

Aus dem Institut für Didaktik und Ausbildungsforschung in der Medizin

Institut der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Martin Fischer

Multiple Choice Fragen in der ärztlichen Fortbildung:

**Analyse von Strukturqualität, Wissens Ebenen und Inhaltskategorien sowie der
Effekte von Qualitätsmaßnahmen in ausgewählten Fachzeitschriften**

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Sabine Drossard

aus

Regensburg

im Jahr 2021

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

Berichterstatter:	Prof. Dr. Martin R. Fischer
Mitberichterstatter:	Prof. Dr. Jörg Schelling Prof. Dr. Matthias Siebeck
Mitbetreuung durch die promovierten Mitarbeiter:	Dr. Lisa Kühne-Eversmann Dr. Daniel Bauer
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	10.11.2021

Wenn du eine weise Antwort verlangst, musst du vernünftig fragen.

- Johann Wolfgang von Goethe



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Promotionsbüro
Medizinische Fakultät



Eidesstattliche Versicherung

Drossard, Sabine

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Titel

Multiple Choice Fragen in der ärztlichen Fortbildung: Analyse von Strukturqualität, Wissens Ebenen und Inhaltskategorien sowie der Effekte von Qualitätsmaßnahmen in ausgewählten Fachzeitschriften

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

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

München, 10.11.2021

Ort, Datum

Sabine Drossard

Unterschrift Doktorandin bzw. Doktorand

Teile dieser Arbeit wurden vorab veröffentlicht:

[a] **Drossard S**, Kühne-Eversmann L, Bauer D, Fischer M. (2013) CME-Artikel in medizinischen Fachzeitschriften: Erfolg von Qualitätsmaßnahmen. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Graz, 26.-28.09.2013. Düsseldorf: German Medical Science GMS Publishing House. DocP05_01. DOI: 10.3205/13gma038

[b] **Drossard S**, Bauer D, Fischer M, Kühne-Eversmann L. (2014) Analyse von Wissensseben und Inhaltskategorien in CME-Aufgaben. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg, 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House. DocP471. DOI: 10.3205/14gma186

[c] **Drossard S**, Bauer D, Fischer M, Kühne-Eversmann L. (2017) Welchen Einfluss haben Autorenrichtlinien der Verlage auf die formale Qualität von CME-Fragen in der ärztlichen Fortbildung? Z. Evid.Fortbild. Qual. Gesundh.wesen (ZEFQ) 121:29–35. DOI: 10.1016/j.zefq.2017.03.015

Hinweise zum Anteil der Autor*innen an den oben genannten Veröffentlichungen:

[a]

Drossard S entwickelte die Fragestellung und die Konzeption der Studie unter Supervision von Fischer M. Drossard S und Kühne-Eversmann L führten die Studie durch und werteten die Daten aus. Bei Unklarheiten wurde die Auswertung zusätzlich von Bauer D kritisch geprüft.

Drossard S gestaltete das Poster und die darauf abgebildeten Grafiken. Das Poster wurde vor Veröffentlichung von Kühne-Eversmann L, Bauer D und Fischer M kritisch geprüft.

Drossard S stellte das Poster im Rahmen der GMA-Jahrestagung 2013 in Graz vor.

[b]

Drossard S entwickelte die Fragestellung und die Konzeption der Studie unter Supervision von Fischer M. Drossard S und Kühne-Eversmann L führten die Studie durch und werteten die Daten aus. Bei Unklarheiten wurde die Auswertung zusätzlich von Bauer D kritisch geprüft.

Drossard S gestaltete das Poster und die darauf abgebildeten Grafiken. Das Poster wurde vor Veröffentlichung von Kühne-Eversmann L, Bauer D und Fischer M kritisch geprüft.

Drossard S stellte das Poster im Rahmen der GMA-Jahrestagung 2014 in Hamburg vor.

[c]

Drossard S entwickelte die Fragestellung und die Konzeption der Studie unter Supervision von Fischer M. Drossard S und Kühne-Eversmann L führten die Studie durch und werteten die Daten aus. Bei Unklarheiten wurde die Auswertung zusätzlich von Bauer D kritisch geprüft.

Drossard S schrieb das Manuskript, welches von Kühne-Eversmann L, Bauer D und Fischer M revidiert wurde, die entscheidende Verbesserungsvorschläge beitrugen.

Inhaltsverzeichnis

1. ZIEL DER ARBEIT / FRAGESTELLUNGEN	8
2. EINLEITUNG.....	9
2.1 Ärztliche Fortbildung / Continuing Medical Education (CME) in Deutschland	9
2.1.1 Hintergrund und Gesetzeslage	9
2.1.2 Zahlen zur ärztlichen Fortbildung in Deutschland	11
2.1.3 Zielsetzung der ärztlichen Fortbildung.....	13
2.2 Formale Qualität von MC-Fragen.....	14
2.2.1 MC-Fragen in der Aus-, Weiter- und Fortbildung	14
2.2.2 Vorgaben für MC-Fragen in der ärztlichen Fortbildung.....	15
2.2.3 Formale Qualität von MC-Fragen und Qualitätssicherungsmaßnahmen.....	17
2.3 Inhaltliche Qualität von MC-Fragen in der ärztlichen Fortbildung.....	17
2.3.1 Praxisbezug und klinische Relevanz von MC-Fragen in der ärztlichen Fortbildung	17
2.3.2 Anspruch der Ärzt*innen an die Fortbildung	18
2.3.3 Faktenwissen und klinische Kompetenz.....	19
2.3.4 Taxonomien zur Einstufung in Wissensniveaus	19
3. MATERIAL UND METHODEN	24
3.1 Qualitative Analyse der MC-Fragen und Auswertung der Autorenhinweise	24
Kapitel 3.1 ist in leicht abgewandelter Form Teil der Publikation [c].	24
3.1.1 Auswahl der Artikel.....	24
3.1.2 Beurteilungskriterien	25
3.1.3 Auswertungsprozess.....	26
3.1.4 Longitudinaler Vergleich.....	26
3.1.5 Bewertung der Qualitätsmaßnahmen	26
3.1.6 Korrelation der formalen Qualität 2012 mit den Autorenhinweisen	27
3.2 Analyse der Wissensebenen und Auswertung der Themen	27
3.2.1 Auswahl der Fragen	27
3.2.2 Erarbeiten der Taxonomie der Wissensniveaus.....	28
4. ERGEBNISSE.....	31
4.1 Qualitative Analyse der MC-Fragen	31
Kapitel 4.1 ist in leicht abgewandelter Form Teil der Publikation [c].	31
4.1.2 Formale Qualität der MC-Fragen im longitudinalen Vergleich 2006-2012	32
4.1.3 Verteilung der Fehlerarten	33
4.2.1. Quantitative und qualitative Auswertung der Autorenhinweise.....	34
4.2.2 Korrelation der Autorenhinweise zur formalen Qualität der Fragen	36
Kapitel 4.2.2 ist in gekürzter Form Teil der Publikation [c].	36
4.3 Inhaltliche Analyse	37
4.3.1 Verteilung der Wissensebenen.....	37
4.3.2 Verteilung der Themen.....	38
4.3.3 Verwendung von Fallvignetten und Bezug zum Wissensniveau.....	39

5. DISKUSSION.....	40
6. ZUSAMMENFASSUNG	45
7. ANHANG.....	46
8. LITERATURVERZEICHNIS.....	52
9. DANKSAGUNG.....	57

1. Ziel der Arbeit / Fragestellungen

Die ärztliche Fortbildung mithilfe von Artikeln und Multiple-Choice (MC) Fragen in Zeitschriften ist weit verbreitet und vielgenutzt. Mehrere Publikationen haben sich bereits mit der Analyse der formalen und teils auch der inhaltlich-didaktischen Qualität der verwendeten Fragen befasst. Seitens der Verlage wurden zunehmend Qualitätssicherungsmaßnahmen eingeführt, darunter die Veröffentlichung von Autorenhinweisen zur Erstellung guter MC-Fragen.

Die vorliegende Studie beschäftigt sich einerseits mit der Frage, wie sich die formale Qualität von Multiple-Choice-Fragen in CME-Angeboten in ausgewählten deutschen medizinischen Fachzeitschriften darstellt und wie sie sich in den vergangenen Jahren entwickelt hat. Außerdem möchte sie mögliche Gründe für die Verbesserung der formalen Qualität aufzeigen, indem die Qualitätsmaßnahmen der Verlage untersucht und den Ergebnissen der Analyse gegenübergestellt werden.

Das Ziel der ärztlichen Fortbildung wird durch die Bundesärztekammer in der (Muster-) Fortbildungsordnung als die Verbesserung der Patientenversorgung sowie die Sicherstellung der Qualität der ärztlichen Versorgung definiert¹.

Auch die Ärzt*innen wünschen sich Praxisbezug und als Resultat der Fortbildung die Weiterentwicklung ihrer klinischen Fähigkeiten. Inwiefern ärztliche Fortbildung mittels Online- und Printmedien diesen Praxisbezug herstellt, ist bisher wenig untersucht worden.

Um diese Fragestellung aufzugreifen, wurden deshalb in der hier vorliegenden Arbeit im zweiten Schritt die untersuchten MC-Fragen auf ihren Inhalt untersucht, wobei zum einen die abgeprüfte Thematik und zum anderen die Art des abgefragten Wissens analysiert wurde. Insbesondere soll gezeigt werden, ob die MC-Fragen ein Niveau erreichen, das klinischen Anwendungsbezug zulässt und über das reine Abfragen von Faktenwissen hinausgeht. Hierzu wurde auch der Anteil der fallbasierten MC-Fragen (Fragen mit Fallvignette) ausgewertet, da insbesondere diese auch höhere Wissensniveaus abfragen können.

2. Einleitung

2.1 Ärztliche Fortbildung / Continuing Medical Education (CME) in Deutschland

2.1.1 Hintergrund und Gesetzeslage

In der Medizin gibt es einen ständigen Wandel sowie einen rasanten Anstieg des vorhandenen Wissens. Gleichzeitig wird Evidenz in der Medizin immer wichtiger. Aus diesen beiden Aspekten ergibt sich die Notwendigkeit zur laufenden Fortbildung von Ärzt*innen, damit ihr Wissen stets am aktuellen Erkenntnisstand entspricht. Die Halbwertszeit medizinischen Wissens wird durch die große Zahl an wissenschaftlichen Erkenntnissen und neuen Behandlungsmöglichkeiten immer kürzer. Schätzungen gehen davon aus, dass das medizinische Wissen sich etwa alle sechs bis acht Jahre verdoppelt². Die Halbwertszeit des Wissens wird von verschiedenen Autor*innen unterschiedlich bewertet und beispielsweise auf sieben Jahre geschätzt³. Eine retrospektive Analyse chirurgischer Publikationen von 1935 bis 1995 ergab eine Abnahme des Wahrheitsgehaltes chirurgischer Lehrmeinungen um 0,75 % pro Jahr und somit eine Halbwertszeit chirurgischen Wissens von 45 Jahren⁴. Studien zeigen, dass das prüfbare Wissen von Ärzt*innen im Laufe der Zeit abnimmt und sie bei Wissenstesten nach einigen Jahren der Berufsausübung signifikant schlechter abschneiden als Kolleg*innen, die zu einem früheren Zeitpunkt geprüft werden⁵.

Somit kann der Wissenserwerb von Ärzt*innen nicht mit dem Ende des Studiums oder der Weiterbildung abgeschlossen sein. Vielmehr muss dieses Wissen im Sinne des lebenslangen Lernens kontinuierlich aktualisiert und ergänzt werden, um eine hohe Versorgungsqualität sicher zu stellen und Patientensicherheit zu gewährleisten⁶. Das lebenslange Lernen wird in der Medizin als Continuing Medical Education (CME) oder auch als Continuing Professional Development (CPD) bezeichnet.

Die Ärztliche Fortbildung ist gemäß §4 der (Muster-)Berufsordnung für alle Ärzt*innen verpflichtend und ihr Nachweis seit dem 01. Januar 2004 für Vertragsärzte der Krankenkassen und Fachärzte im Krankenhaus auch bundeseinheitlich gesetzlich geregelt (§ 95d und § 136b SGB V).

Vor dem Jahr 2003 gab es eine Selbstverpflichtung der Ärzt*innen zur Fortbildung, die in der Berufsordnung festgeschrieben war. 1995 wurde das freiwillige Fortbildungszertifikat der Ärztekammern eingeführt. Im Jahr 2003 wurde schließlich im Rahmen des GKV-Modernisierungsgesetzes die ärztliche Fortbildung sowie ihr Nachweis gegenüber den Krankenkassen für KV-Ärzt*innen und Fachärzt*innen im Krankenhaus gesetzlich verpflichtend eingeführt. In §95 SGB V ist festgehalten, dass Vertragsärzt*innen der KV sich fortbilden müssen, in § 136b SGB V wird die Fortbildungspflicht auch auf Fachärzt*innen im Krankenhaus ausgeweitet.

In der aktuellen Fassung der ärztlichen Berufsordnung aus dem Jahr 2018 wird die Fortbildungsverpflichtung verallgemeinernd beschrieben: „Ärztinnen und Ärzte, die ihren Beruf ausüben, sind verpflichtet, sich in dem Umfange beruflich fortzubilden, wie es zur Erhaltung und Entwicklung der zu ihrer Berufsausübung erforderlichen Fachkenntnisse notwendig ist.“⁷

Grundlage ist die (Muster)Fortbildungsordnung der Bundesärztekammer (BÄK), die zuletzt im Mai 2013 überarbeitet und aktualisiert wurde¹. Die genaue Ausgestaltung der Zertifizierung von Fortbildungsmaßnahmen und Anerkennung von Fortbildungspunkten obliegt den jeweiligen Landesärztekammern auf Basis der Landesheilberufegesetze. Die Fortbildungsveranstaltungen werden hierbei nur zu einem Teil von den Landesärztekammern selbst veranstaltet - der größte Teil wird von externen Veranstaltern (z.B. Kliniken, Kongressorganisations, Zeitschriften und Verlage, private Fortbildungsinstitute) angeboten, die diese Fortbildungsveranstaltungen vorab bei der jeweils zuständigen Landesärztekammer anmelden und die Anerkennung von CME-Punkten beantragen.

Es stehen verschiedene Fortbildungsformate zur Verfügung, die von Selbststudium über Kongresse und Vorträge bis hin zu Hospitationen reichen (Tabelle 1). Ein Punktesystem bildet den Umfang der Teilnahme ab. In der (Muster-) Fortbildungsordnung der Bundesärztekammer sowie in den Fortbildungsordnungen der Landesärztekammern wird explizit darauf hingewiesen, dass Ärzt*innen in der Wahl der

Fortbildungsmethoden frei sind und der Wissenserwerb auf das individuelle Lernverhalten auszurichten ist¹.

Tabelle 1: Fortbildungskategorien. Aus: (Muster-) Fortbildungsordnung der Bundesärztekammer, in der Fassung vom 29.05.2013¹.

Kategorie	Beschreibung	Punktvergabe
<u>Kategorie A</u>	Vortrag und Diskussion	1 Punkt pro 45-minütiger Fortbildungseinheit 1 Zusatzpunkt bei dokumentierter Lernerfolgskontrolle pro Fortbildungsmaßnahme
<u>Kategorie B</u>	Mehrtägige Kongresse im In- und Ausland, welche nicht von anderen Kategorien erfasst werden	3 Punkte pro 1/2 Tag bzw. 6 Punkte pro Tag
<u>Kategorie C</u>	Fortbildung mit konzeptionell vorgesehener Beteiligung jedes einzelnen Teilnehmers (z. B. Workshop, Arbeitsgruppen, Qualitätszirkel, Peer Review, Balintgruppen, Kleingruppenarbeit, Supervision, Fallkonferenzen, Literaturkonferenzen, praktische Übungen)	1 Punkt pro 45-minütiger Fortbildungseinheit 1 Zusatzpunkt pro Maßnahme bis zu 4 Stunden/höchstens 2 Zusatzpunkte pro Tag 1 weiterer Zusatzpunkt bei dokumentierter Lernerfolgskontrolle pro Fortbildungsmaßnahme
<u>Kategorie D</u>	Fortbildungsbeiträge in Printmedien oder als elektronisch verfügbare Version mit nachgewiesener Qualifizierung durch eine Lernerfolgskontrolle in digitaler bzw. schriftlicher Form	1 Punkt pro 45-minütiger Fortbildungseinheit bei bestandener Lernerfolgskontrolle
<u>Kategorie E</u>	Selbststudium durch Fachliteratur und -bücher sowie Lehrmittel	Innerhalb dieser Kategorie werden höchstens 50 Punkte für fünf Jahre anerkannt.
<u>Kategorie F</u>	Wissenschaftliche Veröffentlichungen und Vorträge	Autorentätigkeit: 5 Punkte pro wissenschaftliche Veröffentlichung Referententätigkeit/Qualitätszirkelmoderation/wissenschaftliche Leitung: 1 Punkt pro Beitrag, z. B. Poster/Vortrag, unbenommen der Punkte für die persönliche Teilnahme Die maximale Punktzahl in dieser Kategorie beträgt 50 Punkte in fünf Jahren.
<u>Kategorie G</u>	Hospitationen	1 Punkt pro Stunde, höchstens 8 Punkte pro Tag
<u>Kategorie H</u>	Curricular vermittelte Inhalte, z. B. in Form von curricularen Fortbildungsmaßnahmen, Inhalte von Weiterbildungskursen, die nach der Weiterbildungsordnung für eine Weiterbildungsbezeichnung vorgeschrieben sind, Inhalte von Zusatzstudiengängen	1 Punkt pro 45-minütiger Fortbildungseinheit

<u>Kategorie I</u>	Tutoriell unterstützte Online-Fortbildungsmaßnahme mit nachgewiesener Qualifizierung durch eine Lernerfolgskontrolle in digitaler bzw. schriftlicher Form	1 Punkt pro 45-minütiger Fortbildungseinheit 1 Zusatzpunkt pro 45-minütiger Fortbildungseinheit bei vollständiger Erfüllung der qualitätssteigernden Kriterien eLearning der Bundesärztekammer
<u>Kategorie K</u>	Blended-Learning-Fortbildungsmaßnahme in Form einer inhaltlich und didaktisch miteinander verzahnten Kombination aus tutoriell unterstützten Online-Lernmodulen und Präsenzveranstaltungen	1 Punkt pro 45-minütiger Fortbildungseinheit 1 Zusatzpunkt pro 45-minütiger eLearning-Fortbildungseinheit bei vollständiger Erfüllung der qualitätssteigernden Kriterien eLearning der Bundesärztekammer

Jede Landesärztekammer entwickelt eine jeweils gültige eigene Fortbildungsordnung. Diese gleichen der (Muster-)Fortbildungsordnung der BÄK stark. Lediglich die Landesärztekammer Mecklenburg-Vorpommern weicht in der Definition der Fortbildungskategorien von den oben beschriebenen zehn Kategorien ab und benennt jeweils nur vier Untergruppen, wobei Online- und Printmedien in diesem Fall als Kategorie B bezeichnet werden.

Über einen Zeitraum von fünf Jahren müssen durch jede* Ärzt*in mindestens 250 CME-Punkte nachgewiesen werden. Bei im Krankenhaus angestellten Fachärzt*innen gilt zusätzlich, dass davon 150 Punkte fachspezifisch erworben werden müssen.

Als Sanktionen für das Nichterreichen der vorgegebenen Fortbildungspunkte drohen Honorarkürzungen für die entsprechenden Ärzt*innen bzw. für die betreffenden Krankenhäuser sowie im Extremfall der Entzug der Zulassung.

Ursprünglich gab es für das freiwillige Fortbildungszertifikat eine Beschränkung bezüglich der anzurechnenden Zahl der Fortbildungspunkte in Kategorie D. Die Kategorie E (Selbststudium) ist bei allen Landesärztekammern auf eine Höchstpunktzahl von 50 in 5 Jahren beschränkt, wobei hier kein Nachweis erforderlich ist und diese Punkte pauschal geltend gemacht werden können. Bezüglich der Fortbildung mittels Online- und Printmedien (Kategorie D) gibt es nur bei der Ärztekammer Berlin eine Einschränkung auf höchstens 150 Punkte in 5 Jahren, in der Summe für die beiden Kategorien D (Online- und Printmedien) und I (Online-Fortbildungsveranstaltungen). Das bedeutet, dass bis zu 250 von 250 benötigten CME-Punkten durch die Bearbeitung von Fachartikeln und MC-Fragen erreicht werden können.

Die Befürchtung, dass Ärzt*innen ihre Fortbildungspunkte ausschließlich durch Online- und Printmedien erzielen, hat sich nicht bestätigt. Die Ärztekammer Nordrhein gab 2007 an, dass die Teilnehmer*innen das Angebot themenzentriert und selektiv nutzen⁸.

Tabelle 19 im Anhang zeigt eine Übersicht der von den jeweiligen Landesärztekammern festgelegten Einschränkungen bezüglich des Punkteerwerbs in den unterschiedlichen Kategorien.

2.1.2 Zahlen zur ärztlichen Fortbildung in Deutschland

Im Jahr 2015 gab es laut Ärztestatistik der Bundesärztekammer insgesamt 371.300 berufstätige Ärzt*innen in Deutschland. Davon 150.100 in der ambulanten Versorgung (120.700 als niedergelassene Ärzt*innen, 29.400 als Angestellte) und 189.600 in der stationären Versorgung. Im Jahr 2019 waren der Bundesärztekammer insgesamt 402.119 Ärzt*innen gemeldet, davon waren 159.846 ambulant tätig⁹.

Zur ärztlichen Fortbildung liegen Daten der Bundesärztekammer für die Jahre 2006 - 2012 vor¹⁰. Aus diesen ist ersichtlich, dass pro Jahr weit über 300.000 Fortbildungsveranstaltungen anerkannt, über 5 Millionen Teilnahmen an Fortbildungsveranstaltungen verzeichnet und gut 20 Millionen CME-Punkte gesammelt werden.

Seit der Einführung der Fortbildungspflicht im Jahr 2006 stieg das Gesamtvolumen der erbrachten Fortbildungspunkte bis zum Jahr 2009 stetig an, zwischen 2009 und 2012 ist der Wert stabil bei etwa 20 Millionen Fortbildungspunkten pro Jahr, bei etwa 5 Millionen Teilnahmen an

Fortbildungsveranstaltungen. Somit werden pro Fortbildungsveranstaltung im Durchschnitt 4 CME-Punkte erreicht.

Im Jahr 2011 waren etwa 342.000 Ärzt*innen in Deutschland ärztlich tätig und somit fortbildungspflichtig¹¹, dies entspricht rechnerisch bei 50 Fortbildungspunkten pro Jahr einem zu erreichenden Gesamtvolumen von 17,1 Millionen Punkten, die in der Realität also übertroffen wurden. Der größte Anteil der Punkte wird seit dem Jahr 2006 durchgehend in der Kategorie C (Arbeitsgruppen) erreicht, der Anteil der Punkte lag im Jahr 2006 bei 44 %, in den Jahren 2007-2012 jeweils bei ca. 35 % der Gesamtpunkte. Dies ist vermutlich vor Allem darauf zurück zu führen, dass gerade in den Universitätskliniken und anderen großen Krankenhäusern viele der in den klinischen Alltag integrierten Veranstaltungen wie Tumorkonferenzen, Fallkonferenzen und Besprechungen CME-zertifiziert sind und von den im Krankenhaus angestellten Ärzt*innen ohnehin regelmäßig besucht werden (müssen).

