

Einfluss des Fragenformates in Multiple-choice-Prüfungen auf die Antwortwahrscheinlichkeit

Eine Untersuchung am Beispiel mikrobiologischer Fragen

The effects of phrasing on the answer likelihood in multiple choice based examinations

A case study in microbiology

• Johannes Schulze¹ • Stefan Drolshagen¹ • Frank Nürnberger² • Falk Ochsendorf³ • Volker Schäfer⁴ • Claudia Brandt⁴

Zusammenfassung:

Multiple choice (MC)-Klausuren sind im deutschen Medizinstudium trotz weitgehend fehlender Daten zur Validität dieser Prüfungsform zur Regelprüfung geworden. Darüber hinaus ist unklar, in welchem Ausmaß die Studierenden - auch solche mit guten Prüfungsergebnissen - den geprüften Lernstoff tatsächlich beherrschen. Am Fachbereich Medizin der Johann-Wolfgang-Goethe-Universität Frankfurt wurde am Ende des SS 2003 im Fach Mikrobiologie für die Studierenden des 2. klinischen Semesters eine MC-basierte Abschlussprüfung geschrieben. Die Studierenden des 1. klinischen Semesters hatten - bedingt durch Umstellungen des Curriculums - eine identische Ausbildung. Diese wurde durch eine inhaltlich weitgehend identische, im Format aber andere Klausur abgeschlossen, in der sowohl offene Fragen enthalten waren als auch Fragen, bei denen die Studierenden jede Aussage einzeln auf Korrektheit bewerten mussten. Der Vergleich der Ergebnisse für inhaltlich gleiche Fragen zeigt, dass die Studierenden im MC-Format eine hohe Quote richtiger Antworten erzielen, diese jedoch durch ein geändertes Fragenformat stark reduziert wird. So erreichten nur 20 - 30% der Studierenden ein vollständig richtiges Ergebnis, wenn jede Aussage einzeln bewertet werden musste, während die inhaltlich gleiche Frage im MC-Format 80 - 90% richtige Ergebnisse erzielte. In freien Fragen konnten nur 30 - 40% der Studierenden die richtige Antwort aktiv niederschreiben, während 90 - 99% der Studierenden die richtige Lösung passiv erkannten. Wir interpretieren diese Ergebnisse dahin, dass der Entscheidungszwang in MC-basierten Fragen einen starken Einfluss auf die Quote richtiger Antworten hat, und die Prüfungsergebnisse damit wesentlich durch das Format beeinflusst werden, das Wissen dagegen nicht beherrscht wird. Die Ergebnisse dieser Studie legen nahe, Sorgfalt bei der Auswahl des Prüfungsverfahrens walten zu lassen und der Steuerung des studentischen Lernverhaltens durch das Prüfungsformat wesentlich mehr Aufmerksamkeit zu widmen als bisher.

Schlüsselwörter: Multiple choice Fragen, Validität, offene Fragen, Ratewahrscheinlichkeit, studentisches Lernverhalten, Prüfungserfolg

Abstract:

Multiple choice(MC) questions have become the mainstay medical examinations in Germany, although there are currently few data available to show their validity. It also remains unclear, whether or not students master the knowledge tested, even those students with good test results. Students of the second clinical semester at the Medical Faculty of the Johann Wolfgang Goethe-University in Frankfurt have written an MC-based examination in microbiology. Due to changes in their curriculum students of the first clinical semester took the identical class in microbiology. This class was concluded by an examination identical in content, however in a different format including open questions as well as questions consisting of statements that had to be judged individually whether or not they are true. A comparison of test results between these two groups showed that only 20 - 30% of the students were able to correctly assess five statements, whereas the identical question presented in MC-format yielded 80 - 90% correct answers. For open questions only 30 - 40% of the students were able to actively answer the question correctly whereas 90 - 99% of the students passively recognized the correct answer. These results indicate that the need to choose one among five options as it is inherent for MC-based questions has a strong influence on the ratio of correct answers, and examination results are strongly depended upon the format of the test, while the knowledge presumed to be tested is not necessarily available. Our results suggest that care should be taken when selecting the mode of an examination. Moreover, the faculty should be more aware of the impact of the test format on the students learning behaviour.

