

IAWF
Institut für Aus-, Weiter- und Fortbildung
Medizinische Fakultät Universität Bern

Kompetent prüfen

Handbuch zur Planung, Durchführung und Auswertung
von Facharztprüfungen

im Auftrag der:
Verbindung der Schweizer Ärzte FMH
Österreichischen Ärztekammer ÖÄK

Bern/Wien 1999

INHALT

Vorwort - 9

Einleitung - 11

Zusammenfassung - 15

Kapitel 1 Was prüfen? - 21

Kapitel 2 Wie prüfen? - 39

2.1 Welche Methode wählen? - 43

2.2 Was leisten die Methoden? - 48

Wahlantwortverfahren (MC) - 49

Kurzantwortfragen (KAF) - 51

Strukturierte mündliche Prüfung (SMP) - 52

Strukturierte Beobachtung (SB) - 54

Objektives strukturiertes klinisches Examen (OSCE) - 56

Kapitel 3 Wie wird eine Prüfung mit MC-Fragen entwickelt? - 61

3.1 Welche Anforderungen sollen Fragen erfüllen? - 65

3.2 Worauf zielt das Fragenthema? - 67

3.3 Wie unterscheiden sich die Fragentypen? - 71

3.4 Wie werden Fragen formuliert? - 80

3.5 Wie werden Fragen überprüft? - 89

3.6 Wie wird die Prüfung zusammengestellt? - 91

Kapitel 4 Wie wird eine Prüfung mit Kurzantwortfragen entwickelt? - 97

4.1 Wie werden Prüfungsaufgaben ausgewählt? - 101

4.2 Wie werden Fall und Fragen formuliert? - 102

4.3 Welche Antworten werden erwartet? - 105

4.4 Wie bewerten, korrigieren und auswerten? - 108

Kapitel 5	Wie wird eine strukturierte mündliche Prüfung entwickelt? - 113
5.1	Welche Fragen - welche Antworten? - 117
5.2	Wie werden Prüfungsunterlagen hergestellt? - 121
<hr/>	
Kapitel 6	Wie wird eine Prüfung mit strukturierter Beobachtung entwickelt? - 127
6.1	Was soll beobachtet werden? - 131
6.2	Wie wird die Beobachtung strukturiert? - 132
6.3	Wie wird die Beobachtungssituation standardisiert? - 136
6.4	Wie auswerten? - 138
<hr/>	
Kapitel 7	Wie wird eine objektive strukturierte klinische Prüfung entwickelt? - 141
7.1	Woraus besteht ein OSCE? - 145
7.2	Wie werden Stationen entwickelt? - 150
7.3	Wie werden Examinatoren vorbereitet? - 157
7.4	Wie Standardisierte Patienten rekrutieren und trainieren? - 160
7.5	Wie wird ein Circuit angelegt? - 166
7.6	Wie wird ein OSCE ausgewertet? - 168
<hr/>	
Kapitel 8	Wer besteht die Prüfung? - 179
8.1	Warum ist die Standardsetzung wichtig? - 183
8.2	Wie funktionieren die verschiedenen Verfahren? - 186
8.3	Wie wird geplant? - 192
<hr/>	
Kapitel 9	Was taugt die Prüfung? - 197
<hr/>	
Checklisten - 205	
<hr/>	
Glossar prüfungstechnischer Fachbegriffe - 211	
<hr/>	
Register - 213	

Die Zielgruppe

„Kompetent prüfen“ ist für alle geschrieben, die an der Vorbereitung, Durchführung und Auswertung einer Facharztprüfung mitarbeiten. Würde eine Prüfung streng nach diesem Manual entwickelt, müsste sie nahezu perfekt werden. Die Autoren sind sich aber bewusst, dass die Entwicklung eines Examens einen Prozess darstellt und viele Fachgesellschaften vorerst mit noch nicht optimalen Lösungen arbeiten müssen. „Kompetent prüfen“ möchte diesen Gesellschaften die Richtung aufzeigen und mithelfen, dass ihre Facharztprüfungen ständig überarbeitet, weiterentwickelt und verbessert werden können.

Der Aufbau

„Kompetent prüfen“ kann wie ein Lehrbuch von vorne bis hinten durchgearbeitet werden, es ist aber in erster Linie als Handbuch konzipiert, das man nach Bedarf portionenweise konsultiert. Das erklärt auch die ab und zu vorkommenden Wiederholungen. Um sich schnell zurechtzufinden, dienen die Einführungen an den neun Kapitelanfängen und der Text in den Randspalten.

„Kompetent prüfen“ ist durch verschiedene Darstellungsformen gegliedert: Der Haupttext ist in AGaramond-12-Pt.-Schrift gedruckt. Weiterführende Texte, die zum ersten Verstehen nicht notwendigerweise gelesen werden müssen, stehen in AGaramond-10-Pt.-Schrift. Beispiele aus der Praxis erscheinen ebenfalls in kleiner Schrift, sind aber kursiv gedruckt (AGaramond Italic). Die Beispiele zeigen, wie konkrete Lösungen aussehen. Sie sollen die Themen illustrieren aber auch erweitern. Der Text in der Randspalte (AGaramond Italic, 11 Pt.) liefert Kürzest-Zusammenfassungen und dient dazu, den Inhalt einer Textgruppe schnell zu erfassen. Das vorliegende Manual ist bewusst locker gestaltet und mit leeren Seiten versehen. Die weissen Flächen sind für die Gedanken und Notizen der Leser gedacht.

Im Zusammenhang mit Facharztprüfungen sind Ärztinnen selbstverständlich genau gleich wichtig wie ihre männlichen Kollegen. Das gleiche gilt für Kandidatinnen, Autorinnen, Prüferinnen, Korrektorinnen usw. Im Interesse einer schlanken Formulierung braucht „Kompetent prüfen“ immer nur die männliche Form. Daran möge sich niemand stören, die Gattung ist gemeint und nicht das Geschlecht.

Die Autoren

„Kompetent prüfen“ wurde verfasst von den folgenden Mitarbeitern des Instituts für Aus-, Weiter und Fortbildung (IAWF), Medizinische Fakultät, Universität Bern:

- Bloch, Ralph, Dr. med. et phil., FRCP (C)
(Einleitung, Zusammenfassung, Kapitel 9)
- Hofer, Daniel, Dr. phil., Psychologe FSP (Kapitel 5)
- Krebs, René, dipl. Psychologe FSP (Kapitel 2, 3, 7)
- Schläppi, Peter, Dr. med., Facharzt FMH für Allgemeinmedizin
(Kapitel 1, 4, 6)
- Weiss, Stephan, Dr. med., Facharzt FMH für Sozial- und Präventivmedizin
(Glossar, Register)
- Westkämper, Reinhard, Dr. med., Diplompsychologe (Kapitel 8)

- Jakob, Lorenz (Redaktion)

Der Auftrag für das vorliegende Manual stammt von der

- Verbindung der Schweizer Ärzte FMH
- Österreichischen Ärztekammer ÖÄK

Bürgt die Ärzteschaft für die Kompetenz ihrer Mitglieder?

Der Staat als Garant der Sicherheit seiner Bürger macht die ärztlichen Standesorganisationen verantwortlich für die Qualitätssicherung in der medizinischen Versorgung. In der Tradition der mittelalterlichen Zünfte bürgen die ärztlichen Standesorganisationen denn auch für die kompetente und gewissenhafte Berufsausübung ihrer Mitglieder und verteidigen gleichzeitig deren Sonderrechte. Zu diesen Privilegien gehört nicht nur das Monopol der ärztlichen Berufsausübung, sondern auch die Erlaubnis, die Tabus der Intimsphäre und körperlichen Integrität der Patienten zu brechen. Wenn die Ärzteschaft ihre Standesfreiheit behalten will, muss sie sich - mit kompetenten Mitgliedern - das Vertrauen der Öffentlichkeit sichern. Um die für die Erteilung des Facharzttitels^a notwendige professionelle Kompetenz zu beurteilen, dienen den Standesorganisationen^b Facharztprüfungen und Weiterbildungszeugnisse^c.

Die Verleihung des Facharzttitels ist also mehr als eine intellektuelle Übung. Sie ist letztlich eine politische Entscheidung. Wenn der Titel zu bedenkenlos verliehen wird, lehnt sich die Öffentlichkeit auf. Wenn andererseits kompetenten Ärzten der Titel vorenthalten wird, verliert der Qualifikationsprozess seine Glaubwürdigkeit bei den Kollegen. Das Recht der Bevölkerung auf eine kompetente ärztliche Behandlung muss gegen den Anspruch von langjährig ausgebildeten Ärzten auf wirtschaftliche Unabhängigkeit abgewogen werden.

Ärzte haben durchschnittlich über 24 Jahre mit ihrer Ausbildung (im weiteren Sinne) verbracht, bevor sie sich der Facharztprüfung stellen. Neben dem Zeitaufwand haben sie und ihre Familien auch grosse Geldsummen investiert. Oft sind Kandidaten zudem für den Unterhalt von Ehepartner und Kindern verantwortlich. Deshalb muss eine akzeptable Lösung für die wenigen definitiv Erfolglosen gefunden werden. Dabei darf die qualitätssichernde Funktion der Prüfung aber nicht gefährdet werden. Bis jetzt wurde diesen prinzipiellen Fragen zu wenig Aufmerksamkeit geschenkt. Solange diese Probleme nicht gelöst sind, wird immer Gnade über Recht siegen.

Gegner einer ernsthaften Facharztprüfung argumentieren oft, der Studienabschluss und die Führungszeugnisse der verantwortlichen Chefärzte genügen, um die Bevölkerung vor inkompetenten Ärzten zu schützen. Leider zeigt die Erfahrung, dass dieser Optimismus nur beschränkt berechtigt ist¹. In einer Studie wurden 1991 die Programmdirektoren der amerikanischen Fachgesellschaft für Anästhesiologie befragt, ob und wie weit sie sich selbst von ihren eigenen Schülern behandeln lassen würden.

^a Dazu gehören auch die entsprechend weitergebildeten Allgemeinärzte.

^b FMH (Schweiz) resp. Ärztekammer (Österreich)

^c Unter Weiterbildung verstehen wir die berufliche Bildungsphase, die dem Medizinstudium folgt.

Nur 60 Prozent der Kandidaten genossen das uneingeschränkte Vertrauen ihrer Vorgesetzten. Dennoch wurden alle aufgrund ihrer Führungszeugnisse zur Prüfung zugelassen. Diese korrelierte dann allerdings gut mit der Einschätzung der Programmdirektoren. Man kann sich fragen, wie ethisch es ist, wenn erfahrene Chefarzte für ihre eigene medizinische Behandlung einen höheren Standard fordern als für die Betreuung anderer Patienten.

Eine Facharztprüfung wirkt nur dann qualitätssichernd, wenn sie auch selektioniert. Dies bedeutet, dass Kandidaten, von denen sich die Examinatoren selbst nicht behandeln lassen würden, die Prüfung nicht bestehen dürfen. Also wird es auch gelegentlich Kandidaten geben, welche die Prüfung nie bestehen. Bevor die Facharztprüfungen verbindlich werden, muss über diesen Punkt Konsens unter den Fachgesellschaften bestehen.

Auch der Stellenwert der begleitenden Beurteilung durch die verantwortlichen Vorgesetzten (Rasterzeugnis) ist nicht unbestritten². Als „formative“ Evaluation soll sie die jungen Ärzte formen. Als „summative“ Evaluation muss sie aber gleichzeitig zur Entscheidung beitragen, ob ein Arzt seinen Titel erhält. Eine umfangreiche Literatur belegt den grundsätzlichen Konflikt zwischen formativer und summativer Beurteilung. Während die formative Evaluation subtile Kompetenzmängel so beleuchten soll, dass sie behoben werden können, muss ein summatives Zeugnis robust sein sowohl gegen informelle Proteste wie formelle Rekurse. In dieser Situation kann wohl keine der beiden Aufgaben angemessen erfüllt werden³.

Ralph Bloch

Literatur

- ¹ Slogoff S, Hughes FP, Hug CC Jr, Longnecker DE, Saidmann L. A demonstration of validity for certification by the American Board of Anesthesiology. *Acad Med* 1994;69:740-6
- ² Thompson WG, Lipkin M Jr, Gilbert DA, Guzzo RA, Roberson L. Evaluating evaluation: assessment of the American Board of Internal Medicine Resident Evaluation Form. *J Gen Intern Med* 1990;5:214-7
- ³ Haber RJ, Avins A. Do ratings on the American Board of Internal Medicine Resident Evaluation Form detect differences in clinical competence? *J Gen Intern Med* 1994;9:140-5

Zusammenfassung

Um gültig zu sein, muss eine Facharztprüfung alle wesentlichen Fähigkeiten evaluieren, die zur kompetenten ärztlichen Tätigkeit im entsprechenden Fachbereich notwendig sind. Zu diesen Schlüsselkompetenzen gehört zweifellos ein umfangreiches und sich ständig anpassendes Fachwissen, eine breite Palette klinischer und technischer Fertigkeiten sowie ein solides professionelles Verhalten. Schwer zu erfassen, aber nicht minder wichtig, ist die Bereitschaft, das Wissen und die Fertigkeiten in konkreten Situationen bestmöglich zum Nutzen der Patienten einzusetzen¹.

Ein gültiges Examen erfasst alle wesentlichen Kompetenzen.

Um glaubhaft zu sein, müssen die Schlüsselkompetenzen in einem möglichst breit abgestützten Konsens festgelegt werden. Neben Universitätsprofessoren und Chefärzten grosser Kliniken müssen auch Fachärzte aus kleineren Spitälern und aus Privatpraxen angemessen in den Konsensgremien vertreten sein. Der Prozess muss für jedes Fachgebiet einzeln erfolgen, bevor die Prüfungsformen und die spezifischen Inhalte bestimmt werden können.

Es braucht ein breit abgestütztes Team.

Bei der Festlegung der Schlüsselkompetenzen ist es nützlich, von konkreten Situationen auszugehen, wie sie dem qualifizierten Praktiker oder Kliniker begegnen. Die Konsensgremien bestimmen einerseits sinnvolles und andererseits zwar realistisches aber letztendlich untaugliches fachärztliches Verhalten. Die Schlüsselkompetenzen können aus der Gegenüberstellung der beiden Verhalten identifiziert werden (vgl. Kapitel 1).

Der Praktikeralltag ist Ausgangspunkt um Schlüsselkompetenzen festzulegen.

Es ist klar, dass die Schlüsselkompetenzen eines Fachgebietes nicht flächendeckend geprüft werden können. Ein Examen kann nur eine exemplarische Stichprobe der notwendigen Fähigkeiten und Fertigkeiten beurteilen. Je sorgfältiger eine Prüfung aber geplant wurde, umso zuverlässiger erlaubt sie eine verallgemeinernde Schlussfolgerung von den wenigen tatsächlich geprüften Leistungen auf die gesamte professionelle Kompetenz. Dies bedingt, dass alle wichtigen Bereiche eines Fachgebietes gleichmässig berücksichtigt werden (vgl. Kapitel 1, „Blueprint“, S.28).

Ein zuverlässiges Examen prüft repräsentativ.

Die Schlüsselkompetenzen bestimmen dann weitgehend die Wahl der Prüfungsmethoden. Wahlantwortfragen eignen sich nicht zur Prüfung technischer oder sozialer Fertigkeiten. Aus praktischen Gründen prüfen mündliche Examina die Wissensbreite eines Kandidaten meist nur ungenügend (vgl. Kapitel 2).

Schlüsselkompetenzen bestimmen die Prüfungsmethoden.

Je schlechter eine Prüfung zwischen kompetenten und inkompetenten Ärzten unterscheidet, umso geringer ist die Bereitschaft, einem Kandidaten aufgrund ungenügender Prüfungsergebnisse den Titel vorzuenthalten.

Die Akzeptanz der Resultate korreliert mit der Prüfungsqualität.

Die Gültigkeit (Validität) und die Zuverlässigkeit (Reliabilität) von Facharztprüfungen müssen daher für die Fachgesellschaften und Standesorganisationen von grösster Wichtigkeit sein.

Ein zuverlässiges Examen reduziert Störfaktoren auf ein Minimum.

Eine zuverlässige Prüfung erfordert zudem, dass Störfaktoren so weit wie möglich ausgeschaltet werden. Bei schriftlichen Prüfungen zählen dazu etwa missverständliche Formulierungen oder ungewollte Lösungshinweise (sog. Cues), bei Fertigkeitprüfungen Beobachtungs- resp. Beurteilungsfehler und bei mündlichen Prüfungen die Auswirkungen persönlicher Vorlieben und Interaktionsstile von Examinatoren (vgl. Kapitel 2-7).

Welche Kandidaten bestehen gerade noch und welche nicht?

Summative Prüfungen wie die Facharztprüfung verlangen eine explizite und methodisch abgesicherte Festlegung der Bestehensgrenze. Idealerweise wird diese inhaltsbezogen in einem Konsensverfahren erarbeitet (vgl. Kapitel 8).

Es braucht viel Aufwand für die Entwicklung der Prüfung...

Der gesamte Ablauf der Prüfung - von der Entwicklung bis zur Schlussauswertung - muss sorgfältig geplant werden. Abb.0.1 zeigt notwendige Schritte. In der Vorbereitungsphase müssen Schlüsselkompetenzen bestimmt, Prüfungsmethoden gewählt und Prüfungsinhalte sorgfältig entwickelt werden. Wo Fertigkeiten geprüft werden sollen, haben sich die Examinatoren auf einheitliche Ansprüche zu einigen. Jede Aufgabe und deren Lösung muss von einem zweiten Experten unabhängig überprüft werden. Prüfungsunterlagen und Formulare müssen gedruckt und echte oder Standardisierte Patienten rekrutiert und instruiert werden.

...deren Durchführung...

Der reibungslose Ablauf eines Examens verlangt sorgfältig instruiertes Hilfspersonal. Mögliche Störfaktoren sollten vorausgesehen und notiert werden, sofern man sie nicht eliminieren kann. Der Verlauf aller Prüfungen sollte protokolliert werden, damit Zwischenfälle rekonstruiert werden können.

...und Auswertung.

Nach Abschluss der eigentlichen Prüfung müssen die Resultate so analysiert und interpretiert werden, dass einerseits die Kompetenzen der einzelnen Kandidaten beurteilt und andererseits die Eigenschaften der einzelnen Prüfungsaufgaben, zwischen kompetenten und inkompetenten Kandidaten zu unterscheiden, ermittelt werden können. Nur so lässt sich die Qualität der Facharztprüfung erhalten und ständig verbessern.

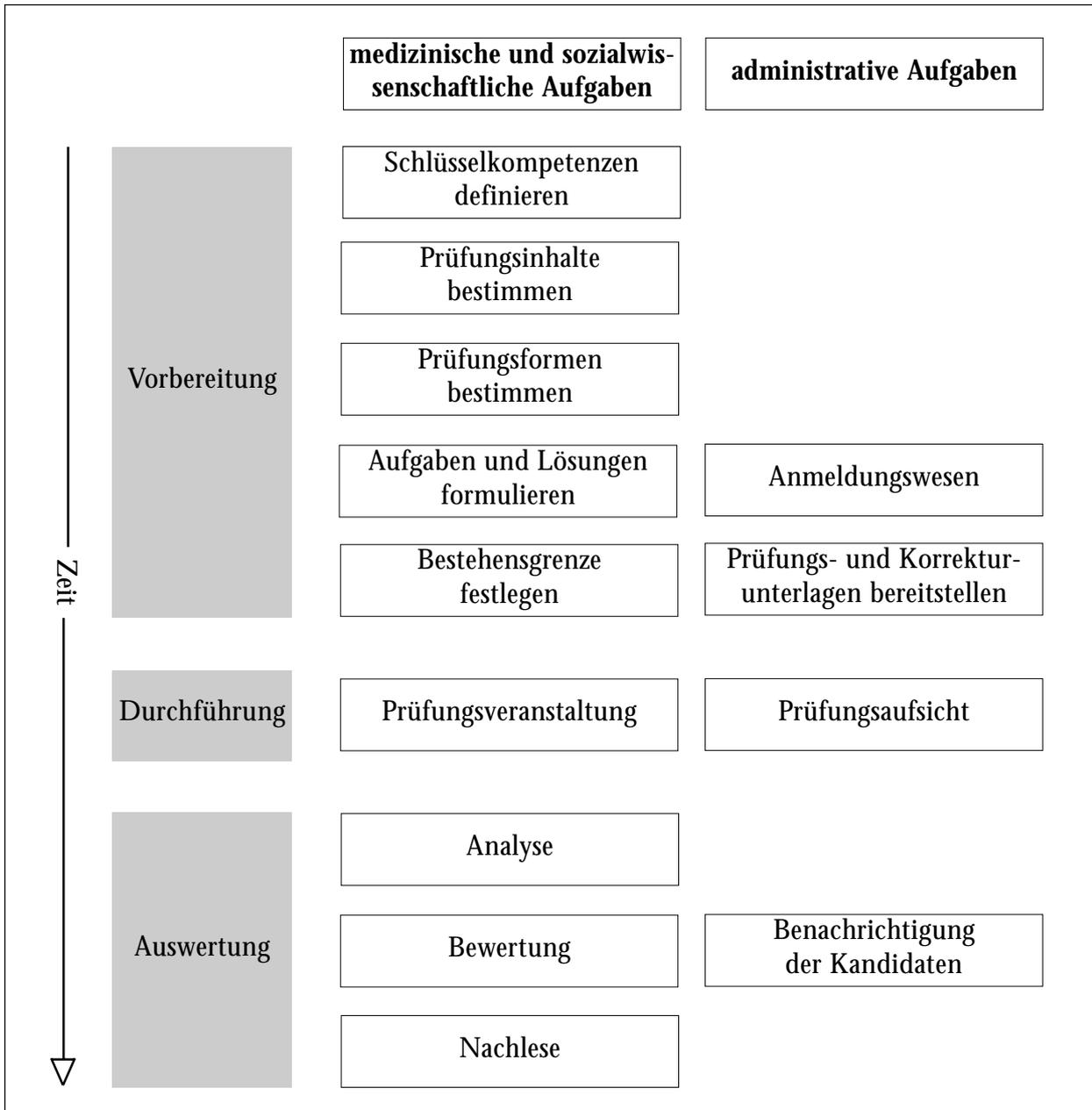


Abb.0.1 Hauptaufgaben bei der Erarbeitung einer neuen Facharztprüfung im zeitlichen Verlauf

Qualität kostet. Hohe Qualitätsansprüche können allerdings nur unter idealen Umständen und mit einem relativ grossen Aufwand befriedigt werden. Dieser bringt entsprechende Kosten für die Arbeit von Experten und Hilfspersonal mit sich. Die Kosten einer idealen Facharztprüfung sind hoch; es sind Kompromisse notwendig. Bei kleinen Kandidatenzahlen ist aus Kostengründen nicht der gleiche Standardisierungsgrad erreichbar wie bei grossen. Entsprechende Kompromisse und Abstriche müssen durchdacht und mit vollem Bewusstsein der Konsequenzen vollzogen werden.

Gewissenhaft und professionell vorbereitete, durchgeführte und ausgewertete Facharztprüfungen können das professionelle Verhalten eines Arztes - ja sogar die Zufriedenheit seiner Patienten - bis zehn Jahre nach der Prüfung voraussagen²!

Ralph Bloch

Literatur

- ¹ Nordquist CR, Danko G, McDermott JF Jr, Streltzer J, Lum KY. Pilot study of explicit grading criteria in the American Board of Psychiatry and Neurology. Part II. Examination. *Am J Psychiatry* 1996;153:1097-9
- ² Ramsey PG, Carline JD, Inui TS, Larson EB, Lo Gerfo JP, Wenrich MD. Predictive validity of certification by the American Board of Internal Medicine. *Ann Intern Med* 1989;110:719-26

1

Was prüfen?

Was muss ein Kandidat in der Facharztprüfung konkret wissen und können? Wenn Fachgesellschaften die Antwort nicht dem zufälligen Entscheid einzelner Exponenten überlassen wollen, müssen sie den Prüfungsinhalt in einem Prozess stufenweise erarbeiten: Zuerst muss ein Berufsbild formuliert werden. Dabei geht es um ein Leitbild, das Berufsvorstellungen der Fachgesellschaft ebenso festhält wie Sichtweisen und Bedürfnisse anderer Gesundheitsberufe und der Gesellschaft. Expertengremien fassen all diese Vorstellungen als Schlüsselkompetenzen des Facharztes zusammen. Über dieses Wissen und Können muss Letzterer am Ende der Weiterbildung verfügen. Da nicht alle Schlüsselkompetenzen in einem Examen geprüft werden können und sollen, muss eine Auswahl getroffen und als Prüfungsziel definiert werden. Das Strickmuster dieses Prüfungsstoffes ist der Blueprint, ein genaues Inhaltsverzeichnis mit gewichteten Kapiteln. Ein solches Inhaltsverzeichnis ist bei jeder Prüfungsform nötig, dient es doch als Vorlage für die repräsentative Stichprobe an Fragen und Aufgaben, die dem Kandidaten schliesslich gestellt werden.

In diesem Kapitel erfahren Sie,

- welche Anforderungen an den Inhalt einer Facharztprüfung zu stellen sind
- warum viele Prüfungsreglemente nicht ausreichen, um ein Examen inhaltlich zu gestalten
- wie Expertengruppen ein gewichtetes Inhaltsverzeichnis erarbeiten
- wie Aufgaben entwickelt werden, die Einzelkompetenzen prüfen

Eine gültige Prüfung entsteht aus einem Prozess

Der Inhalt einer gültigen Facharztprüfung muss folgenden Anforderungen genügen:

- für das Fach repräsentativ
- wissenschaftlich korrekt
- für die fachärztliche Tätigkeit relevant
- an den Bedürfnissen der Bevölkerung orientiert

*Anforderungen an
den Prüfungsinhalt*

Um einen solchen Prüfungsinhalt entwickeln zu können, sind wichtige Punkte zu definieren. Sie sind in Abb. 1.1 aufgeführt und anschliessend erläutert.

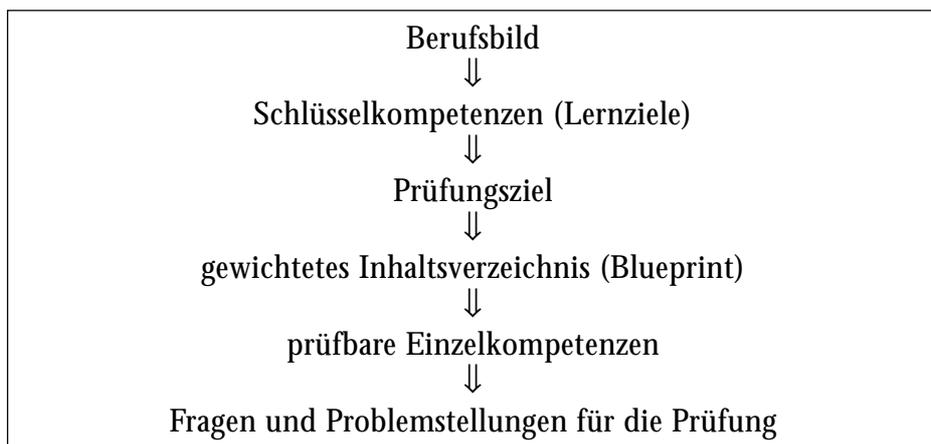


Abb. 1.1 Vom Berufsbild zur Prüfung

Berufsbild und Schlüsselkompetenzen

Jede Fachgesellschaft muss am Anfang der Prüfungsentwicklung explizit ihr Berufsbild formulieren. Aspekte dieses Leitbildes sind:

*Wer sind wir und
was machen wir?*

- fachliches Wissen, Können und Verhalten, wie es die Fachgesellschaft von ihren Mitgliedern verlangt
- Abgrenzungen gegenüber verwandten ärztlichen Fachgesellschaften
- Vorstellungen zur Zusammenarbeit mit anderen Gesundheitsberufen
- Bedürfnisse und Sichtweisen der Gesellschaft, der Patienten, der Versicherer
- Formen der Berufsausübung (Spital, Praxis)

*Schlüsselkompetenzen
definieren in breitem
Konsens*

Das Berufsbild gibt also eine erste allgemeine Übersicht über die Kompetenzen des zukünftigen Facharztes. Was dieser konkret wissen und können soll, wird als Schlüsselkompetenzen (auch als „Weiterbildungsziele“ oder „Lernziele“ bezeichnet) vom Berufsbild abgeleitet. Damit die wichtigen und richtigen Schlüsselkompetenzen gefunden werden, braucht es den Konsens breit abgestützter Gremien. Diese sollten möglichst alle Bereiche und Strömungen eines Faches und dessen Umfeld berücksichtigen. Die Zusammensetzung eines solchen Konsensgremiums ist in Abb. 1.2 skizziert.

Intraprofessionell	Interprofessionell	Gesundheitssystem
Vertreter des Faches (Abdeckung aller Richtungen)	Verwandte Fachdisziplinen Pflegebereich (stationär und ambulant) andere Gesundheitsberufe	Vertreter aus Politik Patientenorganisation Sozialversicherung

Abb. 1.2 Konsensusgremium für die Definition von Schlüsselkompetenzen

*Beispiel American
Board of Internal
Medicine*

Beispiel eines entsprechenden Gremiums ist das „Board of Advisors“ beim American Board of Internal Medicine. Das „Board of Advisors“ hat vor allem Nicht-Ärzte als Mitglieder und steht dem „Executive Committee“, dem zentralen intraprofessionellen Vorstand, zur Seite!

*Welche Unterlagen
helfen beim Definieren
von Schlüsselkompetenzen?*

Es gibt verschiedene Hilfsmittel bei der Definition der Schlüsselkompetenzen. Beispiele sind Erhebungen zu Konsultationsgründen (Beratungsursachen) oder zu Beratungsergebnissen einer Praxis, Praxisalltagsprotokolle, Patientendaten, Operationszahlen, Lehrbuch-Inhaltsverzeichnisse u.ä.

Die nachfolgenden zwei Beispiele aus Kanada und Holland sollen modellhaft aufzeigen, wie Schlüsselkompetenzen - hier im Grundausbildungsbereich - aussehen können. Diese Schlüsselkompetenzen sind dem Umfang entsprechend in Buchform erschienen^{2,3}.

Beispiel aus Kanada

Ein interdisziplinäres Team von Internisten, Pädiatern, Chirurgen, Psychiatern, Gynäkologen/Geburtshelfern sowie Sozial- und Präventivmedizinerinnen hat 1992 in Kanada die „objectives for the qualifying examination“ vorgelegt². Dabei handelt es sich um ein sehr detailliertes Verzeichnis, das Schlüsselkompetenzen beschreibt, über die jeder Arzt nach der Ausbildung verfügen sollte (vgl. Abb. 1.3).

Begründung: Die Differentialdiagnose der Anämie kann ein komplexes Problem darstellen. Eine ungezielte, unstrukturierte Abklärung kann teuer werden und ineffizient sein. Einfache Zusatzuntersuchungen können wichtige Informationen liefern. Manchmal ist die Anämie die einzige Manifestation einer schweren Krankheit.

Klinische Situation: Eine 60-jährige Frau kommt in die Sprechstunde wegen Müdigkeit, Atemnot und Blässe.

Kompetenzen: Kompetente Kandidaten

1. diagnostizieren Anämieursachen entsprechend dem Alter der Patientin
2. erkennen, dass eine Eisenmangelanämie auf eine gastrointestinale Erkrankung hinweisen kann

Teilkompetenzen: Kompetente Kandidaten können

1. erklären, dass ein einfacher Blutausstrich mit Zellzählung der erste Untersuchungsschritt bei einer Anämie ist
2. die Befunde des Eisen- und Vitaminmangels in einem Blutausstrich interpretieren
3. die Anämien nach der Morphologie der roten Zellen klassifizieren
4. wichtige Ursachen der einzelnen Anämieformen aufzählen
5. Erkenntnisse aus Anamnese und körperlicher Untersuchung gewinnen und interpretieren, um die Anämieursache zu bestimmen
6. diagnostische Schlüsseltests aufzählen und interpretieren, um Anämieursachen zu bestätigen
7. die therapeutischen Grundzüge der Eisenmangelanämie erklären

Abb. 1.3 „Anämie“ aus „objectives for the qualifying examination“²

Im Buch „training of doctors, blueprint 1994, objectives of under-graduate medical education in the Netherlands“³ sind von einem breit abgestützten Gremium Berufsbild und Schlüsselkompetenzen festgehalten worden (vgl. Abb. 1.4). Diagnoseorientierte, nach Fächern und Körpersystemen aufgeteilte Fertigungslisten und Listen „klinischer Bilder“ werden in diesem Buch kombiniert mit einer symptomorientierten Problemliste. Zusätzlich sind verschiedene Kompetenzstufen unterschieden.

Beispiel aus den Niederlanden

Einerseits muss ein klinisches Bild erkannt und eingeordnet werden (1), andererseits muss mit dem klinischen Bild gearbeitet werden können (2). Diese 2. Stufe ist weiter unterteilt in eine diagnostische (D) und eine therapeutische (T) Kompetenz. Ein frisch diplomierter Arzt muss beispielsweise in der Situation „Brustkrebs“ sein Wissen zu Diagnostik und Therapie unter Beweis stellen (2). Er muss aber die einzelnen Diagnose- und Therapieschritte nicht selber ausführen können.

<i>Brustkrebs</i>			
<i>Bild / Symptome erkennen und einordnen</i>	<i>Arbeiten in der praktischen Situation</i>		
	<i>Wissen zu Diagnose und Therapie</i>	<i>Diagnose mit eigenen Mitteln</i>	<i>Therapie mit eigenen Mitteln</i>
<i>1</i>	<i>2</i>	<i>D</i>	<i>T</i>

Abb. 1.4 Klinisches Bild „Brustkrebs“³

Prüfungsziel

Welche Schlüsselkompetenzen werden geprüft?

Sind die Schlüsselkompetenzen einmal definiert, werden die Prüfungsziele bestimmt. Diese sollen festlegen, welche Teile der Schlüsselkompetenzen geprüft werden müssen oder können. Diese Auswahl an Wissen und Können bestimmt schliesslich auch die Prüfungsmethode (vgl. Kapitel 2). Teilkompetenzen, die nicht oder zu einem andern Zeitpunkt der Weiterbildung geprüft werden, sind ebenfalls zu bezeichnen.

Viele Fachgesellschaften übernehmen als Prüfungsziel faktisch die Schlüsselkompetenzen^{4,5}. Für die Entwicklung des gewichteten Inhaltsverzeichnisses (siehe unten) ist es aber nötig zu wissen, was wo geprüft werden soll und was nicht.

Das gewichtete Inhaltsverzeichnis (Blueprint)

Ein Blueprint hat Dimensionen,...

Als nächster Schritt hin zum konkreten Prüfungsinhalt entsteht ein möglichst detailliertes Inhaltsverzeichnis zu den Prüfungszielen, ein sog. Blueprint. Dieser Blueprint kann den Prüfungsstoff in verschiedene Ebenen oder Dimensionen aufteilen:

- Fachgebiete
- Körpersysteme
- biologische, psychologische, soziale Aspekte
- Handlungskategorien (Fertigkeiten)
- häufige Konsultationsgründe (Beratungsursachen)
- Alter der Patienten

...Kategorien und Gewichtungen.

Sind die Dimensionen festgelegt, werden sie in Kategorien eingeteilt. Zu wichtigeren Kategorien sollen mehr Fälle oder Fragen in der Prüfung vorkommen als zu weniger wichtigen. Darum müssen die Kategorien von

Experten in einem Konsensprozess gewichtet werden. Jede Kategorie (evtl. Unterkategorie) erhält eine Prozentzahl, resp. eine diesem Prozentsatz entsprechende Anzahl Fragen (vgl. Abb. 1.5).

<i>Kategorie</i>	<i>Gewichtung in %</i>
<i>Säuglinge (0-1)</i>	<i>5</i>
<i>Kinder (1-11)</i>	<i>16</i>
<i>Adoleszente (12-18)</i>	<i>16</i>
<i>Erwachsene (19-64)</i>	<i>47</i>
<i>ältere Menschen (>64)</i>	<i>16</i>
<i>TOTAL</i>	<i>100</i>

*Beispiel einer
 Blueprintdimension*

Abb. 1.5 Blueprintdimension Alter mit Kategorien und Gewichtung⁶

Bei der Gewichtung können Angaben zur Häufigkeit einzelner Blueprintkategorien im Berufsalltag helfen. Solche Angaben liefern beispielsweise Erhebungen über häufige Beratungsursachen einer Praxis oder über die Anzahl durchgeführter Operationen eines Spezialchirurgen. Allerdings sind Häufigkeit und Wichtigkeit nicht immer gleichzusetzen. Auch Inhaltsverzeichnisse von Lehrbüchern können bei der Gewichtung hilfreich sein. Die Kapitellänge entspricht allerdings wohl selten der Relevanz des Themas für die Facharztprüfung.

*Welche Unterlagen
 helfen bei der Gewichtung?*

Gängige Dimensionen sind heute Fachgebiete oder Körpersysteme. Eindimensionale Blueprints weisen oft Anteile verschiedener Dimensionen auf.

Der Blueprint für die MC-Prüfung der schweizerischen Fachgesellschaft für Innere Medizin (vgl. Abb. 1.6) mischt in einer Dimension Fachbereiche (z.B. Onkologie), Körpersysteme (z.B. Knochenkrankheiten), häufige Konsultationsgründe (z.B. Hypertension), Handlungskategorien (z.B. Notfallmedizin), Alter der Patienten (z.B. Geriatrie) und bio-psycho-soziale Dimensionen (z.B. Alkoholismus).

