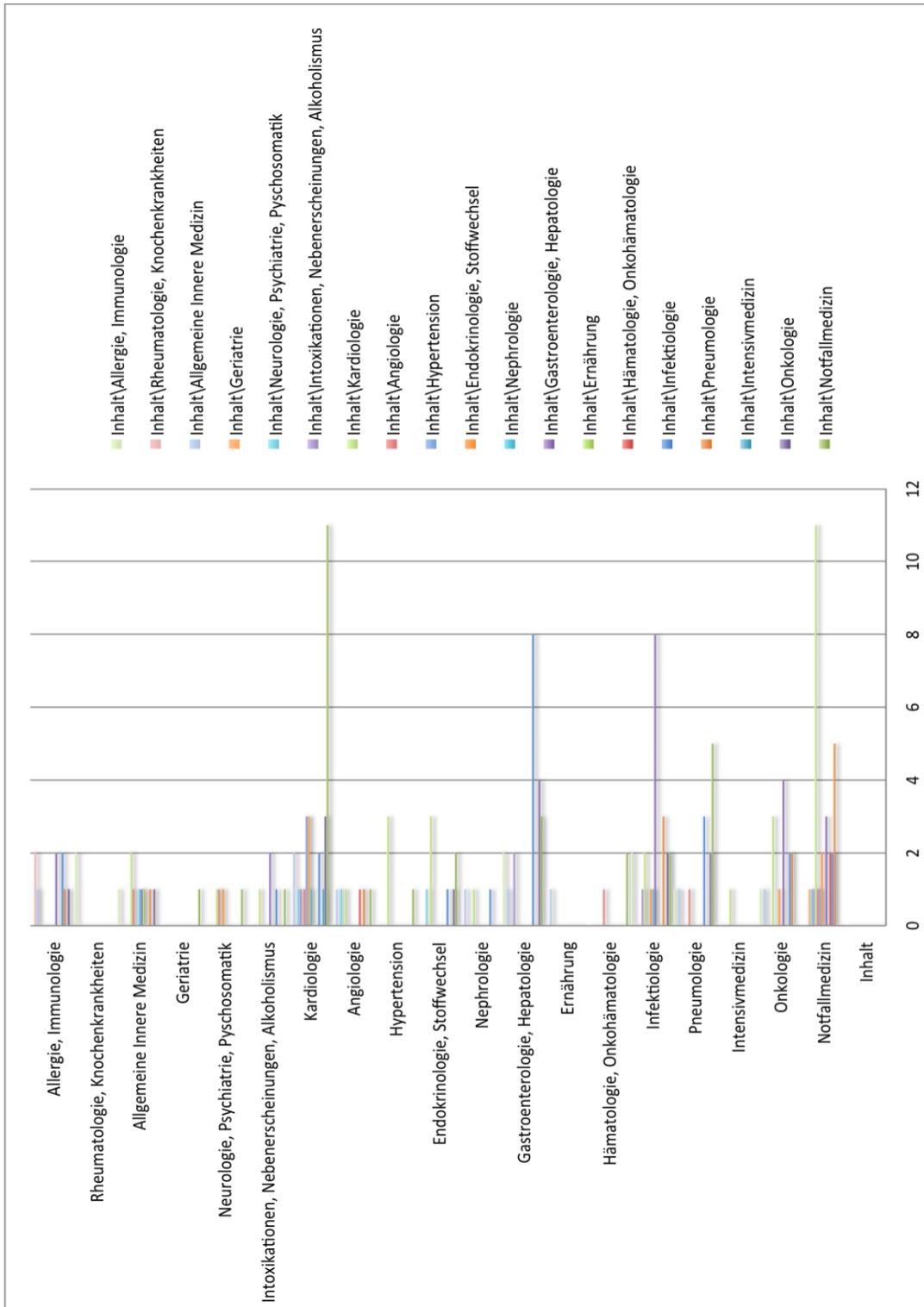


Abbildung 17: Vollständige Abbildung der Überschneidungen von "Fallvorstellungen"