Keywords: multiple choice question, single best choice question, validity, short essay questions, guessing likelihood, learning strategies, examination success

¹ Johann Wolfgang Goethe-Universität Frankfurt, Fachbereich Medizin, Dekanat, Frankfurt/Main, Deutschland

² Johann Wolfgang Goethe-Universität Frankfurt, Fachbereich Medizin, Studiendekanat, Institut für Anatomie III, Frankfurt/Main, Deutschland

³ Johann Wolfgang Goethe-Universität Frankfurt, Fachbereich Medizin, Zentrum für Dermatologie und Venerologie, Frankfurt/Main, Deutschland

⁴ Johann Wolfgang Goethe-Universität Frankfurt, Fachbereich Medizin, Institut für Medizinische Mikrobiologie, Frankfurt/Main, Deutschland

Einleitung

Medizinische Staatsexamina müssen in Deutschland in einem fest vorgegebenen Format durchgeführt werden. Aufgrund der leichten Auswertbarkeit und Objektivität hat sich das Multiple Choice (MC) Fragenformat in den letzten Jahren auch für universitäre Fachprüfungen durchgesetzt. Dieses Fragenformat zeichnet sich durch eine sehr schnelle Auswertbarkeit aus und schon daher die für Prüfungszwecke einzusetzenden Ressourcen.

Bisher ist die Frage ungeklärt, welches Lernverhalten mit diesen stark formalisierten Prüfungen gemessen wird. Zugunsten eines hohen Lernerfolges wird angeführt, dass die Studierenden durch viele "Altfragen" gezwungen werden, das gesamte Gebiet zu erlernen. Der Umfang der Fragensammlungen unterstützt dieses Argument.

Gegen einen Lernerfolg im eigentlichen Sinne, d.h. ob die Studenten das medizinische Wissen in einer Weise erlernen, der der späteren klinischen Realität entspricht, spricht die Notwendigkeit, richtige Antworten nur zu erkennen bzw. zwischen mehreren vorgegebenen Möglichkeiten in einem eingegrenzten Suchraum Entscheidungen zu treffen. Auch der Zwang, sich für eine und nur eine Antwort entscheiden zu müssen, entspricht nicht der späteren Situation, in der die "Lösungen" aktiv erarbeitet und häufig mehrere Alternativen untersucht werden müssen, die mit Wahrscheinlichkeiten, nicht aber mit absoluten Sicherheiten belegt sind. Die Erfahrungen aus mündlichen Prüfungen in Fach- und Staatsprüfungen sprechen ebenfalls dafür, dass die Studierenden zwar ein hohes passives Wissen (Wiedererkennen aus vorgegebenen Möglichkeiten) besitzen, nicht aber dieses Wissen umsetzen und aktiv präsentieren können.

Für Prüfungen wird u.a. gefordert, dass sie valide sind. Dies bedeutet, dass der Prüfungserfolg anzeigt, dass die Studierenden ein dem Lernziel entsprechendes Wissen erworben haben. Wird das Lernziel definiert als eine gute Vorbereitung auf die spätere Tätigkeit im Praktischen Jahr bzw. im Arzt im Praktikum, ist eine Validierung prinzipiell möglich. Bisher sind keine Daten zu einer Validierung von MC-Fragen anhand dieses Standards (oder eines anderen Standards) publiziert. Die vom IMPP vorgenommene Item-Analyse dient der Überprüfung der Konstruktvalidität; Inhalts-, Übereinstimmungs- oder gar prognostische Validität kann sie nicht leisten. Sie kann daher wegen der Fragenstruktur nicht als erfolgsorientierte Validierung angesehen werden.