Beispiel eines eindimensionalen Blueprints

Blueprint Facharztprüfung Innere Medizin		Anteil Fragen %
A Generelle Konzepte und allgemeine Fragen		
1.	1 Prävention 2 Versicherungs- und Sozialwissen 3 Gesundheitsökonomie 4 Ethik 5 Statistik 6 Medizinische Technik inkl. Labor	4
B Fachfragen über spezifische Probleme		
2.	1 Allergie, Immunologie 2 Rheumatologie, Knochenkrankheiten	4 9 13
3.	1 Allgemeine Innere Medizin 2 Geriatrie 3 Neurologie, Psychiatrie, Psychosomatik 4 Intoxikationen, Nebenerscheinungen, Alkoholismus	6 3 5 2 16
4.	1 Kardiologie 2 Angiologie 3 Hypertension	10 2 3 15
5.	1 Endokrinologie, Stoffwechsel 2 Nephrologie	7 4 11
6.	1 Gastroenterologie, Hepatologie 2 Ernährung	8 2 10
7.	1 Hämatologie, Onkohämatologie 2 Onkologie 3 Infektiologie	3 5 7 15
8.	1 Pneumologie 2 Intensivmedizin 3 Notfallmedizin	8 4 4 16
Total		100

Abb. 1.6 Eindimensionaler Blueprint: Schweizerische Facharztprüfung Innere Medizin, schriftlicher Teil mit MC-Fragen (von der Fachgesellschaft freundlicherweise zur Verfügung gestellt)

Abb. 1.7 bis 1.9 zeigen Beispiele mehrdimensionaler Blueprints. Mit mehreren Dimensionen wird ein Fachgebiet besser repräsentiert.

<i>Rahmen → Aufgabe ↓</i>	<i>Notfall- situation</i>	<i>Verdacht auf ernste Störung</i>	<i>akut be- grenzte Störung</i>	<i>psycho- soziales Problem</i>	<i>Prävention Beratung</i>
<i>Anamnese</i>					
<i>körperliche Untersuchung</i>					
<i>Kommuni- kation</i>					

*Beispiel eines
zweidimensionalen
Blueprints*

Abb. 1.7 Zweidimensionaler Blueprint „National Board for Medical Examiners“

Dieser OSCE-Blueprint, mit dem Ärzte nach einjähriger Assistenz qualifiziert werden, weist die Dimensionen „Aufgabe“ und „Rahmen“ auf. Er hat folgende Gewichtungsregeln bei den Aufgabendimensionen: drei Kategorien, je ein Drittel aller Aufgaben und bei den Rahmendimensionen: mindestens eine Station zu jeder Kategorie, Schwergewicht bei akuten, unklaren und chronischen Störungen. Dazu kommen weitere Gewichtungen nach Organsystemen, Fachbereichen, Patientencharakteristika (Alter, Geschlecht, Status). Mindestens ein Fall enthält eine „emotionale Herausforderung“.

Der zweite Teil der Berufszulassungsprüfung des „Medical Council of Canada“ nach ein bis zwei Jahren klinischer Assistenz - ein OSCE mit derzeit 14 Stationen - sieht wie folgt aus:

<i>Dimension</i>	<i>Kategorie</i>	<i>Anzahl Stationen</i>
<i>Fächer:</i>	<i>Medizin</i>	<i>4</i>
	<i>Chirurgie</i>	<i>3</i>
	<i>Pädiatrie</i>	<i>2</i>
	<i>Gynäkologie/Geburtshilfe</i>	<i>2</i>
	<i>Psychiatrie</i>	<i>3</i>
	<i>Sozial-/Präventivmedizin</i>	<i>0</i>
<i>Geschlecht:</i>	<i>F</i>	<i>6</i>
	<i>M</i>	<i>7</i>
	<i>F/M</i>	<i>1</i>
<i>„domain“:</i>	<i>Anamnese</i>	<i>6</i>
	<i>körperliche Untersuchung</i>	<i>3</i>
	<i>Kommunikation</i>	<i>2</i>
	<i>Management</i>	<i>3</i>

*Beispiel eines
dreidimensionalen
Blueprints*

Abb. 1.8 OSCE-Blueprint im „qualifying examination part II“ des „Medical Council of Canada“ 1998

<i>Fragenebene: Dimension Allgemeinmedizin</i>	<i>%</i>	<i>Dimension Fachwissen Allgemeinmedizin</i>	<i>%</i>
<i>Allgemeinmedizinisches Fachwissen</i>	<i>40</i>	<i>Chirurgie</i>	<i>10</i>
<i>Konsultation:</i>	<i>35</i>	<i>Innere Medizin</i>	<i>30</i>
<i>Anamnese</i>	<i>10</i>	<i>Arbeitsmedizin</i>	<i>2</i>
<i>Körperliche Untersuchungen</i>	<i>8</i>	<i>Dermatologie</i>	<i>5</i>
<i>Zusatzuntersuchungen</i>	<i>6</i>	<i>Gynäkologie/Geburtshilfe</i>	<i>6</i>
<i>Therapie, Management, Entscheidungen</i>	<i>8</i>	<i>Neurologie</i>	<i>7</i>
<i>Administration</i>	<i>3</i>	<i>Ophthalmologie</i>	<i>4</i>
<i>Interaktion, Kommunikation, Arzt-Patient-Beziehung</i>	<i>15</i>	<i>Otorhinolaryngologie</i>	<i>7</i>
<i>Ethik</i>	<i>5</i>	<i>Pädiatrie</i>	<i>8</i>
<i>Recht</i>	<i>5</i>	<i>Psychiatrie</i>	<i>9</i>
<i>TOTAL</i>	<i>100</i>	<i>Medizinische Radiologie (Radiodiagnostik)</i>	<i>5</i>
		<i>Rheumatologie</i>	<i>7</i>
		<i>TOTAL</i>	<i>100</i>
<i>Fallebene: Dimension Alter</i>	<i>%</i>	<i>Dimension Konsultationsart</i>	<i>%</i>
<i>Säuglinge (0-1)</i>	<i>5</i>	<i>Sprechstunde</i>	<i>70</i>
<i>Kinder (1-11)</i>	<i>10</i>	<i>Hausbesuch</i>	<i>10</i>
<i>Adoleszente (12-18)</i>	<i>10</i>	<i>Notfall</i>	<i>5</i>
<i>Erwachsene (19-64)</i>	<i>45</i>	<i>Telefon</i>	<i>10</i>
<i>ältere Menschen (>64)</i>	<i>30</i>	<i>Indirekte Konsultation (Akten, Gutachten)</i>	<i>5</i>
<i>TOTAL</i>	<i>100</i>	<i>TOTAL</i>	<i>100</i>

Abb. 1.9 Mehrdimensionaler Blueprint: Schweizerische Facharztprüfung Allgemeinmedizin, schriftlicher Teil mit Kurzantwortfragen (von der Fachgesellschaft freundlicherweise zur Verfügung gestellt)

Via prüfbare Einzelkompetenzen zu konkreten Fragen

Konsens vor dem Fragenschreiben

Es ist sinnvoll, vor der Aufgabenformulierung Konsens über die Richtigkeit und Relevanz von Prüfungsinhalten zu erreichen. Nützlicherweise werden zuerst ein Problem und die entsprechend zu prüfenden Einzelkompetenzen definiert und richtige, allseits akzeptierte Lösungen erarbeitet. Erst dann sollten die Inhalte in Aufgaben verpackt werden.

vom praktischen Problem zu konkreten Prüfungsfragen

Im Folgenden wird beschrieben, wie auf diese Art konkrete Prüfungsinhalte entstehen, die dann in eine Prüfungsform umgesetzt werden können: Bezug nehmend auf ein Blueprint-Kapitel werden praktische Probleme („key feature problems“) aufgelistet. Beispiele: der verwirrte Patient (Neurologie), akutes Abdomen beim Kind (Chirurgie), Eosinophilie (Hämatologie)⁷. Nun werden die Kompetenzen beschrieben, die der Facharzt zur Definition und zur Lösung der Probleme haben muss. Es geht dabei um Kenntnisse und Fertigkeiten in Anamnese, Diagnostik, therapeutischen Verfahren wie auch um Kompetenzen im psychosozial-interaktiven Bereich. Diese einzelnen Kompetenzen dienen als Grundlage, um

Prüfungsfragen oder Beobachtungssituationen zu konstruieren. In Abb. 1.10 - 1.12 ist der Weg vom praktischen Problem bis zur konkreten Prüfungsfrage beschrieben.

*Bestimme ein praktisches Problem aus einer Blueprintkategorie.
Beschreibe typische Situationen für dieses Schlüsselproblem.
Notiere dazu maximal 5 Schlüsselkriterien, die für das Problem wichtig, richtig oder „kritisch“ sind.
Die Kriterien stammen meist aus zwei Bereichen, nämlich erstens aus der Definition des Problems (Symptome, Informationen, Diagnosen, Untersuchungen) und zweitens aus der Problemlösung (wichtige Schritte in der weiteren Abklärung und Therapie, Entscheidungen, Soforttherapie, Patienteninformation).
Schreibe einen praktischen Fall aus dem Schlüsselproblem.
Schreibe Fragen dazu.
Eine Frage deckt meist ein Schlüsselkriterium ab. Ein Fall kann eines oder mehrere Schlüsselprobleme betreffen.*

Abb. 1.10 Die Operationalisierungsschritte eines „key feature problem“⁶

Wenn ein Erwachsener über ein schmerzhaft geschwollenes Bein klagt, sollte der Arzt

- 1. eine tiefe Venenthrombose in seine Differentialdiagnose einbeziehen*
- 2. in der Anamnese des Patienten die Risikofaktoren für eine tiefe Venenthrombose herausarbeiten*
- 3. eine Venenkontrastdarstellung veranlassen als definitiv diagnostische Untersuchung bei tiefer Venenthrombose*

Abb. 1.11 Schlüsselproblem mit Schlüsselkriterien⁶

Paul, 56 Jahre alt, erscheint in Ihrer Sprechstunde, weil ihn sein linkes Bein seit zwei Tagen zunehmend schmerzt. Er erzählt, dass die Wade unterhalb des Knies empfindlich und um den Knöchel herum geschwollen sei. Er habe noch nie so etwas gehabt. Das rechte Bein sei in Ordnung.

1. Welche Diagnosen würden Sie zu diesem Zeitpunkt in Betracht ziehen? Nennen Sie 3.

1. _____

2. _____

3. _____

2. Unter Berücksichtigung Ihrer Diagnosen wollen Sie folgende Elemente in der Anamnese des Patienten besonders beachten (wählen Sie 7 aus):

- | | |
|---|--------------------------------------|
| <i>1. Aktivität bei Symptombeginn</i> | <i>16. Kopfschmerzen</i> |
| <i>2. Alkoholeinnahme</i> | <i>17. kürzliche Halsschmerzen</i> |
| <i>3. Allergien</i> | <i>18. kürzliche Immobilisation</i> |
| <i>4. Angina pectoris</i> | <i>19. kürzliche Operation</i> |
| <i>5. antiinflammatorische Therapie</i> | <i>20. kürzliche Zahnprozedur</i> |
| <i>6. Claudicatio intermittens</i> | <i>21. kürzliches Arbeitsproblem</i> |
| <i>7. Farbe Stuhlgang</i> | <i>22. lumbale Rückenschmerzen</i> |
| <i>8. frühere Harnwegsinfektionen</i> | <i>23. Nykturie</i> |
| <i>9. frühere Knieprobleme</i> | <i>24. Palpitationen</i> |
| <i>10. frühere Neoplasien</i> | <i>25. Parästhesien</i> |
| <i>11. frühere Rückenprobleme</i> | <i>26. parox. nächtl. Dyspnoe</i> |
| <i>12. Hämatemesis</i> | <i>27. Polydypsie</i> |
| <i>13. Hormontherapie</i> | <i>28. Rauchen</i> |
| <i>14. Husten</i> | <i>29. Wunden am Fuss</i> |
| <i>15. Impotenz</i> | <i>30. Wunden an der Hand</i> |

Abb. 1.12 Ein aus dem Schlüsselproblem in Abb. 2.11 entstandener Prüfungsfall⁶

Das obige Beispiel (Abb. 1.10. - 1.12) kann auch als Grundlage für mündliche Prüfungsfragen oder praktische Examen verwendet werden. Bei praktischen Prüfungen ergeben die Schlüsselkriterien die Items, welche die Prüfer beobachten müssen.

von Symptomen zur Frage

Eine andere Möglichkeit, Prüfungsaufgaben zu entwickeln, geht direkt von einem Symptom oder Befund aus⁸. Die Expertengruppe assoziiert zu einem Symptom aus einer Blueprintkategorie möglichst viele passende resp. akzeptierte richtige Diagnosen. Erst jetzt werden Fragen resp. „Stäm-

me“ geschrieben mit einer Diagnose als richtiger Antwort. Die übrigen Diagnosen dienen z.B. bei MC-Fragen vom Typ der erweiterten Zuordnungen („extended-matching“-MC-Fragen) als Ablenker (siehe Abb. 1.13).

Symptom: Müdigkeit

Optionen:

- A) akute Leukämie*
- B) Anämie bei chronischer Krankheit*
- C) Depression*
- D) Eisenmangel*
- E) Epstein-Bar-Virusinfektion*
- F) Folsäuremangel*
- G) Glucose-6-Phosphat-Dehydrogenase-Mangel*
- H) hereditäre Sphärozytose*
- I) Herzinsuffizienz*
- J) Hypothyreose*
- K) Lyme-Krankheit*
- L) mikroangiopathische, hämolytische Anämie*
- M) miliare Tuberkulose*
- N) Vitamin B12-Mangel*

Überleitung: Wählen Sie bei jedem Patienten mit Müdigkeit die wahrscheinlichste Diagnose aus.

Stämme:

1. Eine 19-jährige Frau ist seit einer Woche müde, hat Fieber und Halsweh. Die Temperatur beträgt 38.3° C, es bestehen eine zervikale Lymphadenopathie und Splenomegalie. Eine erste Laboruntersuchung zeigt folgende Befunde: Leukozyten 5000/mm³ (80% Lymphozyten mit vielen Atypien), GOT 200 U/l, normale Werte für alkalische Phosphatase und Serumbilirubin.

richtige Antwort: E

2. Ein 15-jähriges Mädchen klagt seit zwei Wochen über Müdigkeit und Rückenschmerzen. Ihre Haut ist blass und weist Petechien auf. Rücken und beide Oberschenkel sind berührungsempfindlich. Die Blutuntersuchung ergibt ein Hämoglobin von 7.0 mg%, Leukozyten von 2000/mm³ und Thrombozyten von 15'000/mm³.

richtige Antwort: A

Abb. 1.13 „Extended matching-Item“ (R-Typ) auf Basis einer „option list“⁸

*Prüfungsaufgaben
entstehen heute
eher zufällig...*

Heute entstehen Prüfungsaufgaben noch kaum in einem der oben beschriebenen Prozesse. Prüfungsverantwortliche der meisten Medizinal-examen bestimmen die Einzelkompetenzen, die geprüft werden, eher unsystematisch und zufällig. Die Autoren entwickeln Prüfungsaufgaben oft aus ihrer täglichen Arbeit heraus. So entstehen beispielsweise solche zu Fällen von Patienten, die zufälligerweise gerade in die Sprechstunde kommen oder im Spitalbett liegen. Andere Autoren halten sich an Lehrbücher oder veröffentlichte Fragen aus ihrem Fachgebiet. Wieder andere stellen die Prüfung direkt aus einer veröffentlichten fachärztlichen Fragensammlung zusammen.

*...und bieten oft
Schwierigkeiten.*

Es erstaunt nicht, dass bis zu einem Drittel der so vorbereiteten Prüfungsaufgaben den Revisionsprozess nicht überstehen, weil sie als zu kontrovers, zu spitzfindig, zu trivial oder als fachlich falsch beurteilt werden. Oft kommen nicht genügend Prüfungsaufgaben zu allen Blueprintkategorien zusammen. Ein Fragenautor kann eben nicht sämtliche Facetten der Beantwortbarkeit oder der Relevanz eines Problems erfassen. Es ist nachvollziehbar, dass sich Fragenautoren manchmal demotiviert zurückziehen, und Prüfungsverantwortliche Mühe haben, genügend neue Fragen und Fälle zu erhalten.

*Gruppenarbeit
motiviert.*

Werden die Prüfungsaufgaben hingegen wie in diesem Kapitel aufgezeigt entwickelt, kann die qualitative und quantitative Bilanz wesentlich verbessert werden. Wichtiger Aspekt ist dabei die Arbeit in Expertengruppen, die sich regelmässig treffen. Die Teilnehmer erleben diese Art der Prüfungsvorbereitung meist als viel motivierender als die delegierten „Hausaufgaben“ im einsamen Büro.

Peter Schläppi

Literatur

- ¹ American Board of Internal Medicine Resource book. Spring Special Edition, 1997
- ² Objectives for the qualifying examination. Ottawa: The Medical Council of Canada, 1992
- ³ Metz JCM, Stoelinga GBA, Pels Rijcken EH, van den Brand BWM. Blueprint 1994: training of doctors in The Netherlands, objectives of undergraduate medical education. University of Nijmegen, NL, 1994
- ⁴ Ausbildung zum Arzt für Allgemeinmedizin und zum Facharzt. Ärzte-Ausbildungsordnung. BGBl. Nr. 152/1994. Österreichische Ärztekammer Verlag, 1994
- ⁵ Facharzt FMH für Innere Medizin. Weiterbildungsprogramm Schweiz. Ärztezeitung 1995;76:144-8
- ⁶ Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. Acad Med 1995; 70:194-201
- ⁷ Dauphinee D, Fabb W, Jolly B, Langsley D, Wealthall S, Procopis P. Determining the content of certifying examinations. In: Newble D, Jolly B, Wakeford R. The certification and recertification of doctors. Issues in the assessment of clinical competence. Cambridge University Press, 1994:92-104
- ⁸ Case SM, Swanson DB. Constructing Written Test Questions for the Basic and Clinical Science 2nd ed. Philadelphia: National Board of Medical Examiners, 1998

2

Wie prüfen?

INHALT

- 2.1 Welche Methode wählen? - 43
- 2.2 Was leisten die Methoden? - 48
 - Wahlantwortverfahren (MC) - 49
 - Kurzantwortfragen (KAF) - 51
 - Strukturierte mündliche Prüfung (SMP) - 52
 - Strukturierte Beobachtung (SB) - 54
 - Objektives strukturiertes klinisches Examen (OSCE) - 56

Der wichtigste Qualitätsaspekt bei jeder Prüfungsmethode ist die Gültigkeit: ein Examen muss prüfen, ob die Absolventen in der Lage sind, ihre zukünftige Arbeit kompetent auszuführen oder nicht. Dies setzt primär voraus, dass die entscheidenden Teilkompetenzen erfasst werden. Eine Prüfung ist zudem nur dann gültig, wenn die Resultate objektiv und zuverlässig sind. Diese beiden Aspekte sind zugleich wichtige Bedingungen, damit ein Examen in einem Rekurs bestehen kann.

Es gibt verschiedene Möglichkeiten, Kandidaten entsprechend zu prüfen. In diesem Kapitel ist beschrieben, was bei der Wahl der Prüfungsmethoden zu berücksichtigen ist, und welche wichtigsten Vor- und Nachteile die einzelnen Verfahren aufweisen. Für die Wahl der Prüfungsmethoden gibt es allerdings kein allgemeingültiges Rezept. Jede Fachgesellschaft muss in einem Prozess die für sie praktikable beste Lösung entwickeln. Dieses Kapitel bietet Ihnen einige Richtlinien und notwendige Informationen dazu.

In diesem Kapitel erfahren Sie,

- dass sich die Methodenwahl nach den zu prüfenden Kompetenzen richten muss und nicht umgekehrt
- dass eine gute Prüfung in der Regel mehr als eine Methode erfordert, da mit gewissen Vorgehensweisen in erster Linie theoretische und mit andern vor allem praktische Kompetenzen geprüft werden können
- dass der Aufwand, um eine hinreichend objektive und zuverlässige Prüfung zu erreichen, nicht bei jeder Methode gleich hoch ist und dieser Umstand bereits bei der Planung zu berücksichtigen ist
- warum grundsätzlich nur gut erprobte, strukturierte Methoden verwendet werden sollen
- warum bei der Methodenwahl auch die Signalwirkung für das Lernen der Kandidaten zu bedenken ist
- warum es sinnvoll ist, zuerst eine unter den Aspekten der Gültigkeit, der Messzuverlässigkeit und der Signalwirkung ideale Prüfung ins Auge zu fassen, diese dann nach den Rahmenbedingungen auf das Machbare zu reduzieren und anschliessend sukzessive zu optimieren

2.1 Welche Methode wählen?

Die Facharztprüfung soll als Qualitätskontrolle des „Produkts Facharzt“ dienen. Um diesem Anspruch gerecht zu werden, müsste sie im Idealfall messen können, wie sich die Kandidaten im Berufsalltag verhalten. Tatsächlich lässt sich aus Prüfungsergebnissen nur erschliessen, wie sich Personen verhalten können, d.h. über welche Kompetenzen sie verfügen. Wieweit sie im Alltag tatsächlich entsprechend handeln, ist von Faktoren abhängig, die sich in einer Prüfungssituation praktisch nicht evaluieren lassen. Dies müsste man während der Weiterbildung durch längerfristige Beobachtung zu erfassen versuchen.

*prüfbar sind nur
Kompetenzen*

Das Wichtigste an einer Prüfung ist ihre Gültigkeit (Validität). Um gültig zu sein, muss sie die Kompetenzen messen, die erforderlich sind, um die konkreten Aufgaben im Berufsalltag lösen zu können. Wie Sie im Abschnitt 2.2 dieses Kapitels sehen werden, eignen sich nicht alle Prüfungsmethoden gleichermassen zur Erfassung bestimmter Kompetenzen. Die Methodenwahl muss sich also primär nach den zu prüfenden Kompetenzen richten. Die Wahl der am besten geeigneten Methoden garantiert allerdings noch keine valide Prüfung. Wesentlicher ist, dass jeweils die richtigen Prüfungsinhalte gewählt werden¹. Eine Wissensprüfung beispielsweise sollte möglichst ausschliesslich anwendungsorientiertes Wissen messen² und sich auf diejenigen Elemente eines Problems konzentrieren, die für dessen Lösung entscheidend sind^{3,4}.

*Validität durch
Prüfen der richtigen
Kompetenzen und
Inhalte*

Damit die Prüfungsergebnisse verschiedener Kandidaten miteinander verglichen werden können, müssen die Bedingungen für alle möglichst gleich sein. Wenn die Prüfungsaufgaben für alle Kandidaten identisch sind, spricht man von Standardisierung. Diese ist bei praktischen Examen schwieriger zu erreichen als bei schriftlichen Prüfungen. So ist es z.B. echten Patienten kaum zuzumuten, sich von sämtlichen Kandidaten befragen und untersuchen zu lassen. Um dennoch eine hohe Standardisierung zu erreichen, werden bei wichtigen Prüfungen heute vielerorts speziell geschulte sog. Standardisierte Patienten (vgl. Kapitel 7) eingesetzt.

*Objektivität durch
Standardisierung...*

Wenn es nicht möglich ist, allen Kandidaten identische Aufgaben zu stellen, kann die Objektivität der Prüfung durch eine im Voraus geplante Strukturierung verbessert werden. Diese schränkt vermeidbare Störfaktoren, welche die Prüfungsergebnisse verzerren können, ein. Zu den strukturierenden Massnahmen gehören

...und Strukturierung

- eine Auswahl der Prüfungsaufgaben, die gewährleistet, dass alle Kandidaten über verschiedene Themen im gleichen Verhältnis befragt werden

- ein vor der Prüfung festgelegter Ablauf
- ein im Voraus bestimmtes Beurteilungsschema mit entsprechendem Arbeitsblatt
- ein nachvollziehbares Verfahren, wie bei unterschiedlichen Beurteilungen verschiedener Examinatoren vorgegangen wird

Eine geringe Strukturierung weisen traditionelle mündliche Prüfungen auf, vor allem wenn

- der Prüfungsverlauf willkürlich von der Wechselwirkung zwischen Prüfer und Prüfling abhängt
- Patientenfälle und Probleme ad hoc ausgewählt werden
- die Prüfer spontan entscheiden, welche Fragen sie stellen
- die Prüfer ad hoc beurteilen, welche Antworten sie als richtig erachten
- die Notengebung auf einer impliziten, globalen Leistungsbeurteilung (global ratings) beruht

Wie auch mündliche Prüfungen wesentlich strukturierter und damit objektiver gestaltet werden können, zeigt Kapitel 5.

Nur bei hinreichender Strukturierung und Standardisierung sind Leistungen verschiedener Kandidaten miteinander vergleichbar, können inhaltsbezogene Bestehensgrenzen festgelegt und in einem Rekursfall Entscheide fundiert verteidigt werden. Diese Überlegungen dürfen aber nicht dazu verleiten, zu einer MC-Prüfung Zuflucht zu nehmen, wenn unter dem Validitätsgesichtspunkt eine andere Methode angezeigt ist. Mit unterschiedlichem Aufwand lassen sich alle hier beschriebenen Methoden hinreichend objektiv gestalten. Es geht also vielmehr darum, bei der Auswahl und Zusammenstellung von Methoden den Objektivitätsaspekt und den erforderlichen Strukturierungs- und Standardisierungsaufwand in die Rechnung einzubeziehen.

*Reliabilität durch
Reduktion von
Störfaktoren ...*

Im Weiteren ist wichtig, dass ein Examen Kompetenzen zuverlässig misst. Das Mass für die Zuverlässigkeit von Prüfungsergebnissen ist die Reliabilität. Bei einer hohen Reliabilität wird eine Wiederholung der Prüfung weitgehend zu den gleichen Resultaten führen. Die Messzuverlässigkeit wird durch Einflüsse beeinträchtigt, die nichts mit dem zu tun haben, was die Prüfung messen soll. Dazu gehören neben einer mangelhaften Objektivität etwa auch Rateinflüsse oder sprachliche Missverständnisse.

Als wichtigster Störfaktor der Reliabilität wurde in neuerer Zeit die Problem- oder Fallspezifität erkannt¹. Die Kandidatenleistung variiert über verschiedene Probleme und Patientenfälle stark. Erst anhand der Leistungen in einer genügend grossen Stichprobe von Fällen und Problemen lässt sich zuverlässig aussagen, wie ausgeprägt die Kompetenz eines Kandidaten ist. Jede Methode taugt also erst von einem bestimmten Umfang an etwas. Dies ist für die Auswahl und Zusammenstellung der Methoden bedeutsam. Es wäre unsinnig, im Interesse einer breiten Validität ein bisschen Wissen, ein bisschen Fertigkeiten und ein bisschen klinisches Problemlösen zu prüfen. Die Ergebnisse würden so unzuverlässig, dass die Gesamtprüfung auch keine Validität hätte. Ebenfalls bedeutsam ist die Reliabilität für die Rekursfestigkeit einer Prüfung. Ein Misserfolgsentscheid ist schwer zu verteidigen, wenn wegen mangelnder Zuverlässigkeit der Messfehler gross ist.

... und durch eine angemessene Prüfungsdauer

Die Reliabilität kann theoretisch Werte zwischen 0 und 1 annehmen. Der Wert 0 bedeutet, dass die Prüfungsergebnisse völlig zufällig sind. Der Wert 1 steht dafür, dass ein unterschiedliches Resultat zu 100 Prozent auf unterschiedliche Ausprägung der Kompetenz zurückzuführen ist, welche die Prüfung misst. Für Prüfungen, die einschneidende Konsequenzen für die Kandidaten haben können, sollte die Reliabilität sicher nicht unter 0.8 liegen, anzustreben ist ein Wert von 0.9¹.

Reliabilität von 0.8 als untere Limite

Bei der Auseinandersetzung mit dem Prüfungsziel und -inhalt wird sich kaum eine einzelne zu prüfende Kompetenz herauschälen. Wahrscheinlich erlaubt nur eine Methodenkombination, die relevanten Kompetenzen befriedigend zu erfassen. Viele Tätigkeiten setzen voraus, dass Fachwissen verfügbar ist. Selbst für praktische Handlungen muss das Wissen vorhanden sein, wie diese korrekt auszuführen sind. Es ist deshalb nicht verwunderlich, dass Leistungen in Wissenstests gut mit solchen in praktischen Prüfungen übereinstimmen⁵. Allerdings sind auch Fertigkeiten erforderlich, um das Wissen in die Tat umsetzen zu können. Dies gilt für klinische Gesprächsführung und diagnostische Wahrnehmungsleistung ebenso wie für manuelle Handlungen. Vielleicht wollen Sie auch prüfen, ob Ihre Kandidaten komplexe Problemsituationen bewältigen können. Dies erfordert neben Wissensgrundlagen und spezifischen Fertigkeiten Problemlösungs- und Managementfähigkeiten.

Meist genügt eine Prüfungsmethode allein nicht.

Mitberücksichtigen sollten Sie, dass die Methodenwahl als starkes Signal wirkt. Sie signalisiert, welche Kompetenzbereiche leitende Gremien als wichtig erachten und beeinflusst so das Lernverhalten der Kandidaten^{6,7}.

die Prüfung als Signal

- von der idealen ...* Die Wahl der Methoden und deren Umsetzung in konkrete Prüfungen muss ein Prozess sein. Nur eine praktikable und akzeptierte Prüfung hat eine Überlebenschance. Es ist sinnvoll, zuerst eine unter den Gesichtspunkten der Gültigkeit, Messzuverlässigkeit und Signalwirkung ideale Prüfung zu entwerfen und damit die Zielrichtung ins Auge zu fassen. Mit grosser Wahrscheinlichkeit wird ein mehrtägiges Prüfungsverfahren resultieren.
- ...über die realistische...* Erst jetzt sollte aufgrund der Rahmenbedingungen begonnen werden, dieses auf das aktuell Machbare zurückzustützen. Am hartnäckigsten verteidigt werden muss dabei der Aspekt der Validität, resp. der Nähe zum Alltag. Als Anhaltspunkte für eine spätere Verbesserung sollten die Abstriche identifiziert und dokumentiert werden.
- ...zur optimalen Prüfung* Wenn nach einigen Prüfungen die Abläufe eingespielt sind und gewisse Arbeiten leichter von der Hand gehen, sollten die gesammelten Erfahrungen mit den ursprünglichen Idealvorstellungen der Prüfung konfrontiert und gezielte Optimierungen in Angriff genommen werden.
- Wo schränken mich die Rahmenbedingungen ein?* Eine erste Einschränkung stellen oft bereits die Mängel der einzelnen Prüfungsmethoden dar, wie wir sie unter Kapitel 2.2 beschreiben werden. Wichtigere Einschränkungen ergeben sich aber in der Regel aus Rahmenbedingungen der Fachgesellschaft:
- Personelle Bedingungen: Wer kann mithelfen, die Prüfungen zu entwickeln, durchzuführen und auszuwerten? Sind genügend (Standardisierte) Patienten verfügbar?
 - Strukturelle Bedingungen: Sind die räumlichen und materiellen Voraussetzungen der Methode zu erfüllen?
 - Finanzielle Bedingungen: Was darf das Ganze maximal kosten?
 - Akzeptanzbedingungen: Welche Prüfungsdauer ist den Kandidaten und/oder den Prüfern maximal zumutbar? Welche Methoden werden akzeptiert?
- mögliche Vorteile einer Zusammenarbeit nutzen* Gewisse einschränkende Bedingungen können eventuell durch Zusammenarbeit mit anderen Fachgesellschaften überwunden oder zumindest entschärft werden. Vielleicht gibt es überschneidende inhaltliche Bereiche, in denen ein gemeinsamer Fragen-/Fall-/Problem-Pool aufgebaut werden kann. Wenn zwei bis drei kleinere Fachgesellschaften den gleichen Termin und Ort für ihre schriftlichen Prüfungen ansetzen, können sie Kosten bei der Raummiete und der Prüfungsaufsicht sparen usw.

Ein wesentlicher Punkt ist die pro Prüfung zu erwartende Kandidatenzahl. Sie wirkt sich vor allem dort aus, wo Prüfungen individuell durchgeführt, und freie sprachliche Produktionen „von Hand“ ausgewertet werden müssen. Bei grossen Kandidatengruppen wird man deshalb versuchen, Prüfungen einzusetzen, die sich im Gruppenverband durchführen lassen und die hoch standardisiert, am besten sogar maschinell ausgewertet werden können. Bei einer tiefen Kandidatenzahl kann - zumindest unter ökonomischen Gesichtspunkten - sogar eine Prüfung von Faktenwissen im Rahmen eines strukturierten mündlichen Examens erwägt werden.

*die Bedeutung der
Kandidatenzahl*

Für einige Examinatoren mag es verlockend erscheinen, eigene Ideen in die Prüfungsformen einzubringen. Im Laufe der Zeit sind basierend auf einigen wenigen Grundprinzipien zwar unzählige Varianten kreiert, aber nur wenige seriös evaluiert worden. Wir empfehlen dringend, nur gut erprobte Methoden auszuwählen und darauf zu verzichten, diese nach eigenen Ideen abzuwandeln.

*Wählen Sie keine
Exklusivitäten und
Experimente, ...*

Angesichts der grossen Bedeutung der Facharztprüfung aus Sicht der Kandidaten wie auch aus gesellschaftlicher Sicht ist eine hohe Prüfungsqualität erforderlich. Es ist nachgewiesen, dass eine solche nur mit strukturierten Verfahren zu erreichen ist, weshalb wir hier nur solche vorstellen.

*...keine unstrukturierten
Verfahren ...*

Verzichtet haben wir auf die Vorstellung der sogenannten Patient Management Problems (PMPs), die vor allem in den 70er und früheren 80er Jahren intensiv erprobt wurden. Bei dieser Methode bearbeiten die Kandidaten anhand einer Serie von MC-Fragen sukzessive einen Patientenfall. Man hoffte, damit die Problemlösefähigkeit evaluieren zu können. Etliche Untersuchungen führten aber zu ziemlich ernüchternden Ergebnissen. Bei wesentlich grösserem Entwicklungsaufwand als beim gewöhnlichen MC-Verfahren resultierten meist schwächere Reliabilitäten, ohne dass der Nachweis erbracht werden konnte, dass tatsächlich etwas anderes gemessen wurde. Deshalb liess man die Methode weitgehend wieder fallen⁸. Der in Kanada entwickelte Ansatz der Key Feature Prüfung (vgl. Kapitel 4) ist offensichtlich besser geeignet, Fertigkeiten der klinischen Entscheidungsfindung zu prüfen^{3,4}.

...und keine PMPs.

Bewusst stellen wir keine computerbasierten Prüfungsmethoden vor. Das Verhältnis Entwicklungsaufwand/Nutzen stimmt hier unserer Ansicht nach noch nicht. Dies könnte sich allerdings bereits in den nächsten Jahren ändern, da vor allem in den USA intensiv daran gearbeitet wird.

*computerunterstützte
Methoden auf
„stand by“*

2.2 Was leisten die Methoden?

Im Folgenden werden die heute in der ärztlichen Aus- und Weiterbildung international am häufigsten eingesetzten strukturierten Prüfungsmethoden mit ihren Vor- und Nachteilen vorgestellt.

Wie im ersten Teil dieses Kapitels erläutert, hat die Prüfungsgültigkeit (Validität) oberste Priorität. Entsprechend müssen die zu prüfenden Kompetenzen den Ausgangspunkt für die Methodenwahl bilden. Es sei nochmals betont: Die Anwendung einer Methode gewährleistet nicht, dass z.B. die Fähigkeit geprüft wird, Probleme zu lösen. Sie kann eine solche Prüfung höchstens erleichtern, entscheidend sind aber immer die konkret gestellten Aufgaben und Fragen.

Ein wichtiger Punkt ist ferner die Messqualität. Die Aussagekraft einer inhaltlich gültigen Prüfung ist in Frage gestellt, wenn die Resultate nicht hinreichend objektiv und zuverlässig sind. Wiederum gilt, dass die Wahl einer Methode nicht grundsätzlich eine gute Messqualität garantiert oder ausschliesst. Allerdings ermöglichen es einige Methoden, eine gute Messqualität mit weniger Aufwand zu erreichen als andere. Abb. 2.1 ist unter diesem Gesichtspunkt zu lesen.

MC Wahlantwortfragen (MC) KAF Prüfung mit Kurzantwortfragen
 SMP Strukturierte mündliche Prüfung SB Strukturierte Beobachtung
 OSCE Objektives strukturiertes klinisches Examen

Inhaltliche Validität	MC	KAF	SMP	SB	OSCE
Verfügbarkeit (partialisierter) Fachkenntnisse	•••	••	•		•
Interpretieren und Integrieren klinischer Information	••	••	••	•	•
Erklären, Begründen	•	••	••		•
mündlich-sprachlicher Ausdruck			•••	••	••
praktische und interaktive Fertigkeiten				••	•••
Strukturierung und Management von Problemen und Fällen		•	•	•	•
Messqualität					
Objektivität	•••	••	•	•	••
Reliabilität	••	•	•	•	•

Abb. 2.1 Prüfungsmethoden erfassen Kompetenzaspekte unterschiedlich gut und führen zu mehr oder weniger objektiven und reliablen Prüfungen

Für die Methodenwahl dürften auch Prüfungsdauer und -aufwand von Bedeutung sein. Abb. 2.2 gibt eine Übersicht, welche minimale Dauer erforderlich ist, um eine gerade noch ausreichende Messzuverlässigkeit (0.8) zu erzielen und wie die Verhältnisse bezüglich Aufwand für Entwicklung, Durchführung und Auswertung aussehen.

	MC	KAF	SMP	SB	OSCE
Minimal erforderliche Prüfungsdauer (in Stunden), um eine Reliabilität von mindestens 0.8 zu erreichen	3-4	4-5	4-	4-	4-6
Aufwand für die					
Entwicklung	•••	•• (•)	•• (•)	••	••••
Durchführung	•	•	•••	•• (•)	•••
Auswertung	•	••	•	••	•

Abb. 2.2 Minimale Prüfungsdauer und Aufwandverhältnisse für Entwicklung, Durchführung und Auswertung

Nachfolgend werden die fünf Methoden einzeln vorgestellt.