An zweiter Stelle der gemeldeten Fortbildungsveranstaltungen steht die Kategorie A (Vortragsveranstaltungen), die etwa 40 % aller Veranstaltungen ausmacht. Die meisten Teilnahmen an Fortbildungsveranstaltungen betreffen die Kategorie A (im Schnitt 43% aller Teilnahmen), wobei hier etwa 31 % der Punkte gesammelt werden.

Die Kategorie D, Fortbildung mittels Fachartikeln in Online- und Printmedien in Kombination mit einer schriftlichen Lernerfolgskontrolle, die meist in Form von MC-Fragen erfolgt, ist ein weit verbreitetes und häufig genutztes Format.

Fortbildungen dieser Kategorie generieren mit 20-25 % einen beachtlichen Teil der vergebenen CME-Punkte, obwohl sie nur etwa 1 % der Fortbildungsveranstaltungen ausmachen¹⁰.

Eine Umfrage in Sachsen-Anhalt ergab bereits Mitte der Neunziger Jahre - und damit vor Einführung der Fortbildungspflicht, dass 97,1 % aller befragten Ärzt*innen das CME-Angebot in medizinischen Fachzeitschriften nutzen¹². Die Zeitschrift MMW - Fortschritte der Medizin hat laut eigener Aussage pro Jahr mehr als 100.000 Teilnehmer*innen bei den 30-40 angebotenen CME-Fortbildungen¹³.

Eine Auswertung des Deutschen Ärzteblatts der Teilnahmen an den eigenen CME-Beiträgen zeigten für den Zeitraum 09/2004 bis 09/2005, dass 43.000 Mediziner*innen, entsprechend 14 % aller berufstätigen Ärzt*innen in Deutschland an der zertifizierten Fortbildung im Deutschen Ärzteblatt teilnahmen¹⁴. Eine aktuelle Auswertung der CME-Nutzung in den Jahren 2004-2015 bestätigte diese Zahl und ergab, dass in diesem Zeitraum 12-16 % aller in Deutschland tätigen Ärzt*innen jährlich mindestens einmal an der Fortbildung im Deutschen Ärzteblatt teilnahmen¹⁵.

Die Statistik der Bundesärztekammer zeigt, dass der Anteil der in Kategorie D erzielten CME-Punkte am von 7,3 % im Jahr 2006 sich innerhalb kurzer Zeit verdreifacht hat und auf etwa 24 % im Jahr 2009 angestiegen ist. Im Jahr 2012 betrug er 11,22 % - obwohl die Beiträge in Print- und Onlinemedien nur 0,3 % aller gemeldeten anerkannten Veranstaltungen ausmachten. Über die Zeit von 2006 bis 2012 gemittelt machte die Kategorie D etwa jede fünfte Teilnahme an einer Fortbildungsveranstaltung aus¹⁶.

Für Ärzt*innen, die an großen Kliniken oder in der Stadt arbeiten, ist es deutlich einfacher, die erforderliche Anzahl Punkte zu erreichen. Gerade in Universitätskliniken und anderen großen Krankenhäusern sind viele der in den klinischen Alltag integrierten Fortbildungen und Besprechungen wie z.B. Tumorkonferenzen, Röntgenbesprechungen, Morbiditäts- und Mortalitäts-Konferenzen oder Fallvorstellungen CME-zertifiziert. Fortbildungsveranstaltungen finden gehäuft in den größeren Städten statt. Besonders aufwändig ist die Teilnahme an Präsenzveranstaltungen wie Seminaren und Kongressen für niedergelassene Ärzt*innen in eigener Praxis, vor allem aus dem ländlichen Raum. Entsprechend ist vor allem für die Zielgruppe der niedergelassenen Fach- und Hausärzt*innen das CME-Angebot in medizinischen Fachzeitschriften besonders interessant, was sich auch in den Zahlen des Deutschen Ärzteblatts widerspiegelt¹⁵.

Auch international bestätigt sich dies: Das American Journal of Gastroenterology berichtete nach Auswertung der CME-Nutzung, dass es sich bei 91 % der Teilnehmer*innen innerhalb eines Jahres um niedergelassene Ärzt*innen handelte¹⁷. In einer Befragung gaben Internist*innen an, zur Fortbildung bevorzugt Kurse zu besuchen und Fachzeitschriften zu lesen, während Assistenzärzt*innen klinikinterne Veranstaltungen und das Lehrbuchstudium bevorzugten¹⁸.

Bislang erfolgt die Auswahl der Fortbildungsveranstaltungen frei und bedürfnisgesteuert durch die Ärztinnen und Ärzte, d.h. entsprechend dem subjektiv empfundenen Fortbildungsbedürfnis und den persönlichen Interessen, eventuell aber auch entsprechend anderer Kriterien wie Kosten, Zeitaufwand, Verfügbarkeit. Das persönliche Fortbildungsbedürfnis stimmt jedoch nicht unbedingt mit dem tatsächlichen Wissensdefizit überein¹⁹. Im Gegenteil konnte in einer Studie an 16 Hausärzten gezeigt werden, dass diese dazu tendieren, sich insbesondere in den Bereichen fortzubilden, in denen sie sich bereits gut auskennen und jene Bereiche, in denen Wissensdefizite bestehen, eher zu meiden²⁰. Eine Meta-Analyse aus dem Jahr 2006 zeigte, dass Ärzt*innen sehr schlecht darin sind, ihre Kompetenz und ihren persönlichen Fortbildungsbedarf selbst einzuschätzen²¹. Dies lässt ebenfalls einen negativen Einfluss auf die Wahl der Fortbildungsangebote vermuten.

1.1.3 Zielsetzung der ärztlichen Fortbildung

In der (Muster-)Fortbildungsordnung der Bundesärztekammer aus dem Mai 2013 wird das Ziel der Fortbildung in §1 wie folgt festgesetzt: „Die Fortbildung der Ärztinnen und Ärzte dient dem Erhalt und der kontinuierlichen Weiterentwicklung der beruflichen Kompetenz zur Gewährleistung einer hochwertigen Patientenversorgung und Sicherung der Qualität ärztlicher Berufsausübung.“¹

Dies ist eine Weiterentwicklung der (Muster-)Satzungsregelung Fortbildung und Fortbildungszertifikat der Bundesärztekammer in der Fassung vom 20.05.2004, die das Ziel der Fortbildung wie folgt beschreibt: „Fortbildung der Ärzte und der Ärztinnen dient dem Erhalt und der dauerhaften Aktualisierung der fachlichen Kompetenz.“²²

In allen aktuellen Formulierungen werden dagegen die Patientenversorgung und die Qualitätssicherung in den Fokus der Fortbildung gestellt. Die Bundesärztekammer definiert darüber hinaus in ihren „Empfehlungen zur ärztlichen Fortbildung“, aktuell in der 4. Version vom 24.5.2015²³, auf Seite 5 die Kriterien, welche Fortbildungsinhalte unabhängig vom individuellen Fortbildungsbedürfnis erfüllen sollen, wie folgt:

- „Nutzen für Patienten
- Verständlichkeit
- Relevanz und Aktualität
- wissenschaftliche Evidenz/dem aktuellen Stand der Wissenschaft entsprechend
- Anwendbarkeit des Erlernen in der beruflichen Praxis
- Nutzen für den Arbeitsablauf
- Kosten-Nutzen-Verhältnis
- Qualitätsmanagement
- Risikomanagement und Patientensicherheit
- kritische Wertung im Kontext des Themenfeldes
- Unabhängigkeit von ideologischen und wirtschaftlichen Interessen
- Konformität mit den Vorgaben der verfassten Ärzteschaft
- Konformität mit ethischen Grundsätzen (z. B. WHO-Deklaration)“

Manche dieser Kriterien scheinen selbstverständlich, wie z.B. „Verständlichkeit“, „Relevanz und Aktualität“. Andere Kriterien sind dagegen bemerkenswert, da sie gezielt auch den gewünschten Effekt der Fortbildung benennen und damit indirekt Vorgaben für den Inhalt und die didaktische Umsetzung der Fortbildungsveranstaltungen machen. Dies sind die Kriterien „Nutzen für den Patienten“, „Anwendbarkeit des Erlernen in der beruflichen Praxis“ sowie „Nutzen für den Arbeitsablauf“.

Hier wird klar der Fokus auf den Praxisbezug der Fortbildung gelegt, das Kriterium „Nutzen für den Patienten“ fordert sogar einen unmittelbaren Effekt der Fortbildung des Arztes auf die Patientenversorgung. Hieraus lässt sich ableiten, dass der gewünschte Effekt von CME-Fragen im besseren klinischen Handeln der Ärzte besteht. Fortbildung soll nicht nur theoretisches Wissen vermitteln, sondern die praktische Anwendung im Alltag des Arztes verbessern und damit eine Verbesserung der Patientenversorgung sicherstellen.

Des Weiteren gibt die Bundesärztekammer in ihren Empfehlungen auf Seite 9 Hinweise zur didaktischen und formalen Ausgestaltung der Fortbildung:

„Die Zielgruppe ist definiert.

Die zu erreichenden Lernziele sind klar formuliert.

Die Inhalte sind auf die Lernziele abgestimmt, gemäß oben genannten Kriterien ausgewählt und didaktisch in geeigneter Form aufbereitet.

Die Fortbildungsmaßnahme beinhaltet geeignete Methoden zur Fremd- und/oder Eigenbewertung des Lernprozesses bzw. des Lernerfolgs. Hierzu gehören z. B. interaktive Abstimmungssysteme (z. B. Tele-Dialog-Systeme, TED), praktische (z. B. Objective Structured Clinical Examination, OSCE) sowie mündliche oder schriftliche Tests (z. B. Multiple-Choice, MC).

Die mitwirkenden Autoren, Moderatoren, Tutoren, Referenten, Kursleiter und wissenschaftliche Leiter werden aufgrund ihrer Qualifikation ausgewählt; Moderatoren, Tutoren, Referenten und Kursleiter zusätzlich aufgrund ihrer Eignung zur Wissensvermittlung.

Multiple-Choice-Tests erfolgen in Form von 10 Fragen pro Fortbildungseinheit mit 5 Antwortmöglichkeiten, wovon nur eine richtig ist. Die Bestehensgrenze liegt bei mindestens 70 % richtiger Antworten.“

Hier wird von „didaktisch geeigneter Form“ und „geeigneten Methoden“ gesprochen, ohne dass konkrete Hinweise zur bestmöglichen Ausarbeitung gegeben werden. Es werden keine konkreten Lehrformate empfohlen.

Inwiefern die oben genannten Anforderungen an die ärztliche Fortbildung in den aktuellen Formaten erfüllt werden, kann bisher nur gemutmaßt werden. Der Aspekt der Praxisnähe „Nutzen für den Patienten“ bzw. „Nutzen für den Arbeitsablauf“ könnte bei den Fortbildungsveranstaltungen der Kategorie C (Fortbildung mit konzeptionell vorgesehener Beteiligung jedes einzelnen Teilnehmers) am Ehesten erfüllt sein, da die Veranstaltungen in dieser Kategorie oft auch direkt in der Klinik bzw. der Abteilung angeboten werden, zum Großteil in Kleingruppen stattfinden und auch praktische Übungen beinhalten können.

Für die Fortbildungsveranstaltungen der Kategorie D stellt sich die Frage, inwiefern Online- und Printmedien bisher den Praxisbezug der behandelten Themen herstellen und ob durch die Bearbeitung dieser Angebote ein Nutzen für den Patienten erzielt werden kann und erzielt wird. In Hinblick auf die weit verbreitete Nutzung dieses Formates (siehe 1.1.2) ist diese deshalb sehr relevant.

2.2 Formale Qualität von MC-Fragen

2.2.1 MC-Fragen in der Aus-, Weiter- und Fortbildung

Multiple Choice (MC) ist das meistgenutzte Prüfungsformat in der medizinischen Aus-, Weiter- und Fortbildung. Es hat seinen festen Platz in universitären Prüfungen und prägte lange Zeit auch den schriftlichen Teil der Ärztlichen Prüfungen: Die vom Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) in Mainz gestellten schriftlichen Staatsexamensprüfungen bestanden bisher

basierend auf der Ärztlichen Approbationsordnung ausschließlich aus MC-Fragen. Die universitäre Lehre orientiert sich an vielen Fakultäten an den Staatsexamensprüfungen und so findet nach wie vor auch der überwiegende Teil der universitären Zwischenprüfungen in Form von MC-Prüfungen statt. Eine Übersicht über fakultätsinterne Prüfungen in Deutschland aus dem Jahr 2010 zeigt, dass im vorklinischen Abschnitt zwei Drittel aller Prüfungen und im klinischen Abschnitt fast 80 % aller Prüfungen ganz oder teilweise im Multiple-Choice Format abgehalten werden²⁴.

Die Vorteile von MC-Fragen sind ihre schnelle Auswertbarkeit, die objektiv auch von Laien oder vollständig elektrosch erfolgen kann. Es besteht – beispielsweise im Gegensatz zu mündlichen Prüfungen oder Freitext-Aufgaben - eine gute Vergleichbarkeit der Ergebnisse. Jedoch gilt wie für alle Formen der Prüfung, dass das Ergebnis der Prüfung stets nur so gut, reliabel und vergleichbar sein kann, wie das Messinstrument. Die Aussagekraft von MC-Fragen wird von zahlreichen Faktoren beeinflusst und kann beispielsweise durch konstrukt-irrelevante Faktoren beeinträchtigt werden. Dies bedeutet, dass außer dem zu prüfenden Wissen noch andere Aspekte erfasst werden, die die Beantwortung der Frage beeinflussen. Beispielsweise können komplizierte Formulierungen oder Doppelverneinungen dazu führen, dass die Beantwortung der Frage deutlich erschwert wird. In diesem Fall wird nicht nur das medizinische Wissen gemessen, sondern auch Aspekte Textverständnis oder Konzentrationsfähigkeit beeinflussen das Ergebnis der Prüfung²⁵. Probleme ergeben sich hier insbesondere bei Nicht-Muttersprachler*innen. Versteckte Lösungshinweise wie absolute Aussagen (*immer* oder *nie* sind in der Medizin selten zutreffend) oder offensichtlich falsche Distraktoren (Falschantworten) können dazu führen, dass eine Frage auch ohne ausreichendes Wissen korrekt beantwortet werden kann.

Diese Einschränkungen bei Verwendung von MC-Fragen vor dem Hintergrund ihrer großen Verbreitung und der Verwendung insbesondere auch in wichtigen Prüfungen wie den Staatsexamina hat dazu geführt, dass seit vielen Jahren versucht wird, Gütekriterien für MC-Prüfungen zu definieren und ihre Einhaltung bei der Erstellung von Prüfungen durchzusetzen. Es gibt mittlerweile von verschiedenen Autoren publizierte, klare Kriterien für die Erstellung guter MC-Fragen. Zu nennen sind hier die vom US-amerikanischen National Board of Medical Examiners (NBME) veröffentlichten Empfehlungen²⁶ oder auch deutschsprachige Hinweise des Instituts für medizinische Lehre Bern²⁷.

Es ist für den medizinischen Bereich gut belegt, dass sich fehlerhafte Aufgabenformulierungen auf das Ergebnis von Prüfungen auswirken können. Beispielsweise konnte gezeigt werden, dass fehlerhaft formulierte MC-Fragen das Prüfungsergebnis von Studierenden negativ beeinflussen und einen Einfluss auf das Bestehen von Prüfungen haben^{28,29}. Bei der Analyse der Prüfungen an Pflegeschulen konnte gezeigt werden, dass durch formal fehlerhafte MC-Fragen vor allem besonders gute Prüfungsteilnehmer*innen im Verhältnis benachteiligt werden³⁰.

Prinzipiell ist es mit MC-Fragen möglich, auch Wissen höherer Wissensniveaus zu prüfen, die über Faktenwissen (deklaratives Wissen) hinaus gehen³¹. Eine Darstellung der unterschiedlichen Wissensniveaus erfolgt in Abschnitt 2.3.4. Hierzu ist jedoch eine entsprechende Expertise bei der Erstellung von MC-Fragen notwendig. Beispielsweise kann die Verwendung von Fallvignetten Denkprozesse fördern, die die Fähigkeit zu problemlösendem Denken repräsentieren - und somit vermutlich das Prüfen höherer Wissensniveaus fördern³². In der Diskussion der geeigneten Prüfungsformen zur Operationalisierung höherer Wissensniveaus und ärztlicher Kompetenzen geht Miller in einem Review darauf ein, dass eine bewusste Berücksichtigung der Wissensstufen nach Bloom beim Verfassen von MC-Fragen dazu führen kann, dass mit diesen über reines Faktenwissen hinausgehend auch die Fähigkeit zu Analyse oder Interpretation abgeprüft werden kann³³.

2.2.2 Vorgaben für MC-Fragen in der ärztlichen Fortbildung

Der überwiegende Anteil der Zertifizierungen von Fachzeitschriften entfällt auf nur zwei Landesärztekammern, die Bayerische Landesärztekammer (BLÄK) sowie die Ärztekammer Nordrhein (ÄkNo).

Zeitschriftenbeiträge dieses Formates in der Kategorie D haben eine vorgegebene Struktur und bestehen aus einem Fachartikel und jeweils zehn MC-Fragen, die sich auf den Inhalt des Artikels beziehen. Die Landesärztekammern geben in ihren Richtlinien hierbei nur die Struktur der MC-Fragen vor (Einfachauswahl, fünf Antwortoptionen) und verweisen bezüglich des Inhalts bzw. der formalen Qualität

teilweise auf die Empfehlungen des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP)³⁴. Die ÄkNo gibt in ihren „Empfehlungen der Ärztekammern zur Durchführung von ärztlichen Fortbildungsveranstaltungen“ zwar sehr konkrete Hinweise für die Ausarbeitung von Vorträgen, MC Fragen werden jedoch nicht erwähnt. In der überarbeiteten Fassung der Richtlinie zur Fortbildungsordnung der Ärztekammer Nordrhein aus dem Jahr 2016 werden MC-Fragen als Lernerfolgskontrolle grundsätzlich festgelegt, die Richtlinie geht jedoch in diesem Zusammenhang nicht auf Inhalt und Strukturqualität der MC-Fragen ein³⁵.

In den „Empfehlungen zur ärztlichen Fortbildung“ der Bundesärztekammer wird explizit für mediengestützte Fortbildungsmaßnahmen die Durchführung eines Qualitätssicherungsverfahrens (Review-Verfahrens) gefordert²³.

Ob im Rahmen des Zertifizierungsverfahrens eine Überprüfung der gestellten MC-Fragen durch die Landesärztekammern erfolgt, konnte bei einer Recherche zum Zeitpunkt der Manuskript-Verfassung nicht verifiziert werden.

Vorgaben zum Inhalt oder der Ausgestaltung der Fragen werden bisher, abgesehen von den unter 1.1.3 genannten Hinweisen zur Zielsetzung der ärztlichen Fortbildung, in den Fortbildungsordnungen der Landesärztekammern nicht gemacht.

Einzelne Fachzeitschriften weisen jedoch in ihren Autorenhinweisen darauf hin, dass beispielsweise Fallvignetten verwendet werden sollen oder Praxisrelevanz durch den Bezug der Fragen zu konkreten klinischen Situationen erzeugt werden soll. Eine Übersicht über die Vorgaben ausgewählter Zeitschriften im Original-Wortlaut zum Zeitpunkt der vorliegenden Untersuchung gibt Tabelle 2.

Tabelle 2: Vorgaben der Zeitschriften für den Inhalt der CME-Fragen sowie die Verwendung von Fallvignetten.
Quelle: Autorenhinweise der Verlage, die als PDF veröffentlicht und auf der jeweiligen Website der Fachzeitschrift herunterzuladen sind. Letzter Abruf jeweils am 25.03.2013. Die Dokumente liegen als PDF vor.

Deutsches Ärzteblatt	„Mindestens zwei Fragen, die sich auf eine konkrete klinische Situation beziehen“
DMW	„80 % der Fragen sollen durch den Text beantwortbar sein. Maximal 20 % der Fragen sollen Transferfragen darstellen. (d.h. das Gelesene wird auf neue, nicht im Text behandelte Sachverhalte angewandt) Der Leser soll für seine Berufspraxis konkrete Handlungsanweisungen erhalten und diese direkt umsetzen können.“ „Die Quiz-Fragen sollen so formuliert sein, dass der Beantwortende durch seine richtigen Antworten dokumentiert, dass er diagnostisch und therapeutisch richtig handeln wird und die Inhalte im Gesamtzusammenhang des derzeit gesicherten Wissens dieses Fachgebietes beurteilen kann. Reine Wissensreproduktion ist nicht erwünscht.“ „Der Beitrag ist konzentriert auf das „how to do“, also auf direkte Handlungsanweisungen, weniger auf Ätiologie, Pathophysiologie oder die Diskussion der Literatur.“
Der Internist	„Problemorientierte Aufgabenstellungen in Form einer kurzen Fallbeschreibung erhöhen den Lerneffekt“
MMW	„Bitte achten Sie soweit möglich auf Praxisnähe (Zielgruppe: Hausärzte!).“ „Erwünscht sind ein bis drei kurze Schilderungen klinischer Szenarien mit der Frage nach einer konkreten klinischen Entscheidung.“

Thieme (Verschiedene Zeitschriften, z.B. Pneumologie, Intensivmedizin up2date, Notfallmedizin up2date, Die Rehabilitation, Aktuelle Neurologie, Das Gesundheitswesen...)	„Fragen und Antworten sollen das Verständnis des Themas sowie die Kompetenz zur Lösung von Problemen erfragen und fördern.“ „Gerne können problemorientierte Fallbeschreibungen eingebaut werden, die eine konkrete klinische Entscheidung verlangen.“
--	---

2.2.3 Formale Qualität von MC-Fragen und Qualitätssicherungsmaßnahmen

In einer Untersuchung aus dem Jahr 2006 zeigten Stagnaro-Green et al. im New England Journal of Medicine einen hohen Anteil formal fehlerhafter MC-Fragen auf, der bis zu 100 % betrug³⁶. In einer ähnlichen Studie im deutschsprachigen Raum analysierten Kühne-Eversmann et al. 2007 die formale Qualität der MC-Fragen drei deutscher medizinischer Fachzeitschriften. Hierbei lag der Anteil formal fehlerhafter Fragen bei durchschnittlich 68 %³⁷. Diese Einschätzung wurde von Rotthoff et al. 2008 in einer ähnlichen Studie mit vier deutschen Fachzeitschriften bestätigt³⁸.

Zwischenzeitlich wurden, wohl auch als Reaktion auf diese Studien, seitens der Fachverlage verschiedene Qualitätssicherungsmaßnahmen bezüglich der MC-Aufgaben eingeführt. Autorenhinweise wurden nicht nur zum Verfassen eines Artikels, sondern ergänzend auch zur Fragenerstellung veröffentlicht. So veröffentlichte das Deutsche Ärzteblatt im Juni 2008 das Dokument „Regeln zur Konstruktion von MC-Fragen“, in dem die Redaktion die von Kühne-Eversmann et al. vorgelegte Studie zitiert und auf das dort schlechte Abschneiden der Fragen aus der eigenen Zeitschrift Bezug nimmt³⁹.

