

# Die Facharztprüfung Innere Medizin: Eine qualitative und quantitative Analyse von Gedächtnisprotokollen

## The specialist examination in internal medicine: a qualitative and quantitative analysis of minutes from memory



### Autoren

Marc Aurelio Beckers<sup>1,2</sup>, Leah T. Braun<sup>1</sup>, Nurith Epstein<sup>2</sup>, Martin R. Fischer<sup>2</sup>, Ralf Schmidmaier<sup>1</sup>



### Institute

- 1 Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Ludwig-Maximilians-University (LMU) Munich, Germany
- 2 Institut für Didaktik und Ausbildungsforschung in der Medizin, Klinikum der Universität München, Ludwig-Maximilians-University (LMU) Munich, Germany

### Schlüsselwörter

Innere Medizin, Facharztprüfung, Weiterbildung

### Key words

internal medicine, specialist examination, postgraduate training

### Bibliografie

DOI <https://doi.org/10.1055/a-0970-6348>

Online-Publikation: 23.9.2019

Dtsch Med Wochenschr 2020; 145: e10–e17

© Georg Thieme Verlag KG, Stuttgart · New York

ISSN 0012-0472

### Korrespondenzadresse

Marc Aurelio Beckers

Klinikum der Universität München

Medizinische Klinik und Poliklinik IV, Ziemssenstraße 1, 80336 München

Marc.Beckers@campus.lmu.de

### ZUSAMMENFASSUNG

**Einleitung** Die Facharztprüfung berechtigt zur selbstständigen Berufsausübung und ist von besonderer Bedeutung für die Qualität der ärztlichen Versorgung. Sie soll die Inhalte der Weiterbildungsordnung abdecken. Über die tatsächlich abgefragten Inhalte der Prüfung ist wenig bekannt. In dieser Studie soll die Frage beantwortet werden, welche inhaltlichen und strukturellen Charakteristika Facharztprüfungen der Inneren Medizin in Deutschland haben.

**Methoden** 100 Gedächtnisprotokolle von bundesweiten Facharztprüfungen aus den Jahren 2013–2016 wurden hinsichtlich ihrer inhaltlichen Schwerpunkte sowie der Frage-typen quantitativ erfasst und deskriptiv ausgewertet. Die

Ergebnisse wurden mit den Blueprints der Facharztprüfungen der Schweiz sowie der USA verglichen.

**Ergebnisse** In jeder Prüfung werden durchschnittlich 27 (SD = 10) Fragen gestellt; diese lassen sich in 3 Kategorien einordnen: (1) themenbezogene Einzelfragen, (2) fallbezogene Fragen, (3) auf diagnostische Verfahren ausgerichtete Fragen anhand von Medien. Die am häufigsten abgefragten internistischen Themenschwerpunkte sind Kardiologie, Gastroenterologie (jeweils 17%) und Endokrinologie (11%). Bei 50% der Fragen musste Wissen reproduziert, bei den anderen 50% Konzepte oder Prozeduren verstanden und angewendet werden. Beim Vergleich mit amerikanischen und schweizerischen Blueprints wurde eine ähnliche prozentuale Aufteilung der Fragen zu Themengebieten gefunden. In Bezug auf den amerikanischen Blueprint fällt auf, dass mehr Fragen aus anderen Facharztbereichen gestellt werden als in Deutschland. Beim schweizerischen Blueprint wird ein großes Spektrum an fächerübergreifenden Aspekten wie Ethik, Prävention und Wirtschaft abgefragt, die in der deutschen Facharztprüfung kein Prüfungsgegenstand sind.

**Diskussion** In der mündlichen Facharztprüfung Innere Medizin in Deutschland werden ebenso viele Themengebiete wie in ausländischen Facharztprüfungen abgeprüft. Allerdings ist die Varianz zwischen den einzelnen Prüfungen relativ groß. Eine Standardisierung der Prüfung wäre wichtig, um gleiche Prüfungsbedingungen für alle Kandidaten zu schaffen.

### ABSTRACT

**Background** The specialist examination entitles to independent professional conduct and is therefore of great significance for the quality of medical care. It should cover the contents of the continuing education regulations. So far, little is known about the actual content of the exam. In this study, the question was, which content and which structural characteristics do specialist examinations in internal medicine in Germany contain.

**Methods** 100 randomly selected protocols from nationwide specialist examinations from the years 2013–2016 were quantitatively recorded and descriptively evaluated with regards to their main content as well as the various types of questions. The results were compared with the blueprints of the medical examinations in Switzerland and the USA.

**Results** In each exam, an average of 27 (SD = 10) questions are asked. The questions can be categorized into three categories: (1) subject-specific questions, (2) case-related questions, and (3) diagnostic-oriented questions with visual material. Cardiology and Gastroenterology, each with 17 %, and Endocrinology with 11 % are the most frequently requested internal medical topics. For 50 % of the questions, the examinee must reproduce knowledge, while for the other 50 %, concepts and procedures must be understood and used. In comparison with the American and Swiss blueprints, a similar percentual distribution of question contents was found. With regards to the American blueprint, it is noticeable that there

are more questions from other specialist areas such as Urology, Neurology and Psychiatry in comparison to Germany. The Swiss blueprint covers a wide range of interdisciplinary aspects such as ethics, prevention and economics which are not subject to examination in the German specialist examination.

**Conclusion** In the oral specialist examination in internal medicine in Germany as many topics as in foreign specialist examinations are examined. However, the variance between the individual exams is relatively large. A standardization of the exam is important to create equal exam conditions for all candidates.