Am Fachbereich (FB) Medizin der JWG-Goethe-Universität Frankfurt haben wir versucht, die Wertigkeit von MC-Fragen der Mikrobiologie anhand des Kriteriums aktives Wissen bzw. passives Wissen zu validieren. Hierzu wurde 126 Studierenden des 2. klinischen Semesters (Regelsemester Mikrobiologie nach der alten Studienordnung des FB Medizin) mit einer 30 Fragen umfassenden MC-Fragen basierten Abschlussklausur geprüft. Bei 130 Studierenden des 1. klinischen Semesters, die nach der neuen Studienordnung des FB Medizin bereits im ersten klinischen Semester in Mikrobiologie unterrichtet wurden, wurde das mikrobiologische Wissen in einem anderen Fragenformat geprüft, bei der kein Entscheidungszwang zwischen Antwortalternativen bestand bzw. kurze Antworten frei niedergeschrieben werden mussten (short essay question-Format). Beide Kohorten erhielten eine identische Ausbildung; auch die Arbeitsbelastung aus anderen Fächern war

für beide Kohorten fast identisch. Die Ergebnisse dieses Vergleichs werden im Folgenden dargelegt.

Methoden

Bedingt durch die Änderung der Studienordnung des FB Medizin und durch die geforderten Änderungen nach der neuen ÄAppO vom 27. Juni 2002 wurden am Ende des SS 2003 die Studierenden des 2. klinischen Semesters im Fach Mikrobiologie/Virologie mit einer 30 MC-Fragen umfassenden Klausur geprüft. Die Fragen wurden vom Institut für Medizinische Mikrobiologie anhand des Praktikumslehrstoffs formuliert und entsprachen dem bekannten Typ der IMPP-Fragen (Auswahl aus fünf Möglichkeiten, Vorgabe von Statements und Erkennen der richtigen Kombination sowie Begründungsfragen - A weil B). Die Klausuren wurden am Ende der Prüfung wieder eingesammelt. Die Studierenden des 2. klinischen Semesters hatten neben dem Fach Mikrobiologie im gleichen Semester die Scheine in Klinischer Chemie, Pathologie, Pharmakologie und Radiologie zu erwerben sowie sich auf das 1. Staatsexamen (1. SE) im August vorzubereiten. Wegen der Änderungen der neuen ÄAppO werden fast alle Studierenden am 1. SE teilnehmen.

130 Studierende des 1. klinischen Semesters wurden am Semesterende, zwei Wochen nach dem Klausurtermin des 2. klinischen Semesters, in einer schriftlichen Prüfung neben der Mikrobiologie in den Fächern Pathologie, Pharmakologie, Radiologie, Untersuchungskurs der klinischen Fächer und Notfallmedizin geprüft. Die Prüfung wurde als Semesterabschluss-Sammelklausur mit 30 Fragen aus dem Bereich der Mikrobiologie durchgeführt. Als Fragentypen wurden offene Fragen verwendet (die richtige Antwort bzw. richtigen Antworten mussten frei geschrieben werden; 15 Fragen), als Auswahlfragen, wobei den Studierenden nicht bekannt war, wie viele der Möglichkeiten anzukreuzen sind (zwischen null und fünf; 12 Fragen) bzw. als Reihungsaufgaben (vier vorgegebene Aussagen müssen in die richtige Reihenfolge gebracht werden; 3 Fragen). Die Fragentypen sind in Schulze et al. [1] beschrieben. Acht Wochen vor der Semesterabschlussklausur wurden die Studierenden in einer Testklausur mit 100 Fragen- davon 20 Fragen aus dem Bereich der Mikrobiologie- mit dem neuen Fragenformat vertraut gemacht.

Beide Studierendenkohorten absolvierten im SS 2003 ein Praktikum der Mikrobiologie und Virologie mit identischem Inhalt und Umfang. Für die Studierenden des 2. klinischen Semesters enthielt das vorhergehende Semester die Hauptvorlesung in Mikrobiologie. Unterschiede im sonstigen Veranstaltungsumfang bestanden in klinischer Chemie (nur Studierende des 2. klinisches Semesters) sowie in Propädeutik, Biomathematik und Notfallmedizin (2. klinisches Semester: Veranstaltungen des Vorsemesters; 1. klinisches Semester: Veranstaltungen des SS 2003). In Pharmakologie wurde für Studierende des 2. klinischen Semesters im Vorsemester eine Vorlesung angeboten, das Praktikum fand zu Beginn des SS 2003 als Blockpraktikum statt. Die Studierenden des 1. klinischen Semesters hatten eine Vorlesung in Pharmakologie während des Semesters sowie das Praktikum der Pharmakologie als Blockpraktikum am Semesterende.