Zusätzlich wurden Reviewer*innen geschult und Review-Verfahren implementiert. Diese sind qualitativ und quantitativ unterschiedlich ausgestaltet.

Der Effekt dieser Qualitätssicherungsmaßnahmen wurde von verschiedenen Autoren untersucht. Im Jahr 2010 konnten Gutmann et al. bei einer Analyse der formalen Qualität der MC-Fragen, die zwischen 2002 und 2008 in der Fachzeitschrift „Der Nervenarzt“ veröffentlicht wurden zeigen, dass diese sich im Verlauf verbessert haben. Im untersuchten Zeitraum 2002-2005 lag der Anteil formal fehlerhafter Fragen bei 70 %, im Zeitraum 2005-2008 sank er auf unter 50%⁴⁰. Oechsner et al. analysierten neben der formalen Qualität der MC-Fragen zusätzlich auch inhaltliche Aspekte für die im Jahr 2011 in drei unterschiedlichen Fachzeitschriften veröffentlichten Fragen. In dieser Arbeit wurden beispielsweise bei 46 % der Fragen ungewollte Lösungshinweise beanstandet⁴¹.

2.3 Inhaltliche Qualität von MC-Fragen in der ärztlichen Fortbildung

2.3.1 Praxisbezug und klinische Relevanz von MC-Fragen in der ärztlichen Fortbildung

Wie in 2.1.2 und 2.3.2 bereits diskutiert, fordern sowohl die Fortbildungsordnung wie auch die zur Fortbildung verpflichteten Ärzt*innen den Praxisbezug der medizinischen Fortbildung. Insbesondere wird Wert darauf gelegt, dass ein Nutzen für die Patient*innen erzielt und das Gelernte in der täglichen Arbeitspraxis angewendet werden kann. Es stellt sich die Frage, inwiefern dies durch das Lesen von Fachartikeln und die Bearbeitung der zugehörigen MC-Fragen überhaupt geleistet wird und werden kann. Wie unter 2.2.1 ausgeführt, können MC-Fragen sehr wohl geeignet sein, um Anwendungswissen und Analyse- sowie Problemlösefähigkeit zu untersuchen und zu fördern.

Zum Inhalt und der didaktischen Ausgestaltung von MC-Fragen in der CME gibt es abgesehen von der formal-strukturellen Gestaltung bisher noch wenige Untersuchungen.

Nobbe et al. stellten 2008 die Lernziele der CME-Fragen den Kernaussagen der zugehörigen Fachartikel gegenüber, jedoch ohne sich genauer mit der inhaltlich-didaktischen Ausgestaltung der Fragen auseinander zu setzen⁴². Rotthoff et al. untersuchten 2008 MC-Fragen aus drei Zeitschriften nicht nur auf ihre formale Qualität, sondern auch in den Kategorien „Fallvignette“ und „konkrete klinische Entscheidung“³⁸. Eine Fallvignette fand sich in 10 %, die Notwendigkeit für Leser*innen, eine konkrete

klinische Entscheidung zu treffen, wurde 15 % der untersuchten Fragen attestiert. Angaben zur Methodik finden sich nicht, somit bleibt unklar, welche Kriterien für die Einstufung als konkrete klinische Entscheidung zugrunde gelegt wurden.

Oechsner et al. untersuchten 2013 MC-Fragen mehrere Zeitschriften aus dem deutschsprachigen Raum, wobei sie nicht nur die formale Qualität, sondern auch inhaltliche Aspekte bewerteten⁴¹. Die Fragen wurden auf 11 Qualitätsmerkmale hin geprüft, die einerseits formale Kriterien wie z.B. „Beantwortbarkeit ohne Kenntnis der Antwortoptionen“, „Eindeutige Fragenformulierung“ oder „Keine doppelten Verneinungen“, andererseits mit den Kategorien „Konkreter Fall, Problembezug“ und „Analyse, Interpretation, Problemlösung“ aber auch inhaltliche Aspekte der Fragen untersuchte. Auch hier wird jedoch betont, dass keine fachlich-inhaltliche Wertung der Fragen erfolgt. Ein Problembezug bzw. konkreter Fall fanden sich hier in 15 %, eine Problemlösung war aus Sicht der Gutachter in 19 % der Fragen notwendig. Eine genaue Definition der Begriffe Problembezug oder konkrete klinische Entscheidung ist im Rahmen dieser Untersuchung nicht erfolgt.

Die Operationalisierung des Praxisbezugs von MC-Fragen ist jedoch wünschenswert, um die Frage zu beantworten, ob diese den Vorgaben der Ärztekammern sowie den Ansprüchen der Ärzt*innen an die Fortbildung entsprechen. Um die klinische Relevanz der MC-Fragen messbar zu machen, sollte über die Form und den Inhalt hinaus auch das Niveau des abgeprüften Lernziels bestimmt werden.

2.3.2 Anspruch der Ärzt*innen an die Fortbildung

Neben den unter 1.1.3 genannten Anforderungen, die CME erfüllen soll, ist ein ebenso wichtiger Aspekt, welchen Bedarf Ärzt*innen bezüglich der Fortbildung haben. Neben den Formaten ist hier vor Allem die Frage, welche Inhalte von den Ärzt*innen gewünscht sind bzw. in welchem Bereich der größte Bedarf an Fortbildung besteht.

In der Literatur herrscht weitgehend Einigkeit darüber, dass es große Unterschiede im Lernverhalten zwischen Kindern und Erwachsenen gibt. Malcolm Knowles prägte hier den Begriff der „Andragogy“ im Gegensatz zur „Pedagogy“ und beschrieb die Methoden der Erwachsenenbildung als eher lernerorientiert sowie als problembezogen⁴³.

Er arbeitete heraus, dass bei Erwachsenen Erfahrungen (auch aus Fehlern) die zentrale Basis für Lernaktivitäten sind und dass Erwachsene sich bevorzugt für solche Lerninhalte interessieren, die eine unmittelbare Relevanz für und Einfluss auf ihre Arbeit oder ihr Privatleben haben⁴⁴.

Ärzt*innen sehen den größten Fortbildungsbedarf dort, wo ihnen im Arbeitsalltag Probleme bzw. Fragen begegnen. Viele Lernanlässe sind also konkret patientenbezogen oder mit einem bestimmten Fall verknüpft. Unsicherheiten im täglichen klinischen Umgang mit Patienten scheinen ein besonders guter Lernanreiz zu sein⁴⁵.

In einer US-amerikanischen Studie wurden klinische Fragen von Ärzt*innen, die diese sich während der Patientenversorgung stellten, gesammelt und ausgewertet. Hier zeigte sich, dass der Großteil der Fragen in nur fünf Kategorien einzuordnen war: Ursachen eines Befunds, Auswahl eines Tests, Prävention, Auswahl der Therapie sowie Prognose. Die häufigsten Fragen, die identifiziert wurden, waren: „Wie sollte ich einen Befund/eine Erkrankung in dieser Situation behandeln?“, „Ist dieses Medikament in dieser Situation/bei dieser Erkrankung indiziert?“ und „Was ist die Ursache für dieses Symptom?“⁴⁶.

In einer Umfrage unter niedergelassenen Ärzt*innen in Bremen und Sachsen-Anhalt zu positiven und negativen Erfahrungen mit ärztlicher Fortbildung lag der Fokus der Befragten insbesondere auf der Praxisnähe von Fortbildungsveranstaltungen, d.h. diese wünschten sich vor Allem Wissen für den klinischen Alltag. Darüber hinaus wurde der intensive persönliche Erfahrungsaustausch bei Präsenzveranstaltungen als besonders wertvoll betont¹².

Auch in einer Befragung englischer Allgemeinärzt*innen wurden die Fortbildungsbedürfnisse im Kompetenzbereich am Ehesten bei patienten- und beratungsorientierten Fähigkeiten und im Wissensbereich bei klinischen Themen wie Therapien oder häufigen Erkrankungen gesehen⁴⁷.

2.3.3 Faktenwissen und klinische Kompetenz

Faktenwissen stellt in der Medizin, wie in allen Naturwissenschaften, die Grundlage dar, um Kompetenz zu erreichen. Anders als in anderen Fächern benötigen Ärzt*innen jedoch mehr als nur Faktenwissen: sie müssen neben praktischen Fertigkeiten auch klinische Entscheidungsfähigkeiten und Diagnosekompetenz („clinical reasoning skills“) erlangen.

Die Medizin war jahrhundertlang vor Allem eine Erfahrungswissenschaft, bei der das empirisch erlangte Wissen von einem Lehrer an seinen Schüler weitergegeben wurde. Lange Zeit war auch das Medizinstudium von der Vermittlung von Faktenwissen geprägt, was sich auch in den Lernzielkatalogen der Hochschulen sowie den Prüfungen widerspiegelte. In der medizinischen Ausbildung, also dem Hochschulstudium, findet hier jedoch gerade ein Umdenken statt. Die Ausbildung konzentriert sich weniger auf das reine Faktenwissen, sondern es wird mehr Wert auf den umfassenden Erwerb von Kompetenzen gelegt. Im Jahr 2015 wurde vom Medizinischen Fakultätentag (MFT) der Nationale Kompetenzbasierte Lernzielkatalog (NKLM) verabschiedet, in dem festgelegt ist, welche Kompetenzen die Absolvent*innen des Medizinstudiums in Deutschland haben sollen⁴⁸. Im Gegensatz zu vorherigen Lernzielkatalogen stehen hier nicht nur Faktenwissen, sondern auch Fähigkeiten und Haltungen im Fokus. Boshuizen & Schmidt beschreiben ein 3-Stufen-Modell bei der Entwicklung klinischer Expertise. Zunächst erfolgt als erster Schritt die Akquisition von Faktenwissen. Im zweiten Schritt wird dieses durch praktische Erfahrung ergänzt, so dass im dritten Schritt die Integration von theoretischem und praktischem Wissen erfolgen kann, woraus die Wissens-Enkapsulierung folgt⁴⁹.

Die Diagnosefindung wird Medizinstudierenden erleichtert, wenn ihnen kognitive Hilfen zur Verfügung gestellt werden, die sie bei der Kategorisierung und Mustererkennung unterstützen⁵⁰. Die Extraktion erklärender Informationen aus einem Text wird sowohl Novizen als auch Experten vereinfacht, wenn ihr Vorwissen zuvor aktiviert wird⁵¹.

Eine Studie zum Faktenwissen von Internist*innen zeigte, dass diese sieben Jahre nach ihrer Zertifizierung bei einem erneuten schriftlichen Wissenstest schlechter abschneiden. Prädiktiv für das Ergebnis war neben verschiedenen Faktoren wie Dauer zwischen den beiden Prüfungen und Art der Weiterbildung mit 70 % Anteil an der Variabilität vor Allem der Punktwert bei der ersten Prüfung⁵². In einer ähnlichen Studie konnte dieser Zusammenhang bestätigt werden⁵. Die Frage, inwiefern sich die klinische Entscheidungsfindung und letztlich die Patientenversorgung dieser Ärzt*innen über die Jahre entwickelt, kann mit dem Abprüfen medizinischen Faktenwissens alleine jedoch nicht beantwortet werden. Zur Diagnosekompetenz gehört neben der Kenntnis der zugrunde liegenden Fakten auch die Fähigkeit, diese zu organisieren, werten, zu kombinieren und einzuordnen⁵³.

Zusammenfassend kann man sagen, dass Faktenwissen alleine, wie es in schriftlichen Prüfungen abgeprüft werden kann, nicht ausreicht, um ein*e gute*r Kliniker*in zu sein - medizinische Kompetenz ohne zugrundeliegendes Faktenwissen zu erreichen, ist jedoch unmöglich. Somit haben die Lehre und das Abprüfen von Faktenwissen trotz Alledem weiterhin ihre Berechtigung, es sollte jedoch ergänzend auch die Erlangung und Festigung von Kompetenzen gefördert werden.

2.3.4 Taxonomien zur Einstufung in Wissensniveaus

Die bekannten Taxonomien, die sich auf unterschiedliche Wissensdimensionen beziehen, wurden zur Kategorisierung von Lernzielen entwickelt und sind somit nicht direkt für die Bewertung von MC-Fragen anwendbar.

Die wohl bekannteste Taxonomie ist das Modell von Bloom, das die Dimensionen des Wissens und die Dimensionen des kognitiven Prozesses definiert. Bloom beschreibt sechs Stufen: Erinnern, Verstehen, Anwenden, Analysieren, Evaluieren und Erzeugen⁵⁴.

Tabelle 3: Taxonomie des Wissens nach Bloom, aus ⁵⁴

<u>Dimension des Wissens</u>	<u>Dimension des kognitiven Prozesses</u>
1. Erinnern	Wiedererkennen Entsinnen, Abrufen
2. Verstehen	Interpretieren Exemplifizieren Klassifizieren Zusammenfassen Erschließen, Ableiten Vergleichen Erklären
3. Anwenden	Ausführen Implementieren, Anwenden, Durchführen
4. Analysieren	Differenzieren Organisieren Zuschreiben, Zuordnen
5. Bewerten (Evaluieren)	Prüfen Kritisieren
6. Erzeugen	Generieren Planen Produzieren

Um Lernziele besser einordnen zu können, wurde das Grundmodell von Bloom durch Anderson und Krathwohl erweitert und weiterentwickelt. Neben den verschiedenen Wissensdimensionen beziehen sie auch den kognitiven Prozess mit ein. Sie beschreiben vier Dimensionen des Wissens: Faktenwissen, konzeptionelles Wissen, prozedurales Wissen und meta-kognitives Wissen⁵⁵. Diese werden in Tabelle 4 dargestellt und erläutert.

Tabelle 4: Modifizierte Taxonomie des Wissens nach Krathwohl und Anderson, aus ⁵⁵

A. Faktenwissen	Begriffliches Wissen Wissen über spezifische Details und Elemente
B. Konzeptionelles Wissen	Wissen über Klassifikationen und Kategorien Wissen über Prinzipien und Verallgemeinerungen Wissen über Theorien, Modelle und Strukturen
C. Prozedurales Wissen	Wissen über themenspezifische Fertigkeiten und Algorithmen Wissen über themenspezifische Techniken und Methoden Wissen über Kriterien zur Bestimmung und Nutzung geeigneter Verfahren
D. Meta-kognitives Wissen	Strategisches Wissen Wissen über kognitive Aufgaben, inkl. geeignetes kontextuelles und abhängiges Wissen Selbst-Wissen

Außerdem zu erwähnen ist die Miller-Pyramide für die Beurteilung klinischer Kompetenzen, auf die sich u.A. auch der NKLM bezieht. Diese sieht die verschiedenen Ebenen des Wissens als aufeinander aufbauend und benennt die Stufen „weiß“, „weiß, wie“, „zeigt, wie“ und „tut“ (knows, knows how, shows how, does). Miller hat diese Pyramide in Hinblick auf Prüfungen entwickelt und argumentiert, dass in den meisten Prüfungen lediglich Wissen abgeprüft wird, jedoch keine weiter reichenden Kompetenzen. Stufe

eins entspricht dem Faktenwissen, Stufe zwei definiert er als die Anwendung von Wissen, das Sammeln und Zusammenführen von Informationen und die Translation in einen Diagnostik- und Behandlungsplan³³.

Wie bereits beschrieben nimmt die medizindidaktische Ausgestaltung nicht nur des Medizinstudiums, sondern auch der medizinischen Weiterbildung stetig zu, so dass auch die vorhandenen Lernzielkataloge zunehmend nicht nur Lernziele auflisten, sondern auch kategorisieren und einordnen. Die wichtigsten Lernziel-Kataloge, die für diese Arbeit ausgewertet wurden, sind der NKLM aus Deutschland sowie der SCLO (Swiss Catalogue of Learning Objectives for Undergraduate Medical Training) aus der Schweiz, auf den der NKLM sich u.A. bezieht^{48,56}.

Andere etablierte Modelle wie das kanadische CanMEDs wurden für diese Arbeit nicht ausgewertet.

In der Schweiz wurde im Jahr 2001 durch die Schweizerische Medizinische Interfakultätskommission erstmals der Schweizer Lernzielkatalog SCLO veröffentlicht. Dieser enthält die Lernziele, die Studierende der Humanmedizin am Ende des Studiums erreicht haben sollen, und bildet auch die Grundlage für das Examen. 2008 wurde die überarbeitete zweite Version veröffentlicht.

Die Entwickler des SCLO orientierten sich dabei am 1994 in den Niederlanden veröffentlichten „Training of Doctors, Blueprint 1994“ sowie am kanadischen CanMEDs.

Der SCLO unterteilt Lernziele in die drei Kategorien Allgemeine Fertigkeiten (General Skills), Klinische Bilder (Clinical Pictures) und Weiterführendes Wissen (Further Knowledge). In den drei Kategorien sind jeweils zwei Stufen definiert (siehe Tabelle 5)⁵⁶.

Tabelle 5: Einstufung der Lernziele im SCLO, aus⁵⁶

	Level 1	Level 2
General Skills	Some practical experience	Routine
Clinical Pictures	Be able to recognize the possible relevance	Be able to cope with in practice
Further Knowledge	Overview level	Be able to cope with in practice

Der bereits erwähnte NKLM nimmt ebenfalls eine Einstufung der Lernziele in verschiedene Kompetenzebenen vor. Kompetenzen werden definiert als die verfügbaren kognitiven Fähigkeiten und Fertigkeiten sowie Einstellungen, die zur Problemlösung angewendet und integriert werden müssen. Hier werden drei verschiedene Stufen unterschieden, wobei zwei der Stufen kognitive Dimensionen (Faktenwissen, Handlungs- und Begründungswissen) umfassen und die dritte Stufe (Handlungskompetenz) sich auf praktische Fertigkeiten bezieht. Die Einteilung wird in Tabelle 6 dargestellt.

Tabelle 6: Auszug aus dem NKLM, Einteilung der Kompetenzebenen. Aus⁴⁸

1. Faktenwissen	Deskriptives Wissen (Fakten, Tatsachen) nennen und beschreiben.
2. Handlungs- und Begründungswissen	Sachverhalte (Zusammenhänge) erklären und in den klinisch-wissenschaftlichen Kontext einordnen.
3. Handlungskompetenz	3a. Unter Anleitung selbst durchführen und demonstrieren. 3b. Selbständig und situationsadäquat in Kenntnis der Konsequenzen durchführen.

Der NKLM bezieht sich auf die Miller-Pyramide und den SCLO, eine Abbildung zum Vergleich der verschiedenen Kompetenzebenen findet sich in der Einleitung des NKLM⁴⁸.

Abbildung 1: Kompetenzebenen von NKLM und NKLZ im Vergleich zu Miller-Pyramide und SCLO. Aus: ⁴⁸



Weitere im deutschen Raum verwendete Lernzielkataloge sind ebenfalls meist 3-stufig aufgebaut. Die meisten universitären Lernzielkataloge beziehen sich entweder auf die Bloom-Taxonomie (z.B. LMU München Modul 23 AINS⁵⁷) oder aber auf den SCLO (z.B. Hamburg⁵⁸, Kiel⁵⁹) und sind mehr oder weniger stark abgewandelte Formen davon. Die LMU München verwendet beispielsweise im Rahmen des Modul 23 - AINS einen Lernzielkatalog, der drei Ebenen listet: 1. Erinnern, 2. Verstehen und Analysieren, 3. Evaluieren und Erzeugen. Die Universität Lübeck benutzt für den Bereich Innere Medizin eine in Diagnose und Therapie unterteilte dreistufige Taxonomie; die Bloom zwar nicht direkt nennt, sich inhaltlich jedoch an dessen Einstufung anlehnt⁶⁰.

Wie an den beschriebenen Lernzielkatalogen deutlich wird, ist der Erwerb von Wissen im Medizinstudium klar geregelt und stark kategorisiert. Auch in der Weiterbildung zur*m Fachärzt*in zeigt sich in den bestehenden Katalogen eine Definition bestimmter Lernziele und Inhalte, die nicht nur genannt, sondern auch gewichtet werden (beispielsweise Art und Anzahl der Eingriffe im Rahmen der chirurgischen Weiterbildung).

Der Wissenserwerb im Rahmen der ärztlichen Fortbildung dagegen ist relativ frei und wenig reglementiert. Dies ist auch der Tatsache geschuldet, dass es viele, sehr unterschiedliche Szenarien der klinischen Arbeit gibt. Nicht nur in den verschiedenen Fachdisziplinen, sondern auch je nach Art und Weise der Ausübung der ärztlichen Tätigkeit (beispielsweise in der Klinik, niedergelassen in der Praxis, mit oder ohne direkten Patientenkontakt, im öffentlichen Dienst, in der Industrie usw.) unterscheidet sich der Anspruch an Wissen und Fertigkeiten der Ärzt*in enorm. Somit ist es folgerichtig, dass die ärztliche Fortbildung bisher nur einen quantitativen Rahmen vorgibt, die Inhalte sowie Ausgestaltung der

Fortbildung jedoch weitgehend offenlässt. Nichtsdestotrotz ist es wichtig, dass in der ärztlichen Fortbildung die Qualität gesichert sein muss - dies schulden wir den Patient*innen. Sowohl Inhalte als auch Ausgestaltung der Fortbildung müssen deshalb einer ständigen Überprüfung unterliegen. Nicht nur die Inhalte sollten auf Aktualität und Korrektheit untersucht, sondern auch Format und didaktische Ausgestaltung der Fortbildung bewertet und ggf. verbessert werden.

3. Material und Methoden

3.1 Qualitative Analyse der MC-Fragen und Auswertung der Autorenhinweise

Kapitel 3.1 ist in leicht abgewandelter Form Teil der Publikation [c].

3.1.1 Auswahl der Artikel

Insgesamt wurden 720 Fragen aus sechs medizinisch-wissenschaftlichen Fachzeitschriften ausgewertet. Analysiert wurden die drei von Kühne-Eversmann et al.³⁷ untersuchten Zeitschriften „Der Internist“ (Springer-Verlag; INT), „Deutsches Ärzteblatt“ (Deutscher Ärzte-Verlag; DÄB) und „Deutsche Medizinische Wochenschrift“ (Georg Thieme Verlag; DMW), zu denen auch die Rohdaten der Originalstudie für einen Vorher/Nachher-Vergleich vorlagen. Um einen Auswahlbias möglichst gering zu halten, wurden drei weitere Zeitschriften berücksichtigt: MMW - Fortschritte der Medizin (Springer Medizin Urban & Vogel; MMW), Nervenheilkunde (Schattauer Verlag; NHK) und Arzneimitteltherapie (Wissenschaftliche Verlagsgesellschaft, AMT). Hierbei wurden möglichst unterschiedliche medizinische Verlage berücksichtigt und Zeitschriften mit möglichst hoher Auflage ausgesucht, um die Relevanz der Ergebnisse sicher zu stellen (Tabelle 7).

Tabelle 7: Mediadaten der Zeitschriften. Diese Tabelle ist in leicht abgewandelter Form Teil der Publikation [c].