---

## Einleitung

Trotz der differenzierten Weiterbildungsordnung in der Inneren Medizin ist die mündliche Facharztprüfung Innere Medizin in ihrer inhaltlichen und strukturellen Ausrichtung bisher nicht genau reglementiert. Für Prüflinge ist es allerdings wichtig zu wissen, welche Inhalte in welcher Form geprüft werden. Auf der anderen Seite sind gleichwertige Prüfungsbedingungen und eine hohe Testqualität wichtige Faktoren, um nach der Prüfung eine qualitativ homogene und hochwertige Patientenversorgung zu gewährleisten.

Prüfungen in der medizinischen Aus- und Weiterbildung sind von besonderer Relevanz, da sie ein Mittel zur Qualitätssicherung der Patientenversorgung darstellen [1]: Ziel sollte sein, dass nur solche Ärztinnen/Ärzte, welche über eine ausreichende Wissensgrundlage und über alle notwendigen Kompetenzen verfügen, in der Patientenversorgung eigenverantwortlich tätig sind. Verschiedene Prüfungsformate kommen zum Einsatz, wie zum Beispiel die schriftliche, die mündliche oder die praktische Prüfung [2]. Diese Prüfungsformate haben alle ihre Vor- und Nachteile und können verschiedene Kompetenzstufen abprüfen. Im Nationalen Kompetenzbasierten Lernzielkatalog Medizin (NKLM) [3] werden verschiedene Kompetenzebenen in Anlehnung an etablierte Kompetenzmodelle unterschieden: Faktenwissen, Handlungs- und Begründungswissen sowie Handlungskompetenz. Differenzierter ist das ULME-Modell (Klassifikationsmatrix: Untersuchungen der Leistung, Motivation und Einstellungen) [4], welches auf die Bloomsche Taxonomie zurückgeht und neben den 3 Wissensebenen (Fakten, Konzepte und Prozeduren) zusätzlich noch Leistungskategorien (Reproduzieren, Verstehen/Anwenden, Kritisieren/Reflektieren) definiert. Während schriftliche Prüfungen insbesondere zur Reproduktion von Faktenwissen geeignet sind, kann das Reflektieren und Verstehen von Handlungs- und Begründungswissen vor allem in mündlichen Prüfungen abgefragt werden. Die Domäne der praktischen Prüfungen ist die Testung der Handlungskompetenz [5].

Von besonderer Bedeutung für die Qualitätssicherung der ärztlichen Praxis ist die Facharztprüfung, da sie die letzte Prüfung der ärztlichen Weiterbildung ist und bei Bestehen der Prüfung die Ärztin/der Arzt eigenverantwortlich und ohne weitere Aufsicht arbeiten darf. Die Durchfallquoten in Abhängigkeit von der ange-

strebten Facharztbezeichnung werden beispielsweise in Bayern regelmäßig veröffentlicht [6]: In den Jahren 2014/15 legten 230 Ärztinnen/Ärzte die Prüfung zur Fachärztin/zum Facharzt für Innere Medizin ab, von denen lediglich 3 Prüflinge die Prüfung nicht bestanden (Durchfallquote 1,3 %). Bundesweit liegt die Durchfallquote bei 3–5 % (Daten für Bayern siehe [7]), während andere Facharztprüfungen (wie beispielsweise die europäische Facharztprüfung für Notfallmedizin) erheblich höhere Durchfallquoten (bis zu 64 %) aufweisen [8].

Die strukturellen und formalen Rahmenbedingungen zum Erreichen des Facharztstitels sind vorgegeben [9]. So sind die bis zur Facharztprüfung in der Inneren Medizin erworbenen Kompetenzen in der Weiterbildungsordnung (WBO) des jeweiligen Bundeslandes festgelegt. Am Ende der Weiterbildungszeit von 5 (Fachärztin/Facharzt für Innere Medizin) oder 6 Jahren (Internistin/Internist mit Schwerpunktbezeichnung) findet eine mündliche Prüfung statt. Zudem müssen die Weiterbildungsbefugten in Zeugnissen und einem Logbuch, in dem eine Mindestanzahl durchgeführter Prozeduren angegeben sein muss, nach § 9 der Weiterbildungsordnung bescheinigen, dass die Ärztin/der Arzt uneingeschränkt zur Berufsausübung geeignet ist. Die Handlungskompetenz wird im deutschen Facharztsystem durch die Weiterbildungszeit, die im Logbuch dokumentierten durchgeführten Prozeduren und das abschließende Zeugnis der Weiterbildenden abgebildet.

Jeder Prüfling wird einzeln über einen Zeitraum von mindestens 30 Minuten geprüft. Die erworbenen Kenntnisse, Erfahrungen und Fertigkeiten in Bezug auf die vorgeschriebenen Weiterbildungsinhalte werden vom Prüfungsausschuss überprüft (§ 14 Abs. 2, MWBO 2003). Es sind jeweils 3 Prüferinnen/Prüfer, von denen 1 die vorsitzende Protokollführerin/Protokollführer ist, anwesend. Mindestens 2 der Prüfenden müssen die geprüfte Facharzt- bzw. Schwerpunktbezeichnung tragen. Es wird nur über das Bestehen bzw. Nichtbestehen des Prüflings entschieden, der Prüfling erhält keine Note. Der Facharzttitel kann bei bestandener Prüfung ein Leben lang geführt werden.

Über die tatsächlichen Abläufe und abgefragten Inhalte der mündlichen Facharztprüfung Innere Medizin ist wenig bekannt. Einen genauen Blueprint gibt es in Deutschland bisher nicht. In anderen Ländern, wie beispielsweise den USA [10] oder der Schweiz [11], sind die Aufteilung der Fragen nach Fachgebiet ge-

nau festgelegt. Von dieser Aufteilung darf in der tatsächlichen Prüfung – die allerdings in schriftlicher Form stattfindet – nicht abgewichen werden [10, 11]. Hierbei gelten nach Bestehen der schriftlichen Prüfung für die verschiedenen schweizerischen Facharztspezialisierungen jeweils andere Prüfungsmodalitäten mit abweichenden Prüfungsmethoden, unter anderem auch zusätzlich mündliche Prüfungen [12].