Zur Adaptation des Schweregrades wurden 20 Fragen aus der MC-Klausur als zu schwierig angesehen und bei der Fragenformulierung für das alternative Format modifiziert. Zehn Fragen wurden

in identischer oder fast identischer Formulierung geprüft; in drei Fragen war die richtige Antwort frei niederzuschreiben, in 7 Fragen wurde der Lerninhalt ebenfalls als passives Wissen geprüft (Ankreuzen der richtigen Aussagen). Bei 20 Fragen bestanden deutlichere Unterschiede in der Formulierung; für die Fragestellung im alternativen Format wurden aus den MC-Fragen speziellere Aussagen herausgenommen und durch Inhalte mit basalem Wissen ersetzt. Bei 28 der 30 Fragen blieb das der Fragestellung zugrunde liegende mikrobiologische Wissen unverändert.

Für alle MC-Fragen wurde der Prozentsatz der Markierungen für alle Antwortmöglichkeiten errechnet (Studierende des 2. klinischen Semesters). Auswahlfragen (Studierende des 1. klinischen Semesters) wurden in gleicher Weise ausgewertet. Bei offenen Fragen (Studierende des 1. klinischen Semesters) wurden alle Antworten anhand eines Lösungsthesaurus ausgewertet; als richtig wurden in der Regel mehrere Antworten anerkannt. Schreibfehler wurden nur dann berücksichtigt, wenn diese einen wesentlichen Einfluss auf die Lösung hatten, i.e. es andere medizinische Begriffe gibt, die dem Schreibfehler entsprechen. Für alle Fragentypen werden bei vollständig richtiger Beantwortung 5 Punkte vergeben; teilweise richtige Antworten werden mit geringeren Punktzahlen bewertet.

Ergebnisse

Aktives Wissen kann nur durch frei zu formulierende Antworten geprüft werden. Ein Vergleich zeigt, dass die Studierenden die richtige Antwort zwar kennen mögen, aber nicht aktivreproduzieren können. In Abbildung 1 und 2 sind die Ergebnisse für zwei Beispiele dargestellt, bei denen der Zusammenhang zwischen Papillomavirus-Infektionen und Cervixkarzinom (siehe Abbildung 1, Beispiel 1) bzw. sexuell übertragbaren Viren (siehe Abbildung 2, Beispiel 2) erkannt werden mussten. Während 80% der Studierenden (100/127) den Zusammenhang zwischen Papillomaviren und Cervixkarzinom sowie Warzen erkannten und einen Zusammenhang mit Harnblasenkarzinom verneinten (der fehlende Zusammenhang zum Kaposi-Sarkom bzw. M. Whipple war aufgrund der vorgegebenen Kombinationen nicht relevant), konnten nur 42% Papillomaviren als auslösendes Agens benennen (56/130). Noch deutlicher ist der Unterschied im zweiten Beispiel. Während fast 100% der Studierenden *Clostridium perfringens* als Erreger des Gasbrandes erkannten (126/127), waren nur 31,5% der Studierenden in der Lage, *Bordetella pertussis* als Keuchhustenerreger aktiv zu benennen.

Beispiel 1:

MC-Frage:

Welche Erkrankungen können durch Papilloma-Viren verursacht werden?

1. Kaposi-Sarkom
 2. Morbus Whipple
 3. Harnblasenkarzinom
 4. Warzen
 5. Cervixkarzinom
- A: nur 4 ist richtig (0/127 Antworten)
 B: nur 4 und 5 sind richtig (100/127 Antworten)
 C: nur 1 und 3 sind richtig (1/127 Antwort)
 D: nur 2 und 3 sind richtig (0/127 Antworten)
 E: Nur 3, 4 und 5 sind richtig (26/127 Antworten)

Offene Frage:

Welche Viren werden mit der Entstehung des Cervixkarzinoms assoziiert (1 Angabe)?