Zeitschrift	Der Internist (INT)	Deutsches Ärzteblatt (DÄB)	Deutsche Medizinische Wochenschrift (DMW)	MMW – Fortschritte der Medizin (MMW)	Nervenheil-kunde (NHK)	Arzneimitteltherapie (AMT)
Verlag	Springer	Deutscher Ärzteverlag	Thieme	Springer	Schattauer	Apotheker Verlag
Druckauflage (2012)	41.260	414.212	12.329	55.614	30.030	17.600
Zertifizierende Stelle	ÄkNo und LÄKH	ÄkNo	ÄkNo	BLÄK	BLÄK	LÄK BW, BAK

Legende zu Tabelle 7

ÄkNo = Ärztekammer Nordrhein, LÄKH = Landesärztekammer Hessen, BLÄK = Bayerische Landesärztekammer, LÄK BW = Landesärztekammer Baden-Württemberg, BAK = Bundesapothekerkammer

Die ausgewählten Zeitschriften weisen außerdem Unterschiede im Umfang und Elaborationsgrad der Autorenhinweise zur Erstellung von MC-Fragen auf, sodass ein Bezug zwischen diesen und der Qualität der MC-Fragen untersucht werden konnte. AMT nimmt hierbei eine Sonderrolle ein, da hier CME-Aufgaben nicht von den Autoren der Fortbildungsartikel stammen, sondern nach Auswahl eines Artikels als Fortbildungsbeitrag von einer Fachredaktion erstellt werden. NHK steht stellvertretend als auflagenstärkste Zeitschrift für den Schattauer Verlag, die Zeitschrift MMW - Fortschritte der Medizin wurde aufgrund des Elaborationsgrades der Autorenhinweise ausgewählt.

Es wurde pro Zeitschrift jeweils der komplette Jahrgang 2012 (außer Sonderbeilagen) ausgewählt, so dass pro Zeitschrift zwischen 10 und 15 CME-Artikel und somit zwischen 100 und 150 MC-Fragen (durchschnittlich 120 Fragen pro Zeitschrift) ausgewertet werden konnten.

Die ausgewerteten Zeitschriften sind zum Teil von der Nordrheinischen Akademie für ärztliche Fort- und Weiterbildung (DMW, DÄB) und zum Teil durch die Bayerische Landesärztekammer zertifiziert (MMW, NHK). Die Beiträge in AMT - Arzneimitteltherapie sind durch die Landesärztekammer Baden-Württemberg sowie durch die Bundesapothekerkammer anerkannt, das Fortbildungsangebot in INT durch die Landesärztekammer Hessen sowie ebenfalls durch die Nordrheinische Akademie für ärztliche Fort- und Weiterbildung (Tabelle 7).

3.1.2 Beurteilungskriterien

Für den Vorher/Nachher-Vergleich von INT, DÄB und DMW wurde die 10 Items umfassende Fehlerliste von Kühne-Eversmann et al.³⁷, die auf einer ausführlichen Literaturrecherche basiert, unverändert verwendet. Zum Nachweis eines Zusammenhangs von Autorenhinweisen und Aufgabenqualität wurde diese Liste nach Sichtung aktueller Literatur zu Gütekriterien für MC-Fragen^{26,27,61} modifiziert und erweitert. Der Fehler 5 wurde von "negativer Stamm" auf "negative Formulierungen" erweitert, ansonsten gab es keine inhaltlichen Änderungen. Es wurden insgesamt 10 Fehlertypen definiert (Tabelle 8). Im Anhang findet sich die ausführliche Liste mit Erläuterungen und Beispielen (Tabelle 19).

Tabelle 8: Fehlerliste. Diese Tabelle ist Teil der Publikation [c].

Die von Kühne-Eversmann et al. verwendete Fehlerliste wurde weitgehend übernommen. Für den longitudinalen Vergleich wurden die für diese Analyse hinzugefügten und abgeänderten Fehlerkategorien nicht gewertet, um eine möglichst hohe Vergleichbarkeit der Daten zu erreichen.

	Name	Beschreibung
1	K-Typ	Kombinierte Antwortmöglichkeiten
2	AOTA	Antwortoption „Alle Aussagen sind richtig“ („All of the above.“)
3	NOTA	Antwortoption „Keine Antwort ist richtig“ („None of the above.“)
4	Unfokussierter Stamm	Der Fragenstamm ist so ungenau formuliert, dass das Themengebiet gar nicht oder kaum eingeschränkt wird.
5	Negative Formulierungen	5a negativer Stamm und 5b negativ formulierte Antwortmöglichkeiten
6	Hohe Komplexität	Der Fragenstamm oder die Antwortmöglichkeiten sind sehr lang/komplex/ausführlich formuliert.
7	Heterogene Antwortmöglichkeiten	7a inhaltlich heterogene (Unterschiedliche Kategorien der Antwortmöglichkeiten, z.B. Ätiologie, Klinik, Diagnostik, Therapie) und 7b strukturell heterogene Antwortmöglichkeiten
8	Unplausible Distraktoren	Antwortmöglichkeiten, die trivial, unplausibel oder nicht sinnvoll sind.
9	Cues	Unbeabsichtigte Lösungshinweise, die scheinbar die richtige Antwort identifizieren.
9a	Schlüsselwörter	Absolute Begriffe wie immer, nur, nie, ausschließlich, allein.
9b	Grammatikalische Fehler	Ein oder mehrere Distraktoren passen grammatikalisch nicht zum Stamm und fallen so als Falschaussage auf.
9c	Unterschiedlich lange Antworten	Einer der Antworten, meist die richtige, ist ausführlicher/länger/spezifischer als die anderen Antwortmöglichkeiten.
9d	Zwei oder mehr ähnliche Antworten	Die Aufmerksamkeit wird auf zwei oder mehr ähnliche Antworten gelenkt, von denen nur eine richtig sein kann.
9e	Verbale Assoziation zwischen Stamm und Antwort	Wörter oder Satzteile des Stamms werden in einer der Antwortmöglichkeiten wiederholt, die dadurch als die (vermeintlich) richtige identifiziert wird.
9f	Wiederholte Nennung bei Auflistungen	Eine der Antworten beinhaltet Element, die am häufigsten in den anderen Antwortmöglichkeiten genannt wurden.
9g	Cues zwischen Fragen	Der Stamm oder eine Antwortmöglichkeit einer Frage beinhaltet die Information zur Lösung einer anderen Frage der Prüfung.
10	Inhaltliche und grammatikalische Fehler	Inhaltliche Fehler in der Fragenstellung, grammatikalische Fehler im Fragenstamm oder den Antwortmöglichkeiten. Hierzu zählen Tippfehler, Kommafehler, falsche Grammatik usw.

Es besteht bei den meisten der ausgewählten Fehlerkategorien international ein breiter Konsens darüber, dass und warum sie bei der Erstellung von Fragen vermieden werden sollten. Eine gute MC-Frage sollte ein konkretes Lernziel abprüfen, d.h. aus einem fokussierten Stamm und homogenen Antwortmöglichkeiten bestehen. Darüber hinaus sollten ungewollte Lösungshinweise vermieden werden, die die Beantwortung der Frage erleichtern oder erschweren^{26, 27,62}. Andere Fehlerkategorien wie Kombinationsantworten werden eigentlich schon durch die Vorgaben der Ärztekammern ausgeschlossen, fanden sich jedoch trotzdem vereinzelt in den Zeitschriften.

3.1.3 Auswertungsprozess

Alle MC-Fragen wurden von zwei geschulten und im Review von MC-Fragen erfahrenen Untersucherinnen unabhängig voneinander bewertet und im Anschluss diskutiert. In der Bewertung, ob eine Frage als fehlerhaft (Vorliegen mindestens eines Fehlers) einzustufen sei, lag die initiale Übereinstimmung bei 96 %. Strittige Fragen wurden im Anschluss mit einem dritten Experten erneut diskutiert. Dabei konnte in 100 % der Fälle ein Konsens gefunden werden. Das mehrfache Auftreten desselben Fehlertyps (z.B. mehrere Schreibfehler oder mehrere unplausible Distraktoren) bei ein und derselben MC-Frage wurde nur einfach gewertet.

3.1.4 Longitudinaler Vergleich

Die Fragen wurden in die vier Kategorien fehlerfrei, ein Fehlertyp pro Frage, zwei Fehlertypen pro Frage sowie drei oder mehr Fehlertypen pro Frage eingeteilt.

Die Ergebnisse von drei der untersuchten Zeitschriften wurden mit den von Kühne-Eversmann et al. 2007 veröffentlichten Daten verglichen, um einen longitudinalen Vergleich zu ermöglichen.

Für diesen Vergleich wurden die für diese Analyse hinzugefügten und abgeänderten Fehlerkategorien nicht gewertet, um eine möglichst hohe Vergleichbarkeit der Daten zu erreichen.

3.1.5 Bewertung der Qualitätsmaßnahmen

Es wurde für alle Zeitschriften geprüft, ob Autorenhinweise zur Erstellung von MC-Fragen vorliegen. Dazu wurden die Webseiten der Zeitschriften durchsucht. Fanden sich dort keine Hinweise, wurde direkter Kontakt mit Verantwortlichen der Redaktion aufgenommen. Für DMW liegt den Autor*innen neben der online verfügbaren, für die Fragenautor*innen bestimmten Version auch eine interne Version vor, die als Leitfaden für das Fragenreview dient. Bei einer Zeitschrift (INT) wurde der den Autor*innen zur Veranschaulichung bereitgestellte Musterbeitrag ausgewertet.

Die vorhandenen Autorenhinweise der sechs untersuchten Fachzeitschriften wurden anhand der zuvor erstellten Fehlerliste untersucht. Zur Bewertung wurde zuerst für jede Fehlerart geprüft, ob diese in den Hinweisen erwähnt wird. Zur weiteren Quantifizierung der Ausführlichkeit der Autorenhinweise wurden die verwendeten Wörter gezählt, wobei jeweils die Anzahl der Wörter pro Fehlerart ausgewertet wurde. Zusätzlich gab es die Kategorien Inhalt, Form und Hintergrund. Diese Analyse wurde wiederum durch zwei unabhängige Beurteilerinnen einzeln vorgenommen und die Ergebnisse anschließend konsentiert, wobei die initiale Übereinstimmungsrate bei 99,5 % lag.

Zusätzlich zur quantitativen Auswertung wurde zur Illustration eine qualitative Auswertung der Autorenhinweise durchgeführt. Hierzu wurde eine Liste von fünf Kategorien erstellt, um die unterschiedlich ausführlichen Erläuterungen standardisiert vergleichen zu können. Die Erfüllung der fünf Kategorien (1 Negativhinweis, 2 Positivhinweis, 3 Negativbeispiel, 4 Positivbeispiel, 5 Begründung) wurden für alle definierten Fehler in den Hinweisen der Verlage untersucht, wobei eine Kategorie entweder als vorhanden oder als nicht vorhanden gewertet wurde. Hierbei wurden alle Kategorien gleich gewichtet. Somit ergibt sich pro Fehlertyp ein Punktwert von 0 (= kein Hinweis auf den Fehler vorhanden) bis 5 (Hinweis wie etwas nicht formuliert werden sollte, Hinweis wie etwas stattdessen formuliert werden sollte, Beispiel für eine schlechte Formulierung, Beispiel für eine gute Formulierung, Begründung weshalb der Fehler vermieden werden sollte, siehe Tabelle 9). Das erarbeitete Bewertungssystem soll die unterschiedliche Ausführlichkeit der Autorenhinweise abbilden, wobei zusätzlich zur reinen Wortanzahl auch der Inhalt analysiert wurde. Der Höchstwert 5 wurde von keinem der Autorenhinweise der untersuchten Zeitschrift für einen der Fehlertypen erreicht.

Tabelle 9: Kategorien der qualitativen Auswertung

Kategorie	Beschreibung	Beispiel
Negativhinweis	Hinweis, wie etwas nicht formuliert werden sollte.	„Vermeiden Sie verallgemeinernde Fragestellungen ohne konkreten Bezug.“
Positivhinweis	Hinweis, wie etwas formuliert werden sollte.	„Formulieren Sie die Frage möglichst präzise, so dass sie auch ohne Lösungsmöglichkeiten zu beantworten ist.“
Negativbeispiel	Beispiel für eine unerwünschte Formulierung	„Welche Aussage ist richtig?“
Positivbeispiel	Beispiel für eine gute Formulierung	„In welchem Organ finden sich am häufigsten die Metastasen eines Kolon-Karzinoms?“
Begründung	Begründung, warum diese Formulierung vermieden werden sollte.	„Die Frage sollte ein konkretes Lernziel abprüfen.“

3.1.6 Korrelation der formalen Qualität 2012 mit den Autorenhinweisen

Die deskriptiven Daten wurden mittels absoluter und relativer Häufigkeit dargestellt.

Zum Nachweis eines statistischen Zusammenhangs zwischen dem Vorhandensein bzw. der Ausführlichkeit von Autorenhinweisen und der Anzahl fehlerhafter Fragen wurden die Fehlerhäufigkeiten zur Vergleichbarkeit auf 100 Fragen normiert und mittels Chi-Quadrat-Test die Anzahl der Fehler pro Journal gegen das Vorhandensein von Autorenhinweisen verglichen. Mithilfe von Kreuztabellen wurde der mögliche Einfluss der Ausführlichkeit der Autorenhinweise auf die Fehleranzahl bivariat korreliert und mittels Chi-Quadrat-Test auf Unabhängigkeit geprüft. Hierbei beschränkten wir uns auf die sechs am häufigsten vorkommenden Fehlertypen 4, 5, 6, 7, 8 und 9.

Alle statistischen Analysen wurden mit dem Statistikprogramm SPSS, Version 22.0.0 für MacOS durchgeführt (SPSS Inc., Chicago, IL, USA).

3.2 Analyse der Wissensebenen und Auswertung der Themen

3.2.1 Auswahl der Fragen

Um die Fragestellung weiter zu führen, wurden in der hier vorliegenden Arbeit im nächsten Schritt die so untersuchten Fragen zusätzlich auf ihren Inhalt untersucht, wobei zum Einen die abgeprüfte Thematik und zum Anderen die Art des abgefragten Wissens analysiert wurde.

Eine inhaltliche Analyse zur abgefragten Thematik kann nur dann sinnvoll durchgeführt werden, wenn eine Frage ein klar formulierbares Lernziel abprüft. Um die MC-Fragen einem klar definierten Lernziel zuordnen zu können, müssen diese eine fokussierte Fragestellung und möglichst homogene Antwortalternativen aufweisen. Entsprechend der Vorarbeit wiesen nur die Fragen des Deutschen Ärzteblatts entsprechend fokussierte Fragen auf, so dass diese Zeitschrift für die Analyse ausgewählt werde. Zusätzlich zu dem bereits auf formale Qualität untersuchten Jahrgang 2012 wurde der gesamte Jahrgang 2013 untersucht, so dass insgesamt 260 Fragen aus 26 Ausgaben die Stichprobe bildeten.

Beispiel für eine für die inhaltliche Analyse geeignete Frage:

„Welches diagnostische Verfahren kann zur Beurteilung der knöchernen Verhältnisse sowie einer Osteoporose hilfreich sein?

- a) Dopplerultraschall
- b) Messung des Kaliumspiegels im Blut
- c) Kraftmessung
- d) somatosensible Potenziale
- e) lumbale Computertomographie“⁶³

Beispiel für eine nicht geeignete Frage:

„Welche Aussage zu Komplikationen des Morbus Crohn ist falsch?

- a) Eine Thromboseprophylaxe für hospitalisierte Patienten mit chronisch-entzündlichen Darmerkrankungen wird nicht empfohlen.
- b) In der medikamentösen Therapie der Fisteln bei M. Crohn haben Infliximab, Azathioprin und 6-Mercaptopurin den höchsten Stellenwert.
- c) Antibiotika haben einen sehr guten Effekt auf die Ausheilung von Fisteln.
- d) Die größte Sensitivität zur Beurteilung von Fisteln besitzt die Computertomographie.
- e) Intestinale Perforationen lassen sich durch transabdominelle Sonographie ausschließen.“⁶⁴

Da diese Frage sich nicht auf ein klar definiertes Lernziel bezieht, kann weder das Thema der Frage eindeutig bestimmt noch eine Aussage zum Wissensniveau getroffen werden.

3.2.2 Erarbeiten der Taxonomie der Wissensniveaus

Um die Fragen hinsichtlich des abgefragten Wissensniveaus auszuwerten, musste zuerst eine geeignete Taxonomie gefunden werden, die es erlaubt, diese entsprechend einzuordnen.

Es wurde eine Literaturrecherche durchgeführt und die bekanntesten und weit verbreiteten Taxonomien zur Einstufung von Wissen, die in der Medizin Anwendung finden, analysiert.

Die bekannten Taxonomien (Bloom⁵⁴, Krathwohl⁵⁵, siehe 1.4) wurden zur Einteilung von Lernzielen erstellt und lassen sich somit nicht ohne Weiteres auf MC-Fragen anwenden.

Basierend auf theoretischen Überlegungen und den vorhandenen Einteilungen von Krathwohl und Anderson⁵⁵ und Palmers et al.⁶⁵ wurde deshalb nach ausführlicher Literaturrecherche und in Anlehnung an existierende Konzepte (siehe auch 1.4) eine dreistufige Taxonomie erarbeitet, die es erlaubt, die MC-Fragen in drei aufsteigende Niveaus einzuteilen.

Es wurden drei Stufen definiert:

1) Faktenwissen

Erinnern von Informationen.

Die Basiselemente oder Fakten eines Fachgebiets, die bekannt sein müssen, um Problemstellungen aus diesem Bereich lösen zu können.

Für die Beantwortung dieser Frage ist ein reines Erinnern von Fakten ausreichend, ein tiefergehendes Verständnis des Problems ist nicht notwendig.

2) Anwendungswissen

Verstehen und Anwenden.

Verstehen von Informationen und Fähigkeit, diese zu interpretieren und anzuwenden.

Kriterien für die Anwendung von Fertigkeiten, Algorithmen und Methoden.

Wissen über spezifische Fertigkeiten, Techniken, Methoden und Prozeduren.

Wissen über die Kriterien, anhand derer entschieden wird, welche Prozeduren angemessen sind und angewandt werden.

Über das reine Erinnern hinaus ist die Integration verschiedener Fakten notwendig, um die Frage korrekt zu beantworten. Informationen müssen verstanden, interpretiert und angewendet werden.

3) Problemlösendes Wissen

Problemlösung.

Die Anwendung von Wissen und Verständnis in anderem Kontext, um ein Problem zu lösen.

Das Wissen über die Zusammenhänge einzelner Faktoren innerhalb einer größeren Struktur.

Wissen wird implizit vorausgesetzt.

Zur Beantwortung der Frage müssen zuerst die gegebenen Daten interpretiert werden. Außerdem müssen andere Aspekte in die Beantwortung der Frage einbezogen werden, die nicht explizit genannt werden.

Zusammengefasst, übersetzt und modifiziert nach: ^{55,65}

Tabelle 10 fasst die dreistufige Taxonomie zusammen und gibt für jede Stufe ein Beispiel zur Verdeutlichung.

Tabelle 10: Taxonomie der Wissensarten mit Beispielen

<p>1) Faktenwissen</p> <p>Für die Beantwortung dieser Frage ist reines Erinnern von Fakten ausreichend, ein Verstehen des Problems ist nicht notwendig.</p>	<p>Wie hoch ist der prozentuale Anteil von Patienten mit psychosomatischen Störungen in allgemeinmedizinischen Praxen?</p> <p>a) 5 % b) 10 % c) 20 % d) 30 % e) 40%</p>
<p>2) Anwendungswissen</p> <p>Voraussetzung für die Beantwortung der Frage ist das Erinnern von Fakten. Darüber hinaus ist aber die Integration verschiedener Fakten notwendig, um die Frage korrekt zu beantworten. Es ist keine Problemlösung notwendig.</p>	<p>Ein 34-jähriger Patient stellt sich mit Husten und Atemnot in der Notaufnahme vor. Sie auskultieren feinblasige Rasselgeräusche. Laborchemisch zeigt sich eine Leukozytose mit Linksverschiebung und ein erhöhtes CRP (8,9 mg/dl), im Röntgen Thorax sehen Sie eine segmentale Verschattung. Was ist die wahrscheinlichste Diagnose?</p> <p>a) Bronchitis b) Asthma c) Pneumonie d) Infekt der oberen Luftwege e) Lungenkarzinom</p>
<p>3) Problemlösendes Wissen</p> <p>In dieser Frage wird Wissen implizit vorausgesetzt. Zur Beantwortung der Frage müssen zuerst die gegebenen Daten interpretiert werden. Außerdem müssen andere Aspekte in die Beantwortung der Frage einbezogen werden, die nicht explizit in der Frage erwähnt werden. Es ist ein Zwischenschritt, in diesem Fall die Erstellung einer Verdachtsdiagnose, nötig. Verschiedene Optionen müssen gegeneinander abgewogen und gewichtet werden.</p>	<p>Eine 28-jährige Patientin stellt sich mit rezidivierenden Bauchschmerzen und Blähungen bei Ihnen vor. Teilweise treten auch Diarrhoe auf. Was ist Teil der Basisdiagnostik?</p> <p>a) H2-Atemtest b) Gastroskopie mit Biopsien c) Mikrobiologische Untersuchung des Stuhls d) Koloskopie e) Computertomographie des Abdomens</p>

3.2.3 Auswertungsprozess

Zur Festlegung des abgefragten Themas sowie des Wissensniveaus der Frage wurde jede MC-Frage auf ihr Lernziel hin überprüft. Alle Fragen wurden unabhängig von zwei Expertinnen bewertet, strittige Fragen wurden mit einem dritten Experten diskutiert. In allen Fällen konnte ein Konsens bezüglich der Einstufung der Frage erreicht werden.

Zunächst erfolgte für jede Frage die Zuordnung zu einem Wissensniveau. Darüber hinaus wurde für jede Frage bestimmt, ob eine Fallvignette verwendet wurde und wenn ja, ob diese für die Beantwortung der Frage notwendig war oder die Frage auch ohne die Fallvignette beantwortet werden konnte (damit wäre die Fallvignette überflüssig). Hierzu wurde der „Abdecktest“ verwendet: Die Fallvignette wird abgedeckt. Kann die Frage nun auch ohne Kenntnis der Fallbeschreibung beantwortet werden, so wird diese als nicht notwendig eingestuft.

Beispiel für eine Frage mit nicht-notwendiger Fallvignette:

„Bei einem 32-jährigen Mann wurde vor 5 Jahren eine Colitis ulcerosa diagnostiziert, die sich seit > 3 Jahren klinisch und histologisch in Remission befindet. Aktuell berichtet er über Druckschmerzen im rechten Oberbauch. Bei der klinischen Untersuchung fällt ein Hautikterus auf. Welche Aussage bezüglich der Verdachtsdiagnose einer PSC trifft zu?