Faire Prüfungen müssen generell strukturell-formale und inhaltliche Qualitätskriterien erfüllen, um objektiv [13], reliabel [14] und valide zu sein. Die Facharztprüfung sollte inhaltlich eine repräsentative Stichprobe der die fachärztliche Berufsausübung betreffenden Themengebiete umfassen und innerhalb der Themengebiete die verschiedenen Kompetenzebenen abprüfen.

## Offene Frage und Hypothesen

Vor diesem Hintergrund soll die folgende Forschungsfrage beantwortet werden: Welche inhaltlichen und strukturellen Charakteristika haben Facharztprüfungen Innere Medizin in Deutschland? Wir stellten dazu folgende Hypothesen auf:

1. Die Weiterbildungsinhalte werden durch Abfrage vieler Wissensstichproben aus unterschiedlichen Schwerpunktfächern adäquat abgedeckt.
2. Von der WBO geforderte Untersuchungs- und Behandlungsverfahren werden in Form verschiedener Medien angemessen breit geprüft.
3. Die Prüfungsfragen beziehen sich vor allem auf Verständnis und kritisches Reflektieren klinischer Zusammenhänge.
4. Das abgefragte Spektrum inhaltlicher Themengebiete ist in seiner Verteilung dem amerikanischen und schweizerischen „Blueprint“ ähnlich.

## Methoden

### Material

Grundlage der Arbeit sind Gedächtnisprotokolle von Prüflingen, welche an dem 1-wöchigen Intensivkurs Innere Medizin – Vorbereitung auf die Facharztprüfung [15] teilgenommen haben. Der Kurs wird seit 1994 unter der Schirmherrschaft und Qualitätskontrolle der Deutschen Gesellschaft für Innere Medizin (DEGIM) und in Zusammenarbeit mit der Bayerischen Landesärztekammer (BLÄK) durchgeführt. Die Sammlung enthält 196 Prüfungsprotokolle aus den Jahren 2013 bis 2017. Ab einer Stichprobengröße von 50–60 kann von einem Sättigungseffekt (d. h. ab diesem Punkt führt eine weitere qualitative Analyse zu keinem substantiellen weiteren Informationsgewinn) ausgegangen werden [16]. Es wurden 100 Protokolle randomisiert ausgewählt und ausgewertet. Die 100 Protokolle setzen sich aus jeweils 25 Protokollen der Jahre 2013 bis 2016 zusammen, die für jedes Jahr mittels einer zufälligen Sequenzgenerierung ausgewählt wurden [17]. Diese Protokolle umfassten durchschnittlich 720 Wörter (SD = 343).

Die Prüflinge wurden aufgefordert, folgende Angaben zu machen: Datum und Ort der Prüfung, Wohnort, Spezialisierung der Prüfer, Prüfung nach neuer oder alter Weiterbildungsordnung. Sie sollten die Prüfung so ausführlich wie möglich – auch

► **Tab. 1** Beispiele für die verschiedenen Fragetypen.

Fragentyp	Beispiel
themenbezogene Einzelfragen	Welche Erreger für eine interstitielle Pneumonie sind Ihnen bekannt?
fallbezogenen Fragen	Es kommt ein 38-jähriger Patient über die Rettungsstelle, beklagt seit Monaten Abgeschlagenheit, seit Wochen wiederholt Fieber, geringer Gewichtsverlust. Wie gehen Sie vor?
Fragen mit Medien	Hier sehen Sie einen EKG-Streifen von einem jungen Mann (20 Jahre) von 1999. Was sehen Sie?

unter Nennung der eigenen Antworten – wiedergeben. Um möglichst ausführliche Protokolle zu generieren, wurde jedes vollständige Protokoll einer Verlosung zugeführt, bei der ein Geldpreis (in Höhe der Kursgebühr) gewonnen werden konnte.

### Datenauswertung

Zur Erstellung eines Kodiersystems wurden zunächst 25 % der Daten explorativ ausgewertet. Die Inhaltsanalyse erfolgte nach Mayring [18]. Daraufhin wurde ein Kodiersystem entwickelt und auf den gesamten Datensatz angewandt. Im ersten Schritt wurden alle Textpassagen deskriptiv in die folgenden beiden Kategorien eingeordnet: Fragentyp und Themengebiet. Bei inhaltlicher Doppelung einer Frage (z. B. Kardiologie und Notfallmedizin) wurde diese für beide Kategorien codiert. Bei Fragetypen wurden 3 Kategorien unterschieden: fallbezogene Fragen (Beschreibung einer klinischen Situation mit anschließender Frage), themenbezogene Einzelfragen und auf diagnostische Verfahren ausgerichtete Fragen mit Medien (z. B. mit einem Röntgenbild) (► **Tab. 1**).

Die themenbezogenen Einzelfragen wurden nach der Zuordnung zu einem Themengebiet zusätzlich einer Kompetenzstufe zugeordnet. Zur Einordnung der Kompetenzstufen wurde auf das ULME-Kompetenzstufenmodell zurückgegriffen, das sich an der von Anderson und Kathwohl [4] modifizierten Bloomschen Taxonomie [19] orientiert. Dieses Modell ist ein 3 × 3-Modell, welches die folgenden Wissensdimensionen umfasst: Fakten, Konzepte und Prozeduren. Diesen Wissensdomänen können jeweils 1 der 3 Leistungskategorien zugeordnet werden: Reproduzieren, Anwenden/Verstehen und Kritisieren/Reflektieren (siehe ► **Tab. 2**).