Wahlantwortverfahren (MC)

Unter dem Begriff Wahlantwortfragen (Multiple Choice, MC) werden verschiedene Typen standardisierter schriftlicher Aufgaben (Items) zusammengefasst. Die Kandidaten müssen entweder aus mehreren vorgegebenen Antworten eine oder mehrere auswählen, Aussagen einander zuordnen oder beurteilen, ob Aussagen oder kausale Aussagenverknüpfungen richtig sind oder nicht. Die Antwort beschränkt sich immer auf ein Ankreuzen vorgegebener Wahlmöglichkeiten. In einer MC-Prüfung können verschiedene Itemtypen nebeneinander verwendet werden.

Wie funktioniert die Methode?

Eine MC-Prüfung evaluiert primär Umfang und Verfügbarkeit von Fachkenntnissen, in begrenztem Mass zudem die Fähigkeit, mit Hilfe der Kenntnisse klinische Situationen zu beurteilen und Probleme zu lösen. Mehrere Studien zeigen, dass die aus einer gut konstruierten MC-Prüfung resultierende Rangierung der Kandidaten hinsichtlich der Wissensleistung für mindestens 7.5 Jahre stabil bleibt^{9,10,11}. Darüber hinaus ist sie ein respektable Gradmesser für andere Aspekte der klinischen Kompetenz, was auf die grundsätzliche Bedeutung des Wissens hinweist⁹.

Welche Kompetenzen vermag sie zu beurteilen?

Welches sind die Vorteile?

- Der hohe Standardisierungsgrad minimiert subjektive Einflüsse; die Auswertung kann zudem maschinell erfolgen. Die MC-Methode ist daher eine objektive, transparente, faire und vor allem bei grossen Kandidatenzahlen ökonomische Prüfungsform.
- Die hohe mögliche Fragenzahl pro Prüfungsstunde erlaubt eine repräsentative Inhaltsstichprobe und reduziert die störenden Einflüsse von Zufällen (z.B. durch Raten) sowie von Problem- und Fallspezifitäten. Es können zuverlässige Ergebnisse und - im Rahmen des erfassbaren Kompetenzbereichs - inhaltlich gültige Prüfungen erzielt werden.
- Die automatisierte statistische Fragen- und Prüfungsanalyse, wie sie am IAWF zur Auswertung der eidg. Medizinalprüfungen entwickelt worden ist, ermöglicht eine sukzessive Optimierung des Fragenpools.

Wo liegen die Problem- resp. Kritikpunkte?

- Der Evaluationsbereich ist eingeschränkt.
- Die Methode verleitet zum simplen Abfragen von Faktenkenntnissen.
- Das Formulieren formal und sprachlich korrekter MC-Items (z.B. ohne ungewollte Lösungshinweise) muss erlernt werden und ist zeitaufwendig. Besonders anspruchsvoll ist die Kreation von MC-Items, die höhere kognitive Leistungen prüfen.
- Die tatsächliche Kenntnisleistung wird wegen der Lösungsunterstützung durch die vorgegebenen Antworten (sog. cueing-Effekt) überschätzt¹².
- Bei kleinen Kandidatenzahlen können die statistischen Informationen über die Items wegen des Zufallsfehlers nur bedingt genutzt werden. Das Verhältnis Entwicklungsaufwand/Nutzen wird fragwürdig, sofern nicht auf bereits bestehende MC-Fragen zurückgegriffen werden kann.

Mit welchem Aufwand ist zu rechnen?

Um eine gerade noch akzeptable Reliabilität von 0.8 zu erreichen, sind mindestens 120-150 Items erforderlich. Da in der Anfangsphase erfahrungsgemäss rund die Hälfte der Fragen den Revisionsprozess nicht überstehen, als ungenügend angesehen und zurückgewiesen werden, müssen doppelt so viele Rohfragen entwickelt werden, wie die Prüfung enthalten soll.

Wenn die Items klinische Informationen enthalten, welche interpretiert und integriert werden müssen, ist für die Kandidaten eine minimale Bearbeitungszeit von 1^{1/2} Minuten vorzusehen. Eine entsprechende Prüfung dauert 3-4 Stunden.

Die gesamte Prüfungsentwicklung, -durchführung und -auswertung erfordert von den medizinischen Fachexperten einen Aufwand von 300-400 Stunden, von den Prüfungssachverständigen und dem Sekretariat je einen solchen von rund 100 Stunden. Die Anmeldungsadministration ist dabei nicht inbegriffen, für 200 Kandidaten sind 80-100 Stunden zu veranschlagen.

Kurzantwortfragen (KAF)

Gut instruierte Fragenautoren schreiben anhand des möglichst detailliert vorgegebenen Prüfungsinhalts (vgl. Kapitel 1) Fragen. Sie können auch einen Fall beschreiben, zu dem sie mehrere Fragen stellen. Die Fragen rufen nur kurze Antworten (Einzelworte oder kurze Sätze) hervor. Expertengruppen revidieren die Fragen, erstellen vor der Prüfung im Konsens einen Schlüssel richtiger Antworten und korrigieren schliesslich standardisiert nach dem Antwortschlüssel. Die computergestützte Prüfungsauswertung erlaubt eine norm- oder inhaltsorientierte Standardsetzung und eine Rückmeldung an Kandidaten, Prüfer sowie Aus- resp. Weiterbildner. Ein konkretes Beispiel finden Sie in Kapitel 4.

Wie funktioniert die Methode?

Diese schriftliche Prüfungsform wird bei Medizinalprüfungen heute v.a. in angelsächsischen Ländern verwendet und zwar in Kombination mit oder als Alternative zu MC-Fragen. Wir fassen unter dem Begriff „Kurzantwortfragen“ folgende Varianten und Bezeichnungen zusammen, die andernorts gebräuchlich sind: Short-Essay-Questions (SEQ), Short-Answer-Questions (SAQ), Modified-Essay-Questions (MEQ)¹³ und Short-Answer-Management-Problems (SAMPS)¹⁴.

Mit der offenen Antwortform lässt sich beurteilen, wie Kandidaten ihr Wissen zu einem Fall aktiv formulieren können oder wie sie ein Problem lösen. Die Methode kann nicht nur prüfen, ob Faktenwissen vorhanden ist, sondern auch, wie es verstanden, interpretiert, gewichtet und angewendet wird.

Welche Kompetenzen vermag sie zu beurteilen?

■ Facharztkandidaten, aber auch Examinatoren ziehen diese Prüfungsform der MC-Methode oftmals vor, weil sie vordergründig näher bei den Problemen ärztlichen Handelns liegt.

Welches sind die Vorteile?

- Es können Fälle in ihrem ganzen Ablauf dargestellt werden mit wiederholten Fragen wie „Was tun Sie jetzt?“ oder „Welche weiteren Massnahmen ziehen Sie jetzt in Betracht?“.
- In der Regel ist es anspruchsvoller, eine Kurzantwortfrage zu beantworten, als eine inhaltlich äquivalente MC-Frage. Schon beim Prüfen von blossen Faktenwissen muss die Antwort aktiv formuliert statt nur passiv wiedererkannt werden.
- Problemlösungsfähigkeit lässt sich mit Kurzantwortfragen besser prüfen als mit der MC-Methode.

Wo liegen die Problem- resp. Kritikpunkte?

- Um eine akzeptable Gültigkeit und Zuverlässigkeit der Prüfung zu erreichen, muss mit grosser Sorgfalt gearbeitet werden¹⁵. Der Entwicklungsaufwand ist damit kaum geringer als bei einer MC-Prüfung.
- Es ist schwierig, tatsächlich Problemlösungsfragen zu entwickeln und nicht auf blossen Wissensfragen „zurückzufallen“¹⁶.
- Es ist nicht einfach, den Ablauf eines Falles darzustellen, ohne dass eine Frage einen Lösungshinweis (Cue) zu einer anderen Frage gibt.
- Es ist schwierig, die Bewertung über verschiedene Prüfungen hinweg konstant zu halten.

Mit welchem Aufwand ist zu rechnen?

Zum Erreichen einer gerade noch akzeptablen Reliabilität von 0.8 sind mindestens 30 Prüfungsfälle erforderlich (oder ca. 40 Schlüsselprobleme mit insgesamt 60 - 90 Fragen, vgl. Kapitel 1). Dies erfordert mindestens vier Stunden Prüfungszeit. Bei 100 - 200 Kandidaten müssen Fachexperten und Prüfungssachverständige etwa gleich viel Arbeit leisten wie bei einer MC-Prüfung (vgl. S.50f). Bei grösserer Kandidatenzahl steigt der Aufwand wegen der zusätzlichen Auswertungsarbeit weiter an, während er bei einer MC-Prüfung konstant bleibt.

Strukturierte mündliche Prüfung (SMP)

Wie funktioniert die Methode?

Strukturierte mündliche Prüfungen (engl. Structured Oral Examinations, SOE) kennen verschiedene organisierende Massnahmen vor, während und nach dem Examen. Beispiele sind Examinatorenttraining, Verwendung von Protokollformularen und Checklisten sowie auswertungstechnische

Massnahmen. Dadurch lässt sich die Qualität gegenüber unstrukturierten Prüfungen verbessern. Diese haben sich in vielen Untersuchungen als unobjektiv, unreliabel und damit von fraglicher Validität erwiesen^{17,18}.

Strukturierte mündliche Prüfungen können einen rein theoretischen Inhalt haben oder sich auf eine vorher erfolgte Untersuchung eines Patienten beziehen. Sie finden in Form einer Befragung, einer Diskussion, einer Fallvorstellung oder einer Mischung dieser Formen statt. Der Rahmen wird immer vor der Prüfung festgelegt.

Mündliche Prüfungen werden primär eingesetzt, um zu evaluieren, wie Kandidaten ihr Wissen anwenden (Verständnis, Beurteilung, Problemlösung, Therapieplanung). Wesentlich miterfasst wird die sprachliche Ausdrucks- und die Kommunikationsfähigkeit.

- Der Evaluationsbereich ist verglichen mit schriftlichen Prüfungen erweitert.
- Die Methode erlaubt Flexibilität, ein rascher Themenwechsel ist ebenso möglich wie die Vertiefung einer Diskussion. Unter dem Standardisierungsaspekt gilt dies allerdings als problematisch.
- Eine SMP vorzubereiten und durchzuführen ist anspruchsvoll und aufwendig. Es sind über einen langen Zeitraum teure Experten einzusetzen. Die Examinatoren müssen sorgfältig trainiert werden, um die Einflüsse individueller Prüfungsstile und Beurteilungstendenzen zu reduzieren. Ebenfalls aufwendig ist die Festlegung von Evaluationskriterien und deren Bewertung. Zur blossen Prüfung vorhandener Fachkenntnisse ist eine SMP sowohl aus ökonomischen Gründen wie auch wegen der ungenügenden Repräsentativität nicht angebracht.
- Strukturierte mündliche Examen bergen viele schwer kontrollierbare Störfaktoren. Beispiele sind etwa die Fall- resp. Problemspezifität, die Auswirkungen persönlicher Gefühle (Prüfungsangst, Sympathie/Antipathie) oder der Prüfungszeitpunkt.
- Die Ergebnisse von SMPs korrelieren oft nur schwach mit jenen von schriftlichen Examen. Es ist nicht geklärt, wieweit dies darauf zurückzuführen ist, dass tatsächlich etwas anderes gemessen wird, und wieweit die mangelnde Reliabilität der SMPs dafür verantwortlich ist.

Welche Kompetenzen vermag sie zu beurteilen?

Welches sind die Vorteile?

Wo liegen die Problem- resp. Kritikpunkte?

Mit welchem Aufwand ist zu rechnen?

Zum Erreichen einer gerade noch akzeptablen Reliabilität von 0.8 sind mindestens acht Patientenfälle in vier Stunden zu bearbeiten. Dabei sollten die Kandidaten bei jedem Fall von jeweils einem anderen Examinator geprüft werden. Die Reliabilität kann gesteigert werden, wenn zwei Experten pro Fall und Kandidat prüfen. Falls ein Experte einen Kandidaten bei allen Fällen prüft, ergibt sich eine schlechte Prüfungszuverlässigkeit, die durch eine erhöhte Anzahl Fälle nicht wettgemacht werden kann^{1,19}. Voraussetzung für eine hohe Prüfungszuverlässigkeit ist eine Strukturierung und Standardisierung, dies kann z.B. erreicht werden, indem Checklisten verwendet werden²⁰. Der Entwicklungs- und Auswertungsaufwand ist abhängig vom Grad der Strukturierung und der Standardisierung (Fälle, Probleme, Fragen, Checklisten, Verrechnungsblätter, Examinatorentraining) und deshalb nicht allgemein beantwortbar. Der Aufwand, die Prüfung durchzuführen, lässt sich nach der Formel „2 Examinatoren x Prüfungsdauer x Kandidatenzahl“ berechnen.

Strukturierte Beobachtung (SB)

Wie funktioniert die Methode?

Kaum eine Facharztprüfung wird darauf verzichten können, die Fertigkeiten eines Kandidaten zuverlässig zu beurteilen. Das geht nicht ohne direkte Beobachtung. Strukturiert heisst hier v.a., dass entsprechend einem Fertigkeitenblueprint und einer repräsentativen Stichprobe von standardisierten Berufssituationen (Fallblueprint) beobachtet wird. Ferner müssen trainierte Prüfer ein einheitliches Protokoll benutzen.

Die drei Elemente „Prüfer“, „Interaktion Prüfer-Kandidat“ und „Problem“ müssen möglichst konstant sein:

1. Prüfer müssen trainiert werden, damit sie möglichst einheitlich beobachten und protokollieren. Diese Übereinstimmung in der Beurteilung ist in jedem Examen kontinuierlich zu überprüfen (Interraterreliabilität).
2. Die Interaktion Prüfer-Kandidat ist bei der direkten Beobachtung vergleichsweise klein, da keine verbale Interaktion stattfindet. Um auch nonverbale Interaktionen möglichst auszuschalten, kann die Beobachtung mit Video aufgezeichnet und ausgewertet werden.
3. Für die Standardisierung des Problems muss nach Möglichkeit jeder Kandidat in der gleichen Problemsituation beobachtet werden²¹. Vor allem bei komplexeren oder spezifischen Problemsituationen und bei

grösseren Kandidatenzahlen ist das fast nur zu erreichen, wenn Hilfsmittel wie z.B. Standardisierte (simulierte oder echte) Patienten oder Phantome verwendet werden.

Diese Prüfungsmethode wird z.B. bei der Beurteilung von Konsultationsfertigkeiten in der Sprechstunde^{22,23} oder bei der Beobachtung von Reanimationsfertigkeiten²⁴ angewendet.

Beurteilbar sind vor allem verbale, interaktive und praktische Fertigkeiten oder Verhaltensweisen von Kandidaten. In ärztlichen Prüfungen sind das z.B. die Anamneseerhebung, diagnostische oder therapeutische Fertigkeiten, Problemlösungsfähigkeiten oder die Interaktion des Arztes mit einem Patienten.

- Bei guter Strukturierung und Standardisierung sind Fertigkeiten zuverlässig beurteilbar.
- Die Fertigkeitsevaluation kann mit einer strukturierten mündlichen Prüfung kombiniert oder in ein OSCE integriert werden.
- Diese Art, einen Arzt im realen oder simulierten Berufsfeld zu beurteilen, wird von den Kandidaten meist gut akzeptiert²⁵.
- Gewisse Kandidaten ändern ihr übliches Verhalten deutlich, wenn sie wissen, dass sie beobachtet werden. Dies beeinträchtigt die Gültigkeit der Beobachtungsergebnisse.
- Es gibt eine Reihe oft schwer zu kontrollierender Faktoren, welche die Beobachtungszuverlässigkeit beeinträchtigen können, etwa wenn Prüfer untereinander oder Prüfer und Kandidaten voneinander nicht unabhängig sind, oder wenn erwartete Interaktionen zwischen Beobachtern, Kandidaten und Problem auftauchen.
- Ärzte und Patienten akzeptieren Videobeobachtungen nicht immer.

Welche Kompetenzen vermag sie zu beurteilen?

Welches sind die Vorteile?

Wo liegen die Problem- resp. Kritikpunkte?

Die Methode ist aufwendig. Besonders das Training vieler Beobachter, die nötige Beobachtungszeit und die Problemstandardisierung schlagen zu Buche. Um die erforderliche Gültigkeit und Zuverlässigkeit zu erreichen, muss jeder Kandidat mindestens während drei Stunden in 10-20 Problemsituationen beobachtet werden.

Mit welchem Aufwand ist zu rechnen?

Objektives strukturiertes klinisches Examen (OSCE)

Wie funktioniert die Methode?

OSCE (sprich: „oski“) ist keine neue Prüfungsmethode. Es ist vielmehr ein flexibler Prüfungsrahmen, in dem verschiedene Methoden Platz haben. Gemäss einem voraus festgelegten Blueprint werden verschiedene Stationen entwickelt. Diese prüfen je eine bis zwei Kompetenzkomponenten. Dabei geht es vorwiegend um praktische klinische Kompetenzen^{26,27}.

Man unterscheidet Prozedurstationen und Fragestationen. An Prozedurstationen müssen Kandidaten eine praktische Aufgabe lösen. Examinatoren beobachten und bewerten Leistungen mit Hilfe im Voraus entwickelter Checklisten und/oder globaler Beurteilungsskalen. An Fragestationen sind schriftliche Fragen zu beantworten, die häufig im Zusammenhang mit der vorhergehenden Prozedurstation stehen. Die Fragen können sich auf Bild- und Datenmaterial (Röntgenbilder, EKG, Laborergebnisse etc.) beziehen.

Die Stationen werden in Form eines Circuits angelegt, den die Kandidaten durchlaufen. Nach der Prüfung werden sowohl die ausgefüllten Checklisten von den Prozedurstationen wie die Antwortblätter von den Fragestationen nach einem im Voraus festgelegten Schema ausgewertet.

An den Prozedurstationen kommen die Kandidaten mit Patienten in Kontakt, um eine Anamnese zu erheben, eine körperliche Untersuchung durchzuführen oder für ein Gespräch. Dabei ist es oft schwierig, wirkliche, dem vorgesehenen Fall einer Station entsprechende Patienten zu finden. Der Kontakt mit mehreren Kandidaten ist älteren und geschwächten Menschen oft nicht zumutbar. Zudem kann kaum gewährleistet werden, dass sich ein echter Patient gegenüber allen Kandidaten gleich verhält. Aus diesen Gründen arbeitet man meist mit sogenannten Standardisierten Patienten. Es handelt sich dabei teilweise um Menschen mit einer wirklichen Krankheit, häufiger aber um vollkommen gesunde Menschen. Diese wurden geschult und sind fähig, eine ihnen zugeteilte Patientenrolle zu spielen. Damit wird gewährleistet, dass ein „Patient“ während einer Unterhaltung stets gleich antwortet und während einer körperlichen Untersuchung immer das Gleiche mimit^{28,29,30}.

Welche Kompetenzen vermag sie zu beurteilen?

OSCE ist besonders geeignet, um praktische Komponenten der klinischen Kompetenz zu prüfen: Gewandtheit im zwischenmenschlichen Umgang (Kommunikation, Anamnese), Beherrschung technischer Fertigkeiten (körperliche Untersuchungen, Punktionen etc.) sowie Wahrnehmungs- und Interpretationsleistungen (EKG, Röntgenbilder, Labortests). Zusätzlich kann auch Wissen geprüft werden, was allerdings unökonomisch ist.

- OSCE ist durch die Möglichkeit, Fertigkeiten und Handlungskompetenzen zu prüfen, eine wichtige Ergänzung zu einer Wissensprüfung und gibt auch ein entsprechendes Signal an die Lernenden.
- OSCE erlaubt, nicht nur Handlungsergebnisse zu beurteilen, sondern auch den Prozess und die angewandte Technik.
- Die Objektivität wird einerseits durch die für alle Kandidaten identischen Aufgaben und Prüfungsbedingungen gewährleistet und andererseits durch die Verwendung von Checklisten und Beurteilungsskalen.
- Bei hinreichender Länge kann eine gute Reliabilität erzielt werden.
- Die Entwicklung und Durchführung ist sehr zeit- und personalaufwendig und entsprechend teuer.
- Die Durchführung ist störanfällig. Eine kleine Panne an einer einzelnen Station kann den ganzen Circuit zum Erliegen bringen.
- Vor allem bei OSCEs mit vielen kurzen (5 Minuten-) Stationen ist die Durchführung für Examinatoren und Kandidaten sehr stressgeladen und kann z.B. bei Kandidaten zu unnatürlichem Verhalten führen. Es sollten hier unbedingt Pausenstationen eingeplant werden.

Welches sind die Vorteile?

Wo liegen die Problem- resp. Kritikpunkte?

Zum Erreichen einer gerade noch akzeptablen Reliabilität von 0.8 müssen wegen der Problem- resp. Fallspezifität mindestens 20 Stationen integriert werden, wobei die gesamte Prüfungsdauer mindestens vier Stunden betragen muss³¹. Bei der Entwicklung einer OSCE-Prüfung mit zwei standardisierten Patienten pro Station und 120 Kandidaten muss mit Totalkosten bis Fr. 90'000.- gerechnet werden. Das entspricht Fr. 750.- pro Kandidat^{32,33}. 80 Prozent davon sind Personalkosten. Einsparungen sind möglich, wenn weniger standardisierte Patienten eingesetzt werden und wenn institutionell angestellte Mitarbeiter z.T. kostenlos arbeiten. Eine andere Sparmöglichkeit besteht in einem stufenweisen Vorgehen. Anhand der Leistung in einem ersten Circuit mit der Hälfte aller Stationen werden drei Gruppen von Kandidaten ermittelt: a) solche, die sicher erfolgreich sind, b) solche, die bereits durchgefallen sind, c) unklare Mittelgruppe. Nur diese Mittelgruppe hat auch die restlichen Stationen zu absolvieren.

Mit welchem Aufwand ist zu rechnen?

René Krebs

Literatur

- ¹ van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67
- ² Case SM, Swanson DB, Becker DF. Verbosity, window dressing, and red herrings: do they make a better test item? *Acad Med* 1996;71:528-30
- ³ Page G, Bordage G. The Medical Council of Canada's Key Features Project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;70:104-10
- ⁴ Page G, Bordage G. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;70:194-201
- ⁵ van der Vleuten CPM, Van Luyk SJ, Beckers HJM. A written test as an alternative to performance testing. *Med Educ* 1989;23:97-107
- ⁶ Newble DI, Jaeger K. The effect of assessment and examination on the learning of medical students. *Med Educ* 1983;17:165-71
- ⁷ Frederickson N. The real test bias: influences of testing on teaching and learning. *Am Psychol* 1984;3:193-202
- ⁸ Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice questions and patient management problem item formats in assessment of clinical competence. *Med Educ* 1985;19:238-47
- ⁹ Ramsey PG, Carline JD, Inui TS et al. Predictive validity of certification by the American Board of Internal Medicine. *Ann Intern Med* 1989; 110:719-26
- ¹⁰ Solomon DJ, Reinhart MA, Bridgham RG, Munger BS, Starnaman S. An assessment of an oral examination format of evaluating clinical competence in emergency medicine. *Acad Med* 1990;65:S43-4
- ¹¹ Case SM, Swanson DB. Validity of NBME part I and part II scores for selection of residents in orthopaedic surgery, dermatology, and preventive medicine. *Acad Med* 1993;68:51-6
- ¹² Newble DI, Baxter A, Elmslie RG. A comparison of multiple-choice tests and free-response tests in examination of clinical competence. *Med Educ* 1979;13:263-8
- ¹³ Knox JD. What is ... a modified essay question? *Medical Teacher* 1989; 11:51-7
- ¹⁴ Handfield-Jones R, Belle Brown J, Biehn J, Rainsberry P, Brailovsky CA. Certification Examination of the College of Family Physicians of Canada, Part 3: Short-answer management problems. *Can Fam Physician* 1996;42:1353-61
- ¹⁵ Stalenhoef BF, Van Der Vleuten CPM, Jaspers TAM, Fiolet JFBM. The feasibility, acceptability and reliability of open-ended questions. in: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwiestra RP, eds. *Teaching*

- and Assessing Clinical Competence. Groningen: Boekwerk Publ., 1989
- ¹⁶ Feletti GI, Smith EK. Modified essay questions: are they worth the effort? *Med Educ* 1986;20:126-32
- ¹⁷ Muzzin LJ, Hart L. Oral examinations. In: Neufeld VR, Norman GR, eds. *Assessing clinical competence*. New York: Springer, 1985;71-93
- ¹⁸ Yang JC, Laube DW. Improvement of reliability on an oral examination by a structured evaluation instrument. *J Med Educ* 1983;58:864-72
- ¹⁹ Swanson DB. A measurement framework for performance-based tests. In: Hart IR, Harden RM, eds. *Further developments in assessing clinical competence*. Montreal: Can-Heal, 1987:13-45
- ²⁰ van der Vleuten CPM, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: Issues of reliability. *Med Educ* 1991;25:110-8
- ²¹ van der Vleuten C, Newble D, Case S et al. Methods of assessment in certification. In: Newble D, Jolly B, Wakeford R eds. *The Certification and Recertification of Doctors*. Cambridge University Press, 1994:105-25
- ²² Cox J, Mulholland H. An instrument for assessment of videotapes of general practitioners' performance. *BMJ* 1993;306:1043-6
- ²³ Hays RB, Jones BF, Adkins PB et al. Analysis of videotaped consultations to certify competence. *Med J Aust* 1990;152:609-11
- ²⁴ Jansen JJ, Berden HJ, van der Vleuten C, Grol RP, Rethans J, Verhoeff CP. Evaluation of cardiopulmonary resuscitation skills of general practitioners using different scoring methods. *Resuscitation* 1997;34:35-41
- ²⁵ Wakefield J. Direct Observation. In: Neufeld VR, Norman GR, eds. *Assessing Clinical Competence*. Springer Publishing Company; New York, NY, 1985:51-70
- ²⁶ Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13:41-54
- ²⁷ Heard JK, Allen RM, Cason GJ, Cantrell M, Tank PW. Practical issues in developing a program for the objective assessment of clinical skills. *Medical Teacher* 1998;20:15-21
- ²⁸ Barrows HS. *Simulated Patients (Programmed Patients) The development and use of a new technique in medical education*. Springfield, Ill: CC Thomas Publ. 1971
- ²⁹ Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med* 1993;68:443-51
- ³⁰ Swanson DB, Stillmann PL. Use of standardized patients for teaching and assessing clinical skills. *Evaluation & The Health Professions* 1990;13:79-103
- ³¹ van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58-76

- ³² Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real costs of an objective structured clinical examination. *Acad Med* 1993;86:513-17
- ³³ Carpenter JL. Cost analysis of objective structured clinical examinations. *Acad Med* 1995;70:828-33

3

Wie wird
eine Prüfung
mit MC-Fragen
entwickelt?

INHALT

- 3.1 Welche Anforderungen sollen Fragen erfüllen? - 65
- 3.2 Worauf zielt das Fragenthema? - 67
- 3.3 Wie unterscheiden sich die Fragentypen? - 71
- 3.4 Wie werden Fragen formuliert? - 80
- 3.5 Wie werden Fragen überprüft? - 89
- 3.6 Wie wird die Prüfung zusammengestellt? - 91

Wenn Sie bei der Herstellung von MC-Items nach dieser Anleitung vorgehen, sollten die Items in der Regel mehr prüfen als isoliertes Faktenwissen und weitgehend frei sein von formalen Fehlern. Aber auch hier gilt: Übung macht den Meister. Es ist deshalb zu empfehlen, dass neue Fragenautoren in einem mindestens eintägigen Workshop das Herstellen und Revidieren von MC-Items gemeinsam erarbeiten. Zudem sollten Autoren kontinuierlich über das Ergebnis der Revision ihrer Fragen informiert werden, damit sie aus Fehlern lernen können. Ferner gilt: Gut Ding will Weile haben. Auch geübte Autoren rechnen durchschnittlich mit einer Stunde Arbeit, um ein MC-Item herzustellen.

Dieses Kapitel beschreibt im 1. Abschnitt, worauf zu achten ist, wenn MC-Items zur Gültigkeit und Zuverlässigkeit einer Prüfung beitragen sollen. Die Abschnitte 2 - 5 bilden den Hauptteil dieses Kapitels und beschäftigen sich mit der konkreten Entwicklung der Items: Dabei geht es zuerst um die Wahl des Itemthemas und des geeigneten Itemtyps. Es folgen allgemeine und itemspezifische Hinweise zur korrekten Formulierung und zur Vermeidung ungewollter Lösungshinweise. Schliesslich wird der Revisionsprozess beleuchtet. Der 6. Abschnitt beschreibt, worauf zu achten ist, wenn Items zu einer Prüfung zusammengestellt werden.

Wir möchten an dieser Stelle betonen, dass die vorliegende Anleitung weder die Erfahrungsbildung überflüssig machen, noch das zeitliche Engagement zum Verfassen guter MC-Items ersparen kann.

In diesem Kapitel erfahren Sie,

- wie relevante Itemthemen bestimmt werden
- wofür sich welcher Itemtyp eignet
- worauf bei der Itemformulierung zu achten ist
- wie neu verfasste Items zu überprüfen sind
- was bei der Prüfungszusammenstellung zu beachten ist

3.1 Welche Anforderungen sollen MC-Fragen erfüllen?

Wir gehen davon aus, dass mit der geplanten MC-Prüfung ein wichtiger Qualitätsaspekt der künftigen Fachärzte kontrolliert werden soll, nämlich die Verfügbarkeit relevanten Fachwissens und die Kompetenz, dieses in konkreten Problemsituationen anzuwenden.

Die MC-Prüfung als Instrument zur Qualitätssicherung

Damit diese Kontrolle in der Prüfung repräsentativ erfolgen kann, muss das relevante Fachwissen möglichst präzise - idealerweise in Form eines Lernzielkataloges - definiert und ein Blueprint - ein gewichtetes Verzeichnis des Prüfungstoffes - erstellt werden (vgl. Kapitel 1).

Vorbedingung: Inhaltsdefinition und -gewichtung

Letztlich steht oder fällt die Validität und Reliabilität der Prüfung aber mit der Qualität der einzelnen Items. Items sind mit grosser Wahrscheinlichkeit für den Einsatz in einer Prüfung geeignet, wenn sie die nachfolgenden Kriterien erfüllen. Im weiteren Verlauf dieser Anleitung erfahren Sie, wie diese Anforderungen erreicht werden können. Nutzen Sie die Kriterienliste auch zur Qualitätsüberprüfung der von Ihnen verfassten Items.

Hauptbedingung: gute Items

Ein Item trägt dann zur Validität der Prüfung bei,

- wenn das gewählte Fragenthema relevant ist im Hinblick auf die praktische fachärztliche Tätigkeit
Spitzfindigkeiten sind ebenso zu vermeiden wie Trivialitäten.
- wenn das Anspruchsniveau stimmt
Eine Prüfung, die primär Verständnis und Anwendungskompetenz von Wissen kontrollieren soll, darf nicht aus Items bestehen, die lediglich die Verfügbarkeit von auswendig gelernten Einzelfakten erfordern.
- wenn es auf einen klar umschriebenen Inhalt resp. ein Problem fokussiert ist und in sich ein geschlossenes Ganzes bildet
Alle Wahlantworten sollen in die gleiche inhaltliche Kategorie fallen. Abwägen zu müssen, ob eine vorgeschlagene Krankheitsursache wahrscheinlicher richtig ist als eine vorgeschlagene Therapie, hat keinen Bezug zur beruflichen Realität eines Facharztes.
- wenn es eine eindeutig beste Lösung gibt
Inhalte mit kontroversen Lehrmeinungen sind für die MC-Methode ungeeignet, es sei denn, es werde ausdrücklich nach einer bestimmten Lehrmeinung gefragt.

So messen Items gültig...

...und so messen Items zuverlässig.

Ein Item trägt zur Reliabilität der Prüfung bei, wenn es im Sinne der Prüfung differenziert, d.h. wenn es Trennschärfe besitzt. Dies ist der Fall,

- wenn es bezüglich Schwierigkeit der Zielgruppe angemessen ist
Insbesondere problematisch sind zu schwere Items, bei denen auch gute Kandidaten aufs Raten angewiesen sind.
- wenn es sprachlich klar formuliert ist
Es soll Wissen geprüft werden und nicht Sprachverständnis oder Interpretationsglück.
- wenn es keine ungewollten Lösungshinweise (sog. Cues) enthält
Es soll Wissen geprüft werden und nicht MC-Prüfungserfahrung und „Testknackerfähigkeiten“.
- wenn es inhaltlich in die Gesamtprüfung passt

Die ersten drei Punkte sind nicht nur wichtig, um eine gute Reliabilität zu erzielen, sie bilden auch die Voraussetzung, damit die Prüfung gültig messen kann. Das trifft auf den vierten Punkt nur noch bedingt zu. Je homogener die Items einer Prüfung inhaltlich sind, desto höher fällt zwar deren Reliabilität aus, desto stärker wird aber ihr Gültigkeitsbereich eingeschränkt. Je breiter der Kompetenzbereich ist, der mit einer Prüfung erfasst werden soll, umso schwieriger wird es, zugleich eine hohe Reliabilität (≥ 0.9) im Sinne der inneren Konsistenz oder Homogenität (Cronbach alpha) zu erzielen.

Neben diesen allgemeinen gibt es noch einige typenspezifische Anforderungen. Diese werden aufgeführt, wenn die Besonderheiten der einzelnen Typen erläutert werden (S.87f). Es sei hier auch schon darauf hingewiesen, dass es bezüglich Messqualität typenabhängige Unterschiede gibt. Darauf gehen wir im Zusammenhang mit der Typenwahl ein (vgl. Kapitel 3.3).

3.2 Worauf zielt das Fragenthema?

Blueprintkapitel und Prüfungsziel geben Themenbereiche an, zu denen Items kreiert werden können. Ein Fragenthema schränkt auf einen engen Teilaspekt ein, z.B. auf die wahrscheinlichste Ursache einer Störung, auf die Differentialdiagnose oder auf die effizienteste Behandlungsmassnahme. Wenn als Frage zu einem Thema nur formuliert werden kann: „Welche der folgenden Aussagen zu Krankheit X trifft zu?“, ist das Thema mit Sicherheit zu breit, zu heterogen für ein MC-Item.

*vom Themenbereich
zum Itemthema*

Innerhalb eines Themenbereichs ergeben sich relevante Items vor allem aus Teilaspekten,

Kriterium Relevanz

- mit denen ein Facharzt besonders häufig konfrontiert ist
- bei denen Fehler gravierende Folgen haben können
- bei denen Fehlmeinungen verbreitet sind

Überprüfen Sie die Relevanz der vorgesehenen Themen mit der Frage: „Wie wichtig ist es, dass ein Facharzt dieses Problem selbständig lösen resp. die sich ergebende Frage richtig beantworten kann?“

Von eigenen Patientenfällen auszugehen oder von einer konkreten konsultativen Anfrage eines Nicht-Facharztes, kann sehr anwendungsbezogene, relevante Items ergeben. Man muss sich aber hüten, „interessante“ Sonderfälle auszuwählen.

*eigene Fälle als
Inspirationsquelle*

Es ist wichtig, Lehrbücher zu verwenden, um die fachliche Richtigkeit einer Frage abzusichern und zu dokumentieren. Lehrbücher können auch hilfreich sein, um gute Falschantworten (Distraktoren) zu finden. Als Inspirationsquelle für Fragenthemen sind sie aber nicht zu empfehlen. Es entstehen daraus gerne akademische Fragen im negativen Sinn. Ein Beispiel: Ausgehend von einer vorgegebenen Diagnose in der Kapitelüberschrift eines Lehrbuches wird nach dem häufigsten Befund gefragt. Diese Information findet sich in einem Absatz des entsprechenden Kapitels. Mit diesem Vorgehen entsteht keine realistische Frage. In der Praxis läuft der Prozess in umgekehrter Richtung und erfordert in der Regel mehr als eine einfache Assoziation „Diagnose - Hauptbefund“. Die Diagnose muss aus diversen Informationen über einen Patienten erschlossen werden, zu denen auch Untersuchungsbefunde gehören.

*Verwendung von
Lehrbüchern*

Überprüfen Sie die vorgesehenen Themen auf ihre Anwendungsbezogenheit mit der Frage: „Stellt sich das Problem/die Frage einem Facharzt in seiner Praxis auf diese Weise?“

*Kriterium
Anwendungsbezug*

Am einfachsten lassen sich anwendungsbezogene Items sicher zu den Handlungsbereichen Diagnostizieren und Therapieren verfassen. (Beim Letzteren taucht allerdings nicht selten das Problem divergierender Lehrmeinungen auf.) Im Interesse der inhaltlichen Repräsentativität sollten aber nicht nur Themen aus diesen beiden Bereichen gewählt werden.

In der folgenden Liste finden Sie zu sieben Bereichen ärztlichen Handelns je vier mögliche Themenbereiche mit einer konkreten Beispielfrage. Es sei nochmals erwähnt, dass die Themenbereiche aufgrund des Blueprints und des Prüfungsziels gewählt werden müssen. Den meisten Fragen würde eine Fall- oder Problemvignette vorangehen.

Epidemiologie

- **Risikofaktoren**

Für welche Krankheit besteht bei dieser Person ein erhöhtes Risiko?

- **Korrelationen/Assoziationen mit anderen Zuständen**

Welche Krankheit liegt bei diesem Zustandsbild am häufigsten auch noch vor?

- **Auftretenshäufigkeit**

In welchem Ausmass tritt das beschriebene Krankheitsbild in der mitteleuropäischen Bevölkerung auf?

- **gesellschaftliche/volkswirtschaftliche Bedeutung**

Welche Kosten verursacht diese Krankheit im Lande Z?

Prävention und Früherkennung

- **Impfungen**

Welcher Impfstoff sollte in dieser Situation verabreicht werden?

- **Screeningtests**

Welcher Screeningtest ist für dieses Projekt am sinnvollsten?

- **Indikationsstellung für Labortests/Untersuchungen**

Welcher zusätzliche anamnestische Befund lässt Untersuchung Y als sinnvoll erscheinen (ausreichende „pretest probability“)?