- a) Der Nachweis von pANCA ist für die Diagnose einer PSC beweisend.
- b) In 75% der Fälle sind isoliert die intrahepatischen Gallengänge betroffen
- c) Histologisch zeigt sich häufig eine nicht-eitrige destruierende Cholangitis, teilweise mit Nachweis verkäsender Granulome
- d) Pathognomonisch ist eine selektive IgM-Erhöhung.
- e) In der Cholangiographie können Irregularitäten mit Strikturen und Dilatationen als typische perlschnurartige Veränderungen imponieren.“⁶⁶

Außerdem wurde für die ausgewählten Fragen jeweils das Thema des abgeprüften Lernziels untersucht. Anschließend wurden diese in Überkategorien zusammengefasst, wobei folgende Kategorien entwickelt wurden: Ätiologie/Pathogenese, Epidemiologie, Klinisches Bild, Diagnostik/Befunde, Therapie, Pharmakologie, Differentialdiagnose, Vorgehen.

4. Ergebnisse

4.1 Qualitative Analyse der MC-Fragen

Kapitel 4.1 ist in leicht abgewandelter Form Teil der Publikation [c].

4.1.1 Formale Qualität der MC-Fragen für den Jahrgang 2012

Untersucht wurde der Anteil fehlerhafter Fragen in sechs ausgewählten Zeitschriften für den Jahrgang 2012.

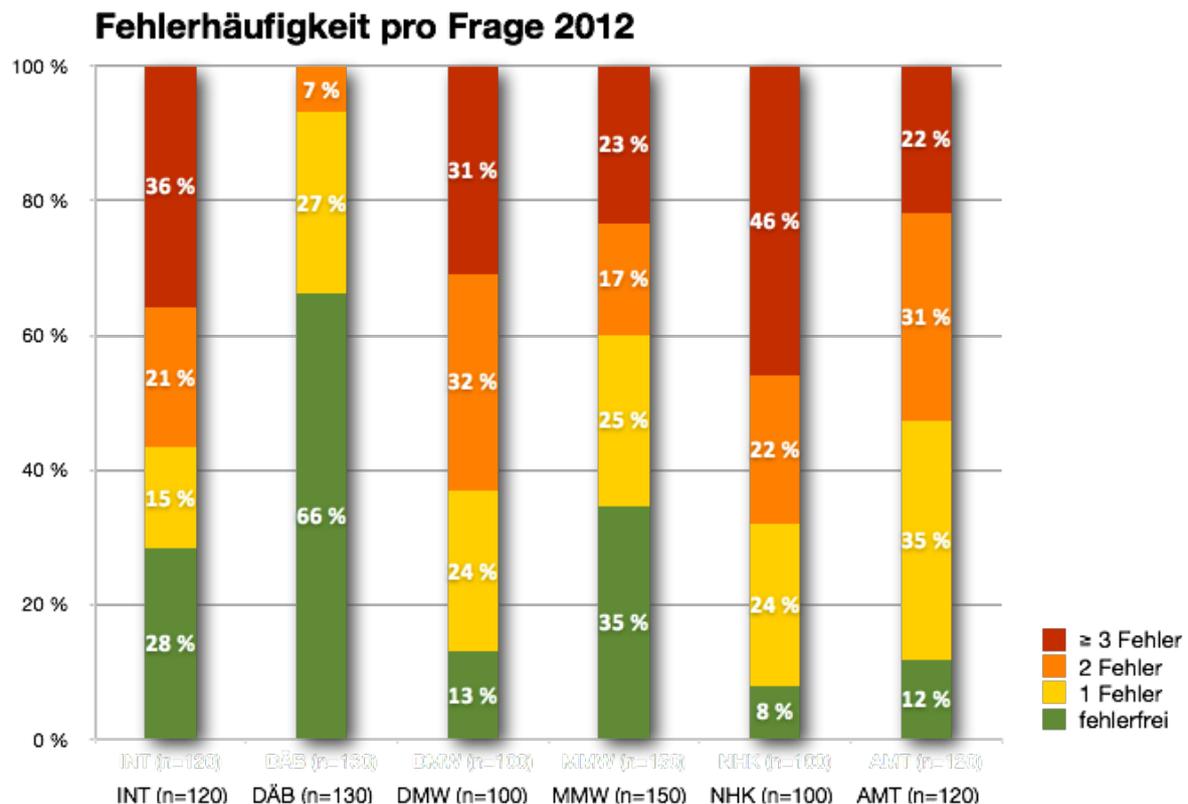
Im direkten Vergleich aller untersuchten Zeitschriften dieses Jahrgangs zeigen sich große Unterschiede in der formalen Qualität der Fragen. Der Anteil fehlerhafter Fragen lag zwischen 33 % und 92 %, der Anteil Fragen mit 3 oder mehr verschiedenen Formfehlern zwischen 0 % und 46 %. Die Ergebnisse sind in Tabelle 11 und

Abbildung 2 dargestellt.

Tabelle 11: Fehlerhäufigkeiten Jahrgang 2012 in Prozent der Fragen. Die Zahlen sind leicht verändert zum Vergleich 2006-2012, da ein Fehlertyp mehr bewertet wurde (5b = Negative Antwortoption). Diese Tabelle ist in leicht abgewandelter Form Teil der Publikation [c].

	INT	DÄB	DMW	MMW	NHK	AMT
Fehlerfrei	28,3	66,2	13,0	34,7	8,0	11,8
1 Fehler	15	26,9	24,0	25,3	24,0	35,5
2 Fehler	20,8	6,9	32,0	16,7	22,0	30,9
3 oder mehr Fehler	35,8	0,0	31,0	23,3	46,0	21,8

Abbildung 2: Anteil formal fehlerhafter Fragen 2012. Diese Abbildung ist Teil der Publikation [a].



4.1.2 Formale Qualität der MC-Fragen im longitudinalen Vergleich 2006-2012

Untersucht wurde in drei ausgewählten Zeitschriften, ob und wie sich der Anteil fehlerhafter Fragen im Jahr 2012 verglichen zum Jahr 2006 verändert hat.

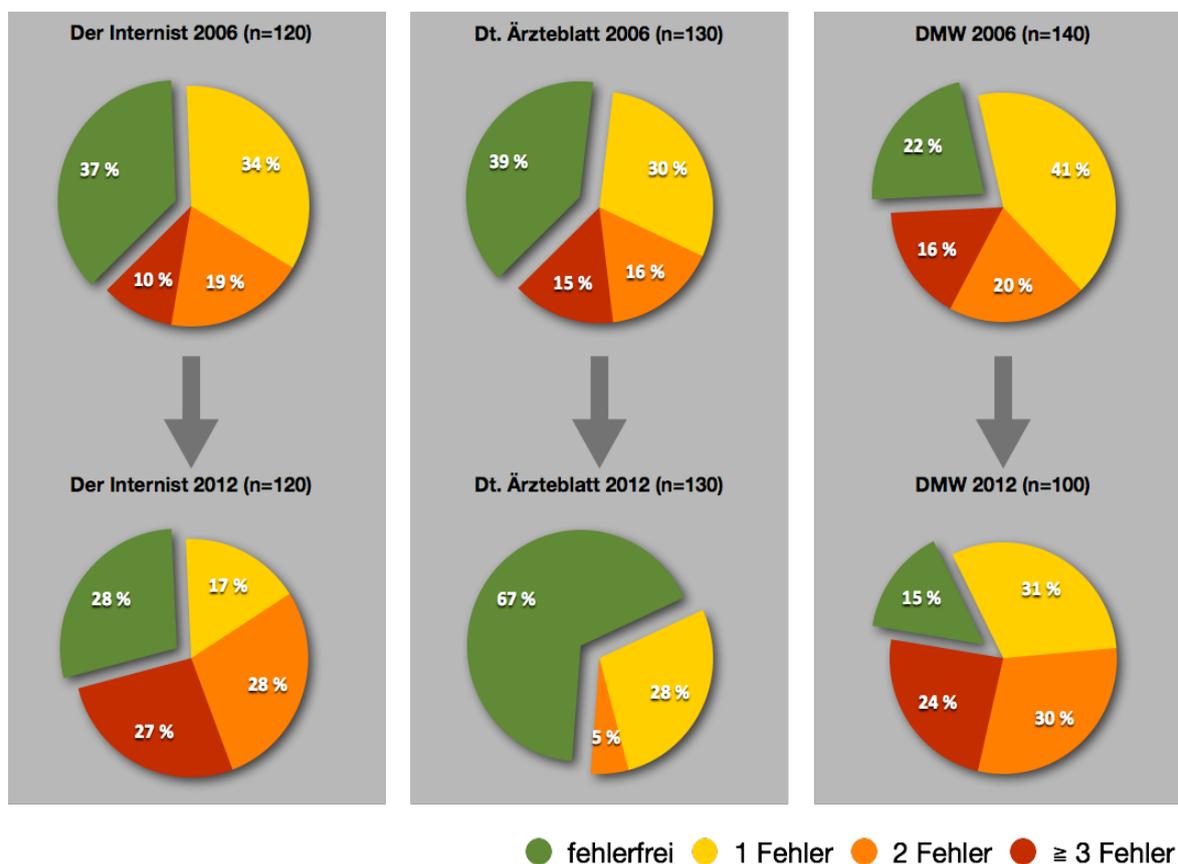
Es erfolgte die Einteilung in fehlerfreie Fragen bzw. Fragen mit einem, zwei, sowie drei oder mehr verschiedenen Formfehlertypen. Tabelle 12 zeigt die Ergebnisse der Analyse. Die Daten für 2006 stammen aus der Analyse von Kühne-Eversmann et al.³⁷.

Tabelle 12: Fehlerhäufigkeiten Jahrgang 2006 und 2012 in Prozent der Fragen (ohne Fehler 5b), Daten für 2006 aus³⁷. Diese Tabelle ist in abgewandelter Form Teil der Publikation [c].

	INT		DÄB		DMW	
	2006	2012	2006	2012	2006	2012
Fehlerfrei	36,7	28,3	39,2	66,9	22,1	15,0
1 Fehler	34,2	16,7	30,0	27,7	41,4	31,0
2 Fehler	19,2	28,3	16,2	5,4	20,0	30,0
3 oder mehr Fehler	10,0	26,7	14,6	0,0	16,4	24,0

Bei zwei der drei untersuchten Zeitschriften (DMW, INT) fand sich ein geringerer Anteil fehlerfreier Fragen im Vergleich zur Analyse 2006 (37 % und 22 % 2006 bzw. 28 % und 15 % 2012), während bei einer der Zeitschriften (DÄB) eine Verbesserung der formalen Qualität der MC-Fragen auffiel: Der Anteil fehlerhafter Fragen lag in dieser Studie 2012 deutlich niedriger als 2006 (61 % vs. 33 %); der Anteil an Fragen mit zwei Formfehlern nahm ebenfalls ab und es fand sich keine Frage mehr, die drei oder mehr Formfehler enthielt (Abbildung 3).

Abbildung 3: Anteil formal fehlerhafter Fragen im Vergleich 2006 – 2012. Diese Abbildung ist Teil der Publikationen [a] und [c].



4.1.3 Verteilung der Fehlerarten

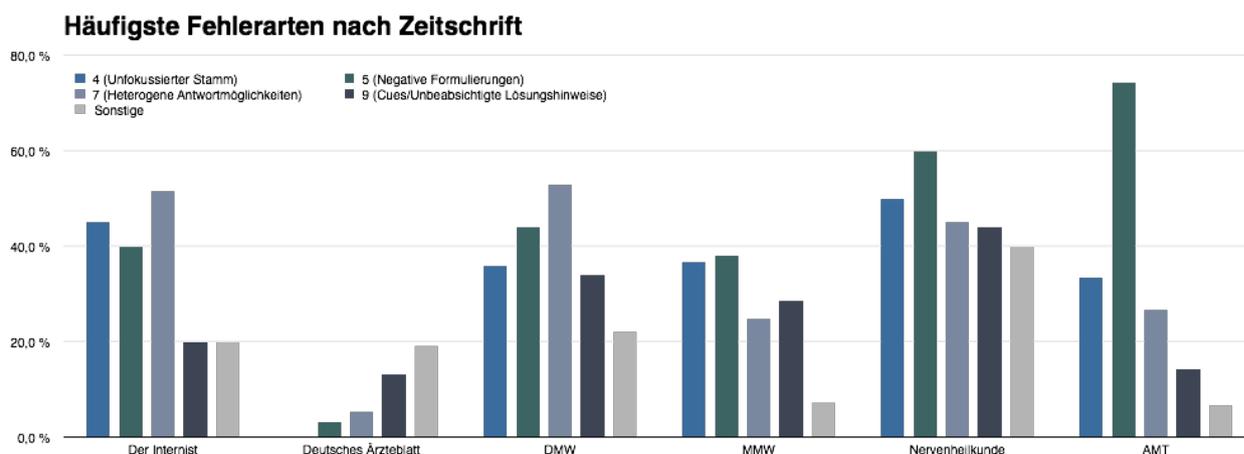
Untersucht wurde für alle sechs Zeitschriften die Häufigkeit und Verteilung der verschiedenen Fehlerarten für den Jahrgang 2012.

Die Ergebnisse sind in Tabelle 13 dargestellt. Es zeigte sich, dass insgesamt vier Fehlerarten (unfokussierter Stamm, negative Formulierungen, heterogene Antwortmöglichkeiten sowie unbeabsichtigte Lösungshinweise) den Großteil der gefundenen Fehler stellen. Allerdings unterscheidet sich die Verteilung der Fehlerarten zwischen den Zeitschriften teils deutlich. Abbildung 4 zeigt die Verteilung der vier häufigsten Fehlerarten.

Tabelle 13: Verteilung der Fehlerarten Jahrgang 2012, normiert pro 100 Fragen. Die 4 häufigsten Fehlerarten sind jeweils hervorgehoben. Diese Tabelle ist in abgewandelter Form Teil der Publikation [c].

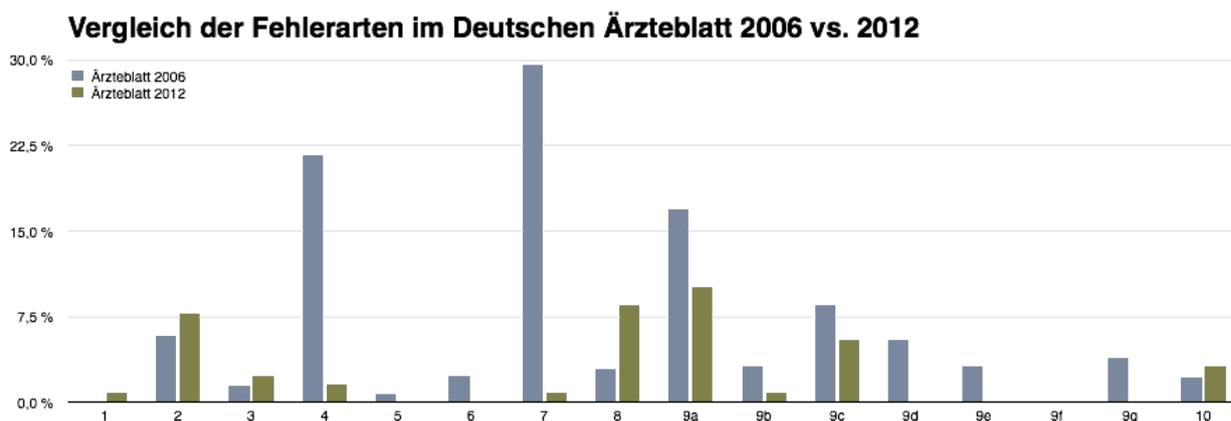
	INT	DÄB	DMW	MMW	NHK	AMT
1	0,0	0,0	0,0	0,0	17,0	0,0
2	0,8	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	0,0	0,0	0,0	0,0
4	45,0	0,0	36,0	36,7	50,0	33,3
5	40,0	3,1	44,0	38,0	60,0	74,2
6	0,8	0,8	0,0	0,0	2,0	0,0
7	51,7	5,4	53,0	24,7	45,0	26,7
8	14,2	10,0	10,0	3,3	13,0	5,8
9	20,0	13,1	34,0	28,7	44,0	14,2
9a	9,2	0,8	26,0	18,7	35,0	8,3
9b	2,5	7,7	4,0	2,7	4,0	0,0
9c	5,8	2,3	2,0	6,7	5,0	4,2
9d	0,8	1,5	1,0	0,7	0,0	0,0
9e	0,8	0,0	1,0	0,0	0,0	0,0
9f	0,8	0,0	0,0	0,0	0,0	0,8
9g	0,0	0,8	0,0	0,0	0,0	0,8
10	4,2	8,5	12,0	4,0	8,0	0,8

Abbildung 4: Auflistung der 4 häufigsten Fehlerarten nach Zeitschrift (2012). Zahlen normiert pro 100 Fragen. Diese Abbildung ist Teil der Publikation [c].



Da das Deutsche Ärzteblatt (DÄB) im Jahr 2012 im longitudinalen Vergleich zum Jahrgang 2006 weniger fehlerhafte MC-Fragen enthielt, wurde für diese Zeitschrift zusätzlich die Verteilung der Fehlerarten für die beiden untersuchten Jahrgänge gegenübergestellt. Hierbei zeigt sich bei der Verteilung der Fehlerarten eine Verschiebung: Die Fehlerart „Unfokussierter Stamm“ stellte 2006 mit über 20 % betroffenen Fragen den zweithäufigsten Fehler, während er 2012 kein einziges Mal vorkam. Der Formfehler „heterogene Antwortmöglichkeiten“, der besonders häufig in Kombination mit unfokussierten Fragestellungen erscheint, wurde 2006 noch bei über 30 % der Fragen bemängelt (häufigster Fehler), in der aktuellen Studie war er nur noch bei knapp 5 % der untersuchten Fragen vorhanden (siehe Abbildung 5).

Abbildung 5: Vergleich der Fehlerarten im Deutschen Ärzteblatt 2006 vs. 2012. Ohne Fehler 5b. Zahlen normiert pro 100 Fragen. Diese Abbildung ist Teil der Publikation [c].



4.2 Bewertung der Qualitätsmaßnahmen

4.2.1. Quantitative und qualitative Auswertung der Autorenhinweise

Die Autorenhinweise der Zeitschriften wurden, soweit vorhanden, bezüglich ihrer Ausführlichkeit sowie bezüglich des Inhalts analysiert.

Schriftliche Hinweise zur Erstellung der MC-Fragen in Form von Autorenhinweisen lagen für das Deutsche Ärzteblatt (DÄB, Deutscher Ärzteverlag), MMW – Fortschritte der Medizin (MMW, Springer) sowie die Deutsche Medizinische Wochenschrift (DMW, Thieme) vor.

Für „Der Internist“ (INT, Springer) lagen keine speziell gekennzeichneten Autorenhinweise zur Erstellung von MC-Fragen vor. In dem den Autoren zur Verfügung gestellten Musterbeitrag finden sich jedoch einige Hinweise zur Fragenerstellung in Form von Merkkästchen, die bei der Auswertung berücksichtigt wurden. Für die Zeitschrift „Nervenheilkunde“ (NHK) aus dem Schattauer Verlag existieren keine Autorenhinweise, die über die Webseite der Zeitschrift zugänglich waren. Nach persönlicher Auskunft der Redaktion existieren solche Hinweise zwar, jedoch nicht in schriftlicher Form.

Die Zeitschrift AMT - Arzneimitteltherapie (Apotheker Verlag) verwendet keine von den einreichenden Autor*innen verfassten Fragen. Die veröffentlichten Fragen werden von einer Fachredaktion der Zeitschrift nach Auswahl des Artikels erstellt. Nach persönlicher Auskunft der Redaktion gibt es zwar gewisse Vorgaben, was das Format der Fragen betrifft, jedoch existieren keine gesonderten Hinweise für die Vermeidung von formalen oder inhaltlichen Fehlern.

Die weitere Ausgestaltung der verlagsinternen Qualitätssicherung wie Schulung der Autor*innen, Review-Prozess, Schulung der Reviewer*innen etc., ist nicht für alle untersuchten Zeitschriften bekannt.

Um die Ausführlichkeit der Autorenhinweise zu objektivieren und einer statistischen Analyse zugänglich zu machen, wurde die Anzahl der zur Beschreibung verwandten Worte ausgewertet und gegenübergestellt. Die Ergebnisse dieser quantitativen Untersuchung sind in Tabelle 14 dargestellt. Die Autorenhinweise des DÄB waren die ausführlichsten der untersuchten Zeitschriften, gefolgt von den Autorenhinweisen der DMW.

Tabelle 14: Auswertung der Ausführlichkeit der Autorenhinweise nach Journal, quantitativ bestimmt nach Anzahl der verwendeten Worte. Diese Tabelle ist Teil der Publikation [c].

Fehlerart	INT	DÄB	DMW	MMW	NHK	AMT
1	0	7	46	13	0	0
2	0	0	0	4	0	0
3	0	0	0	0	0	0
4	0	67	91	29	0	0
5	6	11	50	18	0	0
6	0	10	0	14	0	0
7	5	18	0	8	0	0
8	0	0	5	0	0	0
9	0	313	64	71	0	0
9a	0	68	16	37	0	0
9b	0	31	0	22	0	0
9c	0	38	0	5	0	0
9d	0	57	0	0	0	0
9e	0	34	0	7	0	0
9f	0	42	48	0	0	0
9g	0	43	0	0	0	0
10	0	34	0	0	0	0

Fehlerart	INT	DÄB	DMW	MMW	NHK	AMT
Form	25	41	35	53	0	0
Inhalt	21	152	148	73	0	0
Hintergrund	0	177	0	0	0	0
Gesamt	57	830	439	283	0	0

Zusätzlich erfolgte eine inhaltlich-qualitative Auswertung. Die Autorenhinweise der sechs untersuchten Zeitschriften wurden, soweit vorliegend, anhand des in 3.1.5. beschriebenen Punkteschemas untersucht. Hierbei wurde für jede Fehlerart geprüft, welche der fünf definierten Kriterien erfüllt sind. Somit ergab sich pro Fehlerkategorie ein Wert von 0 bis 5. Die einzelnen Punktwerte wurden zu einer Gesamtsumme pro Zeitschrift addiert.

Die Ergebnisse sind in Tabelle 15 dargestellt. Auch bei dieser Auswertung zeigten sich, analog zur quantitativen Bewertung anhand der Wortanzahl, die Autorenhinweise des Deutschen Ärzteblatts am ausführlichsten.

Tabelle 15: Qualitative Auswertung der Autorenhinweise. Diese Tabelle ist Teil der Publikation [a].