### Software und Statistik

Die Daten wurden mit der Analysesoftware MAXQDA [20] ausgewertet. Die Berechnung von Prozentsätzen sowie die Diagrammerstellung erfolgte mit Excel 2011 für Mac (Version 14.7.3). Mittelwerte und Standardabweichung wurden mit SPSS 24 berechnet. Zur Überprüfung der Reliabilität wurden 12 % der Daten doppelt codiert. Die Intercoder-Reliabilität berechnet mit Krippendorffs alpha beträgt  $\kappa = 0,89$ .

► **Tab. 2** Charakterisierung der Wissenskategorie nach dem ULME-Modell [4].

Wissens-kategorie	Definition
Fakten	Aussagen über Bezeichnungen, Definitionen, Normen
Konzepte	Sacherklärungen, Schlussfolgerungen, Aussagen zu Prozessen und Zusammenhängen zwischen empirischen Phänomenen, Aussagen zu Ursache-Wirkung-Beziehungen
Prozeduren	Bewältigung berufstypischer Handlungssituationen
Leistungskategorie	
Reproduzieren	Wiedergeben von erlerntem Wissen
Verstehen/Anwenden	erlerntes Schema auf neue Situation anwenden
Kritisieren/Reflektieren	Grenzen von Schemata bewusst reflektieren

## Ergebnisse

### Rahmenbedingungen der Prüfung

Es wurden Prüfungsprotokolle aus 11 verschiedenen Bundesländern ausgewertet. In 54 % der Fälle wurde Bayern als Prüfungsort angegeben. Am zweithäufigsten erfolgte die Prüfung in Nordrhein-Westfalen mit 16 %. Die anderen Länder waren mit 1–6 % vertreten.

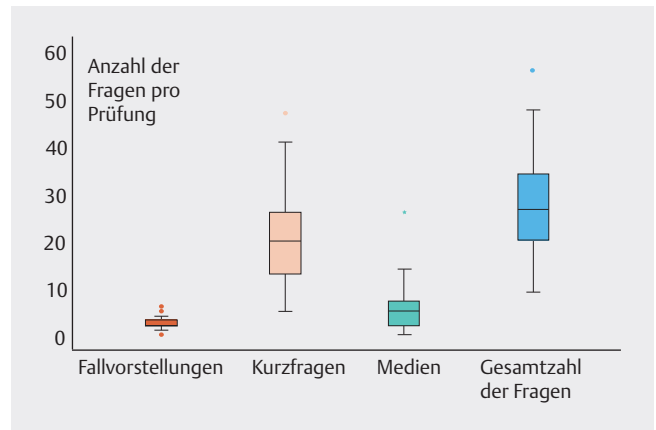
79 % der Prüflinge legten die Prüfung zur Fachärztin/zum Facharzt für Innere Medizin ab, in 17 % der Protokolle wurde keine nähere Bezeichnung angegeben. 4 % der Prüflinge legten die Prüfung zur Fachärztin/zum Facharzt für Innere Medizin mit einer Schwerpunktbezeichnung ab. 32 % der Prüferinnen/der Prüfer hatten die Schwerpunktbezeichnung Gastroenterologie und 17 % Kardiologie. In 21 % der Protokolle wurde die Spezialisierung der Prüferinnen/Prüfer nicht angegeben. Die restlichen 30 % verteilten sich auf die verschiedenen anderen Schwerpunkte.

### Anzahl der Fragen und Fragentypen pro Prüfung

Insgesamt wurden im Schnitt 27,48 (SD (Standardabweichung) = 10,15) Fragen pro Prüfung gestellt. Die Minimalanzahl an Fragen pro Prüfling lag bei 9, während das Maximum 56 war. Von den 3 verschiedenen Fragentypen wurden am häufigsten themenbezogene Einzelfragen gestellt (M = 20,06 (SD = 9,07), 73 %). Häufig wurden Fälle als Einstieg in die Prüfung genutzt (M = 2,35 (SD = 1,17), 8,5 %). Es wurden durchschnittlich 5,07 (SD = 4,05, 18,5 %) Fragen mit Medien gestellt (► **Abb. 1**).

### Themengebiete und Vergleich zum schweizerischen und amerikanischen Blueprint

Die am häufigsten abgefragten Teilgebiete der Inneren Medizin – sowohl bei Fallfragen als auch bei Einzelfragen – waren die Kardiologie und die Gastroenterologie mit 17 %. Auch in allen anderen Bereichen sind die Themengebiete in Fall- und Einzelfragen sehr ähnlich, abgesehen von der Notfallmedizin (siehe ► **Tab. 3**).



► **Abb. 1** Anzahl und Varianz der Fragen.

Hierzu wurden häufig Fälle abgeprüft (12,2 %), seltener Einzelfragen gestellt (4,9 %). Zum Teil überschneiden sich die Themengebiete, so z. B. sehr häufig Kardiologie und Notfallmedizin oder Gastroenterologie und Infektiologie. Von den 2659 Einzelfragen waren 1263 Überschneidungen festzustellen. Von den 320 Fallvorstellungen hingegen wurden 184 2 Themengebieten zugeordnet.

Beim Vergleich der Themenschwerpunkte mit dem Blueprint des American Board of Internal Medicine wird deutlich, dass die Verteilung der Schwerpunkte ähnlich ist (siehe ► **Tab. 3**). Während allerdings die Gastroenterologie in Deutschland mit 17 % schwerpunktmäßig abgefragt wird, sind nur 9 % der Fragen in der amerikanischen Facharztprüfung diesem Bereich zuzuordnen. Auffällig ist zudem, dass in den USA mehr Fragen aus anderen Facharztbereichen wie der Urologie, der Orthopädie oder auch der Neurologie und Psychiatrie gestellt werden [10]. Fragen aus anderen Facharzttrichtungen sind in der deutschen Facharztprüfung zum Internisten kaum bis gar nicht vorhanden.