- **Präventionsmassnahmen**

Welche Massnahme ist am besten geeignet, um in der Altersgruppe des geschilderten Patienten der Entstehung dieses Krankheitsbildes vorzubeugen?

Krankheitsmechanismen

- pathophysiologische Erklärung
Worauf sind diese Befunde am wahrscheinlichsten zurückzuführen?
- wahrscheinlichste Lokalisation
An welcher Körperstelle tritt Befund X bei diesem Krankheitsbild am wahrscheinlichsten auf?
- Erreger/pathogenes Agens
Durch welche Substanz ist dieser klinische Befund am wahrscheinlichsten ausgelöst worden?
- Einfluss auf Laborparameter/physiologische Größen
Welcher Parameter ist in dieser Situation am wahrscheinlichsten erhöht?

Krankheitsverlauf

- übliche Manifestation
Mit welchem klinischen Zeichen tritt die Krankheit X in der Regel in Erscheinung?
- Verlaufsmerkmale
 - *Was tritt im Verlauf dieser Krankheit typischerweise auf?*
- mögliche Komplikationen
 - *Für welche Komplikation besteht bei diesem Patienten ein erhöhtes Risiko?*
- Kriterien für die Prognose
 - *Welcher Befund spricht am ehesten für eine positive Prognose?*

Diagnostik

- Diagnostische Abklärung
Welche weitere Abklärung bringt Sie in dieser Situation diagnostisch am weitesten?
- Stellen der Diagnose
Welches ist die wahrscheinlichste Diagnose?

- **Differentialdiagnose**
Was kommt neben der wahrscheinlichsten Diagnose differentialdiagnostisch am ehesten auch noch in Betracht?
- **Bestätigung oder Ausschluss einer Diagnose**
Welcher zusätzliche Befund stützt diese Diagnose am besten?

Behandlung

- **beste therapeutische Massnahme**
Welches ist die geeignetste Behandlung/medikamentöse Therapie?
- **unerwünschte Effekte der Behandlung**
Auf welche Nebenwirkung sollten Sie gefasst sein, wenn Sie dem geschilderten Patienten Medikament X verschreiben?
- **Diät/physikalische Behandlung**
Welche diätische Massnahme kann das Problem dieses Patienten am ehesten lindern?
- **Wirkungsweise einer Behandlung**
Auf welchem Wirkungsmechanismus beruht die therapeutische Wirkung der Applikation von X?

Strategie (Fällen von Entscheiden)

- **zeitliche Priorität in einer gegebenen Situation**
Welches ist der geeignetste initiale/nächste Schritt in der Betreuung dieses Patienten?
- **generelle Priorität in einer gegebenen Situation (Diagnostik vs. Therapie)**
Welches ist in dieser Situation (z.B. bei Einlieferung in die Notfallstation) die wichtigste Massnahme für den Patienten?
- **Erstellen des Managementplans**
Welche Vorgehensweise/Reihenfolge der Massnahmen ist für die Lösung des Problems dieses Patienten am effektivsten?
- **mögliche Informationsquellen resp. Art ihrer Nutzung als Entscheidungshilfe**
Welches ist die angemessenste Interpretation der Resultate dieser Studie?

3.3 Wie unterscheiden sich die Fragentypen?

Seit Beginn der Entwicklung objektiver schriftlicher Prüfungen gegen Ende des 19. Jahrhunderts sind viele verschiedene Itemtypen kreiert worden. Für die MC-Prüfungen des amerikanischen National Board of Medical Examiners wurde ursprünglich eine Auswahl von über zehn Typen vorgeschlagen². Diese wurden mit Buchstaben etikettiert, welche heute noch (auch in dieser Anleitung) verwendet werden. Zunehmend wurde aber erkannt, dass es unter den Aspekten der Herstellung, der Beantwortung und der Messqualität besser ist, sich auf wenige Typen zu beschränken.

Beschränkung auf wenige Typen

Von der Aufgabenstellung her lassen sich zwei Grundtypen unterscheiden: Best-Antwort-Typen und Richtig/Falsch-Typen. Bei der Gruppe der Best-Antwort-Typen wird die Zahl der auszuwählenden Antworten angegeben; meist ist es nur eine. Die Distraktoren müssen sich nicht schwarz/weiß von der richtigen Antwort oder den richtigen Antworten abheben. Es ist möglich, die Kandidaten eine intellektuell anspruchsvollere und oft auch realitätsnähere Abwägung von Graustufen vornehmen zu lassen. Wir beschreiben folgende Best-Antwort-Typen:

Best-Antwort-Typen..

Typ	Aufgabe
A(pos)	positive Einfachwahl aus fünf Wahlantworten
Aneg	negative Einfachwahl aus fünf Wahlantworten
B	Zuordnung der richtigen aus fünf Wahlantworten zu mehreren Fragen
R	erweiterte Zuordnung (wie Typ B aber mit längerer Wahlantwortliste)
PickN	Mehrfachwahl bei erweiterter Zuordnung (wie Typ R aber mit mehr als einer richtigen Antwort)

Bei der Gruppe der Richtig/Falsch-Typen müssen die Kandidaten für jede einzelne Antwort eine Ja/Nein-Entscheidung treffen. Es sind deshalb nur Antworten tauglich, bei denen dies eindeutig möglich ist. Richtig/Falsch-Typen bergen damit die Gefahr, reines Faktenwissen zu prüfen. Bestantwort-Items sollten in einer Prüfung quantitativ klar dominieren. Wir beschreiben folgende Richtig/Falsch-Typen:

... vor Richtig/Falsch-Typen

Typ	Aufgabe
Kprim	vierfache Entscheidung richtig/falsch
E	kausale Verknüpfung von zwei Aussagen
RF/RF?	Richtig/Falsch-Entscheid evtl. mit „weiss nicht“-Option

Positive Einfachwahl aus fünf Wahlantworten (Typ A)

Definition Auf eine Frage oder unvollständige Aussage folgen fünf Wahlantworten oder Ergänzungen, aus welchen die einzig richtige oder die beste auszuwählen ist.

Ein 53-jähriger Mann leidet seit zwei Tagen unter zunehmender Dyspnoe und produktivem Husten mit eitrigem Sputum. Seit dreissig Jahren raucht er jeden Tag ein Päckchen Zigaretten. Seine Körpertemperatur beträgt 37.2° C. Seine Atemgeräusche sind abgeschwächt mit ein wenig Giemen und Pfeifen. Die Leukozytenzahl beträgt 9000/mm³ mit normaler Differenzierung. Die Gramfärbung des Sputums zeigt zahlreiche Neutrophile und gramnegative Diplokokken. Die Thoraxröntgenaufnahmen zeigen eine Überblähung.

Der Stamm enthält die Informationsgrundlage, z.B. in Form einer Problem- oder Fallvignette...

Welches ist die wahrscheinlichste Diagnose?

oder

Die wahrscheinlichste Diagnose ist:

...und die darauf bezogene Aufgabenstellung in Form einer Frage oder einer Aussage.

- (A) Asthma
 (B) Bronchitis
 (C) Bronchiektasen
 (D) Lungenembolie
 (E) Streptokokkenpneumonie

Unter den fünf Wahlantworten gibt es nur eine richtige oder beste. Die andern vier dienen als Ablenker oder Distraktoren.

richtige Antwort: B, Quelle: Case und Swanson, 1998¹, Niveau: ärztliche Grundausbildung

Eignung Typ A ist nach wie vor der Standardtyp der MC-Methode. Er hat sich international unter verschiedenen Aspekten, insbesondere auch unter messtechnischem Gesichtspunkt, bestens bewährt und sollte jede MC-Prüfung anteilmässig klar dominieren.

Negative Einfachwahl aus fünf Wahlantworten (Typ Aneg)

Definition Auf eine Frage oder unvollständige Aussage folgen fünf Wahlantworten oder Ergänzungen, aus welchen die Ausnahme oder die am wenigsten zutreffende auszuwählen ist. Die Negation ist durch Fettdruck oder Unterstreichung hervorgehoben. Da Aneg-Items gleich strukturiert sind wie die positiven, verzichten wir auf die Wiedergabe eines ganzen Items und zeigen nur einige typische Fragenformen:

- Ein 25-jähriger Patient klagt über X. Er hat Fieber, und Sie stellen mehrere Y fest. Welche Diagnose kommt am wenigsten in Betracht?*
- Welche der folgenden Veränderungen gehört nicht zu den Zeichen einer chronischen X?*
- Bei einer Behandlung mit Z sind folgende Komplikationen zu befürchten, ausser:*

Sicher angezeigt ist dieser Itemtyp in den seltenen Fällen, in denen das Kennen einer wichtigen Ausnahme entscheidend ist. Viel häufiger - so auch in obigen Beispielen - wird er aber anders eingesetzt. Eigentlich soll das Kennen der vier positiven Antworten geprüft werden, die „richtige“ Antwort ist bloss das Abfallprodukt der Lösungen. Von der Prüfungsabsicht her wäre ein PickN-Item logischer, wenn es um eine Abwägung von Graustufen geht und ein Kprim-Item für einen schwarz/weiss-Entscheid. Fragen Sie sich aber zuerst grundsätzlich, ob nicht das Kennen des wichtigsten Zeichens, der gefährlichsten Komplikation, der wahrscheinlichsten Diagnose relevanter und anwendungsnäher, ein positives A-Item also besser wäre. Hinsichtlich Schwierigkeit und Trennschärfe schneiden Aneq-Items im Durchschnitt zwar praktisch gleich gut ab wie die positiven. Unter dem Aspekt der Validität sollte eine Prüfung aber kaum mehr als zehn Prozent Aneq-Items enthalten.

Eignung

Zuordnung (Typ B)

Bei diesem Typ werden eingangs fünf mit den Buchstaben A-E bezeichnete Wahlantworten vorgegeben. Oft eingeleitet durch einen Verbindungssatz, der die Zuordnungsaufgabe präzisiert, folgen zwei bis maximal fünf Fragen oder Aussagen. Jeder davon muss die einzig richtige oder die beste der fünf Wahlantworten zugeordnet werden, wobei eine Antwort mehr als einmal die richtige sein kann. Es handelt sich also um eine Serie von Fragen des Typs A mit den jeweils gleichen Wahlantworten.

Definition

- (A) Becken ap
- (B) Aufnahme nach Dunn, Rippstein, Müller
- (C) Faux profil nach Lequesne
- (D) Computertomographie
- (E) Hüfte axial (cross table lateral)

Welche radiologische Technik eignet sich beim Hüftgelenk am besten zur Ermittlung

1. der seitlichen Überdachung?
(A) (B) (C) (D) (E)
2. der vorderen Überdachung?
(A) (B) (C) (D) (E)
3. des Pfannendachwinkels?
(A) (B) (C) (D) (E)
4. der Ante-/Retroversion der Pfanne?
(A) (B) (C) (D) (E)

richtige Beantwortung: 1 (A), 2 (C), 3 (A), 4 (D) Quelle: Schweiz. Facharztprüfung Orthopädie

- (A) Nüchternblutzucker
- (B) Hämoglobin
- (C) Serumcholesterin
- (D) Serumeisen
- (E) TSH

Die folgenden zwei Personen konsultieren Sie für einen Check-up. Bezeichnen Sie die zur körperlichen Untersuchung am ehesten angezeigte diagnostische Massnahme:

5. 28-jähriger Bäcker, Gewicht 89 kg, Grösse 175 cm. Er raucht seit 11 Jahren täglich ein Päckchen Zigaretten und treibt keinen Sport. Der Vater erlitt 48-jährig einen Herzinfarkt. Die körperliche Untersuchung ergibt keine Abnormalitäten.

- (A) (B) (C) (D) (E)

6. 17-jährige Schülerin, Gewicht 67 kg, Grösse 169 cm. Sie gibt an, sie fühle sich manchmal tagelang sehr müde und habe hie und da Kreuzschmerzen. Sie fahre regelmässig mit dem Fahrrad zur Schule und gehe gerne tanzen. Vor einem Jahr habe sie mit Rauchen angefangen und rauche jetzt etwa ein halbes Päckchen pro Tag. Der Vater habe 50-jährig einen Herzinfarkt gehabt. Die körperliche Untersuchung ergibt ausser einer leichten Kyphose der Brustwirbelsäule keine Auffälligkeiten.

- (A) (B) (C) (D) (E)

richtige Beantwortung: 5 (C), 6 (B), Quelle: modifiziert nach Case und Swanson, 1998', Niveau: ärztliche Grundausbildung

Eignung In der dargestellten Form ist der Typ B sowohl unter dem Validitäts- wie dem Reliabilitätsaspekt gleich gut geeignet wie der positiv formulierte Typ A. Es sollte von der Breite und Bedeutung des zugrunde liegenden Themas her entschieden werden, ob dafür ein einzelnes A-Item ausreicht, oder ob ein B-Komplex mit mehreren Items angemessener ist. Abzuraten ist von Aufgaben, bei welchen als „Fragen“ Definitionen oder Eigenschaftsbeschreibungen vorgelegt werden, denen als „Antwort“ der treffende Begriff zugeordnet werden muss. Solche Kreuzworträtsel-Fragen prüfen Begriffs- und Faktenwissen auf tiefster Stufe.

Erweiterte Zuordnung (Typ R)

Definition Dieser Typ ist gleich strukturiert wie Typ B, allerdings ist die Liste der Wahlantworten wesentlich länger (bis zu 26 Optionen) und alphabetisch geordnet. Er ist für die amerikanischen Medizinalprüfungen (USMLE) entwickelt und erprobt worden und wird dort auch zunehmend eingesetzt¹.

Bei den oben genannten B-Items Nrn. 5 und 6 könnte die Optionenliste wie folgt sinnvoll verlängert werden:

- | | |
|---------------------------|----------------------|
| (A) EKG | (F) Senkungsreaktion |
| (B) Ergometrie | (G) Serumcholesterin |
| (C) Hämoglobin | (H) Serumeisen |
| (D) Nüchternblutzucker | (I) Spirometrie |
| (E) Röntgenthoraxaufnahme | (J) TSH |

Tatsächlich wird für das Verfassen von R-Items ein anderes Vorgehen empfohlen¹. Ausgehend von einer Hauptbeschwerde soll zuerst eine Liste möglicher Antworten (z.B. Diagnosen) aufgestellt werden. Bezogen auf diese Antwortpalette sind dann Patientenfälle als Fragen zu entwickeln (vgl. Abb. 1.12, S.34).

Inhaltlich ist der Typ R wie der Typ B für breitere Themen zu empfehlen, denen mehr als ein Item gewidmet werden sollte. Er scheint sich besonders gut zu eignen, wenn der thematische Ausgangspunkt ein unspezifischer Konsultationsgrund ist wie Atembeschwerden, Bauchschmerzen, Fieber, Müdigkeit usw. Aus der Konstellation weiterer in den einzelnen Fragen gegebenen Befunden soll z.B. die wahrscheinlichste Diagnose abgeleitet werden oder die beste nächste Abklärungsmassnahme oder die aussichtsreichste Behandlungsmassnahme.

Eignung

Die lange Antwortliste soll die Erinnerungshilfe, welche die Wahlantworten bieten (sog. Cueing-Effekt) reduzieren. Dies trägt zu einer Steigerung der Messqualität bei. Man erhofft sich sogar, dass es zu ineffizient und damit unattraktiv wird, die langen Listen durchzuprüfen und dass die Kandidaten die Antworten selbst entwickeln und dann gezielt in der alphabetisch geordneten Liste suchen. Damit würde sich dieser Itemtyp Fragen mit offener Beantwortung annähern.

Mehrfachwahl bei erweiterter Zuordnung (Typ PickN)

Dieser Typ ist in der Regel gleich strukturiert wie der Typ R, denkbar ist auch die Form des A-Typs. Es ist aber nicht nur eine beste oder einzig richtige Antwort zu suchen, die Zahl der auszuwählenden Antworten - in der Regel 2 oder 3 - wird angegeben.

Definition

B-Typ Beispiel Nr. 5 kombiniert mit der längeren R-Typ Antwortliste könnte als Typ PickN lauten:

*Bezeichnen Sie die angezeigten diagnostischen Massnahmen. (Wählen Sie 2)
Richtige Beantwortung: Neben (G) Serumcholesterin wäre angesichts des langjährigen Rauchens auch (E) Röntgenthoraxaufnahme angezeigt.*

Eignung Bei medizinischen Problemen gibt es oft mehr als eine vernünftige Lösung. Eine Störung kann mehrere Ursachen haben, in der Anfangsphase einer Differentialdiagnose müssen mehrere Möglichkeiten in Betracht gezogen werden. Oft ist es sinnvoll, mehrere diagnostische und/oder therapeutische Massnahmen zu ergreifen. Falls zur Auswahl der richtigen Antworten eine Graustufenabwägung vorgenommen werden muss, ist PickN der geeignete Typ. Wird nur bei einer vollständig richtigen Beantwortung ein Punkt zugesprochen, sind PickN-Items deutlich schwerer als Einfachwahlitems. Vor allem bei Items mit mehr als zwei richtigen Antworten ist aus Fairness- und messtechnischen Gründen (bessere Trennschärfe) eine Teilpunktbewertung zu empfehlen (vgl. Typ Kprim).

Vierfache Entscheidung richtig/falsch (Typ Kprim/K')

Definition Auf eine Frage oder unvollständige Aussage folgen vier Antworten oder Ergänzungen. Für jede muss entschieden werden, ob sie richtig oder falsch ist. Um einen Punkt zu erhalten, müssen alle vier Beurteilungen korrekt sein. In einer Auswertungsvariante werden drei korrekte Beurteilungen bereits mit einem halben Punkt honoriert.

1. *Wie lässt sich die Festigkeit von dünnflüssigem Knochenzement verbessern?*

- (A) Abkühlen
- (B) Zentrifugieren
- (C) Zugabe von Antibiotika
- (D) Vakuum-Mischen

richtige Beantwortung: ---, Quelle: Schweiz. Facharztprüfung Orthopädie

2. *Eine 45-jährige Hausfrau leidet seit ca. 20 Jahren unter einer schweren Polyarthrit. Sie ist medikamentös gut eingestellt und wird Ihnen zugewiesen wegen einer progredienten Deformation ihrer Hände, welche das Fassen und Halten von Gegenständen erschwert. Die klinische Untersuchung zeigt eine starke Deformierung der vier Langfinger mit Ulnardeviation. Klinisch sind Handgelenk und Daumen wenig betroffen und aliniert. Die vier Langfinger sind in den Metakarpo-Phalangealgelenken subluxiert, können aber noch reponiert werden. Die Röntgenbilder bestätigen eine antero-mediale Subluxation und eine Destruktion der Metakarpo-Phalangealgelenke. Zur Behandlungsstrategie gehört:*

- (A) Swanson Arthroplastik der Metakarpo-Phalangealgelenke
- (B) Realignement der Strecksehnen über den Metakarpo-Phalangealgelenken
- (C) Transfer des Abductor longus auf den ersten dorsalen Interosseus (Neviaser)
- (D) radio-karpale Arthrodesen in Ulnarstellung des Handgelenkes

richtig Beantwortung: +- - , Quelle: Schweiz. Facharztprüfung Orthopädie

Inhaltlich betrachtet ist der Typ K' angezeigt, wenn es um einen Sachverhalt geht, bei dem mehrere Aspekte bedeutsam sein können, resp. um ein Problem, zu dessen richtiger Lösung mehrere Elemente gehören können. Im Gegensatz zum Typ PickN müssen die Antworten schwarz/weiss beurteilbar sein. Der Typ K' sollte nicht missbraucht werden für heterogene Items, bei denen der Stamm z.B. lautet: „Für die XYitis trifft zu:“ und dann eine Aussage zur Ursache, eine solche zur Lokalisation, eine zur Differentialdiagnose und letztlich eine zur Behandlung gemacht wird.

Eignung

Unter messtechnischem Aspekt sind K'-Items etwas häufiger problematisch als Items der Typen A und B. Eine einzige nicht-funktionierende Teilantwort kann ein Item zu Fall bringen. Hinsichtlich Trennschärfe sind K'-Items mit der Alles-oder-nichts-Auswertung den Typen A und B ebenbürtig. Mit der Auswertungsvariante, bei welcher drei richtige Teilantworten bereits mit $0\frac{1}{2}$ Punkten belohnt werden, sind sie tendenziell gar überlegen. In der Alles-oder-nichts-Form sind sie deutlich schwerer als Einfachwahlfragen und bei den Kandidaten unbeliebt. Beide Punkte werden aber durch die Halbpunkt-Auswertungsversion positiv verändert³.

Beim Typ K, aus dem der oben beschriebene Typ Kprim abgeleitet worden ist, wurden mit den vier Antworten immer folgende fünf Muster vorgegeben, aus denen das zutreffende auszuwählen war:

Kein Typ K mehr!

- (A) 1+2+3 (B) 1+3 (C) 2+4 (D) nur 4 (E) alle vier

Dieser Typ sollte nicht mehr verwendet werden, da er bezüglich Messeigenschaften wesentlich schlechter abschneidet als Typ Kprim. Die Ratewahrscheinlichkeit ist a priori höher (20 % vs. 6.25 %). Wird eine der vier primären Antworten als falsch erkannt, können gleich drei Antwortmuster verworfen werden, und die Ratewahrscheinlichkeit steigt auf 50 % (bei Kprim lediglich auf $12\frac{1}{2}$ %).

Kausale Verknüpfung (Typ E)

Zwei Aussagen sind durch das Wort „weil“ verknüpft. Es sind zuerst unabhängig voneinander die beiden Aussagen als richtig oder falsch zu beurteilen. Wenn beide richtig sind, ist zusätzlich zu entscheiden, ob die weil-Verknüpfung berechtigt ist. Das ergibt fünf Antwortmöglichkeiten:

Definition

- (A) +weil+ Beide Aussagen stimmen, die weil-Verknüpfung ist berechtigt.
 (B) +/+ Beide Aussagen stimmen, deren weil-Verknüpfung ist falsch.
 (C) +/- Die erste Aussage ist korrekt, die zweite ist falsch.
 (D) -/+ Die erste Aussage ist falsch, die zweite ist korrekt.
 (E) -/- Beide Aussagen sind falsch.

Bei Heparin-induzierter Thrombozytopenie kann niedermolekulares Heparin (LMWH) eingesetzt werden,

weil

niedermolekulares Heparin (LMWH) weniger Thrombozytopenie verursacht als unfraktioniertes Heparin.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>+weil+</i>	<i>+/+</i>	<i>+/-</i>	<i>-/+</i>	<i>-/-</i>

richtige Antwort: D, Quelle: Schweiz. Facharztprüfung Angiologie (adaptiert)

Beide Aussagen müssen alle notwendigen Informationen enthalten, um unabhängig voneinander beurteilbar zu sein. In obigem Beispiel muss deshalb im zweiten Teil der Begriff „niedermolekulares Heparin“ wiederholt werden und darf nicht durch das Pronomen „dieses“ ersetzt werden.

Eignung Inhaltlich betrachtet wären Items vom Typ E in Gebieten angezeigt, in denen kausale Zusammenhänge bedeutsam sind, also beispielsweise um die Kenntnis pathophysiologischer Zusammenhänge oder pharmakologischer Mechanismen zu prüfen. Tatsächlich werden sie von Fragenautoren häufig als Ausweg benutzt, wenn sich zu einem Thema keine vier plausiblen Distraktoren für ein Item vom Typ A finden lassen.

Unter messtechnischem Gesichtspunkt sind E-Items ziemlich problematisch. Kausalitäten sind in der Medizin recht selten schwarz/weiß zu beurteilen. (Ist die Kausalität z.B. zu bejahen, wenn die zweite Aussage nur ein Grund unter mehreren ist?) Die Entscheidung zwischen den Antworten A und B ist damit oft auch eine Ermessensfrage und nicht nur vom Fachwissen abhängig. Andererseits ist es schwierig, E-Items zu konstruieren, in denen alles plausibel erscheint, obwohl eine oder gar beide Aussagen falsch sind. E-Items sind damit anfällig auf unfreiwillige Lösungshinweise, welche ihre Trennschärfe beeinträchtigen. Wir empfehlen deshalb, Typ E-Items - wenn überhaupt - nur sehr sparsam einzusetzen und unbedingt die Hinweise auf S.88 zu beachten.

Richtig/Falsch-Entscheid (Typ RF)
evtl. mit „Weiss nicht“-Option (Typ RF?)

Definition Es werden einzelne Aussagen vorgelegt, die als richtig oder falsch zu beurteilen sind. In der Variante mit einer zusätzlichen „Weiss nicht“-Option werden falsche Beurteilungen mit einem Punkteabzug bestraft. Damit soll dem blinden Raten entgegengewirkt werden. Es können auch mehrere Aussagen bezogen auf eine Patientenvignette formuliert werden, deren Richtigkeit zu beurteilen ist.

1. Bei Patienten, die unter ACE-Hemmer-Therapie zu husten anfangen, kann auf Angiotensin 2-Rezeptor-Antagonisten umgestellt werden, um den Husten zum Verschwinden zu bringen.

Ja Nein weiss nicht

2. Bei anurischen Hämodialyse-Patienten kann ein Morbus Conn Ursache der schwer einstellbaren arteriellen Hypertonie sein.

Ja Nein weiss nicht

richtige Antworten: 1 (ja), 2 (nein), Quelle: Schweiz. Facharztprüfung Nephrologie

Typ RF bietet folgende Vorteile: Der Itemtyp ist recht einfach zu kreieren und zu revidieren. Auch eine kleine Examinatorengruppe (z.B. bei einer kleinen Fachgesellschaft) kann die für eine Prüfung erforderliche Itemzahl erstellen. Er erfordert wenig Zeit zur Beantwortung, wodurch in der Prüfungszeit ein breites Inhaltsspektrum abgedeckt werden kann.

Eignung

Die Nachteile sind: Es wird fast ausschliesslich Faktenwissen geprüft. Messtechnisch betrachtet ist der Itemtyp schwach. Es sind sehr viele Items erforderlich, um eine hinreichende Messzuverlässigkeit zu erreichen. Als alleiniger Itemtyp in einer Wissensprüfung sollte der Typ RF nur als letzte Notlösung bei fehlenden Personalressourcen in Betracht gezogen werden.

In der Version ohne „Weiss nicht“-Option ist der grosse Einfluss der Ratewahrscheinlichkeit von 50 Prozent für die schlechte Reliabilität verantwortlich. Mit der Variante mit „Weiss nicht“-Option gerät man vom Regen in die Traufe, da hier neben dem Wissen „Persönlichkeitseigenschaften“ (z.B. Risikobereitschaft) ins Spiel kommen, welche die Zuverlässigkeit der Wissensmessung beeinträchtigen.

3.4 Wie werden Fragen formuliert?

drei Grundregeln: Soll mit MC-Items mehr geprüft werden als Faktenwissen, müssen komplexe Problemsituationen vorgelegt werden, die mehrere zu interpretierende und zu integrierende Informationen enthalten. Gute Items benötigen deshalb häufig einen ausführlichen Stamm. Bei den Items der Typen A, B, R und PickN müssen die Antworten oft gegeneinander abgewogen werden. Dies ist fast nur möglich, wenn sie kurz und übersichtlich sind. Diese Überlegungen führen zur ersten Regel:

1. langer Stamm, kurze Antworten Der Stamm darf lang, die Antworten sollen kurz sein (vgl. Abb. 3.1).

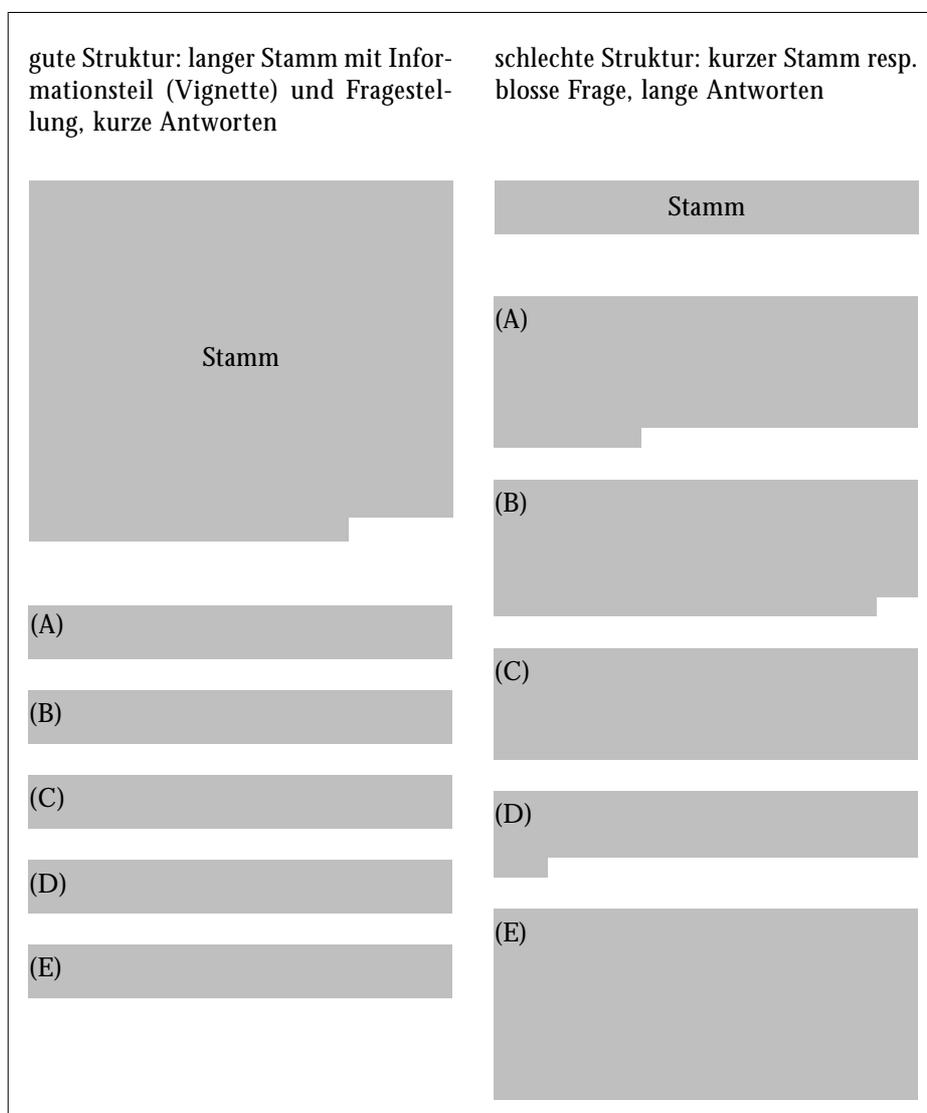


Abb. 3.1 Gute und schlechte Struktur bei MC-Items

Die zwei folgenden Regeln sollen gewährleisten, dass in fairer Weise die Verfügbarkeit und Anwendungskompetenz von Fachwissen geprüft wird.

Items sollen einfach, klar und in einer allgemein akzeptierten Sprache formuliert werden.

2. einfach und klar formulieren

Dazu gehört auch, dass nur geläufige Abkürzungen wie EKG, CT ohne zusätzliche ausgeschriebene Version in Klammern verwendet werden.

Items sollen nicht künstlich kompliziert gemacht oder gar als bewusste Fangfragen konzipiert werden.

3. keine Fangfragen kreieren

Die Schwierigkeit eines Items soll von der Komplexität des zugrunde liegenden Problems, dem geistigen Anspruch (Verständnis, Syntheseleistung, Problemlösung) und der Feinheit der erforderlichen Differenzierung (gegenseitige Nähe der Wahlantworten) bestimmt sein. Es ist unfair und widerspricht mit Sicherheit der Prüfungsabsicht, Items, die triviale Faktenkenntnisse prüfen, durch formale Tricks schwieriger zu machen.

Formulierung des Itemstammes

Die folgenden Ausführungen gelten sinngemäss für alle MC-Itemtypen:

Itemstämme können aus einem einzigen Fragesatz oder einer Aussage bestehen. Mit solchen Items kann durchaus anwendungsrelevantes Wissen geprüft werden, in der Regel handelt es sich allerdings lediglich um Faktenwissen.

Der Stamm kann aus einem Satz...

1. Welche Spüllösung wird für Arthroskopien empfohlen?
2. Das wichtigste klinische Kriterium für die Diagnose einer kompletten Supraspinatusruptur ist:

Mit MC-Items können und sollen möglichst häufig auch die Fähigkeiten geprüft werden, Informationen zu interpretieren und zu integrieren und theoretische Kenntnisse auf ein konkretes Problem anzuwenden. Dafür sind zweiteilige Fragenstämme erforderlich. In einem ersten längeren Teil wird zu einem Patientenfall oder einem anderen Problem aus dem fachärztlichen Berufsfeld informiert. Wir sprechen von einer Fall- oder Problemvignette. Davon abgetrennt erfolgt die kurze Fragestellung.

...oder einer Vignette plus Fragestellung bestehen.

Ein am Termin geborener Säugling wiegt 3090 g. Er ist hypoton und hat eine geringe Saugkraft. Zyanose besteht keine. Er hat einen markanten Epikanthus, eine etwas grosse Zunge, und seine fünften Finger weisen eine Einwärtskrümmung auf. Über sämtlichen präkordialen Auskultationsstellen ist ein lautes holosystolisches Geräusch zu hören. Zudem ist am oberen linken Sternalrand ein Schwirren (Thrill) zu palpieren, welches vom 2. bis

zum 4. Interkostalraum reicht.
Welcher kongenitale Herzfehler liegt am wahrscheinlichsten vor?

Quelle: Case und Swanson, 1998¹

Um die Fragen richtig beantworten zu können, muss vorgängig mit den Informationen diagnostische Integrations- und Interpretationsarbeit geleistet werden.

Patientenvignetten sollen einige oder alle der folgenden Elemente in der angegebenen Reihenfolge enthalten:

1. Alter, Geschlecht, evtl. Beruf, Rasse (*Ein 45-jähriger Wirt ...*)
2. Ort der Behandlung (*...wird in die Notfallabteilung gebracht...*)
3. Konsultationsgrund (*...wegen starker Kopfschmerzen,...*)
4. Dauer (*...die seit zwei Tagen andauern.*)
5. Anamnese, evtl. Familienanamnese (relevante Punkte)
6. Status, physische Befunde
7. evtl. Resultate der diagnostischen Untersuchungen
8. evtl. Anfangsbehandlung, Folgebefunde, etc.

*Abbildungen und
andere Informations-
grundlagen*

Der Informationsteil kann auch Abbildungen beinhalten wie Fotografien, Röntgenbilder, EKG, Audiogramm. Bei Themen der Prävention oder Epidemiologie kann auch ein Literaturzitat oder eine (Verlaufs-)Grafik zugrunde gelegt werden.

Bei der Formulierung sollten Sie darauf achten, dass der Stamm

*Formulierungsregeln
für den Itemstamm*

- alle für die Beantwortung erforderlichen Informationen enthält, so dass in den Wahlantworten keine zusätzlichen gegeben werden müssen
- in der Regel keine überflüssigen Informationen enthält
Eine Ausnahme kann sein, wenn ausdrücklich die Fähigkeit geprüft werden soll, die relevante Information herauszufiltern. Sonst wird aber mit überflüssigem Text wertvolle Prüfungszeit verschenkt.
- möglichst beantwortbar ist, ohne die Wahlantworten zu sehen
- bei Einfachwahlitems möglichst, bei Kprim immer positiv formuliert ist
Aneg-Items sind unter dem Validitätsgesichtspunkt wenig erwünscht. Bei K'-Items entstehen bei negativer Stammformulierung unweigerlich verwirrende doppelte Negationen.

Formulierung der Antworten

Die bisherigen Ausführungen zu den Anforderungen, die Items erfüllen sollen (Fokussierung des Itemthemas, Itemformulierung im Allgemeinen, Formulierung des Itemstammes), gelten praktisch unverändert auch für Items mit offener Beantwortung, sei dies für schriftliche Kurzantwortfragen oder Fragen für eine strukturierte mündliche Prüfung. Das Folgende ist nun ganz MC-spezifisch, für diese Prüfungsmethode aber von grösster Wichtigkeit. Aus einer höchst relevanten, anwendungsbezogenen Problemsituation entsteht durch schlecht gewählte und/oder schlecht formulierte Antworten ein völlig untaugliches Item. Wir konzentrieren uns hier auf die Best-Antwort-Typen und erwähnen abweichende Punkte für den Typ Kprim unter dem nächsten Paragraphen.

MC-spezifische Anforderungen

Damit ein Item zur Validität und Reliabilität der Prüfung beiträgt, sollte das Finden der richtigen Antwort möglichst ausschliesslich davon abhängen, wie weit das zu prüfen beabsichtigte Wissen verfügbar ist. Um dies zu erreichen, müssen die Wahlantworten eine ganze Reihe inhaltlicher, formaler und sprachlicher Kriterien erfüllen.

- Alle Wahlantworten sollen in die gleiche Kategorie fallen, also inhaltlich homogen sein (z.B. alles Diagnosen, alles Ursachen, alles Massnahmen).

inhaltliche Kriterien

Abwägen zu müssen, ob eine vorgeschlagene Krankheitsursache wahrscheinlicher richtig ist als ein Therapievorschlag, hat keinen Bezug zur Berufsrealität eines Facharztes.

- Für Distraktoren sollen klare Gründe bestehen. Es kann sich z.B. um häufige Fehlmeinungen, falsche Konzepte, veraltete Ansichten handeln. Zumindest sollte aber eine klar nachvollziehbare Beziehung zum Itemthema bestehen.

Unplausible, triviale oder gar völlig unsinnige Distraktoren können auch von leistungsschwachen Kandidaten sofort ausgeschlossen werden. Dies erhöht ihre Chance, die richtige Antwort zu raten, womit das Item seine Trennschärfe verliert. Um die Ratewahrscheinlichkeit klein und konstant zu halten, ist es sinnvoll, möglichst durchgehend fünf Wahlantworten vorzugeben. Falls aber nur drei sinnvolle Distraktoren gefunden werden, scheint es uns sinnvoller, nur diese plus die eine richtige anzubieten, als etwas Absurdes aufzunehmen.