Fehlerarten		INT	DÄB	DMW	MMW	NHK	AMT*
1	K-Typ (Kombinationsantworten)	1	2	4	2	0	0
2	AOTA-Typ (All of the above/Alle Antworten sind richtig)	0	0	0	2	0	0
3	NOTA-Typ (None of the above/Keine Antwort ist richtig)	0	0	0	0	0	0
4	Unfokussierter Stamm	0	4	1	3	0	0
5a	Negativ formulierter Fragenstamm	1	2	3	2	0	0
5b	Negativ formulierte Antwortmöglichkeit	0	0	1	1	0	0
6	Hohe Komplexität	0	1	1	2	0	0
7	Heterogene Antwortmöglichkeiten	1	0	0	1	0	0
8	Unplausibler Distraktor (Distraktor ist offensichtlich falsch, trivial, oder beantwortet inhaltlich nicht die Frage im Stamm)	0	0	1	0	0	0
9a	Schlüsselwort (Immer, ausschließlich, nie, in keinem Fall, ...)	0	3	1	3	0	0
9b	Grammatikalischer Fehler (Distraktor passt grammatikalisch nicht zum Stamm)	0	2	0	2	0	0
9c	Unterschiedlich lange Antworten (Eine Antwort ist länger oder ausführlicher)	0	3	0	1	0	0
9d	Zwei oder mehr ähnliche Antworten (von denen nur eine richtig sein kann)	0	2	0	0	0	0
9e	Verbale Assoziation zwischen Stamm und richtiger Antwort	0	2	0	1	0	0
9f	Wiederholte Nennung bei Auflistungen (Optionen werden auffällig häufig genannt)	0	2	2	0	0	0
9g	Cues zwischen Fragen (Eine Frage beinhaltet die Lösung für eine andere)	0	2	0	0	0	0
10a	Inhaltliche Fehler (z.B. Formulierung der Frage, Frage passt nicht zu den Antworten)	0	2	0	0	0	0
10b	Grammatikalische Fehler (Auch Rechtschreibung, Zeichensetzung)	0	2	0	0	0	0
„Gesamtpunkte“		3	29	14	20	0	0

4.2.2 Korrelation der Autorenhinweise zur formalen Qualität der Fragen

Kapitel 4.2.2 ist in gekürzter Form Teil der Publikation [c].

Die formale Qualität der MC-Fragen im Jahrgang 2012 wurde mit der Ausführlichkeit der verfügbaren Autorenhinweise verglichen.

In der Analyse der untersuchten Zeitschriften des Jahrgangs 2012 zeigten sich große Unterschiede in der formalen Qualität der MC-Fragen. Der Anteil formal fehlerhafter Fragen lag zwischen 33 % und 92 %; der Anteil der Fragen mit drei oder mehr verschiedenen Formfehlern zwischen 0 % und 46 % (siehe 4.1.1). Wird die Auswertung der Autorenhinweise der formalen Qualität der MC-Fragen der entsprechenden Zeitschriften gegenübergestellt, so fällt auf, dass die Fragen der Zeitschriften mit besonders elaborierten Autorenhinweisen auch weniger Formfehler aufwiesen. Dieser Zusammenhang war jedoch auf Ebene der Fachzeitschriften weder für die Gesamtfehlerzahl noch für einzelne Fehler statistisch signifikant (Berechnung mittels Chi-Quadrat-Test). Es lässt sich statistisch in der untersuchten Stichprobe also nicht nachweisen, dass das Vorhandensein von Autorenhinweisen einen Einfluss auf die Anzahl der Formfehler hat. Angesichts der mit n=6 Zeitschriften kleinen Stichprobe ist eine statistische Signifikanz eher nicht zu erwarten.

Es zeigte sich jedoch beim Vergleich der Ausführlichkeit der Autorenhinweise mit der Fehleranzahl in einer bivariaten Korrelation eine negative Korrelation zwischen der Anzahl der identifizierten Fehler und der Ausführlichkeit der Autorenhinweise (in Worten) ($r = -0,889$, $p = 0,018$). Werden im Vorhinein die nie oder nur sehr selten vorkommenden und die Form der Frage betreffenden Fehlertypen 1, 2 und 3 sowie der weniger inhaltlich geprägte Fehlertyp 10 (hierunter fallen Tippfehler, Grammatikfehler usw.) ausgeschlossen und die Korrelation nur für die Fehlertypen 4, 5, 6, 7, 8 und 9 gerechnet, so ergibt sich ebenfalls ein signifikantes Ergebnis ($r = -0,832$, $p = 0,04$).

4.3 Inhaltliche Analyse

4.3.1 Verteilung der Wissensniveaus

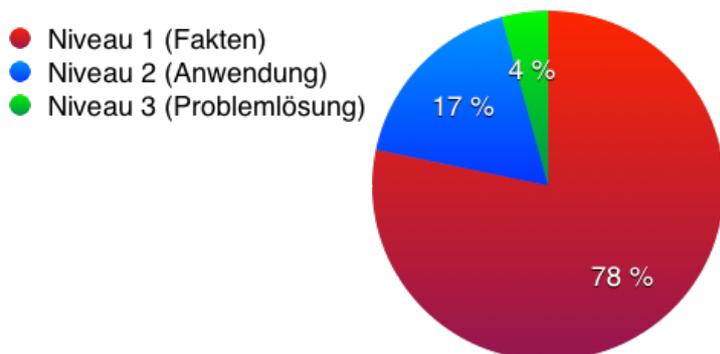
Ausgewählte Fragen wurden zusätzlich zur beschriebenen Untersuchung auf Formfehler auf ihren Inhalt untersucht. Für die inhaltliche Analyse wurden aufgrund oben geschilderter Gründe ausschließlich die Fragen des Deutschen Ärzteblatts (Jahrgänge 2012 und 2013) verwendet.

Im ersten Schritt wurde mit der zuvor erstellten Taxonomie die Art des abgeprüften Wissens analysiert. Bezüglich der Wissensniveaus der Fragen zeigte sich die in Tabelle 16 und **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellte Verteilung. Der überwiegende Teil der Fragen (78 %) prüft reines Faktenwissen ab, die Anwendung von Wissen (17 %) oder problemlösendes Denken (4 %), wie es im klinischen Alltag gefragt ist, ist nur für einen geringen Anteil der Fragen notwendig.

Tabelle 16: Verteilung der Wissensniveaus der MC-Fragen im Deutschen Ärzteblatt für die Jahrgänge 2012 und 2013. Angaben in Prozent, absolute Angaben in Klammern.

	Niveau 1 (Fakten)	Niveau 2 (Anwendung)	Niveau 3 (Problemlösung)	Anteil Fälle	Nicht einordbar
2012 (n = 130)	75,4 % (98)	20,8 % (27)	3,8 % (5)	13 % (17)	0,0 % (0)
2013 (n = 130)	80 % (104)	13,8 % (18)	4,6 % (6)	13 % (17)	1,5 % (2)
Gesamt (n = 260)	77,7 % (202)	17,3 % (45)	4,2 % (11)	13 % (34)	0,8 % (2)

Abbildung 6: Verteilung der Wissensniveaus der MC-Fragen im Deutschen Ärzteblatt für die Jahrgänge 2012 und 2013



4.3.2 Verteilung der Themen

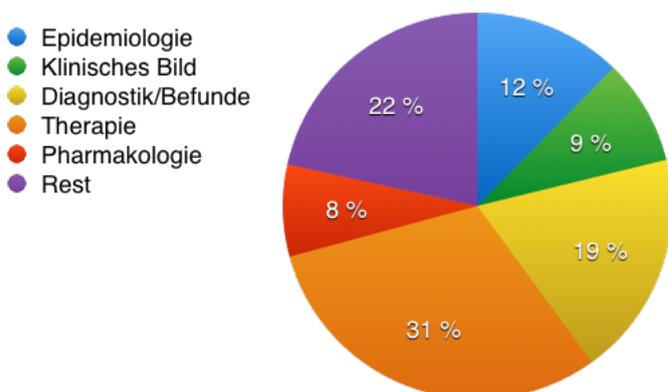
Zur erweiterten inhaltlichen Analyse wurde für jede Frage das Thema des abgeprüften Lernziels analysiert und die Themen in Kategorien eingruppiert. Tabelle 17 und Abbildung 7 zeigen die Themenverteilung der MC-Fragen im Deutschen Ärzteblatt.

Bei der Themenverteilung zeigte sich, dass die dem klinischen Alltag nahestehenden Themen Differentialdiagnose sowie „Vorgehen“ zusammen weniger als 5 % der Fragen und somit nur einen geringen Anteil ausmachten. Zwar liegt das Hauptaugenmerk der Fragen auf den klinisch wichtigen Themen Therapie und Diagnostik, die Fragen gehen jedoch nur selten über reines Faktenwissen hinaus (siehe 4.3.1). Auffallend ist der hohe Anteil an Fragen zur Epidemiologie (13,8 % (2012) bzw. 10,8 % (2013)).

Tabelle 17: Themenverteilung der MC-Fragen im Deutschen Ärzteblatt für die Jahrgänge 2012 und 2013. Diese Tabelle ist in leicht abgewandelter Form Teil der Publikation [b].

	Ätiologie/Pathogenese	Epidemiologie	Klinisches Bild	Diagnostik/Befunde	Therapie	Pharmakologie	Differentialdiagnose	Vorgehen	Physiologie	Sonstiges und N.E.
2012	7	18	14	27	38	13	5	2	0	6
2013	10	14	9	22	42	7	8	4	5	9
Insgesamt	17	32	23	49	80	20	13	6	5	15
Prozentual	6,5 %	12,3 %	8,8 %	18,8 %	30,8 %	7,7 %	5,0 %	2,3 %	1,9 %	5,8 %

Abbildung 7: Themenverteilung der MC-Fragen im Deutschen Ärzteblatt für die Jahrgänge 2012 und 2013



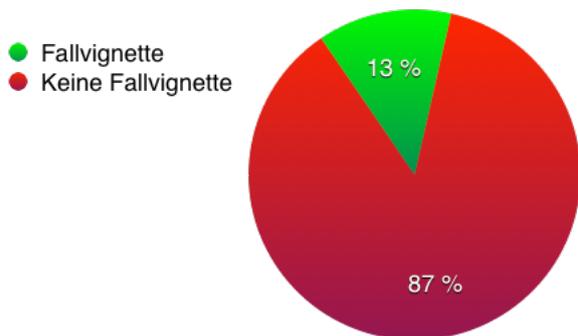
4.3.3 Verwendung von Fallvignetten und Bezug zum Wissensniveau

Zusätzlich wurde für alle Fragen untersucht, ob eine Fallvignette verwendet wurde.

Der Anteil der Fragen mit vorangestellter Fallvignette lag bei 13 % und zeigte sich somit insgesamt gering (Abbildung 8).

Im Gegensatz dazu steht die Anforderung, die viele Journals zur Verwendung von Fallvignetten oder konkreten klinischen Fragestellungen formulieren. Das Deutsche Ärzteblatt fordert in seinen Autorenhinweisen „mindestens 2 Fragen, die sich auf eine konkrete klinische Situation beziehen“ - somit wäre bei 10 Fragen pro Ausgabe ein Anteil Fallvignetten von mindestens 20 % zu erwarten.

Abbildung 8: Verwendung von Fallvignetten in den MC-Fragen im Deutschen Ärzteblatt für die Jahrgänge 2012 und 2013



Zusätzlich fiel auf, dass einige der Fragen, die Fallvignetten verwendeten, dennoch keine Fragestellung mit klinischem Bezug formulierten. Zusätzlich war in vier Fällen die Fallvignette nicht zur Beantwortung der Frage notwendig.

Beispiel für eine Frage mit „scheinbarer“ Fallvignette, bei der die klinische Beschreibung jedoch nicht zur Beantwortung der Frage benötigt wird. Es handelt sich nicht um die Abprüfung eines höheren Wissensniveaus:

Ein sonst gesunder 12-jähriger Junge klagt über Kopfschmerzen. Er besucht die 8. Klasse eines Gymnasiums und treibt viel Sport. Er ist entsprechend der StlKo-Empfehlung geimpft. Die Kopfschmerzen bestehen seit etwa 3 Tagen und treten unabhängig von äußeren Einflüssen auf. In der Familie sind keine neurologischen Erkrankungen bekannt.

Sie möchten den Patienten medikamentös behandeln. Welches Vorgehen ist richtig?

- a) Sie dosieren die Schmerzmittel altersabhängig.
- b) Metamizol ist nebenwirkungsarm und sollte bevorzugt werden.
- c) Triptane sind die erste Wahl.
- d) Paracetamol oder Ibuprofen sind als Schmerzmittel geeignet.
- e) Es sollten frühzeitig Opiate in die Schmerzmedikation einbezogen werden.

In diesem Fall ist die Fallvignette nicht integraler Bestandteil der Frage, der der Sicherstellung des Anwendungsbezugs und der Überprüfung von Handlungswissen dient. Das Vorhandensein von Fallvignetten alleine ist also nicht als ausreichend für den Praxisbezug einer MC-Frage anzusehen.

5. Diskussion

In der vorliegenden Arbeit wurde einerseits untersucht, wie gut die Strukturqualität von in der ärztlichen Fortbildung in Online- und Printmedien eingesetzten MC-Fragen ist und ob die Einführung von Qualitätssicherungsmaßnahmen einen Einfluss auf die formale Qualität der Fragen hatte. Andererseits wurde analysiert, inwiefern eine hohe Strukturqualität der MC-Fragen mit der Qualität der Autorenhinweise korreliert. Hierzu wurden MC-Fragen verschiedener Fachzeitschriften in Hinsicht auf ihre formale Qualität untersucht und Autorenhinweise der Verlage systematisch ausgewertet.

Des Weiteren untersuchte die vorliegende Arbeit die Fragen auf ihre inhaltlich-didaktische Ausgestaltung. Hierzu wurden die Fragen bezüglich ihres Inhalts und des abgeprüften Wissensniveaus sowie der Verwendung von Fallvignetten analysiert. Hierbei wurde erstmalig eine speziell für diese Anwendung entwickelte dreistufige Taxonomie zur Einstufung von MC-Fragen in Wissenskategorien verwendet, für die etablierte Taxonomien adaptiert wurden.

Es zeigte sich ein hoher Anteil formal fehlerhafter Fragen bei allen untersuchten Zeitschriften (min. 34 %, max. 92 %). Bei einer Fachzeitschriften zeigte sich in dieser Studie eine deutliche Verbesserung der formalen Qualität der MC-Fragen seit 2006 (61% fehlerhafte Fragen in 2006 vs. 33 % in 2012) und somit nach der Einführung von Autorenhinweisen. Bei Korrelation des Elaborationsgrades der Autorenhinweise mit dem Anteil der MC-Fragen, die formal fehlerhaft waren, zeigte sich bei Zeitschriften mit ausführlicheren Autorenhinweisen auch eine höhere Strukturqualität der MC-Fragen.

Bezüglich des Inhalts zeigte sich, dass der Großteil der Fragen (78 %) reines Faktenwissen abfragt, während Anwendungswissen (17 %) und problemlösendes Wissen (4 %) einen nur geringen Anteil der Fragen ausmachen. Fallvignetten wurden in nur 13 % der Fragen verwendet.

Die Ergebnisse dieser Analyse von sechs deutschsprachigen Fachzeitschriften stimmen mit den bisher publizierten Zahlen zum Anteil fehlerhafter MC-Fragen in der ärztlichen Fortbildung größtenteils überein.

Deutlich zeigen sich Unterschiede zwischen verschiedenen Zeitschriften, verschiedenen Verlagen und auch zwischen den jeweiligen von verschiedenen Autoren verfassten Beiträgen.

Unterschiede zeigen sich auch bei der Häufigkeit der verschiedenen Fehlerarten, die sich in den bisherigen Untersuchungen voneinander unterscheiden. Die häufigsten Fehler (unfokussierte Fragestellung, negative Formulierung, ungewollte Lösungshinweise) waren jedoch bisher in allen Studien zu diesem Thema ähnlich, auch wenn nicht alle publizierten Studien die gleichen Bewertungsinstrumente zugrunde gelegt und verschiedene Arbeiten den Fokus auch auf andere Eigenschaften der Fragen gelegt haben.

Es zeigte sich autorabhängig eine große Streuung bezüglich der Anzahl fehlerhafter Fragen pro Fachartikel. Dies lässt vermuten, dass Vorkenntnisse bzw. voran gegangene Schulungen zur Erstellung von MC-Fragen einen deutlichen Einfluss auf die formale Qualität der eingereichten Fragen haben.

Letztlich sollten Qualitätssicherungsmaßnahmen diese unterschiedlichen Vorkenntnisse jedoch ausgleichen und autorunabhängig ein gleichbleibend hohes Niveau an die formale Qualität sicherstellen. So konnte z.B. gezeigt werden, dass Relevanz, Trennschärfe und Reliabilitätskoeffizient von MC-Fragen deutlich verbessert werden konnten, indem ein strukturiertes Feedback von Prüfungsexpert*innen an Aufgabenautor*innen erfolgte und Besprechungen zu den Itemkennwerten ein geeignetes Instrument zur Qualitätsverbesserung von MC-Fragen sind⁶⁷.

In der vorliegenden Arbeit konnte eine Korrelation zwischen der Elaboration von Autorenhinweisen und formal hochwertigen MC-Fragen gezeigt werden. Dies legt die Vermutung nahe, dass die Veröffentlichung und Ausarbeitung von Autorenhinweisen einen positiven Einfluss auf die formale Qualität von MC-Fragen hat. Ein Zusammenhang konnte aufgrund der geringen Zahl der untersuchten Journals (n=6) statistisch jedoch nicht nachgewiesen werden. Es bleibt ferner unklar, welchen Einfluss ein verlags- bzw. zeitschrifteninterner Reviewprozess auf den Prozess der Qualitätssicherung hat. Es kann

jedoch vermutet werden, dass solche Verlage, die ausführliche und fundierte Autorenhinweise zur Aufgabenerstellung bereitstellen, tendenziell auch eine systematische interne Qualitätsprüfung durchführen und auch ihre Fragen-Reviewer*innen entsprechend schulen.

Das Deutsche Ärzteblatt bestätigte diese Einschätzung in einem Leserbrief, in dem auf den hohen Einfluss der redaktionellen Bearbeitung hingewiesen wurde. Diese werteten an ihrem eigenen Kollektiv aus, dass 75 % der 300 zuletzt eingereichten Fragen nicht den von der Redaktion vorgegebenen Richtlinien zur Erstellung guter MC-Fragen entsprachen⁶⁸.

Die Auswahl der Fachzeitschriften stellt eine Limitation der Studie dar, da durch die kleine Stichprobe (n=6) nur ein kleiner Teil der Zeitschriften untersucht wurde, die CME-Artikel anbieten. Bei der Auswahl der Fachzeitschriften wurde jedoch die Auflagenstärke berücksichtigt und es wurden Zeitschriften aus allen großen, deutschsprachigen medizinischen Verlagen ausgewählt, so dass eine möglichst repräsentative Auswahl untersucht wurde.

Die Ergebnisse dieser Studie zeigen eine hohe Übereinstimmung mit den bisher publizierten Daten zur formalen Qualität von MC-Fragen in der ärztlichen Fortbildung. Naturgemäß sind diese nicht uneingeschränkt untereinander vergleichbar, da sich die Methodik sowie die zugrunde gelegten Qualitätskriterien in den genannten Studien teilweise unterscheiden. Die hohe Übereinstimmung mit vorherigen Analysen spricht jedoch dafür, dass die Bewertung in der vorliegenden Arbeit mit hoher Objektivität gelungen ist. Da die Daten dieser Arbeit mittlerweile bereits einige Jahre alt sind, wäre eine erneute Untersuchung auf eine weitere Verbesserung der Strukturqualität seit dem Jahr 2012 wünschenswert.

Die Interpretation der Ergebnisse wird dadurch limitiert, dass der Einfluss anderer möglicher Variablen nicht sicher quantifizierbar ist. Hierzu gehören unter anderem das Vorhandensein sowie die Qualität verlags- bzw. zeitschrifteninterner Reviewverfahren, Schulungen für Reviewer bzw. Autoren oder Vorkenntnisse der Autor*innen zu formalen Gütekriterien von MC-Fragen. Für diese Studie lagen nicht für alle untersuchten Fachzeitschriften Informationen über die verlagsinterne Qualitätssicherung vor.

Mit der inhaltlich-didaktischen Ausgestaltung beschäftigen sich bisher nur wenige Untersuchungen. Eine Analyse der Wissensniveaus von CME-Fragen ist bisher noch nicht erfolgt.

Zwei Untersuchungen beschäftigen sich mit der Frage, ob CME-Fragen sich auf einen konkreten Fall beziehen^{38,41}. Der Anteil der als fallbasiert eingestuften Fragen entspricht jeweils etwa dem in dieser Arbeit gefundenen Anteil von Fallvignetten.

Die weitere inhaltliche Bewertung der Fragen bezüglich ihres Inhalts erfolgte jedoch in beiden Fällen nach Maßgabe der medizindidaktischen Gutachter*innen. Ob im Vorhinein eine Systematik zur Einstufung der Fragen erarbeitet wurde, ist dem Methodenteil der Publikationen leider nicht zu entnehmen. Somit ist die Auswertung der Fragen nicht nachvollziehbar, weshalb ein Vergleich zu den Ergebnissen dieser Studie nicht sinnvoll und möglich ist.

Hier liegt eine Stärke dieser Arbeit, die eine eigens für die Beurteilung von abgeprüften Wissensniveaus in MC-Fragen erarbeitete Taxonomie entwickelte, die leicht anzuwenden und gut nachvollziehbar ist. Diese basiert auf einer ausführlichen Literaturrecherche und orientiert sich an den bereits gut validierten Taxonomien von Bloom⁵⁴ sowie von Krathwohl und Anderson⁵⁵.

Limitiert wird die inhaltliche Untersuchung durch die relativ kleine Stichprobe, da für die zweite Analyse nur die Fragen einer einzigen Zeitschrift ausgewertet wurden. Dies ist darin begründet, dass eine sinnvolle inhaltliche Analyse nur bei fokussiert formulierten Fragen möglich ist, so dass die anderen Zeitschriften aufgrund des hohen Anteils unfokussierter Fragen in dieser Arbeit nicht analysiert wurden.

In der vorliegenden Untersuchung war sowohl der Anteil der Fragen mit Fallvignetten als auch der Anteil der Fragen, die höhere Wissensniveaus abprüfen, insgesamt klein. Im Gegensatz dazu steht die Anforderung, die viele Journals zur Verwendung von Fallvignetten oder konkreten klinischen

Fragestellungen formulieren. Zusätzlich fiel auf, dass einige Fragen, die Fallvignetten verwenden, dennoch keine Fragestellung mit klinischem Bezug formulieren bzw. dass in vielen Fällen die Fallvignette nicht zur Beantwortung der Frage notwendig war und somit eher als voran gestelltes Beiwerk, nicht jedoch als integraler Bestandteil der Frage zur Sicherstellung des Anwendungsbezugs und der Überprüfung von Handlungswissen dient. Das Vorhandensein von Fallvignetten alleine ist also nicht als ausreichend für den Praxisbezug einer MC-Frage anzusehen.

Die ärztliche Fortbildung mithilfe von CME-Artikeln in Zeitschriften ist weit verbreitet und vielgenutzt. Es kann allerdings diskutiert werden, ob die Fortbildung mittels MC-Fragen in Fachzeitschriften in ihrer jetzigen Form dem Wunsch und dem Anspruch der Ärzt*innen entspricht, im klinischen Alltag besser und kompetenter Entscheidungen treffen zu können. Zwar ist Faktenwissen Voraussetzung für die konkrete Entscheidung im Einzelfall, das reine Abfragen von diesem fördert jedoch nicht die Problemlöse-Kompetenz.

Die Fortbildungspflicht wurde nicht zuletzt zur Sicherung einer optimalen Patient*innenversorgung eingeführt. Es stellt sich also die Frage, inwiefern ärztliche Fortbildung das klinische Handeln einzelner Ärzt*innen und somit die Patientenversorgung beeinflusst. Alltagsrelevante Themen wie Differentialdiagnostik sowie fallbezogene Entscheidungen im klinischen Kontext sollten in der ärztlichen Fortbildung im Vordergrund stehen.