Papillomavirus 56 Antworten (42%); Adenovirus 9 Antworten; CMV-Virus; Herpes Virus; Herpes simplex Virus, je 6 Antworten; EBV-Virus 4; Coxsackie-Virus 3; Rhinovirus 2; Hepatitis B Virus; Herpes genitalis, Poliovirus; Poliomyavirus; Rotaviren; Rötelnvirus; Warzenvirus, Syphilis je 1 Antwort

Abbildung 1: Fehlendes aktives Wissen trotz guter Ergebnisse in MC-Fragen

Beispiel 2:

MC-Frage:

Wie heisst der Erreger des Gasbrandes

- Corynebakterium jeikeium* (0/127 Antworten)
Corynebakterium parvum (1/127 Antwort)
Clostridium perfringens (126/127 Antworten)
Chlamydia psittaci (0/127 Antworten)
Clostridium difficile (0/127 Antworten)

Offene Frage:

Wie heißt der Erreger des Keuchhustens?

Bordetella pertussis, 41 Antworten (31,5%); *pertussis*, 27; *C. pertussis*, 9; *Clostridium pertussis*, 3; *S. pertussis*, 2; *C. diphtheriae*, 2; *Mycobacterium pertussis*, 2 Antworten; *S. aureus*, *Pseudomonas diphtheriae*, *Staphylococcus pertussis*, *Streptokokous pyogenes*, *Rabies*, *Pseudomonas pertussis*, *Keuchhustenvirus*, *Helicobacter pertussis*, *Zoster*, *Corynebakterium diphtheriae*, *Pseudomonas*, *Campylo-Bakterium*, *Vibrio pertussis*, *Peritus*, *C. botulinum*, *C. pertussis*, *Diphtheriae*, *Bordetella*, *Betelinio pertussis*, *Bacterium pertussis*, *Bacterium diphtheriae*, *B. pertussis*, je 1 Antwort.

Abbildung 2: Fehlendes aktives Wissen trotz guter Ergebnisse in MC-Fragen

Analoge starke Diskrepanzen zwischen dem passiven und aktiven Wissen zeigten sich in drei anderen Beispielen. Während alle Studierenden erkannten, dass HIV sexuell übertragen werden kann und immerhin 42% auch das Herpes simplex Virus 1 als richtige Antwortmöglichkeit erkannten, konnten nur 88,5% (115 von 130 Studenten) HIV aktiv benennen (gefordert wurde die Benennung von 3 Viren, die typischerweise durch Sexualkontakte übertragen werden). Hepatitis B-Virus wurde von 39 Studierenden (49/130) benannt, daneben wurden weitere 23 verschiedene Viren angegeben, u. a. wurden *Candida*, Chlamydien, Gonokokken und Treponemen als sexuell übertragbare "Viren" aufgelistet. Die PCR wurde nur von 30% der Studierenden als Nachweismethode für eine pränatal erworbene HIV-Infektion bei Neugeborenen angegeben, während 35% der Studierenden den ELISA als Methode der Wahl angaben. Dagegen bejahten 87% der Studierenden diese Aussage in einer MC-Frage (Aussage A in einer Weil-Frage).

Große Unterschiede zeigten sich auch beim Vergleich der Ergebnisse zwischen MC-Fragen und Auswahlfragen, bei denen zwischen null und fünf Aussagen als zutreffend markiert werden konnten. In den Abbildungen 3-5 sind drei Beispiele aufgeführt,

in denen identische bzw. sehr ähnliche Lerninhalte geprüft werden. Der Vergleich zeigt deutlich, dass die Anzahl vollständig richtig beantworteter Fragen sehr stark vom Fragenformat geprägt ist. Während im Beispiel 1 die einzig richtige Antwort von 89% der Studierenden angekreuzt wurde (Antwort mit Falsch-Aussage), haben nur 17% der Studierenden in der alternativen Frage die richtigen Möglichkeiten 1, 3 und 4 vollständig angekreuzt. Nur Möglichkeiten 3 und 4 haben 36% der Studierenden angekreuzt (siehe Abbildung 3-5), d.h. das Wissen um den Übertragungsweg war bei einem beträchtlichen Anteil der Studierenden nicht vorhanden. In ähnlicher Weise führte der Entscheidungszwang in den Beispielen 2 und 3 der MC-Fragen zu sehr hohen Quoten richtiger Antworten, während bei der alternativen Fragenform für jede einzelne richtige Aussage die Zustimmungsquote wesentlich niedriger lag, und der Prozentsatz derjenigen Studierenden, die die Frage vollständig richtig beantworteten, nur 30% bzw. 27% betrug. Bedingt durch den Entscheidungszwang in MC-Fragen, liegt die Quote richtiger Antworten wesentlich höher, als dem tatsächlichen, gefestigten Wissen der Studierenden entspricht. Analoge Aussagen lassen sich aus den Ergebnissen der übrigen Fragen ablesen, die in beiden Klausuren identisch oder sehr ähnlich gestellt wurden.