- Die Distraktoren müssen nicht alle völlig falsch, die richtige Antwort muss aber eindeutig die beste sein. Je näher richtige Antwort und Distraktoren inhaltlich und hinsichtlich Wahrheitsgehalt beieinanderliegen, desto schwieriger wird das Item.

- Jede Wahlantwort soll kurz sein und nur eine Aussage enthalten.

- Sich überschneidende Wahlantworten sollten nur verwendet werden, wenn es das zugrundeliegende Problem erfordert. Sie sollten nicht eingesetzt werden, um ein Item künstlich schwieriger zu machen.
- „Alle der obigen“ oder auch Antworten wie „Sowohl B und C sind richtig“ dürfen nicht verwendet werden.
Falls es sich dabei um die vorgesehene Richtigantwort handelt, gibt es mehr als eine richtige Antwort. Falls es eine Falschantwort ist, ist sie geschenkt, sobald nur eines der enthaltenen Elemente als falsch identifiziert werden kann.
- „Keines der obigen“ sollte nur in Ausnahmefällen als fünfte Wahlantwort verwendet werden. Möglich ist die Aussage, wenn sie relevantes Wissen prüft, oder wenn z.B. die richtige Antwort wegen allzu starker (evtl. gar verbaler) Assoziation zum Stamm nicht namentlich erwähnt werden kann. Natürlich müsste „Keines der obigen“ dann auch ab und zu die falsche Antwort sein. Dies ist aber häufig problematisch. Wenn „Keines der obigen“ als falsche Antwort vorgesehen ist, darf die richtige Antwort nicht mehr nur die beste unter den aufgeführten sein. Sobald eine noch bessere, z.B. differenziertere Antwort möglich ist, wählen sehr gute Kandidaten „Keines der obigen“, womit das Item seine Trennschärfe verliert.

formale Kriterien Bei den formalen Kriterien geht es primär darum, unbeabsichtigte Lösungshinweise, sog. Cues, zu vermeiden. Diese erlauben MC-erfahrenen Kandidaten, auch ohne Fachkenntnis die richtige Antwort zu identifizieren oder doch einzelne falsche Antworten auszuschneiden und damit ihre Ratechance zu erhöhen. In zweiter Linie soll der Einfluss bestimmter Beantwortungstendenzen von Kandidaten minimiert werden.

Cues vermeiden Die wichtigsten, häufigsten Cues illustrieren wir an inhaltlichen Nonsense-Items, versuchen Sie also nicht, die folgenden Fragen vom Inhalt her zu beantworten.

- Distraktoren sollen möglichst gleich lang und differenziert sein wie die richtige Antwort.

Bei der Fermierung von Anter mit saurem Gor

- (A) *findet eine Abkühlung statt*
- (B) *entsteht unter der Bedingung einer leichten Erwärmung Anterit im pH-Bereich 2.8-3.2*
- (C) *wird OgI_4 freigesetzt*
- (D) *entsteht Fermantin*
- (E) *bildet sich Gorantoxol*

Für Autoren steht verständlicherweise die richtige Antwort meist klar im Vordergrund und sie bemühen sich, diese möglichst präzise zu formulieren. Den Distraktoren wird als blosses „Hintergrundrauschen“ weniger Aufmerksamkeit geschenkt. Testerfahrene Kandidaten würden in obigem Beispiel deshalb mit guter Erfolgsaussicht Antwort (B) wählen. Es lassen sich nicht immer gleich lange und differenzierte Antworten kreieren. Achten Sie als Autor aber darauf, dass die richtige Antwort nicht regelmässig heraussticht.

- Alle Antworten müssen grammatikalisch zum Stamm passen.

Anter ist eine

- (A) *Legierung*
- (B) *Konglomerat*
- (C) *Verbrennungsrückstand*
- (D) *Spaltprodukt*
- (E) *chemisches Element*

Aus grammatikalischem Grund (weiblicher/männlicher Artikel) kommt nur (A) als richtige Antwort in Frage. Es dürfte selten vorkommen, dass vier Antworten ausgeschlossen werden können, hingegen häufig, dass einzelne ausser Betracht fallen. Lesen Sie als Autor zur Kontrolle dieses Cues jede Antwort in Verbindung mit dem Fragesatz.

- Verbale Assoziationen zwischen Stamm und richtiger Antwort vermeiden.

Hauptmerkmal des KRS (Kognitives Rigiditäts Syndrom) ist/sind

- (A) *ein erhöhter Ferminspiegel im Plasma*
- (B) *zyklische postprandiale Alpträume*
- (C) *häufige Versteifungen der Nackenmuskulatur*
- (D) *eine reduzierte Beweglichkeit im kognitiven Bereich*
- (E) *eine chronische Logo- und Skriptorrhö*

Der Begriff „kognitiv“ taucht sowohl im Stamm wie in Antwort (D) auf und erhöht die Wahrscheinlichkeit, dass es sich dabei um die richtige Antwort handelt. Der Cue kann durch Umschreibung oder Wahl eines Synonyms, hier z.B. „Beweglichkeit des Denkens“ vermieden werden.

- Absolute Begriffe wie „nie“, „immer“ sollten nicht gewählt werden, um Aussagen eindeutig falsch zu machen.

Warum ist bei Trigotie-Patienten die Axosie-Auftretensrate erhöht?

- (A) *Trigote Patienten sind nie berop.*
- (B) *Trigotie führt immer zu Enität.*
- (C) *Alle Trigotie-Patienten werden esophym.*
- (D) *Axosie ist ausschliesslich sequid bedingt.*
- (E) *Trigote Patienten sind gehäuft susmin.*

Wegen den Absoluta in den Antworten A-D werden diese für MC-erfahrene Kandidaten als richtige Antworten unwahrscheinlich. Zudem passt „gehäuft“ qualitativ am besten zu „erhöht“. Der Begriff „immer“ kann oft ohne Änderung der Aussage weggelassen werden. So wäre „weil Trigotie zu Enität führt“ auch ohne Nachdoppelung durch „immer“ bereits eine absolute Aussage.

- Hinweise, welche die Aufmerksamkeit auf 2-3 Antworten einschränken, sind zu vermeiden.

Bei Trigotie

- (A) *handelt es sich um eine Stoffwechselstörung*
- (B) *nimmt die Axosie-Anfälligkeit zu*
- (C) *nimmt die Axosie-Anfälligkeit ab*
- (D) *bleibt die Axosie-Anfälligkeit konstant*
- (E) *ist eine Therapie mit Palismat indiziert*

Die richtige Antwort muss logischerweise B, C oder D sein; die andern beiden brauchen gar nicht in Betracht gezogen zu werden. Bereits wenn sich zwei Antworten auf ein und denselben Aspekt beziehen, und die drei anderen auf drei je unterschiedliche Aspekte, lenkt der MC-erfahrene Kandidat sein Augenmerk sofort auf den doppelt vorkommenden Aspekt. Der formale Fehler rührt wiederum daher, dass sich Autoren stärker auf die richtige Antwort konzentrieren als auf die Distraktoren.

- Die Antworten sollen inhaltlich möglichst logisch angeordnet werden. Wenn bei den Antworten vom Inhalt her eine logische Hierarchie gegeben ist (z.B. bei therapeutischen Massnahmen von „abwarten“ über „medikamentös behandeln“ zu „operativ entfernen“), sollen sie danach angeordnet werden. Falls dies nicht der Fall ist, empfiehlt sich am ehesten eine alphabetische Reihung. Kandidaten, welche die richtige Antwort nicht kennen, werden auch bei der Reihung nach einer Auffälligkeit suchen, die sie als Lösungshinweis nutzen können.

und Beantwortungstendenzen entgegenwirken

- Die richtige Antwort soll nicht überwiegend unter C oder D platziert werden. In der Absicht, die richtige Antwort möglichst „gut zu verstecken“, wurde sie von Autoren früher überdurchschnittlich oft in der Mitte platziert. Dies scheint sich herumgesprochen zu haben. Viele Kandidaten, die raten müssen, bevorzugen bei der Antwortwahl die Mitte (sog. Zentraltendenz). Deshalb platzieren Autoren die richtige Antwort jetzt zunehmend eine Position weiter hinten, unter D. Damit Positions-Wahl Tendenzen von Kandidaten keinen Einfluss gewinnen können, sollen die richtigen Antworten über alle Items einer Prüfung möglichst ausbalanciert unter A-E verteilt sein.

Besondere Anforderungen verschiedener Itemtypen

- Die Negation muss deutlich hervorgehoben werden, sei es durch **Fett-druck** oder Unterstreichung. *Typ Aneg*

- Alle Antworten müssen positiv formuliert sein, da sonst doppelte Negationen zu beurteilen sind.

- In der Regel sollen nicht mehr als drei Fragen zu einem B-Komplex kombiniert werden, um eine inhaltliche Übergewichtungen zu vermeiden und folgenden zwei Gefahren vorzubeugen: *Typ B*
 - Inhaltlich heterogene Fragenkombinationen, bei denen von vornherein bei einzelnen Fragen gewisse Antworten ausser Betracht fallen, sind zu vermeiden.
 - Wenn einer Frage eine bestimmte Antwort zugeordnet werden kann, sollte sie nicht dadurch als richtige Antwort auf andere Fragen von vornherein ausser Betracht fallen.
Bei Verletzung dieser beiden Punkte wird die Ratewahrscheinlichkeit erhöht und damit die Messzuverlässigkeit beeinträchtigt.

- Jede Aussage muss eindeutig richtig oder falsch sein. *für alle
Richtig/Falsch-
Itemtypen
(Kprim, E, RF) gilt:*
- Es sollen nie zwei Aussagen in eine gepackt werden (z.B. Aussage mit Begründung).
Es kann unklar werden, was zu beurteilen ist.
- Negationen sind möglichst zu vermeiden.
Bei der Beurteilung entstehen doppelte Negationen (es ist falsch, dass es falsch ist), die verwirlich sind und die Messabsicht stören können.
- Vage Begriffe wie „gewöhnlich“, „häufig“, „oft“, „assoziiert mit“ sind möglichst zu vermeiden.
Untersuchungen^{4,5} haben gezeigt, dass solche Begriffe mit sehr unterschiedlichen Quantifizierungen verbunden werden. Damit ist es u.U. nicht mehr nur vom Fachwissen abhängig, ob einer solchen vagen Aussage zugestimmt wird.
- Richtige und falsche Aussagen sollten ausbalanciert sein.
Es besteht bei Autoren die Tendenz, auch in der Prüfung noch Wissen vermitteln zu wollen und daher positive Aussagen vorzuziehen. Da die Mehrzahl der Kandidaten dies zu erwarten scheint und bei Nicht-Wissen eher auf eine positive Antwort tippt, weisen diese tendenziell eine schwächere Trennschärfe auf als negative.

- Typ Kprim*
- siehe oben „Aussagen in Richtig/Falsch-Itemtypen“
 - Der Stamm ist immer positiv zu formulieren.
 - Die Formulierung des Stammes muss offen lassen, wieviele der vier Aussagen richtig sind.
Früher wurde dies oft durch leseunfreundliche Formulierungen gewährleistet in der Art von: „welche/r Zusammenhang/Zusammenhänge trifft/treffen zu?“ Heute weisen wir zum einen in der Instruktion für die Kandidaten darauf hin, dass unabhängig von der Einzahl- oder Mehrzahlformulierung 0-4 Aussagen richtig sein können. Zum andern wählen wir möglichst quantitativneutrale Stammformulierungen, wie:

*„Dieses XY manifestiert sich durch: .. kommt vor bei: .. wird behandelt mit:“
„Für XY spricht: .. gilt: .. trifft zu:
„Zu den XY gehört: ... zählt:“*
 - Jede Antwort muss unabhängig von allen andern sein. Insbesondere sind Antworten zu vermeiden, die sich gegenseitig ausschließen.

- Typ E*
- siehe oben „Aussagen in Richtig/Falsch-Itemtypen“
 - Beide Aussagen müssen in sich geschlossen sein und alle notwendigen Informationen enthalten, um unabhängig voneinander auf ihre Richtigkeit hin beurteilt werden zu können.
 - Bei Items mit zwei richtigen Aussagen muss die Kausalität eindeutig richtig oder falsch und nicht Ermessensfrage sein.
 - Unabhängig davon, was richtig und falsch ist, soll die Kausalität für Uninformierte immer plausibel erscheinen.
 - Es soll vermieden werden, dass E-Typen überwiegen, bei denen alles inkl. kausaler Verknüpfung korrekt ist oder bei denen nur der erste Teil richtig ist.

3.5 Wie werden Fragen überprüft?

Der Inhalt und die Form von MC-Items müssen vom Autor und von unabhängigen Experten überprüft werden. Wir empfehlen, ein standardisiertes Itemformular zu verwenden (vgl. Anhang 1, S.93). Dies trägt dazu bei, dass die Autoren alle erforderlichen Angaben machen und die Arbeit der Experten, welche neue Items begutachten, erleichtert wird.

Als Autor sollten Sie nach dem Verfassen eines Items

- dieses nochmals durchlesen und sich vorstellen, dass Sie es selbst beantworten müssen
- anhand der Checkliste eingangs dieser Anleitung überprüfen, ob Sie alle Grundregeln beachtet haben
- im Speziellen überprüfen, ob das Item keine Cues enthält
- angeben, zu welchem Inhaltsgebiet (Blueprintkapitel) das Item gehört und welches die richtige Lösung ist
- den inhaltlichen Fokus des Items in Form einiger Kennworte festhalten und nach Möglichkeit eine oder mehrere Literaturreferenzen angeben, welche die Richtigkeit der „richtigen Lösung“ belegen
- allenfalls begründen, warum gerade diese Distraktoren gewählt worden sind und warum sie falsch sind

Autoren kontrollieren...

...und dokumentieren ihre Items

Legen Sie das Item dann für eine Woche zur Seite und sehen Sie es danach nochmals durch.

Die Revision neuer Items durch unabhängig Experten sollte sinnvollerweise in mehreren Schritten erfolgen.

- 1-2 Inhaltsexperten sollten unabhängig von der Form die fachliche Richtigkeit und die Zugehörigkeit des Itemthemas zum Weiterbildungsprogramm begutachten (vgl. Itemformular).
- Ein weiterer MC-Prüfungsexperte sollte die Items ausschliesslich unter formalem und sprachlichem Gesichtspunkt begutachten.
- Mit den Anmerkungen dieser Experten sollten die Items dem für das Fach repräsentativ zusammengestellten Revisionsgremium gestellt

Experten überprüfen deren Inhalt...

...und Form

werden. In der Sitzungsvorbereitung sollte jedes Mitglied u.a. seine Relevanzeinschätzung festhalten.

*Prüfungskommission
entscheidet*

- In der eigentlichen Revisionsitzung wird über Annahme oder Rückweisung und evtl. Modifikationen entschieden. Die Relevanzeinstufung, welche bei der Prüfungszusammenstellung sehr hilfreich ist, kann als Mittelwert der individuellen Einschätzungen berechnet oder als Konsensurteil ausdiskutiert werden.

3.6 Wie wird die Prüfung zusammengestellt?

- Um von Session zu Session vergleichbar zu messen, sollten Prüfungen nicht nur hinsichtlich der inhaltlichen Gewichtung der Teilgebiete (gemäss Blueprint) konstant gehalten werden, sondern möglichst auch bezüglich Verteilung der Itemtypen. Wir empfehlen deshalb, auch für die Itemtypen einen ungefähren Verteilungsschlüssel zu definieren.
- Das Prüfungsheft muss eine standardisierende Beantwortungsanleitung enthalten, welche das sicherste und ökonomischste Vorgehen beschreibt. Sie muss die Kandidaten auch darüber informieren, wie die Antworten bewertet werden, insbesondere ob falsche Antworten bestraft werden oder ob sie in Zweifelsfällen raten sollen (vgl. Anhang 2, S.94).
- Im Prüfungsheft sollen die Items nach Typen gruppiert werden. Jede Typengruppe soll durch eine konzise Aufgabenbeschreibung eingeleitet werden (vgl. Anhang 3, S.95).
Es ist für Kandidaten in einer Prüfungssituation mühsam, stressfördernd und eine vom Fachwissen unabhängige Fehlerquelle, sich auf dauernd wechselnde Itemtypen einstellen zu müssen.
- Die Prüfung sollte mit 1-2 leichten Items beginnen (sog. Eisbrecher-items).
Jede Prüfung ist eine Stresssituation. Um wirklich die Wissensleistung messen zu können (und nicht z.B. Stressresistenz) ist ein beruhigender Prüfungseinstieg nützlich.
- Die Verteilung der richtigen Antworten auf die Positionen A-E sollte über die gesamte Prüfung etwa ausgeglichen sein. Zudem ist es sinnvoll, Folgen von fünf oder mehr Items mit der gleichen Richtiganwort zu vermeiden.
Die erste Massnahme verhindert den möglichen Einfluss gewisser Antworttendenzen von Kandidaten. (Speziell bekannt ist die Tendenz zur Mitte.) Die zweite Massnahme verhindert, dass Kandidaten angesichts einer langen Folge des gleichen Antwortbuchstabens verunsichert werden, konfuse Überlegungen anzustellen beginnen und evtl. richtige Antworten verschlimmbessern.

René Krebs

Verwendete Quellen

Neben unserer eigenen langjährigen Erfahrung in der Entwicklung, Revision und Auswertung von MC-Prüfungen im Rahmen der ärztlichen Aus- und Weiterbildung in der Schweiz, basieren die Hinweise in dieser Anleitung auf einer Reihe internationaler Publikationen zum Thema „Entwicklung von MC-Items“^(1,6,7,8,9). Besonders verweisen möchten wir auf das vorzügliche Manual von Case und Swanson¹, das über die Homepage des National Board of Medical Examiners, im Untermenü „publications“ heruntergeladen werden kann. (Adresse: <http://www.nbme.org>). Mit freundlicher Genehmigung der beiden Autoren haben wir diverse Hinweise und Itembeispiele sowie im Speziellen die Informationen über die Itemtypen R und PickN dieser Anleitung entnommen.

Literatur

- ¹ Case S, Swanson DB. Constructing Written Test Questions For the Basic and Clinical Sciences 2nd ed. Philadelphia: National Board of Medical Examiners, 1998
- ² Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine. Philadelphia: Lea & Febinger, 1961
- ³ Krebs R. The Swiss way to score multiple true-false items: theoretical and empirical evidence. In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg, eds. Advances in Medical Education. Dordrecht: Kluwer Academic Publishers, 1997:158-61
- ⁴ Case SM. The use of imprecise terms in examination questions: how frequent is frequently? Acad Med 1994;69(suppl):4-6
- ⁵ Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ 1998;32:343-50
- ⁶ Gronlund NE. How to construct achievement tests 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 1987
- ⁷ Linn RL, Gronlund NE. Measurement and assessment in teaching 8th rev. ed. Englewood Cliffs, NJ: Prentice-Hall, 1999
- ⁸ Ellsworth AR, Dunnell P, Duell OK. Multiple-choice test items: What are textbook authors telling teachers? J Educ Res 1990;83:289-93
- ⁹ Aiken LR. Testing with multiple-choice items. J Res Dev Educ 1987;20:44-58

Facharztprüfung XY 2000 MC-Examen „Fachkenntnisse“

durch Autor/in auszufüllen

Revisions-Nr.

Prüfungs-Nr.

richtige Antwort:..... Fragetyp:..... Blueprint 1:..... Blueprint 2:.....

Schlüsselworte:.....

Quelle, Referenz:.....

Frage:

Autor/in:.....

Formulierungs- kontrolle

- gut
 Korrektur
 ungeeignet

Revisions- entscheide

Bezug zum WB-Programm

- gehört zu WB-Progr.
 marginal
 kein WB-Inhalt

- Frage akzeptiert
Relevanz

inhaltliche Vorrevision

wissenschaftlich/ fachlich

- richtig, eindeutig
 fraglich, nicht eindeutig
 falsch

- Überarbeitung Autor
 essentiell mittel

Relevanz

- essentiell
 mittel
 gering

- abgelehnt
 gering

Allgemeine Hinweise zur korrekten Beantwortung von MC-Fragen

Die Kandidaten sollten zu Beginn der Prüfung darauf hingewiesen werden, die folgenden, z.B. auf die Rückseite des Prüfungsheftes gedruckten Hinweise aufmerksam zu lesen. Diese müssen natürlich der konkreten Situation (z.B. kein Typ K' verwendet, kein optisch lesbares Antwortblatt) angepasst werden.

1. Lesen Sie immer die ganze Frage und alle Wahlantworten sorgfältig durch.
2. Bezeichnen Sie Ihre Antworten zunächst im Fragenheft. Bei den Fragetypen mit Wahl der besten Antwort umkreisen Sie die eine Wahlantwort, die Sie für die zutreffende halten. Bei den Fragen vom Typ K' dagegen ist jede Antwort, die zutrifft, mit (+), jede Antwort, die nicht zutrifft, mit (-) zu bezeichnen.
3. Beantworten Sie alle Fragen. Jede richtig beantwortete Frage wird mit einem Punkt bewertet. Falsche Antworten werden nicht bestraft. Wahren Sie sich also die Chance, durch gezieltes Raten Punkte hinzuzugewinnen. Jede nicht beantwortete Frage wird wie eine falsch beantwortete mit 0 Punkten bewertet.
4. Übertragen Sie Ihre Antworten erst auf das Auswertungsblatt, nachdem Sie sich bei allen Fragen definitiv für eine Antwort entschieden und diese im Fragenheft bezeichnet haben. Bei den Fragetypen mit Wahl der besten Antwort darf immer nur ein Feld markiert werden. Für die Antworten auf die K'-Fragen sind die mit „K“ bezeichneten Felder des Auswertungsblattes auszufüllen. Für alle 4 Teilantworten ist hier je das Plus-Feld (+) oder das Minus-Feld (-) zu markieren.
5. Das Auswertungsblatt darf nur mit dem zur Verfügung gestellten Bleistift ausgefüllt werden. Markieren Sie bitte deutlich und exakt und radieren Sie möglichst wenig. Eine undeutliche Markierung oder zusätzliche Bleistiftstriche ausserhalb der vorgesehenen Felder können zu Fehlern beim optischen Lesen durch den Computer und dadurch zu Falschbewertungen führen.
6. Unterschreiben Sie das Fragenheft und das Auswertungsblatt im bezeichneten Feld. Sie bestätigen damit, dass Sie ohne unzulässige Hilfsmittel gearbeitet haben.

Instruktionen zur Beantwortung der einzelnen Typen

Im Prüfungsheft ist jeder Itemgruppe eines bestimmten Typs die entsprechende Beantwortungsinstruktion voranzustellen.

Bezeichnen Sie nur eine Wahlantwort durch Umkreisen des betreffenden Buchstabens:

- bei positiver Formulierung die einzig richtige respektive die am meisten zutreffende Antwort
- bei negativer Formulierung die einzig Ausnahme, die einzig falsche Antwort, resp. die Antwort mit dem am wenigsten zutreffenden Inhalt. (Die Negation ist kursiv und fett geschrieben.)

Typ A
Einfachauswahl

Auf fünf mit den Buchstaben (A) bis (E) bezeichnete Wahlantworten folgt eine Gruppe nummerierter Fragen oder Aussagen. Ordnen Sie jeder davon eine Wahlantwort zu, die einzig richtige resp. die am besten passende, und umkreisen Sie den entsprechenden Buchstaben. Ein und dieselbe Antwort kann dabei mehr als einmal die richtige sein.

Typ B
Zuordnungsaufgabe

Auf eine Liste von maximal 26 Wahlantworten, die alphabetisch geordnet und mit Buchstaben bezeichnet sind, folgt eine Gruppe nummerierter Fragen oder Aussagen. Ordnen Sie jeder davon eine Wahlantwort zu, die einzig richtige resp. die am besten passende, und schreiben Sie den entsprechenden Buchstaben unter die Fragenummer. Ein und dieselbe Antwort kann dabei mehr als einmal die richtige sein.

Typ R
Erweiterte Zuordnung

Auf eine Liste von maximal 26 Wahlantworten, die alphabetisch geordnet und mit Buchstaben bezeichnet sind, folgt eine oder folgen mehrere nummerierte Fragen oder Aussagen. Ordnen Sie jeder davon so viele Wahlantworten zu wie angegeben, und schreiben Sie die entsprechenden Buchstaben unter die Fragenummer. Ein und dieselbe Antwort kann dabei mehr als einmal die richtige sein.

Typ PickN
Mehrfachwahl bei erweiterter Zuordnung

Auf eine Frage oder unvollständige Aussage folgen vier Antworten oder Ergänzungen. Beurteilen Sie bei jeder davon, ob sie richtig oder falsch ist, und bezeichnen Sie sie entsprechend mit (+) oder (-). Unabhängig davon, ob die Frage grammatikalisch im Singular oder Plural formuliert ist, können 1, 2, 3, 4 oder auch gar keine der Antworten richtig sein. Die Lösung wird nur dann mit einem Punkt honoriert, wenn alle vier Antworten korrekt bezeichnet sind. (Alternative bei Halbpunktbewertung: Die korrekte Beurteilung aller 4 Antworten oder Ergänzungen wird mit einem ganzen Punkt honoriert. 3 richtige Beurteilungen erhalten einen halben Punkt.)

Typ Kprim
Vierfache Entscheidung richtig/falsch

Typ E
Kausale Verknüpfung

Umkreisen Sie den entsprechenden Buchstaben:

- (A) (+weil+) Beide Feststellungen stimmen, und die Weil-Verknüpfung ist korrekt.
- (B) (+ / +) Beide Feststellungen stimmen, die Weil-Verknüpfung ist falsch.
- (C) (+ / -) Die erste Feststellung stimmt, die zweite Feststellung ist falsch.
- (D) (- / +) Die erste Feststellung ist falsch, die zweite Feststellung stimmt.
- (E) (- / -) Beide Feststellungen sind falsch.

Typ RF?
Richtig/Falsch-
Entscheid mit „Weiss
nicht“-Option

Für jede richtige Ja- resp. Nein-Beurteilung erhalten Sie einen Punkt, für eine „Weiss nicht“-Antwort null Punkte. Für jede falsche Ja- resp. Nein-Beurteilung wird Ihnen ein Punkt abgezogen. Blindes Raten ist also nicht empfehlenswert. Für die Übertragung der Antworten auf den Antwortbogen gilt:

Ja = A

Nein = B

Weiss nicht = E

4

Wie wird
eine Prüfung mit
Kurzantwortfragen
entwickelt?

INHALT

- 4.1 Wie werden Prüfungsaufgaben ausgewählt? - 101
- 4.2 Wie werden Fall und Fragen formuliert? - 102
- 4.3 Welche Antworten werden erwartet? - 105
- 4.4 Wie bewerten, korrigieren und auswerten? - 108

Die Prüfungsmethode mit Kurzantwortfragen (KAF) ist mit ihren Vor- und Nachteilen in Kapitel 2 beschrieben. Zwei wichtige Vorteile sind, dass Kandidaten ihr Wissen aktiv formulieren müssen und dass beurteilt werden kann, wie Kandidaten Sachverhalte verstehen und wie sie ihre Kenntnisse interpretieren, gewichten oder konkret anwenden.

In diesem Kapitel erfahren Sie,

- dass es eine klar definierte Prüfungsaufgabe braucht, um Kurzantwortfragen formulieren zu können
- welche Tücken beim Formulieren von Prüfungsfällen und -fragen auf Sie warten
- warum vor Prüfungsbeginn ein ausdiskutierter und ausformulierter Antwortschlüssel vorliegen muss
- wie Kurzantwortfragen bewertet, korrigiert und ausgewertet werden

4.1 Wie werden Prüfungsaufgaben ausgewählt?

Wie der Inhalt einer Prüfungsaufgabe erarbeitet wird, ist in Kapitel 1 erläutert. Die wichtigen Arbeitsschritte für die Herstellung, Administration, Durchführung und Auswertung einer Prüfung mit Kurzantwortfragen sind als Checkliste im Anhang aufgeführt.

Ein Beispiel aus der schweizerischen Facharztprüfung Allgemeinmedizin soll die einzelnen Schritte nochmals kurz zusammenfassen:

1. Schlüsselkompetenz:

Wichtige Kenntnisse zur kardiovaskulären Reanimation

2. Blueprintdimensionen:

Allgemeinmedizin:	Therapie, Management, Entscheidungen
Fachwissen:	Innere Medizin
Alter:	Erwachsene
Konsultationsart:	Notfall

3. Schlüsselproblem und Schlüsselkriterien:

Wenn ein dem Facharzt unbekannter Erwachsener akut kollabiert, bei der Erstuntersuchung weder Herzaktionen noch Atmung aufweist und weite, lichtstarre Pupillen hat, weiss der Facharzt, dass

1. sofort eine manuelle kardiopulmonale Reanimation durchgeführt werden muss
2. möglichst Hilfen instruiert oder organisiert werden müssen (anwesende Personen, Ambulanz) für Lagerung, Beatmung, Herzmassage
3. bei Erfolglosigkeit möglichst eine Defibrillation erfolgen soll (200 Joule, mit eigenem Gerät oder einem der Ambulanz)
4. bei weiterer Erfolglosigkeit 1 mg Adrenalin i.v. appliziert und nochmals mit 360 Joule defibrilliert werden soll
5. im Todesfall die Angehörigen betreut werden müssen inkl. Organisation der Leichenpflege
6. es sich bei Ableben des Patienten um einen ausserordentlichen Todesfall mit Meldepflicht an das Untersuchungsrichteramt handelt

Welches Schlüsselproblem sollen der Fall und die Fragen prüfen?

4.2 Wie werden Fall und Fragen formuliert?

Fragenautoren haben die Aufgabe, Schlüsselprobleme und -kriterien in einen Fall aus der Praxis „einzubetten“. Wie das geschieht und wie Kurzantwortfragen in Facharztprüfungen schliesslich aussehen können, ist vielerorts beschrieben worden^{1,2,3,4,5,6}. Im Folgenden wird das eingangs erwähnte Schlüsselproblem modellhaft und schrittweise in einen Prüfungsfall mit entsprechenden Fragen umgesetzt. Häufige Fehler, die beim Formulieren auftreten, werden laufend besprochen und verbessert.

Erinnern Sie sich zuerst an reale Fälle.

Versuchen Sie vor dem Schreiben, sich aufgrund des Schlüsselproblems an konkrete Fälle oder Probleme aus der Praxis zu erinnern. Beachten Sie, dass es kaum einen realen Fall gibt, mit dem sämtliche Schlüsselkriterien geprüft werden können. Andererseits darf der Prüfungsfall, den Sie beschreiben wollen, nicht unrealistisch mit Kriterien vollgepackt werden.

Formulieren Sie jetzt den Fall.

- Fassen Sie sich möglichst kurz, schmücken Sie nicht unnötig aus.
- Beschreiben Sie das Problem oder die Fallgeschichte in einfachen Sätzen möglichst ohne Nebensätze. Es soll nicht das Sprachverständnis des Kandidaten geprüft werden.
- Verwenden Sie präzise definierte Ausdrücke. Benutzen Sie nur gebräuchliche Fremdwörter, Eigennamen, Fachausdrücke, Schlagwörter und Abkürzungen. Vermeiden Sie vage Quantifizierungen wie „gewöhnlich“, „üblich“, „häufig“, „oft“ oder „assoziiert mit“.
- Geben Sie bei Laboruntersuchungen Normwerte an.
- Illustrationen (Graphiken, Tabellen, Bilder, Röntgenbilder usw.) machen oft vieles klarer als Worte es vermögen.
- Wenn der Fall übersetzt werden muss, achten Sie besonders auf einfache Formulierungen und gebräuchliche, verständliche Bezeichnungen.

An einem Samstagmittag - Sie sind eben mit dem Auto in der Praxis angekommen und haben sich vorgenommen, endlich einmal Ihre Buchhaltung in Ordnung zu bringen - ruft eine Ihnen flüchtig bekannte Frau aus dem Nachbardorf völlig aufgeregt an. Ihr 44-jähriger Ehemann sei soeben beim Kaffee auf dem Sofa zusammengesunken, habe kurz gekrampft, atme nicht mehr und sei blau angelaufen.

Verwenden Sie keine unnötigen Ausschmückungen.

Der Einschub im ersten Satz und verschiedene andere Satzteile entsprechen zwar durchaus praxisrelevanten Situationen, sind aber nicht wesentlich, um das Problem lösen zu können:

Eine Frau ruft Sie in Ihrer Praxis völlig aufgeregt an. Ihr 44-jähriger Ehemann sei soeben beim Kaffee auf dem Sofa zusammengesunken, habe kurz gekrampft, atme nicht mehr und sei blau angelaufen.

1. Was tun Sie?

Wenn Sie die Fragen formulieren:

- Stellen Sie pro Prüfungsschritt nur eine präzise Frage, die möglichst kurze Antworten hervorruft.
- Achten Sie darauf, dass die Fragen unabhängig voneinander beantwortet werden können und dass keine versteckten Lösungshinweise („Cues“) vorliegen.
- Teilen Sie mit, was Sie erwarten und wie die Antworten erfolgen sollen („Begründen Sie stichwortartig...“, „Erklären Sie in einer Skizze...“).
- Geben Sie Anforderungen bekannt („Nennen Sie drei Ursachen ...“).
- Nennen Sie die maximal erreichbaren Punkte.

Auf die im Beispiel gestellte offene Frage gibt es viele richtige, aber absurde Antworten wie „erst einmal kurz durchatmen“. Die Erwartungen an den Kandidaten müssen präziser formuliert werden. Zudem sollte die erreichbare Punktzahl genannt werden. Sie signalisiert das Gewicht, das diese Frage innerhalb des Falles und auch im Vergleich zu anderen Fragen hat:

Stellen Sie präzise Fragen...

Eine Frau ruft Sie in Ihrer Praxis völlig aufgeregt an. Ihr 44-jähriger Ehemann sei soeben beim Kaffee auf dem Sofa zusammengesunken, habe kurz gekrampft, atme nicht mehr und sei blau angelaufen.

1. Nennen Sie zwei Massnahmen, die Sie am Telefon veranlassen. Punkte (2)

10 Minuten später - nachdem Sie die Ambulanz angefordert haben - treffen Sie beim Patienten ein und finden diesen leblos, cyanotisch, ohne Herzaktion, ohne Atmung, mit weiten und lichtstarrten Pupillen.

2. Was tun Sie jetzt? Punkte (3)

Mit dem Folgetext wird eine mögliche Antwort auf Frage 1 geliefert (Ambulanz anfordern). Die 2. Frage muss wieder präzisiert werden:

... und vermeiden Sie ungewollte Lösungshinweise.

10 Minuten später treffen Sie beim Patienten ein und finden diesen leblos, cyanotisch, ohne Herzaktion, ohne Atmung, mit weiten und lichtstarrten Pupillen.

2. Was tun Sie jetzt? Listen Sie 3 therapeutische Massnahmen auf, die Sie unternehmen. Punkte (3)

Trotz Ihren Massnahmen stirbt der Patient.

3. Was sagen Sie der zurückgebliebenen Ehefrau und welche Massnahmen treffen Sie jetzt?

*Stellen Sie keine
Doppelfragen.*

Frage 3 ist eine Doppelfrage. Richtig sollte es heissen:

3. *Was sagen Sie der zurückgebliebenen Ehefrau?
Nennen Sie mindestens zwei Dinge.*

Punkte (2)

4. *Welche zwei Massnahmen treffen Sie noch?*

Punkte (2)

*Pro Fall sind mehrere
Schlüsselprobleme
möglich.*

Ein Fall kann mehrere Schlüsselkompetenzen oder -probleme prüfen. Achten Sie auch in diesem Zusammenhang auf versteckte Lösungshinweise. Im vorliegenden Beispiel könnte folgende Schlüsselkompetenz eingefügt werden:

1. *Schlüsselkompetenz: adäquate Abklärung kardiovaskulärer Risikofaktoren*

2. *Blueprintdimensionen:*

Allgemeinmedizin: Therapie, Management, Entscheidungen

Fachwissen: Innere Medizin

Alter: Erwachsene

Konsultationsart: Sprechstunde

3. *Schlüsselproblem und Schlüsselkriterien:*

Wenn sich ein akuter kardialer Todesfall ereignet (Patient im Erwachsenenalter), führt der Hausarzt bei den Geschwistern des Patienten sinnvollerweise durch:

1. *Familienanamnese*

2. *Anamnese zu Gewohnheiten wie Rauchen, Aktivität, Stress, Essen/Gewicht*

3. *körperliche Untersuchung: Blutdruck, Gewicht, Herzauskultation*

4. *Laboruntersuchung: Blutzucker, Lipidstatus*

5. *Ruhe-EKG (primär nicht Ergometrie)*

6. *evtl. Thoraxröntgen, Echokardiographie*

Die weitere Entwicklung des Falles könnte so aussehen:

Später erfahren Sie Folgendes: Der verstorbene Mann war kerngesund, als Angehöriger des Feuerwehripiketts musste er sich vor wenigen Monaten untersuchen lassen, alles war o.k. Er war gut trainiert und ging fast täglich joggen. Hingegen war er ein starker Raucher (ca. 30 Zigaretten pro Tag) und stand am Arbeitsort unter Stress. Ein älterer Bruder des Patienten kam vor Jahren als 33-Jähriger genau gleich ums Leben. Eine Tante war Klosterfrau und starb mit 63 Jahren an einem 3. Herzinfarkt. Auch der Vater des Patienten verschied infolge eines Herzschlages, allerdings erst, als er 80 Jahre alt war. Jetzt leben noch ein 10 Jahre älterer Bruder des Patienten, ferner eine 11 Jahre ältere Schwester (übergewichtig, Raucherin) und eine 50-jährige Schwester. Bei keinem der drei Geschwister bestehen Herzsymptome.