Der schweizerische Blueprint ist ebenfalls ähnlich, allerdings werden auch hier weniger Fragen aus der Gastroenterologie gestellt. Zudem gibt es mit 14 % der Fragen ein großes Spektrum an allgemein- und fächerübergreifenden Fragen, die auch ethische, wirtschaftliche und präventive Aspekte umfassen. Diese Wissensbereiche werden in der deutschen Facharztprüfung gar nicht abgefragt, obwohl sie Gegenstand der Weiterbildungsordnung sind (siehe ► **Tab. 3**).

### Häufigkeit der unterschiedlichen Themengebiete

Insgesamt wurden im Schnitt 10,99 (SD = 3,04) Fragen aus verschiedenen Fachbereichen an jeden Prüfling gestellt. Somit stammen 40 % der Fragen aus verschiedenen Schwerpunktbereichen. Die Varianz ist insbesondere bei den Fragen mit Medien (M = 2,46 (SD = 1,41)) und bei den Fällen (M = 2,73 (SD = 1,45)) hoch. Von den durchschnittlich 20 Einzelfragen pro Prüfung stammen M = 5,80 (SD = 1,92) – also 30 % – aus verschiedenen Fachbereichen (► **Abb. 2**).

► **Tab. 3** Vergleich der Themenvarianz in Deutschland, der Schweiz und den USA [10].

Themengebiet	Deutschland Fälle	Deutschland Einzelfragen	USA	Schweiz
Gastroenterologie und Hepatologie	17,8%	17,8%	9% <sup>1</sup>	9%
Kardiologie	17,2%	17,3%	14%	11%
Endokrinologie und Stoffwechsel	10,9%	11,1%	9%	8%
Infektiologie	7,8%	8,6%	9%	9%
Allergie, Immunologie	4,7%	7,2%	2%	9% <sup>2</sup>
Pneumologie	5,6%	6,3%	9%	8%
Hämatologie	3,4%	6,0%	6%	9% <sup>3</sup>
Onkologie	4,7%	5,2%	6%	9% <sup>4</sup>
Notfallmedizin	12,2%	4,9%	0%	0%
Angiologie	3,8%	4,0%	0%	0%
Allgemeine Innere Medizin	3,1%	3,8%	0%	14% <sup>5</sup>
Intoxikationen/Alkoholismus	1,6%	2,3%	0%	0%
Nephrologie	2,2%	2,0%	6% <sup>6</sup>	8%
Rheumatologie, Knochenkrankheiten	1,3%	1,2%	9% <sup>7</sup>	9% <sup>8</sup>
Neurologie, Psychiatrie, Psychosomatik	0,9%	0,6%	4% <sup>9</sup>	7%
Hypertension	0,9%	0,6%	0%	0%
Ernährung	0,3%	0,5%	0%	0%
Geriatric	0,9%	0,4%	3%	0%
Intensivmedizin	0,6%	0,3%	0%	4%
Dermatologie	0%	0%	3%	4%
Gynäkologie und Geburtshilfe	0%	0%	3%	0%
Sonstiges	0%	0%	2%	0%
Augenheilkunde	0%	0%	1%	0%
HNO und Zahnheilkunde	0%	0%	1%	0%

<sup>1</sup> inklusive Angiologie.

<sup>2</sup> inklusive Rheumatologie.

<sup>3</sup> inklusive Onkologie.

<sup>4</sup> inklusive Hämatologie.

<sup>5</sup> inklusive Ethik, Statistik und Wirtschaftlichkeit.

<sup>6</sup> inklusive Urologie.

<sup>7</sup> inklusive Orthopädie.

<sup>8</sup> inklusive Immunologie

<sup>9</sup> je 4% Neurologie und Psychiatrie.

## Fragen mit Medien

Medien wurden insgesamt 508-mal eingesetzt. Mit Abstand am häufigsten wurden Sonografie-Bilder (50% aller Fragen mit Medien) eingesetzt. Auch Elektrokardiogramme (EKG) (1% dieser Fragen) und Röntgenbilder (8%) wurden abgeprüft. Andere Medien wurde nur selten eingesetzt (► **Abb. 3**).

## Kompetenzstufen

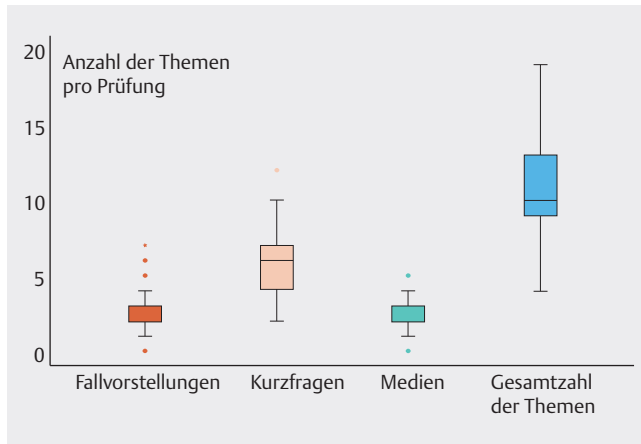
Nach der ULME-Klassifikation wurden die Einzelfragen den verschiedenen Wissens- und Leistungskategorien zugeordnet. Insgesamt bezogen sich 49% der Fragen (1011) auf die Kategorie

Reproduzieren, 50% (1029 Fragen) auf die Leistungskategorie Verstehen und Anwenden und 0,6% (13 Fragen) auf die Kategorie Kritisieren und Reflektieren. Die Aufteilung auf die verschiedenen Wissensdimensionen ist etwas gleichmäßiger: In 43% der Fragen (886) mussten Fakten entweder reproduziert, angewendet oder reflektiert werden; 21% der Fragen (422) bezogen sich auf Konzepte und 36% der Fragen (742) auf Prozeduren. Während die Kategorien Reproduzieren und Kritisieren/Reflektieren sich überwiegend auf Fakten und zu einem geringeren Anteil auf Prozeduren beziehen, ist die Leistungskategorie Verstehen und Anwenden insbesondere bei Konzepten und Prozeduren vorherrschend (siehe ► **Tab. 4**).