Beispiel 1:

MC-Frage:

Welche Aussage zu Polioviren ist falsch?

A. sie werden fäkal-oral übertragen (2/127)
 B. sie gehören zu den Picornaviren (4/127)
 C. die Impfung gegen Polioviren sollte laut Empfehlung der STIKO nur noch mit dem Totimpfstoff Salk durchgeführt werden. (8/127)
 D. sind die Erreger der Kinderlähmung (0/127)
 E. gelten seit 2000 als weltweit ausgerottet. (113/127)

Alternative Frage:

- Welche Aussage/n zu Polioviren ist/sind richtig?
- sie werden fäkal-oral übertragen (48/130)
 die Impfung gegen Polioviren sollte laut Empfehlung der STIKO nur noch mit dem Lebendimpfstoff Sabin durchgeführt werden (17/130)
 die Impfung gegen Polioviren sollte laut Empfehlung der STIKO nur noch mit dem Totimpfstoff Salk durchgeführt werden (79/130)
 sind die Erreger der Kinderlähmung (114/130)
 gelten seit 2000 als weltweit ausgerottet (15/130)

Abbildung 3: Fehlendes Diskriminationsvermögen trotz guter MC-Ergebnisse

Diskussion

Die Ergebnisse der vorliegenden Studie belegen, dass für den Erfolg in MC-basierten Prüfungen neben dem Fachwissen weitere Faktoren einen wesentlichen Einfluss haben. Es kann aus MC-Fragen basierten Prüfungen nicht automatisch abgelesen werden, dass die Studierenden das abgefragte Wissen tatsächlich beherrschen. Aus den in den Abbildungen 3-5 dargestellten Beispielen zeigt sich, dass in MC-Fragen 80 - 90% richtige Antworten erzielt werden können, obwohl durch die alleinige Änderung, dass mehr als eine Antwort richtig sein kann, die Quote der vollständig richtigen Antworten auf 17 - 30% fällt. Unter Zugrundelegung dieser Zahlen - die zwar repräsentativ sein können, jedoch Fragen- und Aussagen-spezifische Variationen beinhalten - ergibt sich, dass der Prüfungserfolg in MC-Fragen basierten Klausuren zum wesentlichen Teil durch das Fragenformat und den damit verbundenen Entscheidungszwang für eine und nur eine richtige Lösung bedingt ist.

Beispiel 2:

MC-Frage

Für den Nachweis virusspezifischer Antikörper können folgende Methoden eingesetzt werden:

1. Polymerase-Kettenreaktion
 2. ELISA
 3. Hämagglutinationshemmtest
 4. Komplementbindungsreaktion
 5. Elektronenmikroskopie

A. nur 1 und 5 sind richtig (0/127)
 B. Nur 1, 2, 3 und 4 sind richtig (22/127)
 C. Nur 3 und 4 sind richtig (1/127)
 D. nur 2 ist richtig (3/127)
 E. nur 2, 3 und 4 sind richtig (101/127)

Alternative Frage:

- Für den Nachweis virusspezifischer Antikörper kann/können folgende Methode/n eingesetzt werden:
- Polymerase Kettenreaktion (48/130)
 ELISA (121/130)
 Hämagglutinationshemmtest (92/130)
 Komplementbindungsreaktion (77/130)
 Elektronenmikroskopie (36/130)

Abbildung 4: Fehlendes Diskriminationsvermögen trotz guter MC-Ergebnisse

Beispiel 3

MC-Frage:

Welche Aussage ist falsch? Methicillin-resistente *Staphylococcus aureus*-Stämme (MRSA)