5. *Welche Abklärungen der drei noch lebenden Geschwister erachten Sie für sinnvoll? Nennen Sie 3.*

Punkte (3)

4.3 Welche Antworten werden erwartet?

Nachdem der Fall formuliert ist, wird er von den Prüfungsverantwortlichen revidiert. Sie verfassen einen Schlüssel, der möglichst alle richtigen Antworten enthält. Diese müssen so formuliert sein, dass später der Korrektor genau weiss, wofür er Punkte geben kann.

Antwortschlüssel erstellen

Richtige Antworten auf Frage 5:

- a) *Risikofaktoren erheben: BD und Rauchen und Gewicht*
- b) *Laborkontrollen: Lipidstatus und Blutzucker*
- c) *Herzauskultation (Geräusche?)*
- d) *Ruhe-EKG (primär nicht Ergometrie)*
- e) *Thoraxröntgen (Herzgrösse)*
- f) *Ultraschallabklärung*

Es genügt z.B. nicht, wenn ein Kandidat schreibt: „Risikofaktoren erheben“, es braucht die Angaben „BD“, „Rauchen“ und „Gewicht“.

Eine für die Fachgesellschaft repräsentative Expertengruppe, die an der Entwicklung der Prüfungsaufgaben nicht beteiligt war, testet die Fragen und den Antwortschlüssel. Es ist auch möglich, die Fragen in einer Facharztprüfung zu testen, ohne dass sie für die Prüfung bewertet werden. Antwortschlüssel können dann wenn nötig ergänzt und unklare Formulierungen verbessert werden. Wenn sich die Fälle und Fragen bewährt haben, kommen sie in den definitiven Prüfungsfragenpool. Der Antwortschlüssel wird nun nicht mehr verändert.

...und diesen testen

In unserem Beispiel sieht der gesamte Antwortschlüssel folgendermassen aus:

1. Nennen Sie 2 Massnahmen, die Sie am Telefon veranlassen. Punkte (2)

Richtige Antworten:

- a) *Instruktion der Ehefrau, was sofort zu machen ist: Lagerung, Beatmung, Herzmassage*
- b) *Ambulanz organisieren*
- c) *Notfallbesuch*

2. Was tun Sie jetzt? Listen Sie 3 therapeutische Massnahmen auf, die Sie unternehmen. Punkte (3)

Richtige Antworten:

- a) *Kardiopulmonale Reanimation CPR mit Beatmung und Herzmassage*
- b) *Falls kein Erfolg: Defibrillation mit eigenem Gerät oder Gerät der Ambulanz (200 Joule)*
- c) *Erneut CPR*
- d) *Falls kein Erfolg: Adrenalin 1 mg i.v. und nochmals Defibrillation mit 360 Joule*

3. Was sagen Sie der zurückgebliebenen Ehefrau? Nennen Sie 2 Dinge. Punkte (2)

Richtige Antworten:

- a) Kondolieren
- b) akutes Herzversagen sehr wahrscheinlich
- c) Autopsie ansprechen

4. Welche 2 Massnahmen treffen Sie noch? Punkte (2)

Richtige Antworten:

- a) wenn möglich Autopsie veranlassen
- b) Information Leichenpflege oder Amtsanzeige oder Versicherungsabmeldung oder Meldung an Pfarrer oder Information der Angehörigen
- c) ausserordentlicher Todesfall, Anzeigepflicht beim Untersuchungsrichter

5. Welche Abklärungen der drei noch lebenden Geschwister erachten Sie für sinnvoll? Nennen Sie 3. Punkte (3)

Richtige Antworten:

- a) Risikofaktoren erheben: BD und Rauchen und Gewicht
- b) Laborkontrollen: Lipidstatus und Blutzucker
- c) Herzauskultation (Geräusche?)
- d) Ruhe-EKG (primär nicht Ergometrie)
- e) Thoraxröntgen (Herzgrösse)
- f) Ultraschallabklärung

Instruieren Sie die Kandidaten und lassen Sie sie üben.

Die Kandidaten werden vor dem Examen schriftlich über die Prüfungsmethode und über wichtige Prinzipien bei der Beantwortung von Kurzantwortfragen informiert^{7,8}. Sie sollten die Gelegenheit haben, mit Beispielen zu üben. Auch im Prüfungsheft⁹ sind Instruktionen festzuhalten wie:

- Gehen Sie von der Annahme aus, Sie führten bereits Ihre Praxis.
- Lesen Sie jede Fallbeschreibung und die dazugehörigen Fragen sorgfältig durch. In Klammern finden Sie oft zusätzliche spezielle Antwortanweisungen.
- Teilen Sie Ihre Zeit gut ein. Am besten gehen Sie die Fragen zügig durch, ohne zu lange bei einer einzelnen zu verweilen.
- Am rechten Rand sind die Punktzahlen angegeben, die Sie für die korrekte Beantwortung der jeweiligen Frage maximal erhalten.
- Antworten Sie mit Einzelworten oder kurzen Sätzen nur auf die gestellte Frage.
- Wenn nach „Abklärung“, „Abklärungsschritten“, „diagnostischem oder therapeutischem Vorgehen“ oder nach „Massnahmen“ gefragt wird, genügen Globalantworten wie z.B. „Anamnese“, „Status“ oder „Labor“ nicht. Es braucht dafür Angaben wie z.B. „Frage nach Gewicht“, „Herzauskultation“ oder „Hämoglobinbestimmung“.
- Schreiben Sie in gut lesbarer Schrift. Unleserliches erhält keine Punkte.
- Liegen Ihrer Meinung nach divergierende Lehrmeinungen, missverständliche Formulierungen oder andere substantielle Mängel der Fragen vor, notieren Sie dies auf dem Kommentarblatt am Schluss des Prüfungsheftes. Solche Bemerkungen werden bei der Auswertung der Prüfung berücksichtigt.

Eine Frau ruft Sie in Ihrer Praxis völlig aufgeregt an. Ihr 44-jähriger Ehemann sei soeben beim Kaffee auf dem Sofa zusammengesunken, habe kurz gekrampft, atme nicht mehr und sei blau angelaufen.

So sieht das ganze Fallbeispiel im Prüfungsheft aus.

1. Nennen Sie 2 Massnahmen, die Sie am Telefon veranlassen. Punkte (2)

1. _____
2. _____

10 Minuten später treffen Sie beim Patienten ein und finden diesen leblos, cyanotisch, ohne Herzaktion, ohne Atmung, mit weiten und lichtstarrten Pupillen.

2. Was tun Sie jetzt? Listen Sie 3 therapeutische Massnahmen auf, die Sie unternehmen. Punkte (3)

1. _____
2. _____
3. _____

Trotz Ihren Massnahmen stirbt der Patient.

3. Was sagen Sie der zurückgebliebenen Ehefrau? Nennen Sie 2 Dinge. Punkte (2)

1. _____
2. _____

4. Welche 2 Massnahmen treffen Sie noch? Punkte (2)

1. _____
2. _____

Später erfahren Sie noch Folgendes: Der verstorbene Mann war kerngesund, als Angehöriger des Feuerwehripiketts musste er sich vor wenigen Monaten untersuchen lassen, alles war o.k. Er war gut trainiert und ging fast täglich joggen. Hingegen war er ein starker Raucher (ca. 30 Zigaretten pro Tag) und stand am Arbeitsort unter Stress. Ein älterer Bruder des Patienten kam vor Jahren als 33-Jähriger genau gleich ums Leben. Eine Tante war Klosterfrau und starb mit 63 Jahren an einem 3. Herzinfarkt. Auch der Vater des Patienten verschied infolge eines Herzschlages, allerdings erst als er 80 Jahr alt war. Jetzt leben noch ein 10 Jahre älterer Bruder des Patienten, ferner eine 11 Jahre ältere Schwester (übergewichtig, Raucherin) und eine 50-jährige Schwester. Bei keinem der drei Geschwister bestehen Herzsymptome.

5. Welche Abklärungen der drei noch lebenden Geschwister erachten Sie für sinnvoll? Nennen Sie 3. Punkte (3)

1. _____
2. _____
3. _____

4.4 Wie bewerten, korrigieren und auswerten?

So kann der Fall bewertet werden

Nach dem Prinzip eines inhaltsorientierten Standardsetzungsverfahrens (vgl. Kapitel 8) bewertet eine Expertengruppe jede Frage nach folgenden Anweisungen:

1. Stellen Sie sich zuerst die Gruppe von Kandidaten vor, die in der Gesamtprüfung an der Grenze zwischen bestanden und nicht bestanden anzusiedeln ist („Borderline-Gruppe“).
2. Schätzen Sie dann bei jeder Frage, wie viele Punkte diese „Borderline-Gruppe“ im Durchschnitt erreichen wird (Zehntelpunktschätzung).

Die Schätzwerte aller Experten werden gemittelt resp. in einem Konsensprozess der Gruppe definitiv festgelegt. Die Summe der definitiven Schätzwerte aller Prüfungsfragen ergibt dann die inhaltsorientierte Bestehensgrenze. In unserem Beispiel könnte das so aussehen:

1. Nennen Sie 2 Massnahmen, die Sie am Telefon veranlassen. Punkte (2)
1.5

- a) Instruktion der Ehefrau, was sofort zu machen ist:
Lagerung, Beatmung, Herzmassage
- b) Ambulanz organisieren
- c) Notfallbesuch

2. Was tun Sie jetzt? Listen Sie 3 therapeutische Massnahmen auf, die Sie unternehmen. Punkte (3)
1.9

- a) Cardiopulmonale Reanimation CPR mit Beatmung und Herzmassage
- b) Falls kein Erfolg: Defibrillation mit eigenem Gerät oder Gerät der Ambulanz (200 Joule)
- c) Erneut CPR
- d) Falls kein Erfolg: Adrenalin 1 mg i.v. und nochmals Defibrillation mit 360 Joule

3. Was sagen Sie der zurückgebliebenen Ehefrau? Nennen Sie 2 Dinge. Punkte (2)
1.2

- a) Kondolieren
- b) akutes Herzversagen sehr wahrscheinlich
- c) Autopsie ansprechen

4. Welche 2 Massnahmen treffen Sie noch? Punkte (2)
1.1

- a) wenn möglich Autopsie veranlassen
- b) Information Leichenpflege oder Amtsanzeige oder Versicherungsabmeldung oder Meldung an Pfarrer oder Information der Angehörigen
- c) ausserordentlicher Todesfall, d.h. Anzeigepflicht bei UR

5. Welche Abklärungen bei den drei noch lebenden Geschwistern erachten Sie für sinnvoll? Nennen Sie 3.

Punkte (3)

2.1

- a) Risikofaktoren erheben: BD und Rauchen und Gewicht
- b) Laborkontrollen: Lipidstatus und Blutzucker
- c) Herzauskultation (Geräusche?)
- d) Ruhe-EKG (primär nicht Ergometrie)
- e) Thoraxröntgen (Herzgrösse)

Die Prüfungshefte der Kandidaten und die Antwortschlüssel können den Korrektoren zugeschickt werden. Wenn das aus logistischen Gründen oder wegen grossen Kandidatenzahlen nicht möglich ist, treffen sich die Korrektoren „am grossen Tisch“: Ein Korrektor bewertet bei allen Kandidaten einen Fall oder eine Frage. Jedes korrigierte Heft reicht er seinem Kollegen weiter, bis es die Runde gemacht hat. Die Antworten werden dabei auf ein wenn möglich computerlesbares Erfassungsblatt übertragen.

Wie korrigieren?

Während der Prüfung darf der Antwortschlüssel nicht mehr verändert werden. Es soll nicht vorkommen, dass die Meinung weniger Korrektoren den Validierungsprozess unterläuft, den die Prüfungsaufgaben durchgemacht haben. Auch logistische Gründe verlangen einen konstanten Antwortschlüssel. Würde dieser während der Korrektur geändert, müssten sämtliche bis zum entsprechenden Zeitpunkt bereits korrigierten Prüfungshefte erneut bewertet werden. Die Korrektoren sind aber aufgefordert, Auffälliges auf einem Kommentarblatt schriftlich festzuhalten. Die Prüfungskommission kann dann diese Kommentare bei der Bewertung ebenso berücksichtigen wie die Bemerkungen, welche die Kandidaten während dem Examen im Prüfungsheft angebracht haben.

*Antwortschlüssel
nicht mehr abändern*

Die Korrektoren sind gut zu instruieren. Sie können z.B. folgende Anweisungen erhalten:

*Instruktionen für die
Korrektoren*

- Lesen Sie die Antworten durch und beurteilen Sie nach dem Antwortschlüssel.
- Markieren Sie alle richtigen Antworten einer Frage zuerst im Prüfungsheft.
- Es gibt nur Punkte für Antworten, die im Antwortschlüssel vorkommen. Über evtl. weitere richtige Antworten eines Kandidaten wird nicht mehr diskutiert.
- Bei Unsicherheiten oder Fragen wenden Sie sich an den Prüfungsvorsitzenden.
- Auffälligkeiten zu einzelnen Fragen notieren Sie bitte auf das beiliegende Kommentarblatt.
- Markieren Sie alle richtigen Antworten einer Frage unabhängig von der maximal möglichen Punktzahl.
- Übertragen Sie die richtigen Antworten auf das Erfassungsblatt.
- Wenn Sie bei einem Kandidaten mit der Korrektur fertig sind, reichen Sie das Prüfungsheft zusammen mit dem Erfassungsblatt Ihrem Korrektornachbarn weiter.

Wie lange dauert die Korrektur?

Die Korrekturdauer ist vor allem von der Kandidatenzahl, der Fragenform und der Qualität des Antwortschlüssels abhängig. Offenere Fragen, die viele und ausführliche Antworten erfordern, verlängern die Korrekturzeit ebenso wie z.B. wenig präzierte Antwortschlüssel. Bei unserem Prüfungsbeispiel brauchte der Korrektor zwei bis drei Minuten um den ganzen Fall bei einem Kandidaten zu bewerten.

Im Rahmen des OSCE des „qualifying examination part II“ in Kanada hatte es 1998 sieben Fragestationen mit Kurzantwortfragen gegeben. 22 Personen korrigierten die Antwortblätter der 1800 Kandidaten zentral an einem Wochenende. Das ergibt umgerechnet auf ein Antwortblatt eines Kandidaten pro OSCE-Station eine Korrekturzeit von etwa zwei Minuten.

erste Auswertung

Eine erste Auswertung einer Prüfung mit Kurzantwortfragen

- analysiert die einzelnen Prüfungsaufgaben (Schwierigkeit und Trennschärfe der Fälle, Fragen und Antworten unter Berücksichtigung der maximal möglichen Punkte pro Frage)
- beschreibt die Messeigenschaften der Prüfung
- erfasst alle Kommentare der Beteiligten (Korrektoren, Kandidaten, Prüfungsaufsicht)
- schlägt vor, welche Fragen aufgrund der ersten Auswertung bei der definitiven Bewertung nicht berücksichtigt werden sollen

Die Prüfungskommission muss nun die definitiven Fragen und Antworten bezeichnen, mit denen die Kandidaten beurteilt werden sollen.

Die Prüfungskommission entscheidet.

Eine zweite Auswertung der Prüfung

- errechnet die definitiven Kandidatenleistungen
- stellt die Kandidatenleistungen gesamthaft dar
- berechnet die Bestehensgrenze aufgrund des gewählten Standardsetzungsverfahrens

Eine unabhängige Expertengruppe überprüft die Auswertung, im speziellen diejenige der „Grenzkandidaten“. Die Prüfungskommission bestimmt nun aufgrund aller Auswertungsschritte diejenigen Kandidaten, welche die Prüfung nicht bestanden haben und diskutiert Härtefälle. Die inhaltsorientierte Bestehensgrenze gerät dabei oft unter Druck. Was es in dieser Situation alles zu bedenken gibt, ist in Kapitel 8 ausgeführt.

Peter Schläppi

Literatur

- ¹ Knox JDE. How to...use modified essay questions. *Med Teach* 1980;2:20-4
- ² Irvine WG, Bamber JH. The cognitive structure of the modified essay question. *Med Educ* 1983;16:326-31
- ³ Webber RH. Structured short-answer questions: an alternative examination method. *Med Educ* 1992;26:58-62
- ⁴ Preparatory course booklet on the MICGP examination. The Irish College of General Practitioners, 1997
- ⁵ Guide to the certification examination, The College of Family Physicians of Canada, Mississauga, Ontario, 1998
- ⁶ Neighbour R. The MRCGP modified essay questions. *Persönliche Mitteilung*, 1996
- ⁷ Holden NL. Answering the short-answer question paper. *Br J Hosp Med* 1994;51:44-6
- ⁸ Trigwell P, Williams C, Yeomans D. MRCPsych examination technique: the short-answer question paper. *Br J Hosp Med* 1996;55:135-8
- ⁹ Prüfungsheft Facharztprüfung Allgemeinmedizin. Schweizerische Gesellschaft für Allgemeinmedizin SGAM, 1999

5

Wie wird
eine strukturierte
mündliche Prüfung
entwickelt?

INHALT

- 5.1 Welche Fragen - welche Antworten? - 117
- 5.2 Wie werden Prüfungsunterlagen hergestellt? - 121

Wenn Sie eine strukturierte mündliche Prüfung (Structured Oral Examination, SOE) erstellen, müssen Sie sich im Wesentlichen das Gleiche überlegen wie bei der Herstellung von Kurzantwortfragen (vgl. Kapitel 4). Zuerst muss definiert werden, was genau mit den Aufgaben geprüft werden soll. Als nächstes müssen die erwarteten Antworten eindeutig formuliert werden. Schliesslich ist festzulegen, wie die Aufgaben dem Kandidaten gestellt werden.

Der wesentliche Unterschied zu Kurzantwortfragen besteht darin, dass Prüfer und Prüfling direkt miteinander interagieren. Dieser Kontakt kann die Prüfungssituation und die Beurteilung der Kandidatenleistung negativ beeinflussen, d.h. die Objektivität, Reliabilität und damit auch die Validität des Examens können gefährdet werden. Dagegen hilft nur eine entsprechende Strukturierung und Standardisierung sowohl der Prüfungssituation als auch der Prüfungsaufgaben. Der optimale Strukturierungs- und Standardisierungsgrad muss für jede Prüfung einzeln bestimmt werden. Jedes Examen ist daher in diesem Zusammenhang kontinuierlich zu evaluieren.

Dieses Kapitel ist in zwei Abschnitte gegliedert. Im ersten geht es um die Zusammenstellung geeigneter Prüfungsaufgaben, im zweiten um die Herstellung strukturierender Examensunterlagen.

In diesem Kapitel erfahren Sie,

- was die Aufgaben prüfen sollen
- welche Antworten in Frage kommen
- welche Fragetypen zur Verfügung stehen
- wie Unterlagen für den Kandidaten aussehen
- wie ein Befragungs- und Bewertungsraster für den Experten erstellt wird
- wie ein Punkte- bzw. Notenblatt aussieht

5.1 Welche Fragen – welche Antworten?

Die Wahl einer Prüfungsaufgabe und das Formulieren der entsprechenden Fragen orientiert sich an den Vorgaben im jeweiligen Fach. Das gewichtete Inhaltsverzeichnis (Blueprint) beschreibt, wie die Bereiche in der Prüfung vorkommen müssen und wie sie gewichtet werden. (Weitere Ausführungen zu diesem Thema können in Kapitel 1 nachgelesen werden.)

Was sollen die Aufgaben prüfen?

Strukturierte mündliche Prüfungen können einen rein theoretischen Inhalt haben oder sich auf eine vorher erfolgte Untersuchung eines Patienten beziehen. Sie finden in Form einer Befragung, einer Diskussion, einer Fallvorstellung oder einer Mischung dieser Formen statt. Dieser Rahmen wird immer vor der Prüfung festgelegt. Die einzelnen Fragen werden je nach Strukturierungs- und Standardisierungsgrad (vgl. Kapitel 2, S.43) der mündlichen Prüfung nur stichwortartig festgehalten oder vollständig ausformuliert.

Wie werden die Aufgaben präsentiert?

Der Strukturierungsgrad einer mündlichen Prüfung kann auf einem Kontinuum von wenig bis hoch angesiedelt werden:

Es wird lediglich festgelegt, zu welchen Blueprintkategorien die Kandidaten befragt werden. Wie diese Vorgaben in konkrete Prüfungsaufgaben umgesetzt werden, bleibt den Examinatoren (Experten) überlassen. Ebenso fehlen Richtlinien, welche Antworten als gültig anerkannt werden und welche nicht. Dieser Entscheid bleibt ebenfalls im Ermessen der Experten.

Wenig strukturiert bzw. standardisiert heisst...

Eine so strukturierte Prüfung ist weder reliabel, noch objektiv, noch valide. Die Reliabilität ist darum nicht gewährleistet, weil mit grosser Wahrscheinlichkeit jeder Experte andere Prüfungsaufgaben aus den vorgegebenen Blueprintkategorien entwickelt und damit die Gefahr der Fallspezifität in Kauf nimmt. Objektiv kann ein wenig strukturiertes mündliches Examen deshalb nicht sein, weil die erwarteten Antworten ad hoc aufgrund der Befragung zustande kommen. Verschiedene Examinatoren werden daher nur selten zu einer gleich differenzierten Beurteilung der Kandidatenleistung kommen. Diese wird eher auf einem globalen Eindruck beruhen. Damit ist auch die Validität der Prüfung gefährdet.

...weniger reliabel, objektiv und valide...

Zu jeder Blueprintkategorie wird genau definiert, was gefragt werden soll. Ebenso vorgegeben und für jeden Kandidaten gleich ist die Ausgangslage für die Befragung (Fallbeschreibung, Röntgenbild etc.). Die einzelnen Fragen sind ausformuliert. Ferner ist genau bestimmt, welche Antworten erwartet werden, bzw. gültig sind. Das Ermessen des Examinators ist damit stark eingeschränkt. Das Expertenwissen fliesst in die Prüfungsvor-

...als hoch strukturiert bzw. standardisiert.

bereitung ein, bei der Definition der Aufgaben und der richtigen Antworten. Ein Examinatorenttraining stellt sicher, dass jeder Beurteiler die Bewertungskategorien gleich interpretiert und die vom Kandidaten gegebenen Antworten gleich bewertet. Eine so strukturierte mündliche Prüfung ist deutlich objektiver, reliabler und valider. Ein Beispiel einer entsprechenden Prüfungsaufgabe ist in Abb. 5.2 (S.122) gegeben.

*reibungsloser
Prüfungsablauf*

Ein strukturiertes und standardisiertes Verfahren stellt im Weiteren sicher, dass die Prüfung reibungslos abläuft. Da alle Prüflinge die gleichen Aufgaben erhalten, muss ausgeschlossen werden können, dass bereits geprüfte Kandidaten mit noch nicht geprüften Kontakt haben.

Welche Antworten werden erwartet?

Je eindeutiger mögliche Antworten vor der Prüfung formuliert worden sind, desto objektiver (verschiedene Experten kommen zur gleichen Beurteilung) und zuverlässiger ist ein Examen. Wie Sie zu prüfungstauglich ausformulierten Antwortschlüsseln kommen, ist in Kapitel 4 beschrieben. Analoges gilt für die strukturierte mündliche Prüfung:

Die Prüfungsverantwortlichen formulieren für jede Frage einer Aufgabe einen Schlüssel, der möglichst alle richtigen Antworten enthält. Ebenso bestimmen sie, wie viele korrekte Antworten ein Kandidat geben muss, um eine bestimmte Punktzahl zu erhalten. Anschliessend testet eine für die Fachgesellschaft repräsentative und an der Entwicklung der Prüfungsaufgabe nicht beteiligte Expertengruppe die Fragen und den Antwortschlüssel.

Welche Fragetypen stehen zur Verfügung?

Es gibt unterschiedliche Typen von Fragen, die in verschiedenen Situationen hilfreich sein können. Generell gilt: Fragen sollen klar und verständlich gestellt werden. Bei den zum voraus definierten und schriftlich festgehaltenen Fragen können folgende Typen unterschieden werden:

- Eröffnungsfragen
- Konsultationsfragen
- Handlungsfragen

Bei einem strukturierten mündlichen Examen interagieren Prüfer und Prüfling stark. Deshalb lassen sich zusätzlich die beiden folgenden Fragetypen, die nicht im voraus definiert werden können, anwenden:

- kurze Zwischenfragen
- punktuelle Kurzfragen

Solche Fragen dienen dazu, Unklarheiten zu beseitigen. Sie sollten gezielt und sparsam angewendet werden, da sie unerwünschte Nebeneffekte haben können.

Normalerweise beginnt jede Prüfungsaufgabe mit einer Eröffnungsfrage. Diese sollte eher einfach zu beantworten sein und dem Kandidaten helfen, sich an die Prüfungssituation zu gewöhnen. Die Eröffnungsfrage entscheidet nicht über die Leistung des Kandidaten.

am Anfang die Eröffnungsfrage

- *Gegen welche Krankheiten soll gemäss nationalem Impfplan grundsätzlich jede Person immunisiert werden?*
- *Welche HIV-Therapien empfiehlt die Eidgenössische Kommission für Aidsfragen?*
- *Welches sind die Kontraindikationen für die Verwendung von Thiopenton?*
- *Was bedeutet „nüchtern“ in einer pädiatrischen Verordnung?*

Konsultationsfragen beziehen sich auf ein klinisches Problem und stellen die beraterische Funktion des Arztes in den Mittelpunkt. Sie lassen Zeit für eine Diskussion, die der Prüfer nach eigenem Ermessen und gemäss den zu prüfenden Inhalten vorantreiben kann.

Konsultationsfragen ermöglichen die Diskussion.

- *Ein schwerer Asthmatiker tritt ein für eine Cholezystektomie. Es handelt sich um eine Wahloperation.*
 1. *Wie klären Sie diesen Patienten vorgängig ab?*
 2. *Wie betreuen sie ihn?*
- *Während der Operation hat der Patient erhöhten inspiratorischen Druck und einseitigen Lufteintritt. Die Sauerstoffsättigung beträgt 89 %. Was tun Sie?*

Handlungsfragen beziehen sich auf eine Reihe von Entscheidungen und Handlungen in einem bestimmten klinischen Umfeld. Die Antworten sollen ohne grössere Diskussion erfolgen.

Handlungsfragen, um ein klinisches Szenario schnell durchzuarbeiten

- *Sie werden dringend in den Gebärsaal gerufen zu einer Mutter mit ausgedehnter Blutung ante partum. Die Frau befindet sich im Schockzustand, die fetalen Herztöne sind schwach.*
Was tun Sie als erstes?
- *Ein Patient hat vor zwei Tagen einen Myokardinfarkt gehabt und ist noch auf der Intensivstation. Sie werden zu ihm gerufen zwecks Kardioversion. Gerade jetzt, eine Stunde nach dem Mittagessen, hat er eine ventrikuläre Tachykardie bekommen.*
Was veranlassen Sie sofort?

Zwischenfragen, um dem Kandidaten auf die Sprünge zu helfen

Kurze Zwischenfragen bzw. Zwischenbemerkungen liefern dem Kandidaten mit Schwierigkeiten Informationen, die ihm bei der Beantwortung der ursprünglichen Frage (z.B. einer Handlungsfrage) einen Schritt weiter helfen. Solche Fragen sollten vorsichtig verwendet werden, da sie unter Umständen den Gedankengang des Kandidaten unterbrechen. Die Prüfung kann dann den Charakter eines Kreuzverhörs erhalten.

punktueller Kurzfragen, um die Leistung besser beurteilen zu können

Auch punktuelle Kurzfragen werden dann eingesetzt, wenn ein Kandidat Schwierigkeiten zeigt. Sie stecken ein Problem mosaikartig ab und ermöglichen dem Examinator, besser zu beurteilen, in welcher Tiefe ein Kandidat die Prüfungsaufgabe verstanden hat.

Wechselwirkungen im Auge behalten

Kandidat und Prüfer stehen in Wechselwirkung. Die beiden können sich gegenseitig beeinflussen, beispielsweise durch ihr Auftreten oder durch ihre Fähigkeit zu formulieren. Seien Sie sich als Prüfer bewusst, dass auch Ihre nonverbalen Verhaltensweisen die Leistungsfähigkeit des Kandidaten beeinflussen können. Bereits subtile verbale oder nonverbale Hinweise vermögen den Kandidaten zu motivieren bzw. zu demotivieren. Zur eigenen Kontrolle und zur Verbesserung der Prüfungsqualität empfiehlt es sich, die Prüfung auf Video aufzunehmen. Diese Aufnahmen können später analysiert und für weitere Examinatorentrainings verwendet werden.

5.2 Wie werden Prüfungsunterlagen hergestellt?

Vor der Prüfung müssen drei Instrumente bereit stehen.

1. Unterlagen für den Kandidaten (Fallbeschreibungen, Röntgenbilder, Szenarien usw.)
2. ein Befragungs- und Bewertungsraster für den Examinator, das als Protokoll der Prüfung dient
3. ein Punkteblatt für die Bewertung der einzelnen zu prüfenden Aspekte, das als Beleg für die weitere Auswertung der Prüfung dient

Der Kandidat soll genau wissen, was von ihm verlangt wird und wieviel Zeit ihm pro Aufgabe zur Verfügung steht. Diese Informationen sollte er zu Beginn der Prüfung bzw. vor jeder neuen Aufgabe schriftlich erhalten (vgl. Abb. 5.1). Daraus ergeben sich folgende Vorteile:

Warum und wie weise ich Kandidaten an?

- Alle Kandidaten erhalten die gleichen Informationen. Subtile Hinweise auf erwartete Antworten, wie sie beim mündlichen Formulieren der Aufgaben entstehen können, werden minimiert.
- Im Rekursfall kann belegt werden, was vom Kandidaten verlangt wurde.

Facharztprüfung Rheumatologie

Prüfungsaufgabe: Sie sind als Rheumatologe in einer Praxis tätig. Nun wird Ihnen eine beschwerdefreie 46-jährige Patientin zugewiesen, bei der vor einem Jahr als Zufallsbefund beim Hausarzt ein Antistreptolysin-Titer um 1000 i.E./ml festgestellt wurde. Eine probatorische Penicillin-Therapie hatte keinen Einfluss auf den Antistreptolysin-Titer. Weitere Kontrollen zeigen Werte jeweils zwischen 500 und 700 i.E./ml.

Sagen Sie uns, wann Sie die Aufgabe durchgelesen haben. Wir werden Sie während 10 Minuten befragen.

Abb. 5.1: Unterlage für Kandidaten (adaptiertes Beispiel aus der schweizerischen Facharztprüfung Rheumatologie)

Im integrierten Befragungs- und Bewertungsraster sind die konkreten Fragen und Antworten einer Prüfungsaufgabe aufgeführt (vgl. Abb. 5.2). Der Raster dient den Examinatoren zur Protokollführung und als Grundlage zur Kandidatenbewertung. Es werden alle Punkte aufgeführt, die bewertet werden sollen, letztlich also alle erwarteten Antworten. Der Examinator kann angeben, ob der Kandidat die Fragen korrekt beantwortet hat. Je nach dem können die einzelnen Antworten gewichtet werden. Der Raster sollte zudem Gelegenheit bieten, weitere Beobachtungen zu notieren. Während der Prüfung sollte jeder Examinator für sich einen Raster ausfüllen. Das unterschriebene Dokument dient bei Rekursen als Beweismittel.

Wozu dient der Befragungs- und Bewertungsraster?

<i>Facharztprüfung Rheumatologie</i>	
<i>Name:</i>	<i>Vorname:</i>
<i>Examinator:</i>	
<i>Schlüsselproblem: AST-Erhöhung bei beschwerdefreier Patientin</i>	<i>Blueprintkategorien: Labor Rheumatisches Fieber</i>
<i>Prüfungsaufgabe: Sie sind als Rheumatologe in einer Praxis tätig. Nun wird Ihnen eine beschwerdefreie 46-jährige Patientin zugewiesen, bei der vor einem Jahr als Zufallsbefund beim Hausarzt ein Antistreptolysin-Titer um 1000 i.E./ml festgestellt wurde. Eine probatorische Penicillin-Therapie hatte keinen Einfluss auf den Antistreptolysin-Titer. Weitere Kontrollen zeigen Werte jeweils zwischen 500 und 700 i.E./ml.</i>	
<i>Fragen:</i> 1. Sind persistierende ASO-Titer bei Gesunden möglich? 2. Muss weiter nach einem Focus gesucht werden? 3. Liegt allenfalls ein persistierendes rheumatisches Fieber vor? 4. Was wissen Sie über den Titerverlauf bei Infektionen und Parameter, die ihn beeinflussen?	<i>Richtige Antworten:</i> <input type="checkbox"/> Ja (5%) <input type="checkbox"/> Nein <input type="checkbox"/> Nein. Diagnose rheumatisches Fieber richtet sich nach der Klinik (Jones-Kriterien). Bei 15 bis 20% kann ASO-titer negativ sein. <input type="checkbox"/> Titerhöhe ausgedrückt in internationalen Einheiten <input type="checkbox"/> nur ein wenigstens zweifacher Titeranstieg belegt kürzlich vorausgegangene Infektion <input type="checkbox"/> „Oberer Normwert“ variiert von Person zu Person, Ort und Jahreszeit <input type="checkbox"/> Falsch-positive Titer möglich durch Lebererkrankung, Serumbetalipoproteine und Faktoren einiger Bakterienarten
<i>Bemerkungen:</i>	
<i>Jede richtig gegebene Antwort ergibt einen Punkt (maximal mögliche Punkte: 7)</i>	
<i>Total Punkte:</i>	
<i>Datum:</i>	<i>Uhrzeit:</i> <i>Visum:</i>

Abb. 5.2 Befragungs- und Bewertungsraster für Examinator (adaptiertes Beispiel aus der schweizerischen Facharztprüfung Rheumatologie)

WIE WERDEN PRÜFUNGSUNTERLAGEN HERGESTELLT?

<i>Facharztprüfung Rheumatologie</i>						
<i>Name:</i>			<i>Vorname:</i>			
<i>Prüfungsaufgaben (Blueprintkategorien)</i>	<i>Examinat. 1 (Punkte)</i>	<i>Examinat. 2 (Punkte)</i>	<i>Durchschn. Exam. 1 u. 2</i>	<i>max. mögliche Punkte</i>	<i>Bestehens- grenze</i>	<i>Aufgabe erfüllt (+) nicht erfüllt (-)</i>
<i>1. Tarsaltunnelsyndrom (Anatomie, Physiologie)</i>				<i>9</i>	<i>6</i>	
<i>2. Selektive Cox 2 Inhibition (Pathophysiologie, Aetiologie)</i>				<i>10</i>	<i>7</i>	
<i>3. Bursitis (Klinisch. Symptomatologie)</i>				<i>4</i>	<i>3</i>	
<i>4. AST-Erhöhung (Labor)</i>				<i>7</i>	<i>5</i>	
<i>5. Enger Spinalkanal (Bildgebende Verf., Röntgen)</i>				<i>6</i>	<i>4</i>	
<i>6. Physiotherapiemethoden (Therapie[methoden], Management)</i>				<i>6</i>	<i>4</i>	
<i>7. IV-Fall, Begutachtung (Sozialwesen, Recht, Ethik)</i>				<i>5</i>	<i>4</i>	
<i>8. Leistenschmerzen, Gicht (Differential Diagnose, Diagnostik)</i>				<i>11</i>	<i>8</i>	
<i>9. Untersuchung des HWS (Prinzipien einer klinische Untersuchung)</i>				<i>8</i>	<i>6</i>	
<i>TOTAL erfüllter Prüfungsaufgaben</i>						
<i>Prüfungstermin:</i>			<i>bestanden: (ja, nein)</i> <i>Die Prüfung ist bestanden, wenn 7 der 9 Aufgaben erfüllt sind.</i>			
<i>Ort:</i>			<i>Datum:</i>			
<i>Prüfungsleiter:</i>			<i>Präsident:</i>			

Abb. 5.3 Beispiel eines Punkteblattes (adaptiertes Beispiel aus der schweizerischen Facharztprüfung Rheumatologie)

Wozu dient das Punkteblatt? Auf dem Punkteblatt sind alle Prüfungsaufgaben des gesamten Examens aufgelistet. Hier werden die Punkte der Befragungsraster der einzelnen Examinatoren übertragen. Die Prüfungskommission entscheidet, wie die Punkte der einzelnen Examinatoren miteinander verrechnet werden. Normalerweise wird das arithmetische Mittel berechnet. Weichen die Punktzahlen der Examinatoren zu weit voneinander ab, so können die Differenzen aufgrund des Befragungsrasters bereinigt werden.

Auf dem Punkteblatt ist für jede einzelne Prüfungsaufgabe angegeben, welches das Punktemaximum ist und wie viele Punkte ein Kandidat erreichen muss, um die entsprechende Prüfungsaufgabe zu bestehen. Diese Grenzen werden für jede Prüfungsaufgabe von einem Expertenteam mittels eines erprobten Verfahrens im Voraus bestimmt (vgl. Kapitel 8). Ebenfalls im Voraus wird die Bestehensgrenze der Gesamtprüfung festgelegt. In unserem Beispiel müssen die Kandidaten 7 der 9 Prüfungsaufgaben erfüllen, um die mündliche Facharztprüfung Rheumatologie zu bestehen.

Fragen im Zusammenhang mit der Bewertung

1. Sollen die Examinatoren wissen, wie viele Punkte der Kandidat bei einer einzelnen Prüfungsaufgabe erreichen muss, damit sie als erfüllt bewertet wird?

Was spricht dafür? Wenn die Examinatoren die Grenze kennen, sind sie in der Lage, bei Borderline-Kandidaten genauer nachzufragen, um zu einer sicheren Bewertung zu kommen.

Was spricht dagegen? Mündliche Prüfungen sind anfällig auf unterschiedliche Störvariablen. Das Wissen um die Grenze kann die Beurteilung negativ beeinflussen, wenn der Examinator primär das Erreichen der Schwelle im Auge hat. So kann passieren, dass der Kandidat weit kritischer beurteilt wird, solange er die Schwelle noch nicht erreicht hat. Nach Erreichen der Grenze (und damit Erfüllen der Prüfungsaufgabe) wird er nur noch partiell befragt.