Dies bedeutet, dass auch an die Prüfungen im Kontext der ärztlichen Fortbildung der Anspruch gestellt werden muss, nicht nur reines Faktenwissen zu prüfen, sondern darüber hinaus auch die Anwendung des Wissens im klinischen Alltag/Kontext und insbesondere auch problemlösendes Wissen abzufragen.

Es besteht im Bereich der Fortbildung mithilfe von Online- und Printmedien weiterhin eine Diskrepanz zwischen den Fortbildungsthemen bzw. den CME-Fragen, die sich zumeist auf wissenschaftlich definierte medizinische Themen konzentrierten, und den von den Ärzt*innen gewünschten Inhalten, die sich Praxisnähe wünschen und klinische Entscheidungen, Differentialdiagnostik und Wissen für die Praxis als zentral ansehen^{12,46}.

Ob MC-Fragen im Rahmen der ärztlichen Fortbildung sinnvoll und notwendig sind, wird immer wieder diskutiert. Auch die Frage, ob die in der Literatur allgemein anerkannten Gütekriterien für gute MC-Fragen im Rahmen von Prüfungen auch in diesem Kontext berücksichtigt werden sollten, wird von einigen Autoren in Frage gestellt⁴². Broclain et al. konnten jedoch bereits 1998 zeigen, dass die Beantwortung von MC-Fragen im Anschluss an die Lektüre eines Fachartikels das Erinnern von Faktenwissen fördert und Inhalte des Artikels hierdurch besser im Gedächtnis bleiben⁶⁹.

Es stellt sich die Frage, welche Funktion MC-Fragen in der ärztlichen Fortbildung erfüllen und ob sie überhaupt geeignet sind, eine Qualitätssicherung zu gewährleisten. Von vielen Autor*innen und Ärzt*innen werden sie vermutlich weiterhin als reine Lesekontrolle und Pflichtübung angesehen, das abgeprüfte Wissen rückt in den Hintergrund. Das Deutsche Ärzteblatt gibt an, dass die von ihm veröffentlichten MC-Fragen Anlass für kritische Rückmeldungen sind. Gleichzeitig betont die Redaktion jedoch, dass aufgrund der hohen Teilnehmerzahlen aus Kapazitätsgründen keine alternative Prüfungsform angeboten werden kann⁷⁰.

Die MC-Fragen in der ärztlichen Fortbildung in Fachzeitschriften wurden in erster Linie als Lesekontrolle eingeführt, um einen personenbezogenen Nachweis der „Teilnahme“ an der Fortbildung zu ermöglichen⁴².

Die Wahrnehmung als Lesekontrolle wird dadurch bekräftigt, dass der Anteil der korrekten Antworten unter den Einsendungen bei den meisten Zeitschriften bei etwa 90% liegt. Beim Deutschen Ärzteblatt waren beispielsweise im Jahr 2008 nach dreieinhalb Jahren durchschnittlich 95% der eingeschickten Antworten korrekt⁷⁰.

Diese Einschätzung hat sich im Laufe der Zeit gewandelt, mittlerweile stehen Lernziele im Vordergrund und die Fragen nehmen immer mehr Prüfungscharakter an. Die Bundesärztekammer bezeichnet die Fragen in den aktuellen Empfehlungen zur ärztlichen Fortbildung explizit als „Lernerfolgskontrolle“.

Die vorliegende Untersuchung zeigt, dass es bereits Fragen gibt, die klinisches Wissen prüfen. Ihr Anteil sollte jedoch deutlich erhöht werden. Hierbei kann die Verwendung von Fallvignetten hilfreich sein. Die gute formale Qualität der MC-Fragen ist stets Voraussetzung für das Abprüfen höheren Wissens. Wenn tatsächlich die Anwendbarkeit des Erlernten das primäre Ziel der Fortbildung ist, müssen auch die Lernziele entsprechend formuliert sein. (z.B. "Der Arzt kann die verschiedenen Therapieoptionen in der Praxis mit dem Patienten diskutieren sowie die beste Therapie für seinen Patienten auswählen und durchführen" statt "Der Arzt kennt die verschiedenen Therapieoptionen".)

Die Tatsache, dass sich die formale Qualität der MC-Fragen im Deutschen Ärzteblatt im Gegensatz zu der Untersuchung aus dem Jahr 2006 deutlich verbessert hat, ist ein Indiz dafür, dass durch eine strukturierte Qualitätssicherung der Verlage eine strukturelle Verbesserung der in der ärztlichen Fortbildungen genutzten MC-Fragen erreicht werden kann. Dieser Effekt kann auf unterschiedliche Aspekte zurückgeführt werden. So empfiehlt es sich, detaillierte Autorenhinweise zur Erstellung guter MC-Fragen zu veröffentlichen, um die unterschiedlichen Vorkenntnisse der Fragenautor*innen auszugleichen. Des Weiteren sollten die Fragenautor*innen gezielt darauf achten und dafür geschult werden, klinisches Wissen abzufragen. Die eingereichten Fragen sollten vor Veröffentlichung von einem in der Erstellung von MC-Fragen geschulten Team gesichtet und überarbeitet werden, wobei Verbesserungsvorschläge an die Autor*innen zurückgemeldet werden.

Die Bundesärztekammer definiert die kontinuierliche Verbesserung der Behandlungsqualität und somit die Gewährleistung einer hohen Versorgungssicherheit für die Patient*innen als Ziel der ärztlichen Fortbildung²³. Um dieses Ziel zu erreichen, müssen die zur Überprüfung des Fortbildungserfolges verwendeten MC-Fragen den definierten und publizierten Qualitätskriterien entsprechen.

In den Richtlinien der Bundesärztekammer wird bei den Ausführungen zum Ziel der ärztlichen Fortbildung der Fokus auf den Praxisbezug gelegt. Insbesondere das explizit festgeschriebene Kriterium „Nutzen für den Patienten“ fordert klar einen unmittelbaren Effekt der Fortbildung der Ärzt*innen auf die Patientenversorgung.

Dementsprechend können die Fortbildungsveranstaltungen sowie die verwendeten Formate einer kritischen Prüfung unterzogen werden. In dem Dokument der BÄK werden zwar neben mündlichen und praktischen Prüfungen explizit MC-Fragen als mögliches Format zur Überprüfung von Wissenserwerb genannt, Qualitätskriterien für deren Erstellung oder Hinweise zur Einbettung in den klinischen Kontext werden jedoch nicht genannt. Es wird pauschal gefordert, dass "geeignete Lernerfolgskontrollen" stattfinden sollen.

Für den großen Anteil der Ärzt*innen in der Niederlassung (siehe 1.1.3), aber auch für die Kolleg*innen an kleineren Krankenhäusern ist aus Zeit- und Kostengründen die Fortbildung im Rahmen von Online-Veranstaltungen besonders attraktiv. Jedoch haben Online-Fortbildungsformate noch weitere Vorteile: So können verschiedene Medienformate eingesetzt werden, auch können die Lernenden sich aus verschiedenen Formaten die jeweils für ihr individuelles Lernverhalten passenden Angebote auswählen⁷¹.

Letztlich ist für die Qualität der Fortbildung weniger das Medium (Web-basiert vs. Druckprodukte wie Zeitschriften, Diskussionen über Chat, E-Mail oder persönlich, Fallkonferenz als interaktive Online-Veranstaltung oder als Kleingruppentreffen) sondern die inhaltliche und didaktische Ausgestaltung der jeweiligen Veranstaltung entscheidend. Eine Ausweitung des Angebots an Online-Fortbildungen ist somit insgesamt wünschenswert, wobei hier vor Allem die Schaffung qualitativ hochwertiger Angebote wie Online-Fallbesprechungen (Case Discussion), Problem-Based Learning, interaktive Angebote usw. im Vordergrund stehen sollte.

Für das Abprüfen praktischer Fähigkeiten sind schriftliche Prüfungen weiterhin nicht das Format der Wahl bzw. gänzlich ungeeignet. In diesem Sinne kann gefordert werden, dass für den Nachweis der

Fortbildungspunkte unterschiedliche Fortbildungsveranstaltungen besucht und beispielsweise auch Nachweise praktischer Fähigkeiten erbracht werden müssen.

MC-Fragen sind ein viel genutztes Format in der ärztlichen Fortbildung. Aufgrund der leichten und günstigen Anwendbarkeit von MC-Fragen und insbesondere aufgrund der schnellen, computergestützten und eindeutigen Auswertungsmöglichkeit, bleiben MC-Fragen ein wichtiger Baustein für Prüfungen im medizinischen Bereich, sei es in der Aus-, Weiter- oder Fortbildung.

Dennoch bleiben Limitationen bezüglich der Anwendung von Multiple Choice Fragen, die jedoch mit anderen Formaten abgedeckt werden könnten. Es sollte weiterhin kritisch diskutiert werden, ob MC-Fragen als alleiniges Format geeignet sind, eine Qualitätssicherung in der ärztlichen Fortbildung zu gewährleisten sowie letztlich auch wie gefordert die Behandlungsqualität zu verbessern. Andere, interaktivere Prüfungsformate wie z.B. Key Feature-Fälle sind geeignet, um Problemlösungen im klinischen Kontext abzufragen^{72,73} und könnten in Anbetracht der zunehmenden Verbreitung von Online-Angeboten auch in der ärztlichen Fortbildung eingesetzt werden.

Farmer et. al. konnten bereits 2005 nach einem Beobachtungszeitraum von 5 Jahren zeigen, dass Key Feature-Fälle geeignet sind, um bei der Zertifizierung australischer Hausärzt*innen klinisches Entscheidungswissen zu überprüfen⁷⁴.

Die Frage, ob die Veröffentlichung von Autorenhinweisen alleine eine Verbesserung der Fragenqualität erreichen kann, oder ob andere Maßnahmen der Qualitätssicherung durch die Verlage ergriffen werden sollten, ist noch nicht abschließend beantwortet. Hier ist weitere Forschung notwendig. Auch existieren bisher keine Daten, inwiefern die für den Artikel genannten Lernziele anschließend in den MC-Fragen abgebildet sind.

Letztlich sollte auch eine Untersuchung zu alternativen Prüfungsformen, wie beispielsweise der Verwendung von Key Feature-Fällen in der ärztlichen Fortbildung mittels Online- und Printmedien erfolgen, um deren Praktikabilität und Effekt auf den Lernerfolg zu untersuchen. Da solche alternativen Formate aktuell noch nicht in Verwendung sind, wäre eine Erweiterung der Prüfungsformate durch die Verlage wünschenswert.

Es stellt sich die Frage, inwiefern eine Neubewertung und Weiterentwicklung der ärztlichen Fortbildung in Deutschland erfolgen sollte und muss - insbesondere der Nutzen der ärztlichen Fortbildung bezüglich der Verbesserung der Patientenversorgung ist bisher zu wenig wissenschaftlich untersucht.

6. Zusammenfassung

Die ärztliche Fortbildung (CME) mithilfe von Artikeln und MC-Fragen in medizinischen Fachzeitschriften ist weit verbreitet und vielgenutzt. Mehrere Publikationen haben sich bereits mit der Analyse der formalen und teils auch der inhaltlich-didaktischen Qualität der verwendeten Fragen befasst. Seitens der Verlage wurden zunehmend Qualitätssicherungsmaßnahmen eingeführt, darunter die Veröffentlichung von Autorenhinweisen zur Erstellung guter MC-Fragen. Inhaltlich-didaktische Untersuchungen von MC-Fragen in der CME sind bisher nur in beschränktem Umfang erfolgt.

Die vorliegende Studie beschäftigt sich einerseits mit der Frage, wie sich die formale Qualität von MC-Fragen in CME-Angeboten in ausgewählten deutschen Fachzeitschriften darstellt, wie sie sich in den vergangenen Jahren entwickelt hat und welcher Zusammenhang zu Qualitätsmaßnahmen der Verlage im zeitlichen Verlauf besteht.

Hierzu wurden die MC-Fragen der CME-Artikel der Fachzeitschriften Deutsches Ärzteblatt, Der Internist, Deutsche Medizinische Wochenschau, MMW – Fortschritte der Medizin, Nervenheilkunde und Arzneimitteltherapie anhand einer vorher erstellten Fehlerliste bezüglich ihrer formalen Qualität untersucht und Autorenhinweise der Verlage systematisch ausgewertet. Die Ergebnisse wurden miteinander korreliert. Es zeigte sich ein hoher Anteil formal fehlerhafter Fragen bei allen untersuchten Zeitschriften (zwischen 34 und 92 %). Bei einer der untersuchten Zeitschriften (DÄB) konnte nach Einführung von Autorenhinweisen eine deutliche Verbesserung der formalen Qualität der MC-Fragen festgestellt werden (Der Anteil von 61% fehlerhafter Fragen in 2006 sank auf 33 % in 2012). Bei Korrelation der Elaboration der Autorenhinweise mit dem Anteil formal fehlerhafter MC-Fragen zeigte sich bei Zeitschriften mit ausführlicheren Autorenhinweisen auch eine höhere Strukturqualität der MC-Fragen.

Des Weiteren untersuchte die vorliegende Arbeit die Fragen auf ihre inhaltlich-didaktische Ausgestaltung. Hierzu wurden die Fragen bezüglich ihres Inhalts und des abgeprüften Wissensniveaus sowie der Verwendung von Fallvignetten analysiert. Hierbei wurde eine speziell für diese Anwendung entwickelte dreistufige Taxonomie zur Einstufung von MC-Fragen in Wissenskategorien verwendet.

Inhaltlich decken die MC-Fragen die erwartungsgemäß für den klinischen Alltag relevanten Themengebiete ab, jedoch zum überwiegenden Teil nur auf dem Niveau des reinen Faktenwissens (78%), zu einem kleineren Teil auf dem Niveau des Anwendungswissens (17 %). Fragen, die problemlösendes Denken erfordern, sind sehr selten (4 %). Fallvignetten wurden in 13 % der Fragen verwendet.

Es gibt bisher keine Empfehlungen dazu, welcher Anteil verschiedener Wissensniveaus sowie welche inhaltlichen Kategorien durch die Fragen abgedeckt werden sollten. Damit der Alltag der Zielgruppe besser abgebildet und somit die klinische Relevanz gesteigert wird, aber auch, um die Qualität der Fortbildung zu steigern, sollte der Anteil höherer Wissensniveaus in den Fragen deutlich angehoben werden. Hier können Fallvignetten hilfreich sein. Auch die Verwendung von Key Feature-Fällen in der ärztlichen Fortbildung sollte hierzu als Alternative oder Ergänzung zu MC-Fragen angeregt werden.

7. Anhang

Zusätzliche Materialien:

Tabelle 18: Vorgaben der Landesärztekammern, Einschränkungen bezüglich der Anerkennung von Fortbildungspunkten nach Kategorien.

Tabelle 19: Ausführliche Liste der Fehlerarten mit Beschreibung und Beispielen

Tabelle 18: Vorgaben der Landesärztekammern, Einschränkungen bezüglich der Anerkennung von Fortbildungspunkten nach Kategorien. Quelle: jeweils die Fortbildungsordnung der Ärztekammer in der Fassung des genannten Datums, PDF online abrufbar auf der Homepage der entsprechenden Landesärztekammern am 10. Oktober 2016.

Ärztekammer	Einschränkung (jeweils für den Zeitraum 5 Jahre)	Stand
Landesärztekammer Baden-Württemberg	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte (ab 01.01.2019)	18.12.2013
Bayerische Landesärztekammer	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	13.10.2013
Ärztekammer Berlin	Kategorie D und I: in der Summe höchstens 150 Punkte Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte Kategorie G: Höchstens 150 Punkte	09.04.2014
Landesärztekammer Brandenburg	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	14.12.2013
Ärztekammer Bremen	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	23.09.2013
Ärztekammer Hamburg	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	02.12.2013
Landesärztekammer Hessen	Kategorie A: Maximal 13 Punkte pro Tag Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte Kategorie G: Höchstens 12 Punkte pro Tag	16.09.2014
Ärztekammer Mecklenburg-Vorpommern	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	04.12.2013
Ärztekammer Niedersachsen	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	10.05.2014
Ärztekammer Nordrhein	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	23.11.2013
Landesärztekammer Rheinland-Pfalz	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	19.11.2015
Ärztekammer des Saarlandes	Kategorie E: Höchstens 50 Punkte	23.06.2004
Sächsische Landesärztekammer	Kategorie E: Höchstens 50 Punkte	30.11.2009

Ärztammer Sachsen-Anhalt	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte Kategorie G: Höchstens 50 Punkte	09.11.2013
Ärztammer Schleswig-Holstein	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	20.11.2013
Landesärztammer Thüringen	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	14.04.2014
Ärztammer Westfalen-Lippe	Kategorie E: Höchstens 50 Punkte Kategorie F: Höchstens 50 Punkte	29.03.2014

Tabelle 19: Ausführliche Liste der Fehlerarten mit Beschreibung und Beispielen. Die von Kühne-Eversmann et al.³⁷ veröffentlichte Tabelle wurde modifiziert und erweitert.

Nr	Fehlerart und Beschreibung	Beispiel
1	K-Typ Kombinierte Antwortmöglichkeiten	a) 1 und 3 sind richtig b) 4 und 5 sind richtig c) 1, 2 und 5 sind richtig d) Alle sind richtig e) Keine ist richtig.
2	AOTA Antwortoption „Alle Aussagen sind richtig“ ohne, dass ein K-Typ vorliegt	a) Alle Aussagen sind richtig.
3	NOTA Antwortoption „Keine Antwort ist richtig“ ohne, dass ein K-Typ vorliegt	a) Keine der Aussagen ist richtig.
4	Unfokussierter Stamm Der Fragenstamm ist so ungenau formuliert, dass das Themengebiet gar nicht oder kaum eingeschränkt wird. Der Stamm sollte möglichst beantwortbar sein, ohne die Wahlantworten zu sehen und die Frage klar auf ein Themengebiet einschränken (z.B. Therapie von Asthma)	Welche Aussage trifft zu? Welche Aussage zu Asthma bronchiale trifft zu? Welche Aussage zu Therapie und Prävention von Asthma trifft zu? (Hier werden mehrere Themen genannt)
5	Negative Formulierungen	
5 a	Negativer Stamm Der Fragenstamm ist negativ formuliert.	Welche Aussage ist falsch? Welche der folgenden Nebenwirkungen ist nicht typisch für Antiarrhythmika?
5 b	Negative Antwortmöglichkeiten Eine oder mehrere der Antwortmöglichkeiten sind negativ formuliert.	a) Männer sind nicht häufiger betroffen als Frauen. b) Ein CT bringt keinen diagnostischen Nutzen.
6	Hohe Komplexität Der Fragenstamm oder die Antwortmöglichkeiten sind sehr lang/komplex/ausführlich formuliert. Formulierungen, die eine hohe Konzentration beim Lesen der Frage erfordern.	
7	Heterogene Antwortoptionen	

7 a	<p>Inhaltlich heterogen</p> <p>Unterschiedliche Kategorien der Antwortmöglichkeiten, z.B. Ätiologie, Klinik, Diagnostik, Therapie. Dieser Fehler kommt meist in Kombination mit Fehler 4 vor.</p>	<p>Welche Aussage zum Asthma bronchiale trifft zu?</p> <ul style="list-style-type: none"> a) Es sind mehr Frauen als Männer betroffen. b) Budesonid hat in der Therapie keinen Stellenwert. c) Zur Diagnose sollte eine Lungenfunktionsprüfung mit Reversibilitätstest durchgeführt werden. d) Grundlage ist die bronchiale Obstruktion. e) Im Gegensatz zur COPD ist die Obstruktion beim Asthma reversibel.
7 b	<p>Strukturell heterogen</p> <p>Die Antwortmöglichkeiten unterscheiden sich in ihrer Struktur</p>	<p>Wie viele Schwangerschaften entstehen trotz Einnahme einer Mikropille?</p> <ul style="list-style-type: none"> a) 10 pro Jahr b) 3 pro 1000 Frauenjahre c) 10 pro 1000 Frauenjahre d) 15 pro 1000 Frauenjahre e) 25 pro 1000 Frauenjahre
8	<p>Unplausible Distraktoren</p> <p>Antwortmöglichkeiten, die trivial, unplausibel oder völlig unsinnig sind. Für Distraktoren sollten klare Gründe bestehen (z.B. häufige Fehlmeinungen, falsche Konzepte, veraltete Ansichten.)</p> <p>Als unplausible Distraktoren wurden auch Distraktoren gewertet, die zwar in sich plausibel sind, jedoch die Frage nicht beantworten, da sie das Thema verfehlen.</p>	<p>Der Stich welches Insektes führt am häufigsten zur Anaphylaxie?</p> <ul style="list-style-type: none"> a) Schmetterling <p>Welche Aussage zur Therapie des HCC mittels interventioneller Ablation trifft zu?</p> <ul style="list-style-type: none"> a) Die Therapie ist immer nebenwirkungsfrei. <p>Welche Aussage zur Therapie des Asthma bronchiale ist richtig?</p> <ul style="list-style-type: none"> a) Männer sind häufiger betroffen als Frauen.
9	<p>Cues</p> <p>Unbeabsichtigte Lösungshinweise, die scheinbar die richtige Antwort identifizieren. Es kann sich jedoch auch um eine Falschantwort handeln, die durch ihre Form richtig erscheint und den Prüfling so „auf die falsche Fährte“ lockt.</p>	
9 a	<p>Schlüsselwörter</p> <p>Absolute Begriffe wie immer, nur, nie, ausschließlich, allein.</p> <p>In der Medizin sind solche absoluten Aussagen nur sehr selten zu treffen, deshalb identifizieren sie die Antwort in denen sie vorkommen als eindeutig</p>	<p>Fieber ist immer das erste Symptom der Mononukleose.</p> <p>Morbus Crohn tritt nie gleichzeitig mit einer anderen Autoimmunerkrankung auf.</p>

	falsch (Aussagen über rechtliche Zusammenhänge ausgenommen).	
9 b	<p>Grammatikalische Fehler</p> <p>Ein oder mehrere Distraktoren passen grammatikalisch nicht zum Stamm und fallen so als Falschaussage auf</p>	<p>Welche Symptome sind typisch für die Mononukleose?</p> <ul style="list-style-type: none"> a) Ausschlag b) Fieber und Erschöpfung c) Lymphknotenschwellungen d) Hepato- und Splenomegalie e) Einschränkung der Nierenfunktion <p>Welcher Parameter sollte bei Verdacht auf Hypothyreose bestimmt werden?</p> <ul style="list-style-type: none"> a) TSH b) T3 und T4
9 c	<p>Unterschiedlich lange Antworten</p> <p>Einer der Antworten, meist die richtige, ist ausführlicher/länger/spezifischer als die anderen Antwortmöglichkeiten.</p>	<p>Welches Symptom ist eine typische Nebenwirkung von Seroquel?</p> <ul style="list-style-type: none"> a) Kopfschmerzen b) Fieber c) Ekzem d) Leberfunktionsstörungen (Anstieg der Leberenzyme, Ikterus, hepatische Enzephalopathie) e) Neurologische Ausfälle
9 d	<p>Zwei oder mehr ähnliche Antworten</p> <p>Die Aufmerksamkeit wird auf zwei oder mehr ähnliche Antworten gelenkt, von denen nur eine richtig sein kann. Die anderen Antwortmöglichkeiten können außer Betracht gelassen werden.</p>	<p>Bei Hepatitis...</p> <ul style="list-style-type: none"> a) Handelt es sich um eine Stoffwechselstörung b) Steigt der GOT-Wert an c) Nimmt der GOT-Wert ab d) Bleibt der GOT-Wert gleich e) Ist eine Therapie mit Betablockern indiziert
9 e	<p>Verbale Assoziation zwischen Antwort und Stamm</p> <p>Wörter oder Satzteile des Stamms werden in einer der Antwortmöglichkeiten wiederholt, die dadurch als die (vermeintlich) richtige identifiziert wird.</p>	<p>Welcher Test eignet sich zum Ausschluss eines Cushing-Syndroms?</p> <ul style="list-style-type: none"> a) Dexamethasonhemmtest b) Aldosteron-Renin-Quotient c) Cortisol-Bestimmung d) Abdomensonographie e) Katecholamine im 24h-Urin
9 f	<p>Wiederholte Nennung bei Auflistungen</p> <p>Eine der Antworten beinhaltet Elemente, die am häufigsten in den anderen Antwortmöglichkeiten genannt wurden.</p>	<p>Was benötigt man für die Berechnung der GFR?</p> <ul style="list-style-type: none"> a) Kreatinin, Harnstoff, Geschlecht b) Kreatinin, Alter, Geschlecht c) Kreatinin, Alter, Gewicht d) Kreatinin, Albumin, Geschlecht e) Kreatinin, Alter, Urinmenge
9 g	<p>Cues zwischen Fragen</p>	<p>Frage 1: Nach welchem Klassifikationssystem wird die Refluxösophagitis eingeteilt?</p>