## Diskussion

### Wissensstichproben im internationalen Vergleich

Es wurden alle Schwerpunkte der Inneren Medizin in den Prüfungen abgefragt, wobei allerdings Kardiologie und Gastroenterologie deutlich überrepräsentiert waren. Allerdings ist dieses Resul-



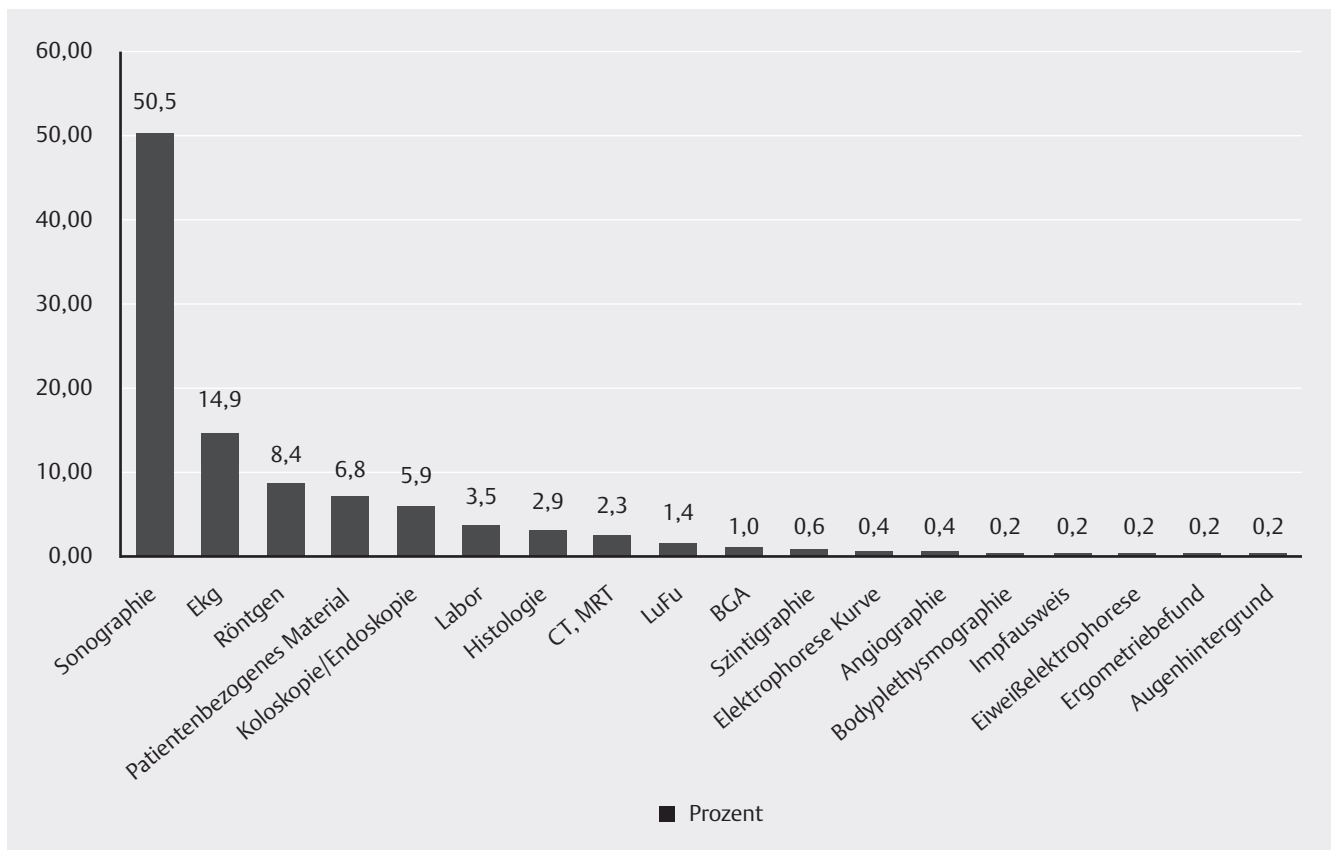
► **Abb. 2** Anzahl und Varianz der verschiedenen Themen pro Prüfung.

tat vor dem Hintergrund, dass auch die Prüferinnen/Prüfer schwerpunktmäßig Kardiologinnen/Kardiologen oder Gastroenterologinnen/Gastroenterologen waren, durchaus verständlich. Bei durchschnittlich 27 Fragen pro Prüfung kann davon ausgegangen werden, dass viele Wissensproben aus den unterschiedlichen Richtungen abgefragt wurden. Problematisch ist allerdings, dass die Anzahl der Fragen und die Varianz der Themengebiete stark zwischen den einzelnen Prüfungen variiert. Dies könnte natürlich methodisch bedingt sein, dass sich nicht alle Teilnehmerinnen/Teilnehmer in gleichem Maße an die Fragen erinnern konnten. Es könnte allerdings auch darauf hinweisen, dass aufgrund der geringen Standardisierung eine große interindividuelle Variabilität besteht.

Die deutsche, die amerikanische sowie die schweizerische Facharztprüfung sind sich – trotz unterschiedlichem Prüfungsformat – in der Verteilung der Themengebiete sehr ähnlich, obwohl die deutsche Facharztprüfung keiner inhaltlichen Reglementierung unterliegt. Diese Verteilung ist in der deutschen Prüfung allerdings nur im Durchschnitt so, einzelne Prüfungen weichen von diesem Blueprint erheblich ab.