2. Kann die Bestehensgrenze der Gesamtprüfung auch auf der Basis der total erreichten Punktzahl der einzelnen Prüfungsaufgaben errechnet werden?

Was spricht dafür? Wenn das Total der erreichten Punktzahlen bei den einzelnen Aufgaben über das Bestehen der Gesamtprüfung entscheidet, können Kandidaten

schlecht gelöste Prüfungsaufgaben (wenig Punkte erreicht) mit gut gelösten kompensieren. Das steigert u.U. die Akzeptanz der Prüfung.

Geht man von der Annahme aus, dass alle in einer Prüfung gestellten Prüfungsaufgaben relevant sind und ein Minimum an Kompetenzen prüfen, wird man die Möglichkeit des Kompensierens von einzelnen Prüfungsaufgaben nicht zulassen. Grössere Kompetenzlücken in bestimmten Teilbereichen werden so nicht in Kauf genommen.

Was spricht dagegen?

Daniel Hofer

6

Wie wird
eine Prüfung mit
strukturierter
Beobachtung
entwickelt?

INHALT

- 6.1 Was soll beobachtet werden? - 131
- 6.2 Wie wird die Beobachtung strukturiert? - 132
- 6.3 Wie wird die Beobachtungssituation standardisiert? - 136
- 6.4 Wie auswerten? - 138

Eine Beobachtungsprüfung muss möglichst strukturiert und standardisiert ablaufen, damit die Resultate objektiv, zuverlässig und gültig werden. Wichtige Aspekte dazu sind in Kapitel 2 (S.43) beschrieben. Strukturierte Beobachtung heisst im Besonderen Folgendes:

Die Prüfungsstruktur ist vor dem Examen klar festgelegt (Blueprints für Fertigkeiten und Beobachtungssituationen, Beobachtungsform, Anzahl Beobachter, Beobachtungsdauer). Der Beobachtungsprozess wird objektiv und zuverlässig dokumentiert. Die Kandidatenresultate werden aus diesen Dokumenten korrekt abgeleitet.

Die Beobachtungssituation sollte zudem so gut wie möglich standardisiert werden. Beispiele dafür sind Phantome für die Beurteilung manueller Fertigkeiten oder Standardisierte Patienten für die Bewertung interaktiven Verhaltens oder von Problemlösungsfähigkeiten.

Strukturierung und Standardisierung haben zum Ziel, dass die Resultate der Kandidaten möglichst nur von deren Prüfungsleistung abhängen. Dies ist zu überprüfen, indem das Beobachtungsexamen kontinuierlich evaluiert wird.

In diesem Kapitel erfahren Sie,

- wie Beobachtungsprüfungen aufgebaut sein müssen, damit die Prüfer möglichst objektiv und zuverlässig die Fertigkeiten eines Kandidaten beurteilen können
- wie Checklisten und Protokolle aussehen, die in der Beobachtung eingesetzt werden
- welche Rahmenbedingungen zu beachten sind, wenn Fertigkeiten und Verhalten beobachtet werden
- wie bewertet und ausgewertet werden kann

6.1 Was soll beobachtet werden?

Mit der Methode der strukturierten Beobachtung werden die verbalen, interaktiven und praktischen Fertigkeiten eines Facharztes beurteilt (vgl. Kapitel 2). Um zu bestimmen, welche entsprechenden Kompetenzen geprüft werden sollen, braucht es einen

- Fertigkeitsblueprint. Dieser ermöglicht, dass in der Prüfung eine repräsentative Stichprobe von Fertigkeiten beobachtet wird. Daneben ist ein
- Fallblueprint notwendig, da die Wahl des Problems bzw. des Falls die zu beobachtenden Fertigkeiten beeinflussen kann (Fallspezifität). Der Fallblueprint definiert für das Fach repräsentative Problemsituationen, in denen die Fertigkeiten beobachtet werden.

Das Prinzip, wie Prüfungsinhalte definiert werden, ist in Kapitel 1 beschrieben.

Als Beispiel eines Fallblueprints sind die Kriterien angefügt, die in Holland verwendet wurden, um reale, auf Video aufgenommene Hausarztkonsultationen für die Beurteilung der Hausärzte zu selektionieren⁴. Die dafür herangezogenen 16 Fälle resp. Konsultationen mussten so weit wie möglich folgende Auflagen erfüllen:

- *Repräsentation von mindestens acht Kapiteln der ICPC (International Classification of Primary Care)*
- *Repräsentation der fünf ICPC-Kapitel mit der höchsten Prävalenz (Verdauung, kardiovaskuläres System, Bewegungsapparat, respiratorisches System, Haut)*
- *in mindestens acht Fällen müssen zu den Problemlösungen akzeptierte „guidelines“ bestehen*
- *mindestens 14 der 16 Konsultationen müssen zwischen 5 und 15 Minuten dauern*
- *die Spanne des Patientenalters muss von „unter 18“ bis „über 65“ reichen*
- *es müssen sechs bis zehn Patientinnen vorkommen*
- *mindestens acht Fälle müssen Erst-, mindestens vier Folgekonsultationen sein*
- *in höchstens zwei Fällen dürfen Patienten mehr als zwei Gründe haben, den Arzt aufzusuchen*
- *in höchstens zwei Fällen dürfen Patienten ausschliesslich psychosoziale Probleme haben*
- *die ersten fünf Videokonsultationen werden für die Beurteilung nicht berücksichtigt*

6.2 Wie wird die Beobachtung strukturiert?

Checklisten Zu beobachtende und zu bewertende Fertigkeiten können in Checklisten geordnet werden. Damit lassen sich z.B. eine spezifische körperliche Untersuchung oder die technische Ausführung einer therapeutischen Massnahme verlässlich beobachten. Die Kandidatenleistung wird nach dem Ja/Nein-Prinzip (erfüllt/nicht erfüllt) beurteilt (vgl. Abb. 6.1).

	<i>nicht oder unkorrekt ausgeführt</i>	<i>korrekt ausgeführt</i>
<i>Diagnostik</i>		
<i>1. Beurteilung des Bewusstseins</i>		
<i>Kandidat spricht den Patienten laut an</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>verabreicht einen starken Schmerzstimulus</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>2. Beurteilung der Zirkulation</i>		
<i>Kandidat überprüft den Karotispuls einseitig während mehr als 4 Sekunden</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>3. Beurteilung der Luftwege</i>		
<i>Kandidat schaut, ob die Luftwege frei sind und ob der Patient atmet</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>4. Das diagnostische Vorgehen wird in der richtigen Reihenfolge ausgeführt</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>5. Kandidat schliesst Diagnostik in 30 Sekunden ab</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Ausführung</i>		
<i>6. Kandidat beginnt mit Thoraxkompression</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>7. Kandidat zeigt richtige Kompressionsposition</i>		
<i>richtige Position der Schultern</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>richtige Position der Hände</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>8. Kandidat zeigt adäquate Kompressionstechnik</i>		
<i>hält die Frequenz bei 80-100 Kompressionen pro Minute</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>führt Zyklen von 15 Kompressionen und 2 Beatmungen aus</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>9. Kandidat zeigt adäquate Beatmungstechnik</i>		
<i>führt ein korrektes Kopf-Kinn-Manöver durch</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>bedeckt mit eigenem Mund den Mund des Patienten völlig</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>verhindert Luftaustritt aus der Nase des Patienten</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>beobachtet den Thorax während der Beatmung</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>10. Der Thorax hebt und senkt sich während der Beatmung</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Allgemeiner Eindruck (1-10): ...</i>		

Abb. 6.1 Beobachtung von Hausärzten bei der kardiopulmonalen Reanimation in Holland¹

Skalierte Protokolle werden vor allem dann eingesetzt, wenn kommunikative Fertigkeiten, Verhalten oder Problemlösungsfähigkeiten beobachtet werden sollen. Die Beobachtungskriterien werden dabei mit Hilfe einer abgestuften Skala bewertet.

*abgestufte
Beobachtungs-
protokolle*

Als Beispiel (Abb. 6.2) sind auszugsweise Beobachtungskriterien aufgeführt, wie sie in der schweizerischen Facharztprüfung für Allgemeinmedizin verwendet werden. Sie wurden teilweise in den Niederlanden entwickelt². Die Kandidatenleistungen werden in einer zweistündigen realen Sprechstunde mit Unterstützung von Videoaufnahmen beurteilt.

(...)							
■ <i>beschwerdegerichtetes Fragen: passive Gesprächstechnik</i>							
<i>zielgerichtetes Fragen, Führen, Nachfragen</i>	1	2	3	4	5	6	0
■ <i>Arzt-Patienten-Verhältnis bei der körperlichen Untersuchung</i>							
<i>Erklärungen vor Untersuchung, Erklärungen während Untersuchung, respektvolle Begegnung</i>	1	2	3	4	5	6	0
■ <i>Zusammenfassen der Diagnostik</i>							
<i>Befunde, Zusammenhang mit Konsultationsgrund, Schlussfolgerungen</i>	1	2	3	4	5	6	0
■ <i>Zusammenfassen der Diagnostik auf patientenorientierte Art</i>							
<i>kleine Schritte, Wiederholen, Anpassen ans Wertsystem des Patienten, Anpassen ans Informationsbedürfnis des Patienten</i>	1	2	3	4	5	6	0
■ <i>Abschätzen der Reaktion des Patienten auf das Zusammenfassen der Diagnostik</i>							
<i>Abwarten der Reaktion, Besprechen der Reaktion</i>	1	2	3	4	5	6	0
(...)							
<i>1 sehr schlecht 2 schlecht 3 ungenügend 4 genügend 5 gut 6 sehr gut 0 nicht beurteilbar</i>							

Abb. 6.2 Beobachtungskriterien in der schweizerischen Facharztprüfung für Allgemeinmedizin

Beobachtungskriterien können auch globaler formuliert werden. Immer wichtiger wird aber dann das Training der Beobachter, damit diese Gleiches möglichst gleich bewerten.

globalere Raster

Als Beispiel sind in Abb. 6.3 Beobachtungskriterien aufgeführt, wie sie für die Konsultationsbeobachtung einer Sprechstunde in Schottland entwickelt wurden³.

1. Bestand ein offensichtlicher Fehler in der Diagnostik oder im Management?	<input type="checkbox"/> ja	<input type="checkbox"/> nein
2. Wie gut erkannte der Arzt die Gründe des Patienten, die zur Konsultation führten?	1	2 3 4 5 6
3. Wie klar definierte der Arzt das vorliegende Problem?	1	2 3 4 5 6
4. Wie gut passte der Arzt seine Erklärungen den Bedürfnissen des Patienten an?	1	2 3 4 5 6
5. Wie gut löste der Arzt das vorliegende Problem?	1	2 3 4 5 6
6. Wie effektiv setzte der Arzt dabei Zeit, Untersuchungen und Personal ein?	1	2 3 4 5 6
7. Wie gut war die Arzt-Patient-Beziehung?	1	2 3 4 5 6
1 sicher ungenügend 2 ungenügend 3 knapp genügend 4 genügend 5 gut 6 sehr gut		

Abb. 6.3 Kriterien einer Konsultationsbeobachtung in Schottland

*mindestens zehn
Beobachtungssituationen in drei
Stunden*

Erfahrungen zeigen, dass für eine gültige Beobachtungsprüfung 10 - 20 verschiedene Beobachtungssituationen nötig sind. Dabei sollte jeder Kandidat mindestens drei Stunden lang beobachtet werden^{2,4,5}. Die Evaluation jedes einzelnen Examens muss zeigen, wie viele Examinatoren wie lange und in wie vielen Situationen beobachten müssen, damit die Prüfung zuverlässig ist.

In Holland wurden bei der direkten Beobachtung ärztlicher Interview- und Beratungsfertigkeiten befriedigende Reliabilitäten mit 8-10 Prüfungsfällen und 2-2¹/₂ Std. Prüfungszeit pro Kandidat erreicht². Als niederländische Hausärzte in der Sprechstunde mit Video beobachtet wurden, ergaben 16 beobachtete Fälle zuverlässige Beobachtungen. Die Fälle wurden nach spezifischen Kriterien (vgl. S.131) aus einer einwöchigen Periode mit Videoaufzeichnung ausgewählt und bewertet⁴.

Das „simulated office orals“-Examen des „certification examination for family physicians“ in Kanada⁵ weist aktuell fünf Fälle auf, wie sie weiter unten beschrieben sind (Examenszeit 1¹/₄ Std. pro Kandidat). Damit wird ein Reliabilitätskoeffizient um 0.6 erreicht. Für einen wünschbaren Wert von 0.8 müssten 15 - 20 solcher Beobachtungssituationen etabliert werden (3³/₄ bis 5 Std.), was in Kanada zurzeit am Aufwand scheitert. Die Evaluatoren weisen aber auf die vergleichsweise höhere Validität des Examens hin.

*mindestens zwei
Prüfer...*

Jede Situation sollte grundsätzlich von mindestens zwei Prüfern beobachtet werden. Es ist besser, wenn jeweils zwei Beobachter alle Kandidaten in ein und derselben Problemsituation beobachten und bewerten, als wenn zwei Beobachter einen Kandidaten durch alle Problemsituationen der Prüfung begleiten.

Wenn aus ökonomischen oder logistischen Gründen die Ressourcen knapp sind, ist es günstiger, die Zahl der Prüfer statt die Zahl der Beobachtungssituationen zu verkleinern: Es ist besser, wenn ein Prüfer alleine in zwei Fällen beobachtet als wenn dies zwei Prüfer zusammen in nur einem Fall tun. Es ist aber auch möglich, dass Personen, die nicht Fachspezialisten sind, für gewisse Beobachtungen geschult werden. Dies zeigen z.B. Erfahrungen aus Neuseeland⁶ oder Kanada⁷. Auch Standardisierte Patienten können die Beobachtung übernehmen (vgl. Kapitel 7).

Zu einer objektiven Beobachtung gehören unabhängige Prüfer. Befreundete Examinatoren oder solche aus derselben Klinik sind ebenso wenig unabhängig wie ein Prüfer, der den Kandidaten aus der Weiterbildungszeit gut kennt. Im Weiteren muss eine Beobachtung so ablaufen, dass die Prüfer unabhängig voneinander zu ihrem Urteil kommen können. Ein allenfalls nötiges Konsensprozedere bei stark divergierenden Urteilen ist von einer unabhängigen Drittperson oder -instanz zu supervisionieren.

....die unabhängig arbeiten...

Damit Prüfer zuverlässig beurteilen, müssen sie geschult werden. Entsprechende Workshops können folgende Inhalte haben:

... und trainiert worden sind

- Beobachten mit Checklisten und Protokollen (simulierte Prüfungssituationen)
- Vergleichen der eigenen Beobachtung mit derjenigen anderer
- Formulieren von Unsicherheiten
- Gestaltung einer möglichst stressfreien Beobachtungssituation
- Erleben, wie es ist, beobachtet zu werden

Auch die Kandidaten sollen die Prüfungssituation üben können. Eine Möglichkeit dazu bieten solche Prüfer-Trainingsveranstaltungen.

Bei einer direkten Beobachtung hält sich der Prüfer unauffällig und ruhig im Hintergrund. Er sitzt oder steht möglichst ausserhalb des Gesichtsfeldes des Kandidaten oder eines allenfalls eingesetzten Standardisierten Patienten. Um den Beobachter noch besser zu „verstecken“, bieten sich die verdeckte Beobachtung hinter der Einwegscheibe oder die Videoaufzeichnung an. Es ist auch möglich, dass ein Standardisierter Patient Beobachter-Funktion übernimmt.

Wie verhält sich der Beobachter?

6.3 Wie wird die Beobachtungssituation standardisiert?

Die Situation, in welcher Kandidaten beobachtet werden, sollte für alle Prüflinge wenn immer möglich gleich sein. Standardisieren lassen sich solche Situationen beispielsweise mit Hilfe von Modellen oder speziell präparierten Phantomen. Im Fall der in Abb. 6.1 erwähnten Reanimationsbeobachtung protokolliert das Phantom - ergänzend zur Beobachtung der Prüfer - automatisch gewisse Aktionen, wie z.B. das Beatmungsvolumen.

Eine weitere Möglichkeit ergibt sich mit dem Einsatz von Standardisierten Patienten. Diese Patienten können echt sein oder durch Ärzte oder Schauspieler simuliert werden.

Alle Mitwirkenden müssen vor ihrem Prüfungseinsatz gründlich instruiert, trainiert und evaluiert werden. Was dabei alles zu unternehmen und zu bedenken ist, beschreibt Kapitel 7.

In der Postgraduate-Hausarztprüfung in Kanada wird die Prüfungssituation (eine 15-minütige Sprechstunde) mit simulierenden Patienten („gespielt“ von Ärzten) standardisiert⁸. Die Standardisierten Patienten üben die Simulationen anhand ausführlicher Anweisungen in einem sog. Fallbuch (vgl. Abb. 6.4).

(...)

Instruktionen an die Darstellerin:
Sie präsentieren sich als Amelia Martin, 50-jährig. Sie fühlen sich seit einem Jahr gar nicht mehr gut wegen Müdigkeit, Kopfschmerzen und Völlegefühl.

Sie haben bereits mehrere Ärzte aufgesucht, die Behandlungen aber stets abgebrochen, weil die verschriebenen Medikamente nicht zu helfen schienen. Eine gewisse Anzahl Medikamente haben sie aber eingenommen (...es folgen detaillierte Angaben zur gesamten Anamnese...).

(...)

Spezielle Anweisungen: Sie sind sehr gesprächig und intensiv geschminkt. Ihr Äusseres entspricht Ihrem Alter, nicht aber Ihre Kleidung. Sie tragen grelle Jogging-Kleider. Sie haben eine grosse Handtasche bei sich, die unter anderem Ihre Medikamente und die Laborresultate enthalten.

Abb. 6.4 Verhaltensanweisungen für Standardisierte Patienten (Fallbuch)

Die simulierende Ärztin (vgl. Abb. 6.4) beurteilt die Leistung der Kandidaten anhand von Beobachtungskriterien, wie sie in Abb. 6.5 auszugsweise ausgeführt sind.

(...)

Instruktionen an die Beobachter:
Bewerten Sie nach folgenden Beurteilungskategorien und -kriterien:

1. Identifikation der medizinischen Probleme

<i><u>medizinische Probleme (3 Kriterien):</u></i>	<i><u>erlebte Symptome (10 Kriterien):</u></i>
<i>Hypokaliämie, Hypothyreose, schlecht kontrollierte Hypertonie</i>	<i>Gefühle: Konfusion, Angst, Frustration. Vorstellungen: hat etwas Schwerwiegendes. Konsequenzen: fühlt sich weniger funktionsfähig, müde nach Anstrengungen. Erwartungen: will Sofortlösung, bessere Medikamente („der Arzt kann sowieso nicht helfen“).</i>

Weitere analoge Bewertungskategorien und entsprechend fallspezifische Kriterien:

2. Identifikation der Unfähigkeit, eine therapeutische Beziehung mit den vorgängig konsultierten Ärzten zu etablieren (23 ausformulierte Kriterien)

3. sozialer und familiärer Kontext (10 ausformulierte Kriterien)

4. Management der medizinischen Probleme (14 ausformulierte Kriterien)

5. Management des Aufbaus einer therapeutischen Beziehung (9 ausformulierte Kriterien)

6. Strukturierung und Ablauf der Konsultation (4 ausformulierte Kriterien)

(...)

Abb. 6.5 Beobachtungsanweisungen für Standardisierte Patienten

6.4 Wie auswerten?

Bestanden oder nicht bestanden?

Die Grundsätze der Standardsetzung, wie sie in Kapitel 8 dargelegt sind, gelten auch für eine Prüfung mit strukturierter Beobachtung. Ein kriterienorientiertes Verfahren drängt sich insbesondere auch darum auf, weil Kriterien bereits für die Beobachtung festgelegt worden sind. Vor der Prüfung müssen aber noch die für das Bestehen minimal geforderten Fertigkeiten oder Fertigungsanteile definiert werden.

Im Beispiel der simulierten Sprechstunde (vgl. Abb. 6.5) werden in jeder Kategorie die Anzahl Kriterien definiert, die zum Urteil „gut bestanden“, „bestanden“ oder „nicht bestanden“ führen. Das Standardsetzungsverfahren definiert dann weiter die zu bestehende Anzahl Kategorien pro Prüfung (vgl. Kapitel 8).

In der ersten Kategorie des erwähnten Beispiels („Identifikation der medizinischen Probleme“) sieht die Standardsetzung wie folgt aus⁸:

	<i>medizinische Probleme</i>	<i>erlebte Symptome</i>
<i>gut bestanden</i>	<i>erfasst 1., 2. und 3.</i>	<i>Kandidat exploriert das Erleben der Symptome aktiv, um ein tiefes Verständnis zu erreichen. Braucht dazu verbale und nonverbale Techniken inkl. einen wirksamen Fragebogen. Hört aktiv zu.</i>
<i>bestanden</i>	<i>erfasst 1. und 2.</i>	<i>Kandidat informiert sich, um ein genügendes Verständnis zu erreichen. Braucht dazu passende Fragen und nonverbale Techniken.</i>
<i>nicht bestanden</i>	<i>erfasst 1. und 2. nicht</i>	<i>Kandidat zeigt nur wenig Interesse am Erleben der Symptome. Erreicht nur ein minimales Verständnis. Erfasst verbale und nonverbale Hinweise des Patienten kaum, unterbricht ihn sogar.</i>

Abb. 6.6 Bewertungskriterien (vgl. Abb. 6.5)

Geprüfte, Prüfer und Prüfung evaluieren

Die Auswertung der protokollierten Kandidatenleistungen soll nicht nur die Qualifikation der Fachärzte ermöglichen, sondern auch den Geprüften ihre individuellen Leistungen rückmelden.

Die Beurteilungsübereinstimmung der Prüfer (Interraterreliabilität) und die Messeigenschaften der Prüfung (Reliabilität, Validität) sind regelmässig zu eruieren.

Abschliessend müssen die Auswertungen an Prüfungsverantwortliche, Prüfungskonstrukteure und Prüfer weitergeleitet werden. Die Weiterbildner sollten über allfällige Kompetenzlücken der Kandidaten informiert werden.

Peter Schläppi

Literatur

- ¹ Jansen JJ, Berden HJ, van der Vleuten C, Grol RP, Rethans J, Verhoeff CP. Evaluation of cardiopulmonary resuscitation skills of general practitioners using different scoring methods. *Resuscitation* 1997; 34(1):35-41
- ² van Thiel J, Kraan HF, van der Vleuten CP. Reliability and feasibility of measuring medical interviewing skills: the Maastricht history-taking and advice checklist. *Med Educ* 1991;25(3):224-9
- ³ Campbell LM, Howie JGR, Murray TS. Summative assessment: a pilot project in the west of Scotland. *Br J Gen Pract* 1993;43:430-3
- ⁴ Ram P, van der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: comparison of observation in a multi-station examination using standardized patients with observation of consultations in daily practice. *Acad Med* 1999;74:62-9
- ⁵ Handfield-Jones R, Belle Brown J, Biehn J, Rainsberry P, Brailovsky CA. Certification Examination of the College of Family Physicians of Canada, Part 2: Conduct and general performance. *Can Fam Physician* 1996;42:1188-95
- ⁶ Thomson AN. Reliability of consumer assessment of communication skills in a postgraduate family practice examination. *Med Educ* 1994; 28(2):146-50
- ⁷ Bullock G, Kovacs G, Macdonald K, Story BA. Evaluating procedural skills competence: interrater reliability of expert and non-expert observers. *Acad Med* 1999;74:76-8
- ⁸ Examin de certification en médecine familiale , exemple d'un cas (N°4), le collège de médecins de famille du Canada, 1994

7

Wie wird
eine objektive
strukturierte
klinische Prüfung
entwickelt?

INHALT

- 7.1 Woraus besteht ein OSCE? - 145
- 7.2 Wie werden Stationen entwickelt? - 150
- 7.3 Wie werden Examinatoren vorbereitet? - 157
- 7.4 Wie Standardisierte Patienten
rekrutieren und trainieren? - 160
- 7.5 Wie wird ein Circuit angelegt? - 166
- 7.6 Wie wird ein OSCE ausgewertet? - 168

Dieses Kapitel beschreibt detailliert die Entwicklung eines objektiven strukturierten klinischen Examens (OSCE). Wir empfehlen Ihnen, sich vor der Lektüre anhand der kurzen Beschreibung am Ende des 2. Kapitels einen Überblick über die Funktionsweise sowie die Vor- und Nachteile dieses Prüfungsansatzes zu verschaffen.

Der erste Abschnitt dieses Kapitels befasst sich mit den Voraussetzungen für die Durchführung eines OSCE. Er beschreibt, welche Personen mit welchen Aufgaben beteiligt sind, welches die Hauptschritte der Entwicklung darstellen und welche Stationentypen zu empfehlen sind. Der zweite Abschnitt beschreibt das Vorgehen bei der Entwicklung einzelner Stationen. Die nächsten beiden Abschnitte geben Auskunft über die Vorbereitung der Examinatoren und der Standardisierten Patienten (SPs). Der fünfte Abschnitt behandelt die Planung und den praktischen Aufbau des Stationencircuits während der sechste Teil Aspekten der Auswertung eines OSCE gewidmet ist. Ein fünfteiliger Anhang illustriert, wie Unterlagen und Dokumente für Entwicklung, Durchführung und Auswertung aussehen können.

Wir beschreiben aus den folgenden zwei Gründen das OSCE-Verfahren ausführlicher als andere Methoden: Die Prüfungsform ist im deutschsprachigen Raum noch wenig bekannt. OSCEs sind störanfällig. Ein Fehler in einer einzelnen Komponente kann bereits den Gesamterfolg gefährden.

In diesem Kapitel erfahren Sie,

- welche Aufgaben die Entwicklung und Durchführung eines OSCE für die Beteiligten umfasst
- welche Stationentypen zu empfehlen sind und wie eine Station entworfen, ausgearbeitet und praktisch erprobt wird
- wie viele Examinatoren und SPs mit welchen Eigenschaften nötig sind und wie ihr Training aufgebaut werden kann
- was zur Planung und Installation eines Circuits gehört
- welche Aspekte bei der OSCE-Auswertung zu klären sind

7.1 Woraus besteht ein OSCE?

Für das vorliegende Kapitel haben wir vor allem Informationsmaterial aus dem Bereich der ärztlichen Grundausbildung verwendet und zwar zum einen aus dem 2. Teil des „Qualifying Examination“ des Medical Council of Canada und zum andern aus dem „Standardized Patient Project“ des amerikanischen National Board of Medical Examiners. Da diese beiden OSCEs zur selben Zeit an verschiedenen Orten genau gleich (standardisiert) durchgeführt werden müssen, wurde das Vorgehen bei der Entwicklung und Durchführung sehr detailliert dokumentiert. Das heisst aber nicht, dass OSCEs nur in der Grundausbildung eingesetzt werden. Auch für die Evaluation in der Postgraduierten-Ausbildung wurde in verschiedenen Fachbereichen über OSCEs publiziert (z.B. Allgemeinmedizin, Innere Medizin, Pädiatrie, Chirurgie, Anästhesie).

Vorbemerkungen

Falls eine Gesellschaft für ihre Facharztprüfung diese Prüfungsform in Erwägung zieht, ist es sicher sinnvoll, international nach evtl. schon bestehenden OSCEs zu suchen. Wer die Möglichkeit hat, als Beobachter an einem OSCE in seinem eigenen oder einem verwandten Fachbereich teilzunehmen und von den Erfahrungen der Konstrukteure und den an der Durchführung Beteiligten zu profitieren, sollte dies unbedingt tun.

Welche Personen spielen eine Rolle?

Die Prüfungskommission, welche für die Fachgesellschaft repräsentativ sein sollte, ist für folgende Aufgaben verantwortlich:

Prüfungskommission

- Festlegung des Blueprints und der Prüfungsziele (vgl. Kapitel 1)
- Planung der Prüfungsentwicklung
- Rekrutierung der Autoren
 - der Patiententrainer
 - der Examinatoren
 - des Durchführungs-Verantwortlichen
- Verwaltung des OSCE-Stationenpools
- Auftragserteilung für das Erstellen neuer Fälle
- Revision der erstellten Fälle
- Auswahl der Stationen für eine Prüfung
- Standardsetzung
- Instruktion der Examinatoren
- Erstellung der schriftlichen Prüfungsunterlagen (Sekretariat)
- Anmeldungsadministration (Sekretariat)
- Prüfungsauswertung und Prüfungsevaluation

Um diese Aufgaben zu erfüllen, kann es sinnvoll sein, mit einer Fachperson resp. einer Fachinstitution für Prüfungsentwicklung, -auswertung und -evaluation zusammenzuarbeiten. Sicher benötigt wird ein Sekretariat.

Autoren Als Autoren von Fällen resp. Stationenszenarien mit den entsprechenden Beurteilungskriterien (Checklisten) sind Fachärzte erforderlich. Sinnvollerweise kreieren diese neue Fälle in kleinen Teams in einem Workshop.

Patiententrainer Patiententrainer haben die Aufgabe, mit SPs deren Rollen einzuüben. Der hauptverantwortliche Patiententrainer kann zudem die Aufgabe erhalten, die SPs zu rekrutieren. Er sollte bei der Prüfungsdurchführung anwesend sein. Die Trainer brauchen nicht Fachärzte zu sein. Sie müssen von solchen aber genau erfahren, wie sich eine bestimmte Gesundheitsstörung manifestiert. Wichtig ist, dass sie selbst gut „schauspielern“ und dies auch gut vermitteln können.

Examinatoren Examinatoren beurteilen anhand von Checklisten die Kandidatenleistungen an den Prozedurstationen und werten die Antwortblätter von den Fragestationen aus. Für diese Aufgaben kommen in erster Linie wieder Fachärzte in Frage. Wenn genau definierte Checklisten vorliegen, können für gewisse Stationentypen - nach entsprechendem Training - auch Nicht-Fachärzte eingesetzt werden¹. Die Anzahl erforderlicher Examinatoren hängt von der Zahl der Stationen und der Kandidaten ab. Es ist nicht möglich, länger als einen halben Tag praktisch ununterbrochen konzentriert zu beobachten und zu beurteilen. Deshalb braucht es, wenn ein OSCE während eines Vor- und eines Nachmittags läuft, pro Station zwei Examinatoren.

Durchführungsverantwortlicher Der Durchführungsverantwortliche

- findet und bucht geeignete Prüfungslokalitäten
- plant den Circuit
- beschafft die erforderlichen Einrichtungs- und Prüfungsmaterialien
- erstellt Zeitpläne und Checklisten für die Materialkontrolle
- organisiert das akustische Signalsystem
- organisiert die Verpflegung
- richtet die Stationen ein
- beschildert den Circuit
- überwacht den Prüfungsablauf

Um alle diese Aufgaben erfüllen zu können, braucht es verständlicherweise etliches Hilfspersonal.

SPs spielen während der Prüfung echte Patienten. Sie werden für Anamnese-, Untersuchungs- und Beratungsstationen benötigt. Die Zahl ist von den ausgewählten Stationen und der Zahl der Kandidaten abhängig. Auch SPs können kaum während eines ganzen Tages eingesetzt werden.

Standardisierte Patienten

Es ist offensichtlich, dass OSCE ein sehr personalintensives Prüfungsverfahren ist. Man kann davon ausgehen, dass bei Kandidatenzahlen zwischen 100 und 250 ebensoviel Personal für die Entwicklung, Durchführung und Auswertung erforderlich ist.

wieviel Personal?

In welchen Schritten erfolgt die Entwicklung?

In einer ersten Phase geht es um die Entwicklung neuer Fälle. Folgende Schritte sind nötig:

Entwicklung neuer Fälle

- Feststellen des Bedarfs gemäss Blueprint und Abklären allfälliger Lücken im Pool der vorhandenen Fälle
- Rekrutieren von Autoren, Aufträge erteilen
- Kreieren neuer Fälle (wenn möglich im Rahmen eines Workshops)
- inhaltliche und formale Vorrevision
- Revision im Rahmen einer Sitzung der Prüfungskommission
- Erfassen der akzeptierten (revidierten) Fälle in den OSCE-Pool

In der zweiten Phase wird die konkrete nächste Prüfung entwickelt. Dazu gehört vor allem:

Entwicklung der Prüfung

- Auswahl der Fälle für die nächste Prüfung
- Besprechung derselben mit Patiententrainern
- Rekrutierung und Instruktion der Examinatoren
- Rekrutierung und Training der SPs
- Buchung der Prüfungslokalitäten, Planung des Circuits und Einrichtung der Prüfungsräume (vgl. Aufgaben des Durchführungsverantwortlichen)
- Kandidatenadministration

In der dritten Phase wird die Prüfung durchgeführt, ausgewertet und evaluiert.

Durchführung und Auswertung

Die Planung eines OSCE sollte mindestens ein Jahr vor dem Prüfungstermin beginnen, angesichts des Milizsystems sind Zyklen von anderthalb Jahren sinnvoller.

Welche Stationentypen sind zu empfehlen?

Beschränkung auf zwei Standardlängen

Bei der Planung der Stationen muss bereits daran gedacht werden, dass diese in einem Circuit angelegt werden. Dieser sollte möglichst flüssig absolviert werden können. Wir empfehlen deshalb, sich auf zwei Standardlängen zu beschränken. Kurzstationen von beispielsweise fünf oder sieben Minuten Dauer reichen aus, um zu beurteilen, wie fokussierte Teilanamnesen oder gezielten Untersuchungen an SPs sowie klar umschriebenen Handlungen an Phantomen, Modellen oder Präparaten ausgeführt werden. Sollen die Leistungen bei umfassenderen Anamnesen, Untersuchungen und Behandlungen oder bei Fähigkeiten und Fertigkeiten im Umgang mit Patienten und komplexeren Problemsituationen evaluiert werden, drängen sich Stationen von doppelter Länge auf.

Vor- und Nachteile von Kurz- und Langstationen

Eine repräsentative und damit hinreichend zuverlässige Prüfung erfordert in der Regel eine genügend grosse Zahl von Stationen (ca. 20), die sowohl inhaltlich verschieden sind als auch unterschiedliche Kompetenzen evaluieren. Ist die Gesamtprüfungszeit z.B. auf 4 Stunden limitiert, können maximal 15 Stationen von 14-minütiger Dauer untergebracht werden aber etwa 36 Stationen von 5 Minuten Dauer. Unter dem Reliabilitätsgesichtspunkt ist es ratsamer, eine grössere Zahl kurzer Stationen zu entwickeln als wenige lange. Andererseits gibt es Kompetenzen (z.B. Management einer komplexeren Situation), die sich nicht innerhalb von 5 Minuten gültig beurteilen lassen. Es besteht die Gefahr, dass die komplexe Ganzheit einer echten Arzt/Patient-Interaktion in einem Fertigkeiten-Hickhack vieler kurzer Stationen verlorengelht. Verschiedentlich wurde zudem beobachtet, dass Kandidaten in kurzen Stationen tendenziell hektisch befragen und keine natürlichen Pausen zulassen. Der ideale Mix von Kurz- und Langstationen muss fachspezifisch unter den beiden Gesichtspunkten, dem der Validität und dem der Reliabilität, abgewogen werden.

Doppelstationen: Prozedurstationen kombiniert...

Gut bewährt haben sich Doppelstationen (couplet stations). In einer Prozedurstation findet zuerst eine Begegnung mit einem SP statt. Der Kandidat muss bei diesem z.B. innert 5 Minuten eine gezielte Teilanamnese erheben oder eine spezifische Untersuchung durchführen. Er wird dabei von einem Arzt als Examinator anhand einer Checkliste beurteilt.

...mit Fragestationen

Nach dem akustischen Schlusszeichen erhält der Kandidat vom Examinator die schriftlichen Aufgaben, die er in der nachfolgenden Fragestation zu bearbeiten hat. Die Pause dauert eine Minute. Die Aufgaben der Fragestation sollen sich auf den eben befragten und/oder untersuchten Patienten beziehen. Typische Aufgaben sind:

- die Befunde aus der Prozedurstation aufschreiben
- eine Differentialdiagnose aufstellen
- Röntgenbilder, CT, Laborbefunde etc. interpretieren
- einen Untersuchungs- oder Behandlungsplan vorschlagen

Am ehesten empfiehlt sich für die schriftliche Befragung die Form der Kurzantwortfragen (vgl. Kapitel 4). Für gewisse Aufgaben ist auch eine MC-Antwortform denkbar (vgl. Kapitel 3).

In der Fragestation ist die Anwesenheit eines Examinators nicht erforderlich. Nach der akustischen Ankündigung des Stationendes verlässt der Kandidat den Raum und lässt sein Antwortblatt zurück. Dieses wird von einer Hilfsperson herausgeholt, die auch kontrollieren muss, ob die nötigen Unterlagen (z.B. Röntgenbilder) noch vorhanden und unbeschädigt sind. Es ist z.B. vorgekommen, dass ein Kandidat die verlangte Lokalisierung auf einem Röntgenbild nicht nur im Schema auf seinem Antwortblatt markiert hat sondern auch direkt auf dem Originalröntgenbild, das im Prüfungsraum hing. Da die nachfolgenden Kandidaten diese Information hatten, musste die ganze Fragestation nachträglich aus der Bewertung eliminiert werden.

Bei langen Prozedurstationen von 10-15 Minuten Dauer ist es denkbar, in den letzten 2 Minuten einige standardisierte mündliche Fragen einzuplanen, die der Examinator stellt (vgl. Kapitel 5). Damit die Bedingungen für alle Kandidaten gleich sind, muss dieser mit der Befragung unbedingt auf das entsprechende akustische Signal warten, auch wenn der Kandidat seine Interaktion mit dem SP vorher abgeschlossen hat. Ein Kandidat hat damit auch die Möglichkeit, sich nochmals dem Patienten zuzuwenden, um etwas Zusätzliches zu fragen oder zu untersuchen, das ihm nachträglich einfällt.