A. sind gegen alle β -Laktam-Antibiotika resistent (6/127)
 B. haben eine besondere krankenhaushygienische Bedeutung (0/127)
 C. können im Nasen-Rachen-Raum von gesunden Personen (Keimträgern) nachgewiesen werden (9/127)
 D. führen immer nur zur Kolonisation des Patienten, niemals aber zu invasiven Infektionen (110/127)
 E. sind bisher selten resistent gegen Glykopeptidantibiotika (Vancomycin, Teicoplanin) (2/127)

Alternative Frage

- Welche Aussage/n zu Methicillin-resistenten *Staphylococcus aureus*-Stämmen (MRSA) ist/sind richtig?
- alle β -Laktamantibiotika gelten als unwirksam (73/130)
 haben eine besondere krankenhaushygienische Bedeutung (121/130)
 können im Nasen-Rachen-Raum von gesunden Personen (Keimträgern) nachgewiesen werden (92/130)
 führen immer nur zur Kolonisation des Patienten, niemals aber zu invasiven Infektionen (6/130)
 sind bisher selten resistent gegen Glykopeptidantibiotika (Vancomycin, Teicoplanin) (92/130)

Abbildung 5: Fehlendes Diskriminationsvermögen trotz guter MC-Ergebnisse

Auch eine Abschätzung aus dem Vergleich zwischen MC-Fragen und Fragen mit freien Antworten zeigt, dass die richtige Antwort nur von einer Minderheit aktiv (d.h. ohne Wahlantwortmöglichkeiten) benannt werden kann, obwohl sie bei MC-Fragen von fast allen Studierenden erkannt wird, obwohl die Studierenden frühzeitig auf das geänderte Prüfungsformat vorbereitet wurden. Dem entspricht auch die Erfahrung aus mündlichen Prüfungen, in denen selbst bei einfachen Fragen nur wenige Studierende schnell und prompt die richtige Antwort aktiv wissen; die meisten Studierenden müssen häufig länger überlegen und bieten unpräzise, zum Teil widersprüchliche bzw. offensichtlich geratene Aussagen als Antwort an.

Diese Ergebnisse sollten auch hinsichtlich der Validität von Prüfungen beachtet werden. Als ein Kriterium einer guten Ausbildung kann beispielsweise dienen, dass die Ausbildung dazu befähigen

soll, wichtige Sachverhalte zu beherrschen (aktiv wiedergeben werden können) oder richtig zu interpretieren (richtige oder falsche Aussagen ohne weitere Entscheidungshilfen zu identifizieren,). Dieses Kriterium bedingt, dass das Prüfungsformat angemessen ist, wenn die Prüfungsfragen und geforderten Antworten den späteren beruflichen Anforderungen entsprechend gestaltet werden.

Die Forderung einer frei hinzuschreibenden Antwort entspricht dabei häufigen klinischen Situationen, z.B. bei der Anamnese oder der Anforderung diagnostischer Maßnahmen. Sie bedingt aktives Wissen für ein Handeln ohne wesentliche Vorgaben. Mögliche Vorgaben hierzu wären lediglich Leitlinien zum strukturierten klinischen Vorgehen (z.B. Standard-Vorgehensweisen), die zur Zeit in verstärktem Maße entwickelt werden. Das MC-Fragenformat erscheint nach den Ergebnissen dieser Studie nicht geeignet, diesen Wissensbereich adäquat abzudecken.

Die Beurteilung der Richtigkeit einer Aussage entspricht am ehesten der Interpretation von Befunden und Ergebnissen; sie ist im klinischen Alltag in der reinen Form nur selten zu identifizieren und findet sich am ehesten in der Bewertung der Untersuchungsergebnisse, d.h. entspricht der erhaltene Befund dem Erwartungsergebnis bei der vermuteten Krankheit. Auch hierfür ist zu fordern, dass die Entscheidungsfindung richtig/falsch nicht durch Formatprobleme beeinflusst wird. MC-Fragen enthalten durch den Entscheidungszwang zu einer und nur einer richtigen Angabe zu einem hohen Masse Entscheidungshilfen, die nicht auf medizinischem Wissen basieren.