### Medien

In jeder Prüfung werden im Schnitt 5 Fragen mit Medien gestellt, also ca. 25% der Fragen. Allerdings zeigt die Analyse, dass Sonografie-Bilder deutlich überrepräsentiert sind im Vergleich zu ande-



► **Abb. 3** Auf diagnostische Verfahren ausgerichtete Fragen mit Medien. EKG = Elektrokardiogramm; LuFu = Lungenfunktionsprüfung; BGA = Blutgasanalyse; CT = Computertomografie; MRT = Magnetresonanztomografie.



► **Tab. 4** Verknüpfung von Leistungskategorie und Wissensdimension (vgl. ► **Tab. 2**).

	n = 2053 Fragen	Wissensdimension			gesamt
		Fakten	Konzepte	Prozeduren	
Leistungskategorie	Reproduzieren	41 %	0 %	8 %	49 %
	Anwenden/Verstehen	2 %	20 %	28 %	50 %
	Kritisieren/Reflektieren	0 %	0 %	0 %	1 %
	gesamt	43 %	20 %	36 %	100 %

ren Medien. Auf andere für den Internisten wichtige Untersuchungsverfahren wird eher selten eingegangen.

Es wäre hilfreich, die Anzahl und Art der abgeprüften Medien zu standardisieren, damit auch andere wichtige technische Untersuchungen, wie die Lungenfunktionsprüfung oder das Röntgenbild, nicht zu kurz kommen. Es ist bereits bekannt, dass Assistenzärztinnen/-ärzte und auch Fachärztinnen/-ärzte zum Teil unzureichende Fähigkeiten hinsichtlich der Befundung von technischen Untersuchungen haben [21]. Deshalb ist es umso wichtiger, dass diese Kompetenz in der Facharztprüfung ausreichend abgeprüft wird.

## Kompetenzebenen

Aus der Analyse der Wissensdomänen und Kompetenzstufen ergibt sich, dass es sich bei etwa der Hälfte der Fragen um Konzepte und Prozeduren handelt, was dem Wesen einer mündlichen Prüfung entspricht. Kritische Auseinandersetzungen und Reflexionen finden praktisch gar nicht statt, sollten aber bei einem akademischen Beruf einen signifikanten Stellenwert einnehmen. 41 % aller Fragen beziehen sich auf das Reproduzieren von Faktenwissen. Das mündliche Prüfungsformat ist allerdings nicht die effizienteste und nicht die zuverlässigste Prüfungsform für das Abfragen von Faktenwissen, hierfür sind schriftliche Formate eindeutig geeigneter. Die Stärken der mündlichen Prüfung sind hingegen, dass insbesondere das tiefgreifende Verständnis für klinische Zusammenhänge abgefragt werden könnte.

Im CanMEDs Framework [22] werden unter der Bezeichnung „Medizinischer Experte“ die folgenden ärztlichen Rollen subsummiert: Kommunikator, Teamarbeiter, Manager, Gesundheitsforscher, Lehrer und Lerner, die/der professionelle Ärztin/Arzt [22]. Natürlich kann in einer mündlichen Prüfung nicht auf jeden Aspekt eingegangen werden; es ist allerdings auffällig, dass der Fokus der mündlichen Facharztprüfung vor allem auf dem reinen medizinischen Wissen liegt. Im klinischen Alltag, in der Patientenversorgung und insbesondere für den eigenverantwortlich handelnden Facharzt sind die anderen ärztlichen Rollen aber ebenfalls relevant. Die anderen Teilaspekte könnten in Form eines anderen Prüfungsformats abgeprüft werden.

## Limitationen

In dieser Studie wurden erstmals die Inhalte der Facharztprüfung Innere Medizin analysiert. Allerdings stammte die Hälfte der Protokolle aus bayrischen Prüfungen, deswegen können nur eingeschränkt allgemeine Schlüsse auf die bundesweiten Prüfungen

gezogen werden. Zudem spiegeln diese Protokolle nur die Wahrnehmung der Prüflinge wider; es wäre hilfreich, auch Protokolle der Prüferinnen/Prüfer oder Audioaufnahmen der Prüfungen zu analysieren. Es bleibt unklar, wie genau sich die Prüflinge an die Prüfung erinnern konnten und wie sehr dadurch ggf. die Analyse verzerrt wird.

## KERNAUSSAGEN

- Es wäre sinnvoll, durch inhaltliche Blueprints eine möglichst breite und vom Schwerpunkt der Prüferin/des Prüfers unabhängige Auswahl der Themenschwerpunkte zu unterstützen.
- Es wäre sinnvoll, durch strukturelle Blueprints bzgl. Fragetypen und Kompetenzstufen den Vorteil der organisatorisch aufwendigen mündlichen Einzelprüfungen konsequent auszuschöpfen: Prüfen, Konzepte/Prozeduren anzuwenden/zu reflektieren. Das Reproduzieren von Faktenwissen sollte durch effizientere und zuverlässigere Prüfungsformate wie Multiple-Choice-Questions erfolgen.
- Es wäre sinnvoll, die Prüferinnen/Prüfer, die ihre Zeit und Expertise für die nächste Medizinergeneration einbringen, bzgl. standardisierter mündlicher Prüfungen zu schulen, wie es im Hochschulbereich für Prüferinnen/Prüfer des mündlichen Staatsexamens schon etabliert ist.
- Da die Facharztprüfung als letzte institutionalisierte Kompetenzkontrolle in Deutschland einen besonderen Stellenwert besitzt, ist nicht nur das Erreichen einer optimalen Variabilität im Sinne gleichwertiger Prüfungsbedingungen von großer Bedeutung. Letztlich muss auch die Qualität der einzelnen Prüfungen den hohen Erwartungen, die an die ärztliche Kompetenz im realen Alltag gestellt werden und von der Landesärztekammer als Standesorganisation gefordert sind, gerecht werden. Die Prüfungsbedingungen sollten also nicht nur für alle Prüflinge möglichst gleich sein, sondern auch inhaltlich möglichst gut die geforderten Kompetenzen abprüfen.