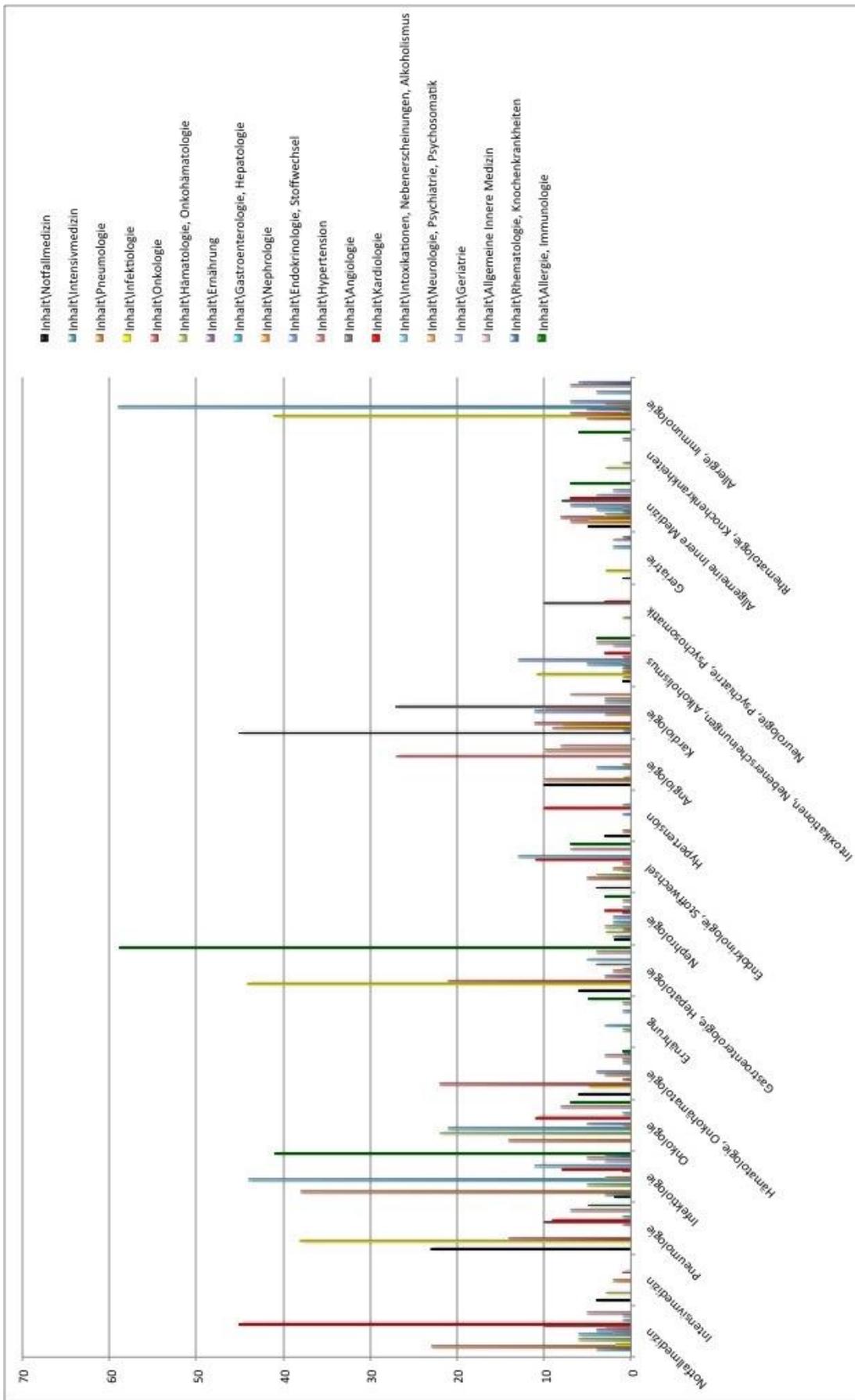


Abbildung 18: Vollständige Abbildung der Überschneidungen der Themengebiete in der Kategorie "Kurzfragen"

11.2 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Ausführungen, die anderen veröffentlichten oder nicht veröffentlichten Schriften wörtlich oder sinngemäß entnommen wurden, habe ich kenntlich gemacht.

Die Arbeit hat in gleicher oder ähnlicher Fassung noch keiner anderen Prüfungsbehörde vorgelegen.

Düsseldorf, 25.03.2022

Marc Beckers

Ort, Datum

Unterschrift