Große systematische Untersuchungen zur Validität von MC-Fragen liegen für Deutschland nicht vor; lediglich eine Studie von Gebert [2] zeigt, dass Studierende im Physikum zwar in der Lage sind, die Fragen zur Physik in ausreichendem Maße richtig "anzukreuzen", dass jedoch das den Fragen zugrunde liegende Wissen nur selten in hinreichendem Maße verfügbar oder gar gefestigt ist. Diese Ergebnisse unterstützen in hohem Maße unsere Interpretation, dass nur eine Minderheit der Studierenden in der Lage ist, medizinische Sachverhalte zuverlässig zu beurteilen. Die hohe Quote richtiger Ergebnisse in MC-basierten Prüfungen beruht vermutlich auch darauf, sich entscheiden zu müssen und Altfragen oder vorgegebene Aussagen als richtig wiederzuerkennen. Die Ergebnisse unserer Studie zeigen, dass der Anteil der durch den Entscheidungszwang richtig getroffener Entscheidungen wahrscheinlich hoch ist.

Offen bleibt das Problem der Validität einer Prüfung. Für MC-Fragen existieren bisher keine Daten, die zeigen, dass MC-Prüfun-

gen valide sind. Hierbei soll nicht bestritten werden dass gute Examensergebnisse positiv mit guten Ergebnissen in mündlichen Prüfungen korrelieren. In dieser Hinsicht sind die Staatsexamensergebnisse wertvoll, sie scheinen aber nicht unbedingt anzuzeigen, dass die Prüflinge den Lernstoff beherrschen, der den Prüfungsfragen zugrunde liegt. Offensichtlich zeigen gute MC-Prüfungsergebnisse eher an, dass erfolgreiche Studierende in der Lage sind, "richtige" Entscheidungen zu treffen (welches ebenfalls ein Ausbildungsziel unter vielen sein sollte). Sie belegen nicht, dass der Lernstoff im Sinne der o.a. Kriterien beherrscht wird.

Alle in der Ausbildung Involvierten sollten sich dieses Dilemmas bewusst sein. Der Einfluss der Prüfungsgestaltung auf das Lernverhalten der Studierenden kann gar nicht hoch genug eingeschätzt werden. Es ist daher unumgänglich, dass durch eine entsprechende Prüfungsgestaltung versucht wird, die angestrebten Lernziele einer Fakultät auch tatsächlich zu erreichen. Auch wenn die ideale Prüfungsform noch nicht gefunden ist, können alternative Wege beschritten werden. Die am Fachbereich Medizin der JWG-Universität Frankfurt eingeschlagene Richtung ist anderenorts genauer dargelegt (Schulze et al. [3]). Das bisher vorherrschende MC-Format allein erscheint nach den Ergebnissen dieser Studie nicht geeignet, die von Fakultäten oder dem Gesetzgeber geforderten Ausbildungsziele zu prüfen. Es sollte durch andere Modelle ersetzt bzw. ergänzt und damit zu einer Prüfungsform unter vielen werden. Auf keinen Fall sollte, wie es bisher der Fall ist, dass Lernverhalten der Studierenden in wesentlichem Ausmaß von der Vorbereitung auf MC-Prüfungen gesteuert werden.

Korrespondenzadresse:

• PD Dr. med. Johannes Schulze, Johann Wolfgang Goethe-Universität Frankfurt, Fachbereich Medizin, Dekanat, Theodor-Stern-Kai 7, 60590 Frankfurt/Main, Deutschland, Tel.: 069/6301-5681, Fax: 069/6301-5922
j.schulze@em.uni-frankfurt.de

Literatur:

- [1] Schulze J, Drolshagen S, Nürnberger F, Siegers CP, Syed Ali S. Prüfen und Prüfungen nach der neuen Approbationsordnung - Grundsätze und Rahmenbedingungen. *Med Ausbild.* 2004;21:30-34.
- [2] Gebert G. Medizinstudium: Naturwissenschaftliche Grundkenntnisse nach der Vorklinik. *Dtsch Arztebl.* 2002;99:A-252.
- [3] Schulze J, Kersken-Nülens U, Drolshagen S, Nürnberger F. Struktur des Medizinstudiums nach der neuen Ärztlichen Approbationsordnung (ÄAppO). in: *Quo vadis medicæ? - Neue Wege in der Medizinerbildung.* Bonn: Stifterverband für die Deutsche Wissenschaft; 2004. S. 36-38.