## Interessenkonflikt

Alle Autoren geben an, dass kein Interessenkonflikt besteht.  
Ethik: Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

## Danksagung

---

Wir danken Katharina Borrmann für ihre Hilfe bei der Codierung der Daten.

## Literatur

---

- [1] Norcini J, Anderson B, Bollela V et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical teacher* 2011; 33: 206–214
- [2] Harden RM, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical education* 1979; 13: 39–54
- [3] Fischer MR, Bauer D, Karin Mohn N. Finally finished! national competence based catalogues of learning objectives for undergraduate medical education (NKLM) and dental education (NKLZ) ready for trial. *GMS Zeitschrift für medizinische Ausbildung* 2015; 32: 3–4
- [4] Anderson LW, Bloom BS. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's*. Pearson Education. 2014
- [5] Bloch R, Hofer D, Krebs R. *Handbuch ‚Kompetent prüfen‘*. Bern: Abteilung für Ausbildungs- und Examensforschung (AEE). 1999. Im Internet: <http://www.iawf.unibe.ch>; Stand: 02.02.2018
- [6] Bayerische Landesärztekammer (blaek). Tätigkeitsbericht der Bayerischen Landesärztekammer 2014/15 dem 74. Bayerischen Ärztetag vorgelegt (09.2015). Im Internet: [http://www.blaek.de/werwaswo/bay\\_aerztetag/berichte/TB\\_2014\\_2015\\_neu.pdf](http://www.blaek.de/werwaswo/bay_aerztetag/berichte/TB_2014_2015_neu.pdf); Stand: 04.05.2018
- [7] Bayerische Landesärztekammer (blaek). Tätigkeitsbericht der Bayerischen Landesärztekammer 2016/2017 (09.2017). Im Internet: [http://www.blaek.de/werwaswo/bay\\_aerztetag/berichte/TB\\_2016\\_17\\_X3\\_NEU.pdf](http://www.blaek.de/werwaswo/bay_aerztetag/berichte/TB_2016_17_X3_NEU.pdf); Stand: 26.07.2018
- [8] Dißmann P. Europäische Facharztprüfung für Notfallmedizin. *Notfall+ Rettungsmedizin* 2014; 17: 697–700
- [9] Bayerische Landesärztekammer (blaek). Richtlinien über den Inhalt der Weiterbildung (24.11.2018). Im Internet: <http://www.blaek.de>; Stand: 09.12.2018
- [10] American Board of Internal Medicine (aboim). Certification Examination Blueprint (Januar 2018). Im Internet: <http://www.abim.org>; Stand: 15.06.2018
- [11] Schweizerische Gesellschaft für Allgemeine Innere Medizin (SGAIM). Guidance for the Swiss Board Examination in General Internal Medicine (Dezember 2017). Im Internet: [http://www.sgaim.ch/fileadmin/user\\_upload/Facharztpruefung/2018\\_GuidanceFAPAIM\\_2018\\_definitv\\_sto.pdf](http://www.sgaim.ch/fileadmin/user_upload/Facharztpruefung/2018_GuidanceFAPAIM_2018_definitv_sto.pdf); Stand: 26.07.2018
- [12] Montagne S, Meng-Hentschel J, Huwendiek S et al. Prüfungen: Schweizer Perspektiven (Vorträge). Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bern; 14.09.2016–16.09.2016; 121
- [13] Vleuten C, Norman G, Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Medical education* 1991; 25: 110–118
- [14] Norcini JJ. Standards and reliability in evaluation: when rules of thumb don't apply. *Academic medicine: journal of the Association of American Medical Colleges* 1999; 74: 1088–1090
- [15] Deutsche Gesellschaft für Innere Medizin (DGIM). Intensivkurse Innere Medizin (2018). Im Internet: [http://www.klinikum.uni-muenchen.de/Medizinische-Klinik-und-Poliklinik-IV/de/Ueber-uns/download/Intensiv\\_Kurse\\_Innere\\_Medizin/IIM\\_2019/index.html](http://www.klinikum.uni-muenchen.de/Medizinische-Klinik-und-Poliklinik-IV/de/Ueber-uns/download/Intensiv_Kurse_Innere_Medizin/IIM_2019/index.html); Stand: 26.07.2018
- [16] Mason M. Sample size and saturation in PhD studies using qualitative interviews. In: *Forum qualitative Sozialforschung/Forum: qualitative social research* 2010
- [17] Haahr M. Random.org: True Random Number Service. Im Internet: <http://www.random.org>; Stand 09.02.2018
- [18] Mayring P. Combination and integration of qualitative and quantitative analysis. *Forum: Qualitative Social Research*. 2001; 2
- [19] Bloom BS Hrsg *Taxonomy of Educational Objectives: The Classification of Educational Goals: By a Committee of College and University Examiners*. New York: David McKay. 1971
- [20] VERBI Software Consult. Sozialforschung GmbH. MAXQDA, Software für qualitative Datenanalyse (1989–2019). Im Internet: <https://www.maxqda.de/faq/wie-zitiert-man-maxqda>; Stand: 02.07.2018
- [21] Christiansen JM, Gerke O, Karstoff J et al. Poor interpretation of chest X-rays by junior doctors. *Dan Med J* 2014; 61: A4875
- [22] Frank J, Snell L, Sherbino J. CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada 2015. Im Internet: <https://www.acgme.org>; Stand 05.06.2018