	<p>Der Stamm oder eine Antwortmöglichkeit einer Frage beinhaltet die Information zur Lösung einer anderen Frage der Prüfung.</p>	<p>Frage 8: Bei einem Patienten mit einer leichten Refluxösophagitis (Los Angeles B) wird...</p>
<p>10 a</p>	<p>Inhaltliche Fehler</p> <p>Inhaltliche Fehler in der Fragenstellung.</p> <p>Zu Beispiel 1:</p> <p>Gemeint ist wohl nicht die Therapie von Frauen allgemein, sondern speziell von Frauen im gebärfähigen Alter bzw Schwangeren.</p> <p>Zu Beispiel 2:</p> <p>Die richtige Antwortmöglichkeit E enthält nicht nur diagnostische Maßnahmen, das Legen eines venösen Zugangs fällt nicht darunter. Also ist die Fragestellung falsch formuliert und es handelt sich somit um einen inhaltlichen Fehler.</p>	<p>„Welche Aussage zur antikonvulsiven Therapie bei Frauen ist falsch?</p> <ul style="list-style-type: none"> a) Das Fehlbildungsrisiko unter Valproat bis 1 000 mg/d beträgt ca. 4 bis 5%. b) Für Lamotrigin besteht erst bei höheren Dosierungen ein Fehlbildungsrisiko. c) Bei chronischer Einnahme von Antikonvulsiva sind orale gegenüber alternativen Verhütungsmethoden zu bevorzugen. d) Zum Schutz vor einer Spina bifida sollte Folsäure frühzeitig ergänzt werden. e) Medikamentenumstellungen bei eingetretener Schwangerschaft sollten vermieden werden.“⁷⁵ <p>„Welche diagnostischen Maßnahmen werden in der präklinischen Akutversorgung empfohlen?</p> <ul style="list-style-type: none"> a) Schweregrad der Defizite erfassen und Steroide geben. b) Zeitpunkt des Ereignisbeginns erfragen und Acetylsalicylsäure geben. c) Blutglukose bestimmen und zuerst ein EEG veranlassen. d) Vitalparameter erfassen und Acetylsalicylsäure geben. e) Venösen Zugang am nichtparetischen Arm legen und Vitalparameter erfassen.“⁷⁶
<p>10 b</p>	<p>Grammatikalische Fehler und Sonstiges</p> <p>Grammatikalische Fehler im Fragenstamm oder den Antwortmöglichkeiten. Hierzu zählen Tippfehler, Kommafehler, falsche Grammatik usw.</p>	<p>Welche Aussagen trifft zu?</p> <p>Die Therapie der Therapie der Tuberkulose...</p>

8. Literaturverzeichnis

1. Bundesärztekammer. (Muster-)Fortbildungsordnung 2013 der Bundesärztekammer (Arbeitsgemeinschaft der deutschen Ärztekammern), in der Fassung vom 29.05.2013. Online abrufbar unter <http://www.bundesaerztekammer.de>, letzter Zugriff am 26.09.2020.
2. Kaufmann, C. R. Computers in surgical education and the operating room. *Ann Chir Gynaecol* **90**, 141–146 (2001).
3. Masic, I., Novo, A., Deljkovic, S., Omerhodzic, I. & Piralic, A. How to assess and improve quality of medical education: lessons learned from Faculty of Medicine in Sarajevo. *Bosn J Basic Med Sci* **7**, 74–78 (2007).
4. Hall, J. C. & Platell, C. Half-life of truth in surgical literature. *Lancet* **350**, 1752–1 (1997).
5. Ramsey, P. G., Carline J. D., Inui T. S., Larson E. B., LoGerfo J. P., Norcini J. J. & Wenrich M. D. Changes over time in the knowledge base of practicing internists. *JAMA* **266**, 1103–1107 (1991).
6. Levine, H. G., Moore, D. E. & Pennington, H. C. Continuing education for health professionals: developing, managing and evaluating for maximum impact on patient care. In: Green JS, editor. *Evaluating continuing education and Outcomes*. San Francisco, CA: Jossey Bass. (1984).
7. Bundesärztekammer. (Muster-)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte, in der Fassung der Beschlüsse des 121. Deutschen Ärztetages 2018 in Erfurt. Online abrufbar unter www.bundesaerztekammer.de. Letzter Zugriff am 26.09.2020.
8. Griebenow, R. Zertifizierte Fortbildung in Printmedien 2007 - Eine Standortbestimmung. *Dtsch med Wochenschr* **132**, 1937–1938 (2007).
9. Homepage der Bundesärztekammer: Die ärztliche Versorgung in Deutschland. Ärztestatistik der Bundesärztekammer, online abrufbar <https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/>, letzter Zugriff am 15.06.2020.
10. Bundesärztekammer. Quantitative Auswertungen von anerkannten Fortbildungsveranstaltungen auf Grundlage des EIV-Datenbestandes. Zahlen abgerufen am 19.12.2012. Bundesärztekammer, Dezernat I. Große Senatssitzung, Januar 2013. Zur Verfügung gestellt durch Prof. Weidringer, Bayerische Landesärztekammer.
11. Bundesärztekammer. Ärztestatistik zum 31. Dezember 2011, erstellt von der Bundesärztekammer. Online abrufbar unter www.bundesaerztekammer.de, letzter Zugriff am 18.10.2020.
12. Gerlach, F. M. & Beyer, M. Ärztliche Fortbildung aus der Sicht niedergelassener Ärztinnen und Ärzte-repräsentative Ergebnisse aus Bremen und Sachsen-Anhalt. *Z Arztl Fortbild Qualitatssich* **93**, 581–590 (1999).
13. MMW - Fortschritte der Medizin, online abgerufen am 18.10.2016 unter <http://www.springer.com/medicine/journal/15006>
14. Marx, C. cme im Deutschen Ärzteblatt--/43 000 Teilnehmer--sehr gute Bewertung. *Dtsch Arztebl* **102**, 2453–2456 (2005).
15. Christ, H., Franklin, J., Griebenow, R. & Baethge, C. An Analysis of 2.3 Million Participations in the Continuing Medical Education Program of a General Medical Journal: Suitability, User Characteristics, and Evaluation by Readers. *J Med Internet Res* **19**, e49 (2017).
16. Bundesärztekammer. Quantitative Auswertungen von anerkannten Fortbildungsveranstaltungen auf Grundlage des EIV-Datenbestandes. Große Senatssitzung, Januar 2013. Zur Verfügung gestellt durch Dr. Weidringer, Bayerische Landesärztekammer.
17. Scolapio, J. S., Deluise, W., Al-Haddad, M., Bianchi, L. & Cesario, K. Continued Medical Education Provided by Medical Journals: The 'Red Journal's' Experience. *Am J Gastroenterology* **102**, 1590–1593 (2007).

18. Seelig, C. B. Changes Over Time in the Knowledge Acquisition Practices of Internists. *South Med J* **86**, 780–783 (1993).
19. Norman, G. R., Shannon, S. I. & Marrin, M. L. The need for needs assessment in continuing medical education. *BMJ* **328**, 999–1001 (2004).
20. Sibley, J. C., Sackett D. L., Neufeld V., Gerrard B., Rudnick K. V., Fraser W. A randomized trial of continuing medical education. *N Engl J Med* **306**, 511–515 (1982).
21. Davis, D. A., Mazmanian P. E., Fordis M., Van Harrison R., Thorpe K. E., Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* **296**, 1094–1102 (2006).
22. Bundesärztekammer. (Muster-)Satzungsregelung Fortbildung und Fortbildungszertifikat, Stand: 20.05.2004. Online abrufbar unter www.bundesaerztekammer.de, letzter Zugriff am 26.09.2020.
23. Bundesärztekammer. Empfehlungen zur ärztlichen Fortbildung. (2015). PDF abrufbar unter https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Fortbildung/EmpfFortb_20150424.pdf. Letzter Zugriff am 13.12.2019.
24. Möltner, A., Duelli, R., Resch, F., Schultz, J. H. & Jünger, J. Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten. *GMS Z Med Ausbild* **27 (3)**, Doc44. (2010).
25. Tamir, P. Positive and negative multiple choice items: How different are they? *Stud Educ Evaluation* **19**, 311–325 (1993).
26. Case, S. M. & Swanson, D. B. Constructing written test questions for the basic and clinical sciences. 3rd Edition (rev). National Board of Medical Examiners, Philadelphia, PA. (2002).
27. Krebs, R. Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung. Bern: Institut für Medizinische Lehre IML, Abteilung für Ausbildungs- und Examensforschung AAE (2004).
28. Downing, S. M. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Acad Med* **77**, S103–S104 (2002).
29. Downing, S. M. The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education. *Adv Health Sci Educ Theory Pract* **10**, 133–143 (2005).
30. Tarrant, M. & Ware, J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ* **42**, 198–206 (2008).
31. Wass, V., Van der Vleuten, C., Shatzer, J. & Jones, R. Assessment of clinical competence. *Lancet* **357**, 945–949 (2001).
32. Schuwirth, L. W. T., Verheggen, M. M., Van Der Vleuten, C. P. M., Boshuizen, H. P. A. & Dinant, G. J. Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ* **35**, 348–356 (2001).
33. Miller, G. E. The assessment of clinical skills/competence/performance. *Acad Med* **65**, S63–7 (1990).
34. Koch, H. Fortschreibung der Richtlinie zum Erwerb des freiwilligen Fortbildungszertifikates. *Bayer Aztebl* **3**, 90 (2009).
35. Empfehlungen der Ärztekammern zur Durchführung von ärztlichen Fortbildungsveranstaltungen, Nordrheinische Akademie für ärztliche Fort- und Weiterbildung, Stand 2013. Online abrufbar unter www.aekno.de. Letzter Zugriff am 18.10.2016. In der Folge ersetzt durch Richtlinie zur Fortbildungsordnung der Ärztekammer Nordrhein vom 06.07.2016. Letzter Zugriff am 26.09.2020
36. Stagnaro-Green, A. S. & Downing, S. M. Use of flawed multiple-choice items by the New England Journal of Medicine for continuing medical education. *Med Teach* **28**, 566–568 (2006).

37. Kühne-Eversmann, L., Nussbaum, C., Reincke, M. & Fischer, M. R. CME-Fortbildungsangebote in medizinischen Fachzeitschriften: Strukturqualität der MC-Fragen als Erfolgskontrollen. *Med Klin* **102**, 993–1001 (2007).
38. Rotthoff, T., Fahron, U., Baehring, T. & Scherbaum, W. A. Die Qualität von CME-Fragen in der ärztlichen Fortbildung – eine empirische Studie. *Z Arztl Fortbild Qualitatssich* **101**, 667–673 (2008).
39. Deutsches Ärzteblatt: Regeln zur Konstruktion von MC-Aufgaben, Stand Januar 2017. Online abrufbar unter www.aerzteblatt.de. Letzter Zugriff am 26.09.2020.
40. Gutmann, A., Degirmenci, Ü., Kreil, S., Kornhuber, J. & al, E. Verbesserung der formal-didaktischen Qualität der CME-Fragen aus Der Nervenarzt. *Nervenarzt* **81**, 1363–1367 (2010).
41. Öchsner, W. & Böckers, A. Ärztliche Fachzeitschriften als implizite Vorbilder für gute Multiple-Choice-Fragen—in welchem Maß erfüllen die Tests der continuous medical education inzwischen Qualitätskriterien? *Z Evid Fortbild Qual Gesundheitswes* **107**, 468–474 (2013).
42. Nobbe, H., Lösche, P. & Griebenow, R. Struktur der Fortbildungsartikel und Fragen zur Lernerfolgskontrolle am Beispiel der Zeitschriften Der Orthopäde und Der Unfallchirurg. *Med Klin* **103**, 14–19 (2008).
43. Knowles, M. S. *The Adult Learner*. Houston, TX. Gulf Publishing Company, Book Division, (1973).
44. Knowles, M. S. *Andragogy in Action: Applying Modern Principles of Adult Learning*. Jossey Bass. (1984).
45. Duffy, F. D. & Holmboe, E. S. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *JAMA* **296**, 1137–1139 (2006).
46. Ebell, M. H., Cervero, R. & Joaquin, E. Questions asked by physicians as the basis for continuing education needs assessment. *J Contin Educ Health Prof* **31**, 3–14 (2011).
47. Branthwaite, A., Ross, A., Henshaw, A. & Davie, C. Continuing education for general practitioners. *J R Coll Gen Pract Ocas Pap* **38**, 1–39 (1988).
48. Nationaler Kompetenzbasierter Lernzielkatalog Medizin 2015. (2015). Abrufbar unter www.nklm.de. Letzter Zugriff am 26.09.2020.
49. Boshuizen, H. & Schmidt, H. G. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cogn Sci* **16**, 153–184 (1992).
50. Elieson, S. W. & Papa, F. J. The effects of various knowledge formats on diagnostic performance. *Acad Med* **69**, S81–3 (1994).
51. Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C. & Patel, V. L. Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *J Educ Psychol* **81**, 610–619 (1989).
52. Carline, J. D., Inui, T. S., Larson, E. B., LoGerfo, J. P. & Ramsey, P. G. The knowledge base of certified internists. Relationships to training, practice type, and other physician characteristics. *Arch Intern Med* **149**, 2311–2313 (1989).
53. Glaser, R. Education and thinking: The role of knowledge. *Am Psychol* **39**, 93–104 (1984).
54. Bloom, B. S., Englehart, M. D., Furst E. J., Hill W. H. & Krathwohl, D. R. *The Taxonomy of Educational Objectives. Handbook I: The cognitive domain*. New York: David McKay Co. Inc. (1956).
55. Krathwohl, D. R. A revision of Bloom's taxonomy: An overview. *Theory Into Practice* **41**, 212–218 (2002). Quelle der Übersetzung: http://ep.elan-ev.de/wiki/Lernziel#cite_note-anderson-2. Letzter Zugriff am 26.09.2020
56. Bürgi, H. *et al.* Swiss catalogue of learning objectives for undergraduate medical training. (2008). Abrufbar unter http://sclo.smifk.ch/downloads/sclo_2008.pdf. Letzter Zugriff am 26.09.2020.

57. Medizinische Fakultät der Ludwig-Maximilians-Universität München. Lernzielkatalog AINS. (2011). Online abgerufen unter <https://www.mecum.med.uni-muenchen.de>, letzter Zugriff am 15.02.2016.
58. Medizinische Fakultät der Universität Hamburg, Prodekanat für Lehre. Hamburger Lernzielkatalog KliniCuM, Stand 23.10.2009. PDF online verfügbar unter http://www.uke.de/dateien/zielgruppen/studium-lehre/dokumente/lernzielkatalog_091104_mat.pdf, letzter Aufruf am 13.12.2019.
59. Karsten, G. Lernzielkatalog Medizin, zweiter Studienabschnitt. Medizinische Fakultät der Christian-Albrechts-Universität zu Kiel. 2. Aktualisierte Auflage (2011). PDF online verfügbar unter https://www.uni-kiel.de/anaesthesie/docs/Studenten/Lernzielkatalog_Medizin_CAU-Kiel_2te-Auflage_Januar-2011.pdf, letzter Aufruf am 13.12.2019.
60. Weitz, G et. al. Lernzielkatalog Innere Medizin der medizinischen Universität zu Lübeck. 3. Auflage (2014). PDF online verfügbar unter https://www.uni-luebeck.de/fileadmin/uzl_med/pdf/studierende/PraktischesJahr/Lernziele/Innere_Medizn.pdf, letzter Aufruf am 13.12.2019.
61. Haladyna, T. M. & Rodriguez, M. C. & Downing, S.M. Developing and validating multiple choice test items. NY: Routledge. (2013).
62. Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ* **15**, 309–333 (2002).
63. Kalf, R., Ewald, C., Waschke, A., Gobisch, L. & Hopf, C. Degenerative lumbar spinal stenosis in older people: current treatment options. *Dtsch Arztebl* **110**, 613–23– quiz 624 (2013).
64. Koch, A., Tacke, F., Streetz, K. L. & Trautwein, C. Behandlung chronisch entzündlicher Darm-erkrankungen in der Intensivmedizin. *Dtsch med Wochenschr* **137**, 1107–18– quiz 19–20 (2012).
65. Palmer, E. J. & Devitt, P. G. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Med Educ* **7**, 49 (2007).
66. Pannicke, N., Schramm, C. & Lohse, A. W. Autoimmunerkrankungen der Leber. *Internist* **53**, 943–956 (2012).
67. Kropf R, Krebs, R., Rogausch, A. & Beyeler, C. Auswirkungen angeleiteter Itemanalysebesprechungen mit Dozierenden auf die Qualität von Multiple Choice Prüfungen *GMS Z Med Ausbild.* **27 (3)**, Doc46.(2010).
68. Marx, C. & Baethge, C. Leserbrief zu ‘Welchen Einfluss haben Autorenrichtlinien der Verlage auf die formale Qualität von CME-Fragen in der Ärztlichen Fortbildung?; 121:29-35. *Z Evid Fortbild Qual Gesundheitswes* **126**, 98 (2017).
69. Broclain, D., Mura P., Nivet-Carre D., Barbot J., Nory F., Bardelay G. Quasi-experimental study on the effectiveness of the readers’ test in the medical journal La revue Prescrire. *J Contin Educ Health Prof* **18**, 47–57 (1998).
70. Baethge, C. & Marx, C. CME im deutschen Ärzteblatt - eine Auswertung nach dreieinhalb Jahren. *Dtsch Arztebl* **22**, 229–232 (2008).
71. Wutoh, R., Boren, S. A. & Balas, E. A. eLearning: a review of Internet-based continuing medical education. *J Contin Educ Health* **24**, 20–30 (2004).
72. Page, G., Bordage, G. & Allen, T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* **70**, 194 (1995).
73. Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F. & Jünger, J. A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach* **27**, 450–455 (2009).
74. Farmer, E. A. & Hinchy, J. Assessing general practice clinical decision making skills: the key features approach. *Aust Fam Physician* **34**, 1059–1061 (2005).
75. Klamer S, Weber YG, Lerche H. Epilepsie. Wichtige Therapieempfehlungen für Praxis und Klinik. *Nervenheilkunde.* **31**, 263-269 (2012).

76. Nolte CH, Endres M. Akutversorgung des ischämischen Schlaganfalls. *Internist*. **53**, 585–94 (2012).

Teile dieser Arbeit wurden vorab veröffentlicht:

[a] **Drossard S**, Kühne-Eversmann L, Bauer D, Fischer M. (2013) CME-Artikel in medizinischen Fachzeitschriften: Erfolg von Qualitätsmaßnahmen. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Graz, 26.-28.09.2013. Düsseldorf: German Medical Science GMS Publishing House. DocP05_01. DOI: 10.3205/13gma038

[b] **Drossard S**, Bauer D, Fischer M, Kühne-Eversmann L. (2014) Analyse von Wissens Ebenen und Inhaltskategorien in CME-Aufgaben. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg, 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House. DocP471. DOI: 10.3205/14gma186

[c] **Drossard S**, Bauer D, Fischer M, Kühne-Eversmann L. (2017) Welchen Einfluss haben Autorenrichtlinien der Verlage auf die formale Qualität von CME-Fragen in der ärztlichen Fortbildung? *Z. Evid.Fortbild. Qual. Gesundh.wesen (ZEFQ)* 121:29–35. DOI: 10.1016/j.zefq.2017.03.015

9. Danksagung

Ich danke meinem Doktorvater Herrn Prof. Martin Fischer für die Zeit der Betreuung, die konstruktive Kritik, die Freiheit bei der Auswahl und die Unterstützung bei der Bearbeitung meiner Forschungsfragen.

Mein besonderer Dank gilt meiner Betreuerin Frau Dr. Lisa Kühne-Eversmann für die jederzeit sehr gute, zuverlässige und sorgfältige Betreuung. Sie stand mir stets mit großer Motivation zur Seite und hatte immer ein offenes Ohr, einen guten Rat und eine Tasse Kaffee für mich.

Genauso großer Dank gilt meinem Co-Betreuer Herrn Dr. Daniel Bauer, der mich unermüdlich mit Anregungen, kritischen Nachfragen und noch kritischeren Korrekturen unterstützt und motiviert hat und dabei stets dazu beigetragen hat, diese Arbeit auf Kurs zu halten.

Außerdem danke ich Anja Härtl und dem gesamten Team des Instituts für die freundliche Aufnahme, die positive Unterstützung, Rückmeldung und Motivation. Christian Günther Strobel danke ich besonders für die Hilfe bei der statistischen Auswertung und das Gegenrechnen meiner Analysen.

Weiterhin danke ich den Mitarbeiter*innen der Verlage sowie der Ärztekammern, die mir ihre Daten zur Verfügung gestellt haben und mir für Auskünfte zur Verfügung standen.

Im Jahr 2019 überarbeitete der Springer-Verlag seine Autorenhinweise zur Erstellung von MC-Fragen und zitiert in diesen auch die aus dieser Arbeit entstandene Publikation, worüber ich mich sehr freue.

Nicht zuletzt danke ich meiner Familie für die uneingeschränkte und unermüdliche Unterstützung während des Studiums und der Promotion. Meinen Eltern danke ich für ihren liebevollen und geduldigen Beistand sowie den moralischen und finanziellen Rückhalt. Meiner Schwester danke ich für ihre Vorbildfunktion und ihre motivierenden Worte. Ohne euch wäre diese Arbeit niemals möglich gewesen. Vielen Dank!