*lange Stationen
ergänzt durch struk-
turierte Befragung*

Durch die Kombination der Beobachtung praktischer Fertigkeiten einerseits und der schriftlichen oder mündlichen Prüfung von Wissens- und Denkleistungen andererseits wird die Prüfung inhaltlich heterogener. Man erweitert damit den Validitätsanspruch, muss aber in Kauf nehmen, dass die Reliabilität im Sinne der inneren Konsistenz abnimmt¹.

*Validität vs.
Reliabilität*

7.2 Wie werden Stationen entwickelt?

Die Entwicklung von OSCE-Stationen ist zeitaufwendig. Eine gute Ausbeute wird zum einen erreicht durch klare Aufträge und zum andern durch ein strukturiertes, mehrstufiges Vorgehen bei der Entwicklung.

klarer Auftrag Die Auftragserteilung und die Koordination der Stationenentwicklung obliegt der Prüfungskommission. Den Ausgangspunkt bilden dabei der Blueprint (vgl. Kapitel 1) und eventuell formulierte Prüfungsziele. Für OSCEs ist es sicher angezeigt, mit mindestens zweidimensionalen Blueprints zu arbeiten. Die eine Dimension enthält den Inhaltsbereich oder eine Typologie von Konsultationsgründen (vgl. Abb. 1.7, S.31), die zweite die praktischen Aufgaben (Umgang/Kommunikation mit Patienten, Anamnese, körperliche Untersuchung, manuelle Fertigkeit, Problemmanagement usw.). Sofern bereits ein Pool von früher erarbeiteten und z.T. schon erfolgreich verwendeten Stationen besteht, können die Aufträge auf spezifische Lücken ausgerichtet werden.

Beispiel Im Interesse der allgemeinen Verständlichkeit wählen wir zur Illustration ein Auftragsbeispiel für das Niveau der ärztlichen Grundausbildung. Ein gezielter Auftrag könnte wie folgt lauten:

Patientenproblem: *Bauchschmerzen*
Rahmen: *Notfallsituation*
Plausible Diagnosen: *Appendizitis*
Gastroenteritis
Morbus Crohn
Nierenstein
Situation: *Erstkonsultation eines 18-35-jährigen Fabrikarbeiters, der wegen akuten Bauchschmerzen auf die Notfallstation kommt*
Primäre Aufgabe: *Untersuchung 75 %*
Sekundäre Aufgabe: *Anamnese 25 %*
Prüfungsziele für *Körperliche Untersuchung - Abdomen*
Prozedurstation: *U1: zeigt angemessene gezielte Untersuchung resp. erforderliche Techniken*
U2: erkennt wichtige körperliche Befunde
U3: vermeidet unnötige Unannehmlichkeiten
Anamnese
A1: erkennt Zeichen und Symptome, welche eine sofortige ärztliche Intervention erfordern (Notfallsituation)
A2: erfragt mit dem Konsultationsgrund verbundene Symptome
A3: erhebt Familienanamnese
A4: erhebt Informationen bezüglich früherer Krankheiten und Operationen
Prüfungsziele für *U2: erkennt wichtige körperliche Befunde*
Fragestation: *U4: interpretiert wichtige körperliche Befunde*

Wir konzentrieren uns hier auf die von der Konstruktion her anspruchsvollste Stationenform, diejenige mit einem SP. In Anlehnung an das „Standardized Patient Project“ des National Board for Medical Examiners in den USA schlagen wir ein dreistufiges Vorgehen zur Entwicklung von OSCE-Stationen vor:

1. Entwurf
2. Ausarbeitung
3. praktische Erprobung

Erster Schritt: Entwurf

Es hat sich bewährt, die konkreten Fälle für OSCE-Stationen im Rahmen eines Workshops durch kleine Teams von z.B. drei Personen (2 Ärzte, 1 Patiententrainer) entwickeln zu lassen. Jedes Team denkt sich zum Auftrag einen konkreten Patientenfall aus, an welchem zumindest einige der Examensziele geprüft werden können, und fasst diesen zusammen.

Zur Fallzusammenfassung gehören:

- Angaben zur Person (Alter, Geschlecht, Beruf oder Branche/Funktion)
- Ort des Patientenkontaktes
- Konsultationsgrund
- Aufgabe, die der Arzt an diesem Patienten erfüllen soll
- Falls sinnvoll, zuweisende/r Institution/Arzt, evtl. Zuweisungsdiagnose
- Effektive Diagnose/Befunde
- Liste der Differentialdiagnosen
- Ausschlusskriterien (z.B. darf eine Person, die Appendizitis simulieren soll, keine Appendektomie haben)
- Begleitdokumente (Überweisungsschreiben, krankengeschichtliche Dokumente): ja/nein; falls ja, in ein oder zwei Sätzen deren Inhalt skizzieren

Es kann hilfreich sein, den Teams checklistenartige Schemata vorzugeben, die gewährleisten, dass beim Zusammentragen der Informationen zur Anamnese und zur Untersuchung nicht Wesentliches vergessen wird. Die Teams brauchen dabei nur zu den Punkten Angaben zu machen, die für den Fall relevant sind. Es können schon die Punkte bezeichnet werden, die in der Checkliste auftauchen sollten. Im Anhang 1 (S.171) finden Sie je ein mögliches Schema für die Strukturierung der Angaben zur Anamnese und zur körperlichen Untersuchung.

Abschliessend sollen stichwortartig mögliche Beurteilungspunkte notiert werden, aus denen später eine Checkliste entwickelt werden kann.

Vorgehen bei der Entwicklung der Stationen

Fallzusammenfassung

Angaben zu Anamnese und Untersuchung

mögliche Beurteilungskriterien

Zum Prüfungsziel U1 „zeigt angemessene gezielte Untersuchung resp. erforderliche Techniken“ könnten die Stichworte etwa lauten:

- palpiert Abdomen
- beurteilt Entlastungsschmerz
- palpiert Leber und Milz

Mögliche Stichworte zum Prüfungsziel U3 „vermeidet unnötige Unannehmlichkeiten“ wären etwa:

- positioniert den Patienten bequem
- palpiert mit Rücksicht auf die Schmerzen des Patienten
- erklärt sein Vorgehen vor oder während der Ausführung

*Begutachtung durch
die Prüfungs-
kommission*

Die bisherigen Arbeiten der Teams werden von der Prüfungskommission und/oder im Plenum anhand folgender Fragen begutachtet: Sind sie angemessen für eine Facharztprüfung? Entsprechen sie den gesetzten Vorgaben? Lassen sie sich durch SP verkörpern? Sind sie realistisch? Können sie in der gesetzten Stationenlänge dargestellt und von den Kandidaten bewältigt werden?

Eine alternative Möglichkeit ist, den Autoren je einige Aufträge bereits vor dem Workshop zu übermitteln. Sie können dann in ihren archivierten Patientendaten nach passenden Fällen suchen und eventuelle Unterlagen wie Röntgenbilder und konkrete Labordaten zum Workshop mitbringen.

Zweiter Schritt: Ausarbeitung

Wenn ein Fall als grundsätzlich geeignet taxiert worden ist, geht es darum, fallspezifische Informationen für die Auswahl und das Training des SP oder der SPs zusammenzutragen, die Kandidateninstruktion zu verfassen und eine konkrete Checkliste für die Beurteilung zu erstellen.

*Informationen für die
Rekrutierung von SPs*

Als erstes ist das demographische Profil der SPs festzulegen: Alter, Geschlecht, Beruf, schulischer Hintergrund. Dann ist die Erscheinung zu beschreiben: Statur, Kleidung, Gepflegtheit usw. Schliesslich erfolgen Angaben über erforderliche körperliche Besonderheiten, die sich nicht simulieren lassen sowie unverträgliche Merkmale z.B. Operationsnarben.

*Welches Profil und
welche Krankenge-
schichte hat der SP?*

Als Grundlage für das Training der SP muss als nächstes das Profil, das in den Anamnese- und Untersuchungsschemata des 1. Schrittes skizziert wurde, verfeinert werden. Es soll ein möglichst plastisches Bild entstehen.

Wurden im ersten Schritt unter Punkt V des Anamneseschemas (vgl. Anhang 1) drei Medikamente aufgelistet, die der Patient einnimmt, sollte jetzt spezifiziert werden, wer diese verschrieben hat, wie lange sie eingenommen werden, ob irgendwelche Nebenwirkungen festgestellt wurden und ob der Patient sich erinnert, die Medikamente wie verschrieben eingenommen zu haben. Solche Informationen erlauben dem SP, die meisten Fragen zu beantworten, welche von Kandidaten gestellt werden könnten.

Besonders wichtig sind detaillierte Angaben über den Gesundheitszustand und den Konsultationsgrund, da diese bei der Befragung und Untersuchung im Vordergrund stehen werden. Ist der Konsultationsgrund ein bestimmter Schmerz, sind Angaben erforderlich, die dem Patienten ermöglichen auf Fragen zu antworten wie: Wann hat der Schmerz begonnen? Wie lange dauerte er? Wie fühlt er sich an? Was mildert, was verstärkt ihn? Gab es bereits frühere Episoden? Die Angaben müssen in Laiensprache formuliert werden, damit eine realistische Simulation möglich wird.

Bei der medizinischen Vorgeschichte sowie der Sozial- und Familiengeschichte sollte man sich auf Informationen beschränken, die für das Prüfungsziel der Station bedeutsam sind. Jedes unnötige Detail macht das Training der SPs schwieriger.

Ebenfalls erforderlich sind Angaben über das Verhalten des Patienten während der Konsultation

Wie verhält sich der SP?

Patient ist freundlich, etwas gleichmütig. Akzeptiert, dass Untersuchungen durchgeführt werden müssen. Akzeptiert resigniert die Notwendigkeit einer Operation, scheint dies schon geahnt zu haben. Beantwortet Fragen, führt aber nicht von sich aus weiter aus. Patient sollte zu Beginn auf der Untersuchungsliege sitzen, nicht unter qualvollen Schmerzen aber mit der Tendenz, sich möglichst wenig zu bewegen. Zieht gerne die Knie hoch. Spricht den Arzt als „Herr Doktor“ an.

und die Reaktion auf unterschiedliches Kandidatenverhalten (interessiert, gleichgültig, distanziert, schroff).

Patient achtet den Arzt, ohne eingeschüchtert zu sein. Wenn er ohne angemessenes Engagement behandelt wird, kann er zeigen, dass er dies nicht schätzt, indem er etwas unfreundlicher wird.

Dann sind alle Fragen festzuhalten, die der SP den Kandidaten während der Konsultation stellen soll, um sie auf Checklistenpunkte hinzuführen. Ebenfalls hilfreich sind einige neutrale Fragen, die er stellen kann, um ein natürliches Gespräch in Gang zu halten.

Welche Fragen stellt der SP?

Wie lange werde ich im Spital bleiben müssen? Wann werde ich wieder arbeiten gehen können?

Schliesslich braucht es Hinweise, wie der Patient auf verschieden Antworten des Arztes reagieren soll, im Besonderen auch auf die Schlussfragen und -kommentare des Arztes.

Wie reagiert der SP?

Wenn der Kandidat entscheidet, ihn nach Hause zu schicken, fragt er: „Sind Sie sicher, dass ich heimgehen kann? Ich fühle mich wirklich ziemlich schlecht.“

Kandidateninstruktion

Die Kandidateninstruktion wird ausserhalb des Prüfungsraumes neben der Tür aufgehängt. Die Kandidaten lesen sie in der Pause vor der Station. Die Instruktion soll folgende Informationen enthalten: Name und Alter des Patienten, Konsultationsort, Konsultationsgrund, evtl. zuweisende Person oder Instanz, Aufgabe und zur Verfügung stehende Zeit. Bei Doppelstationen soll auch schon auf die Aufgabe in der Fragestation hingewiesen werden (vgl. Abb. 7.1).

Hans Ramseier
Kandidatenanweisung

Herr Hans Ramseier, 30-jährig, kommt auf die Notfallstation wegen Bauchschmerzen, die seit 4 Tagen andauern.

Untersuchungsbefunde:
Temperatur 37.9 °C
Blutdruck 128/84 mmHg
Atmung 16/Min.
Puls 92/Min., regelmässig

Sie haben 10 Minuten Zeit, um bei diesem Patienten eine Anamnese zu erheben und eine körperliche Untersuchung durchzuführen, die dem Problem angemessen sind.

In der folgenden Station werden Sie über dieses Problem einige Fragen zu beantworten haben.

Lassen Sie bitte diese Instruktion hier.

Abb. 7.1 Kandidatenanweisung zur Prozedurstation einer Doppelstation

Überweisungsbrief, Informationsblatt

Falls den Kandidaten ein Überweisungsbrief oder ein Blatt mit Informationen über den Patienten abgegeben werden soll, muss das entsprechende Dokument vor der Checkliste verfasst werden. Der Zweck eines Informationsblattes kann sein, die Kandidaten mit gewissen standardisierten Grundinformationen zu versehen.

erforderliche Einrichtung und Gegenstände

Es ist sinnvoll, dass die Autoren v.a. für die Prozedurstationen festhalten, welche Einrichtungsgegenstände (Untersuchungsliege, Tisch, Stühle) und Utensilien (Blutdruckmessgerät, Reflexhammer usw.) erforderlich sind.

Als letzter Schritt kann die Checkliste mit eventuellen Erläuterungen erstellt werden. Damit sie übersichtlich bleibt, sollte eine solche Liste nicht mehr als 20-25 Items enthalten. Es ist also wichtig, sich auf die Items zu konzentrieren, die für die Beurteilung der entsprechenden Fertigkeit/en entscheidend sind. Wenn in der Zielsetzung der Station der Untersuchung ein dreimal so grosses Gewicht gegeben wurde wie der Anamnese, soll sich dies in der Verteilung der entsprechenden Items widerspiegeln. Jedes Item soll sich lediglich auf eine Kandidatenfrage oder -handlung beziehen. Bei der Formulierung ist darauf Rücksicht zu nehmen, ob ein ärztlicher Examinator oder der SP selbst die Checkliste ausfüllen wird.

Zum ersten Item des Prüfungsziels U1 „palpiert Abdomen“ könnte eine Itemformulierung aus Patientensicht lauten:

„Kandidat führte in vier Quadranten meines Bauches eine tiefe Abtastung durch (drückte Hände tief hinein), wobei er entfernt von der Schmerzregion begann und diese zuletzt abtastete.“

Für ärztliche Examinatoren genügt:

„tiefe Palpation in vier Quadranten, wobei Schmerzregion zuletzt“

Wenn SP anamnestische Fragen beurteilen sollen, ist es sinnvoll, primär die für die Vergabe von Punkten gültige SP-Antwort als Item vorzugeben. Zur Erläuterung sollten aber auch noch mögliche Kandidatenfragen festgehalten werden, welche die Antwort hervorrufen können.

Zum Prüfungsziel A2 „erfragt mit dem Konsultationsgrund verbundene Symptome“ könnte ein Item mit Erläuterung lauten:

SP-Antwort: „Ich habe die letzten Tage, seit dies begonnen hat, keinen Appetit mehr gehabt.“

Kandidaten-Fragen: „Was haben Sie gegessen?“, „Waren Sie in letzter Zeit hungrig?“, „Haben Sie einen guten Appetit?“

Unter dem Aspekt der Gültigkeit ist denkbar, besonders wichtigen Items (Schlüsselkriterien) ein höheres Gewicht zu geben. Dem ist zu entgegen, dass man sich möglichst auf wichtige Items beschränken sollte. Die Messzuverlässigkeit kann nachgewiesenermassen durch eine Gewichtung nicht verbessert werden. Eine Fehlentscheidung bezüglich der Erfüllung eines doppelt gewichteten Items verdoppelt diesen Fehler. Wir raten nicht grundsätzlich von einer Gewichtung ab, weisen aber darauf hin, dass bei einer sorgfältig konstruierten Checkliste der Nettogewinn bescheiden ist.

Checkliste mit Items

Itemgewichtung

Diverse Untersuchungen haben gezeigt, dass globale Beurteilungen der Leistung bei einer strukturierten und standardisierten praktischen Aufgabe ebenso zuverlässig oder gar zuverlässiger sein können als diejenige anhand einer detaillierten Checkliste^{2,3}.

Globalbeurteilungen

Eine Möglichkeit ist, 4 bis 6-stufigen Lickert-Skalen einzusetzen. So wird etwa für die Beurteilung der technischen Fertigkeiten kanadischer Chirurgieassistenten anhand von Modellen (bench models) zusätzlich zu aufgabenspezifischen Checklisten die in Anhang 2 (S.172) gezeigte 5-stufige Skala verwendet⁴.

Ebenfalls sinnvoll kann es sein, von den Examinatoren abschliessend ein globales Gesamturteil zu verlangen, ob der Kandidat die Aufgabe insgesamt erfüllt hat, also in dieser Station bestehen soll oder nicht.

*Kommunikation und
Gestaltung der Arzt/
Patienten-Beziehung*

Einheitliche globale Skalen können auch für die Beurteilung diverser allgemeiner Verhaltensaspekte eines Kandidaten in der Begegnung mit einem Patienten verwendet werden. Diese sind erfahrungsgemäss wenig fall- oder problemspezifisch. Kommen bei einem bestimmten Aspekt verschiedene Examinatoren in verschiedenen Stationen zu sehr ähnlicher Beurteilung, kann das Urteil als zuverlässig erachtet werden. Als mögliches Beispiel zeigen wir in Anhang 3 (S.173) die entsprechende Checkliste, welche aktuell im kanadischen Qualifizierungsexamen verwendet wird.

In Anhang 4 (S.174) finden Sie eine vollständige Dokumentation zu einer Anamnesedoppelstation mit einem Anämie-Patienten. Es handelt sich dabei um ein leicht modifiziertes Beispiel aus der kanadischen Qualifikationsprüfung.

Dritter Schritt: praktische Erprobung

Die Hauptrolle im dritten Schritt spielen die Patiententrainer. Die neu kreierten Patientenszenarien werden mit SPs eingeübt, sofern bereits solche vorhanden sind, notfalls auch mit andern Anwesenden. Es können Videoaufnahmen gemacht werden. Die beim Training gemachten Erfahrungen und die gemeinsame Analyse der Videoaufnahme kann zu Modifikationen des Patientenporträts und/oder Checklistenitems führen. Falls die Station und die Darstellung als gut erachtet werden, kann das Video zu Instruktionzwecken dienen, z.B. für das Training von SPs oder Examinatoren.

Alle schriftlichen Unterlagen werden am Schluss durch die Prüfungskommission redigiert.

7.3 Wie werden Examinatoren vorbereitet?

Examinatoren werden sowohl zur Beurteilung in den Prozedurstationen benötigt wie auch zur Auswertung der Antwortblätter von den Fragestationen, sofern dort mit Kurzantwortfragen gearbeitet wird. Grosse Fachgesellschaften mit vielen Kandidaten pro Prüfungssession rekrutieren für diese beiden Funktionen sinnvollerweise getrennt Examinatoren. Bei kleineren Kandidatenzahlen ist es zumutbar, dass Examinatoren, welche während der Prüfung an Prozedurstationen gearbeitet haben, einige Tage nach der Prüfung nochmals zusammenkommen, um die Antwortblätter von den Fragestationen auszuwerten.

Beurteiler von praktischen und schriftlichen Leistungen

Nachgewiesenermassen ist es nicht nötig, die Kandidatenleistung an einer Prozedurstation gleichzeitig von zwei unabhängigen Examinatoren beurteilen zu lassen. Der Gewinn ist wesentlich höher, wenn mehr Stationen mit je einem Examiner besetzt werden als weniger Stationen mit je zweien¹. Zu berücksichtigen ist aber, dass die Arbeit hohe andauernde Konzentration, gute Hörleistungen und Stressresistenz erfordert. Es ist nicht möglich, länger als einen halben Tag entsprechend zu arbeiten. Wenn ein OSCE wegen der Kandidatenzahl einen ganzen Tag dauert, müssen deshalb für jede Station zwei Examinatoren vorbereitet werden.

Wie viele und mit welchen Eigenschaften?

Untersuchungen haben gezeigt, dass für gewisse Beurteilungen nach entsprechendem Training auch Personen ohne ärztliche Ausbildung eingesetzt werden können, so z.B. SPs¹.

Bei den Fragestationen sollte jeder Auswerter nur die Antwortblätter einer einzigen Station bearbeiten, um den Einfluss systematischer Beurteilungstendenzen möglichst klein zu halten. Es sind also mindestens so viele Auswerter erforderlich wie es Fragestationen gibt.

Der hauptverantwortliche Auswerter muss

hauptverantwortlicher Auswerter

- sicherstellen, dass am Prüfungstag alle erforderlichen Unterlagen vorhanden und in einwandfreiem Zustand sind
- die Auswerter über ihre Aufgabe informieren
- die Auswertungsarbeit laufend überwachen
z.B. Werden fälschlich richtige Antworten nicht gewertet oder nicht vorgesehene Antworten gewertet?

- die von den Auswertern bearbeiteten Antwortblätter kontrollieren, wenn diese am Ende zu ihm gebracht werden
- die Geheimhaltung des Prüfungsmaterials z.B. Antwortschlüssel, Prüfungsunterlagen sicherstellen

*Instruktions-
veranstaltung*

Aus Sicherheitsgründen sollten sowohl die Examinatoren für Prozedur- wie für Fragestationen erst am Prüfungstag mit ihrer konkreten Station vertraut gemacht werden. Es ist aber sinnvoll, sie im Voraus, z.B. an einer 1^{1/2} - 2-stündigen Abendveranstaltung, auf ihre Aufgaben vorzubereiten.

*Instruktionen für
Beurteiler an
Prozedurstationen*

Wichtige Punkte, die Examinatoren an Prozedurstationen wissen müssen, sind:

- Wie sieht eine Checkliste aus und wie wird sie ausgefüllt?
Ideal ist es, wenn ein Demonstrationsvideo einer Station existiert, die nicht in der Prüfung eingebaut ist und anhand derer die Beurteilung mittels Checkliste durchgespielt werden kann.
- Welche Stationentypen, Checklisten und Beurteilungsskalen gibt es?
Stationen, bei denen zum einen die spezifische Aufgabe (z.B. Anamnese, Beratung) anhand einer Checkliste zu evaluieren ist und zusätzlich die Kommunikation/Interaktion anhand einer Beurteilungsskala; Doppelstationen, bei denen am Ende die Fragen für die Fragestation ausgehändigt werden müssen; Langstationen; Langstationen, bei denen am Schluss mündlich Fragen gestellt werden müssen usw.
- Die Bestehensanforderungen für verschiedene Stationen sind unterschiedlich (evtl. sind auch die einzelnen Items verschieden gewichtet). Examinatoren können (und sollen) nicht wissen, wie viele (und welche) Items erfüllt sein müssen, damit ein Kandidat eine Station bestanden hat.
- Falls eine globale Beurteilung „nicht bestanden/bestanden“ abgegeben werden soll, muss dies unbedingt bei allen Kandidaten erfolgen. Das Urteil soll nicht aus der Summe der erfüllten Checklistenpunkte abgeleitet werden, sondern wirklich den persönlichen Gesamteindruck des Examinators über die Kandidatenleistung wiedergeben.
- Es ist sinnvoll, vor Beginn der Prüfung die Checkliste gemeinsam mit dem SP durchzugehen; dies verbessert sowohl die Beurteilung wie die SP-Leistung.

- Für das Ausfüllen der Checkliste ist das Kandidatenverhalten relevant und nicht dasjenige des SP.

Stellt ein Kandidat eine sehr offene Frage und der SP gibt darauf unnötigerweise Informationen preis, welche nach der Checkliste eine gezielte Frage des Kandidaten erfordert hätten, darf das entsprechende Item nicht als erfüllt gewertet werden.

- Während der laufenden Kandidaten/SP-Interaktion darf der Examinator nicht eingreifen. Dies kann eine kritische Situation nur verschlimmern, da zum Störeinfluss des SP noch derjenige des Examinators hinzukommt.

Wenn ein Examinator beim ersten Kandidaten das Gefühl hat, der SP gebe ohne gezielte Kandidatenfragen zuviel Information preis, kann er diesen in der Pause darauf hinweisen. Andere Rollenänderungen sind aber nicht zulässig. Gibt es diesbezüglich ein ernstes Problem, muss der zuständige Patiententrainer beigezogen werden.

Instruktionen für die Auswertung von Kurzantwortfragen sind in Kapitel 4 gegeben. Zur Illustration mag das Beispiel einer Fragestation in Anhang 4 dienen. Instruktionen dieser Art sollten den Auswertern in schriftlicher Form ausgehändigt aber auch im Rahmen einer gemeinsamen Information mit diesen besprochen werden. Falls auch bei der Fragestation eine Globalbeurteilung verlangt wird, muss zusätzlich darüber instruiert werden.

*Instruktionen für
Auswerter der
Antwortblätter von
Fragestationen*

Am Tag der Prüfung soll sowohl für Examinatoren wie für Auswerter von Antwortblättern noch einmal eine ca. halbstündige Instruktion stattfinden, um die wichtigsten Punkte in Erinnerung zu rufen.

7.4 Wie Standardisierte Patienten rekrutieren und trainieren?

Was sind SP? Bereits anfangs der 60er Jahre wurde begonnen, für Prüfungs- und Ausbildungszwecke gesunde Personen auf eine oder mehrere Patientenrollen zu trainieren^{5,6}. Für solche SPs sprechen vor allem drei Gründe: Zum einen bietet es in der Regel erhebliche Probleme, am Prüfungstermin echte Patienten mit den erforderlichen Krankheitsbildern zur Verfügung zu haben. Zum andern ist es Patienten kaum zumutbar, sich nacheinander von einer grossen Zahl von Kandidaten befragt oder untersucht zu lassen. Schliesslich - und für die Prüfungsabsicht besonders wichtig - ist ein konstantes, standardisiertes Verhalten echter Patienten gegenüber allen Kandidaten kaum zu gewährleisten.

Im Unterschied zu einem Rollenspiel „verkörpern“ SPs einen bestimmten Patiententyp vollständig realistisch und zuverlässig. Das heisst, Alter, Geschlecht und körperliche Erscheinung stimmen, die Anamnese, die Persönlichkeit, die emotionalen Reaktionsweisen und sogar die körperlichen Befunde werden getreu wiedergegeben. Alternativ zur Bezeichnung „Standardisierte Patienten“ findet man auch die Begriffe „simulierte Patienten“ oder „programmierte Patienten“.

Was ist simulierbar? Zu den simulierbaren Befunden gehören:^{5,6}

Aphasie	Ganganomalie	Muskelschwäche
Babinskireflex	Gelenkbewegungseinschränkung	Parkinson
Bauchschmerzen	Hörverlust	Petechien
Bewusstlosigkeit	Hypomanie	Pneumothorax
blauer Fleck, Druckstellen	Kernigsches Zeichen	Schwitzen
Brudzinkizeichen	Kosto-vertbrale Gelenkschmerzen	Steifigkeit
Cheyne-Stokes-Atmung	Kurzatmigkeit	Tremor
Druckschmerz	Lichtempfindlichkeit	Verwirrung
Fazialisparese	Loslass-Schmerz	Visusverlust

Bei der Rekrutierung von SPs muss auf die erforderlichen äusseren Merkmale und eventuell vorausgesetzte nicht simulierbare körperliche Befunde geachtet werden. Alles Weitere wird im Training erarbeitet. In diversen Untersuchungen wurde nachgewiesen, dass SPs von Ärzten nicht als solche erkannt werden können⁷.

Wer eignet sich als SP? Allgemein eignen sich Leute mit guter Kommunikationsfähigkeit einschliesslich der Fähigkeit gut zuhören zu können, die eine gewisse Einsicht in eigene Verhaltensweisen haben. Zudem sollten sie ein Interesse an medizinischen Fragen haben und es als wertvoll ansehen, einen Beitrag zur Qualitätskontrolle künftiger Fachärzte leisten zu können. Weite-

re wichtige Qualitäten eines SP sind Bereitschaft, sich ohne Hemmungen untersuchen zu lassen sowie eine Portion Humor. Hilfreich, aber nicht gezwungenermaßen notwendig sind schauspielerische oder medizinische Erfahrungen.

Schauspieler konzentrieren sich manchmal zu sehr aufs Spielen und nehmen zuwenig Rücksicht auf die Checkliste und die unterschiedlichen Verhaltens- und Vorgehensweisen der Kandidaten. Medizinisches Personal hat manchmal Mühe, die Sicht der Patienten zu verstehen und sich unvoreingenommen zu verhalten. Primär wichtig ist auch hier, dass die oben erwähnten persönlichen Qualitäten stimmen, sonst nützt der berufliche Hintergrund nichts oder wirkt sich gar negativ aus.

Bei der Auswahl von SPs müssen auch die Sicherheits- und Vertraulichkeitsaspekte des Exams im Auge behalten werden. Es ist z.B. sicher nicht zu empfehlen, jemanden als SP einzustellen, der eine persönliche Beziehung zu einem Kandidaten hat.

Rekrutierung

Die Selektion von SPs ist praktisch ebenso wichtig wie das spätere Training derselben. Das Such- und Auswahlprozedere ist zeitaufwendig, deshalb sollte damit begonnen werden, sobald die erforderlichen Patientenprofile bekannt sind.

Erfolgreiche Suchstrategien sind:

- persönliche, mündliche Kontakte, „Mund zu Mund Propaganda“
- Ausschreibung in einer Spital- oder Arzt-Zeitschrift oder Anschläge an Steckwänden
- Anschreiben spezieller Gruppen: Schauspielgruppen, Seniorenvereinigungen, freiwillige Spitalhelfer usw.
- kurze Ausschreibung in einer Tageszeitung; dabei sind viele Reaktionen zu erwarten und eine sorgfältige Vorauswahl am Telefon zu planen.

*erfolgreiche
Suchstrategien*

Potentielle Interessenten müssen erfahren,

- wer SPs sucht und wo man sich anmelden muss
- worin kurz gesagt die Aufgabe besteht
- dass ein 4-8-stündiges Training erforderlich ist, das flexibel organisiert wird
- wann und wo die SPs verfügbar sein müssen (Prüfungsdaten)
- wie Trainings-, Einsatzzeiten und Spesen vergütet werden
- welche Art von Leuten gesucht werden (Geschlecht, Alter usw.)
- dass keine Erfahrung erforderlich ist, aber Personen gewünscht sind, die gut kommunizieren können und bereit sind, etwas Neues zu versuchen

*notwendige Informa-
tionen an SPs*

notwendige Informationen von SPs

Die schriftlich von den Interessenten um eine Anstellung als SP erhobenen Informationen können die Vorauswahl erleichtern. Wichtige Fragen sind:

- Name, Adresse, Telefon
- Geburtsdatum, Geschlecht
- Wie haben Sie von dieser Arbeit als SP erfahren?
- Was interessiert Sie daran?
- Kennen Sie Kandidaten, welche diese Prüfung absolvieren werden?
- Welche Verhaltens- oder Gefühlsart denken Sie am besten simulieren zu können?
- Wie würden Sie Ihre persönliche Art beschreiben (z.B. gelassen, redselig, ruhig, gebärdereich)
- Geben Sie bitte Ihre momentane Anstellung und relevante Arbeitserfahrungen an.
- Beschreiben Sie kurz andere frühere Erfahrungen, von denen Sie denken, dass sie Ihnen helfen würden, ein Patientenproblem zu simulieren.
- Um uns bei Ihrer Zuordnung zu einem Patientenproblem zu helfen, geben Sie uns bitte folgende Informationen:
Grösse: Gewicht:
- Führen Sie alle wahrnehmbaren körperlichen Befunde auf wie Herzgeräusche, Operationsnarben, Diabetes, Hör- oder Sehverlust.

Informationsveranstaltung für grössere Gruppen

Wenn etliche SPs gefunden werden müssen und sich auch viele auf die Ausschreibung gemeldet haben, ist es sinnvoll, zu einer abendlichen Informationsveranstaltung einzuladen. Daran sollte möglichst auch ein SP-Trainer teilnehmen.

In einem ersten Teil müssen Informationen über die Prüfung, das Training und die Rolle von SPs gegeben und Fragen beantwortet werden. Hilfreich ist natürlich, wenn eine SP-Demonstration gegeben werden kann (evtl. auch ab Video). Im zweiten Teil sollten von Personen, die sich ernsthaft für die Tätigkeit interessieren, gewisse grundlegende Informationen erhoben und ein Foto aufgenommen werden. Abschliessend muss informiert werden, wie es weitergeht.

Auswahlinterview

Unabhängig davon, wie potentielle SPs gefunden werden, muss mit allen vor Beginn des Trainings ein Gespräch geführt werden. Dabei sind obige Fragen aus dem Bewerbungsformular zu vertiefen. Weitere Fragen können sein:

- *Gibt es Themenbereiche, in die Sie nicht einbezogen werden möchten (z.B. Abtreibung, terminale Erkrankung)?*
- *Wie fühlen Sie sich, wenn Sie körperlich untersucht werden?*

Diese Frage ist vor allem wichtig für die Fallzuteilung, evtl. wird sie darum besser noch zurückgestellt. Es ist in der Regel weniger bedrohlich, wenn sich die Frage auf eine ganz bestimmte Untersuchung bezieht, als wenn sie allgemein beantwortet werden soll.

Aufschlussreich können kurze „Spielproben“ der Bewerber sein. Man macht einen kurzen Trainingslauf mit einem einfachen Problem, z.B. Kopfschmerzen. Die Anwärter erhalten eine schriftliche Anweisung und ein wenig Zeit, um sich vorzubereiten und spielen dann die Patientenrolle mit dem Interviewer in der Kandidatenrolle. Dies erlaubt zu beobachten, wie sie einsteigen, ob sie fähig sind, die Kopfschmerzen darzustellen, wie gut sie die Instruktionen befolgen, kurz, wie leicht sie trainierbar sein werden.

kurze Testläufe

Sind mehr (gute) Bewerber vorhanden als SPs benötigt werden, ist es sinnvoll, einige als Reserve-SPs oder als Hilfspersonal einzusetzen. Wichtig ist, dass jeder neue SP die Unterlagen zu den Aufgaben, Rechten und Pflichten (z.B. bezüglich Vertraulichkeit sämtlicher Examensinformationen) vor dem ersten Training erhält und unterzeichnet.

überzählige SPs

SP-Training

Im SP-Training wird nach einführenden Erläuterungen zuerst ein Trainingsplan festgelegt. Sinnvoll sind in jedem Fall zwei Proben und eine Hauptprobe kurz vor dem Prüfungstermin.

Trainingsplanung

Um eine standardisierte Darstellung zu gewährleisten, müssen die SPs genügend Gelegenheit haben, ihre Rolle zu üben, zu spielen und Feedback zu erhalten. Für komplexe Fälle ist es sinnvoller, mehr und evtl. kürzere Proben vorzusehen, da viel Information besser verteilt aufgenommen werden kann. Je nach Fall sollten insgesamt 4 - 8 Std. ausreichen.

Beim Training des konkreten Falles geht es um die Vermittlung dreier Komponenten: die persönliche Eigenart und Gefühlslage des Patienten, die klinischen Informationen und Symptome und die Übereinstimmung der Darstellung mit den Beurteilungspunkten der Checkliste.

Was wird trainiert?

Wenn von der Zahl der Stationen und der zu prüfenden Kandidaten her entschieden wurde, die Stationen auf zwei Circuits aufzuteilen, die an zwei aufeinanderfolgenden Tagen stattfinden, können SPs für zwei verschiedene Stationen vorbereitet werden. Es muss aber darauf geachtet werden, dass sich die Stationen möglichst stark unterscheiden (z.B. eine Anamnese- und eine Untersuchungsstation aus unterschiedlichen Inhaltsbereichen). Dies ist schlecht möglich, wenn die beiden Circuits am gleichen Tag durchgeführt werden. Ein SP kann kaum länger als einen halben Tag eine volle, konzentrierte Leistung bieten.

Fragen nach Zeichen und Symptomen, die im Skript nicht vorgesehen sind, soll der SP negativ beantworten, im Sinne von „nicht vorhanden“. Bei Fragen zur Familienanamnese soll, sofern nicht im Skript etwas Konkretes vorgegeben ist, etwas völlig Unspektakuläres, Langweiliges angegeben werden. Auf offene Fragen soll allgemein sehr zurückhaltend reagiert werden.

allgemeine Regeln für SPs

Es kann eine Information gegeben werden, die der Kandidat schon in der schriftlichen Anweisung erhalten hat, oder der Kandidat kann gebeten werden, näher zu erklären, was er genau wissen möchte. Wenn keine dieser Reaktionen passt, sollte der SP auf jeden Fall nur Information geben, die ein einziges Item der Checkliste betrifft. Fragt der Kandidat z.B., ob der Patient in letzter Zeit irgendwelche weitere Veränderungen bemerkt habe und es einige gibt, soll der SP nur eine angeben. Sinnvoll ist natürlich, wenn Kriterien, wie zu reagieren ist, bereits im Skript vorgegeben sind, z.B.:

„Dies sind Symptome, welche Sie bemerkt haben, die Sie aber nicht mit der Hauptbeschwerde in Beziehung bringen. Geben Sie diese Auskunft nur auf gezielte Fragen.“

erste Trainingssession

In der ersten spezifisch auf den darzustellenden Fall ausgerichteten Trainingsphase erhalten die SPs vor dem „Spielen“ ihrer Rollen das entsprechende schriftliche Skript und besprechen dieses mit dem Trainer.

Es geht darum, dass sie sich die ganze Krankheitssituation von der Entstehung der Symptome bis zum Moment der Konsultation genau vorstellen und sich in den Patienten einfühlen können. Besonders hilfreich ist es, wenn SPs gewisse Zeichen und Symptome aus Erfahrung am eigenen Leib oder bei einem Bekannten kennen. Beim Erarbeiten der Darstellung ist es sinnvoll, jeweils nur einen kleinen Ausschnitt, z.B. Antworten auf einige wenige Checklistenpunkte zu proben. Dies hilft den SPs, die Informationen gemäss der Checkliste aufzugliedern. Wenn zwei SPs für die gleiche Rolle vorbereitet werden, können sie sich im Spielen abwechseln. Sie können dabei voneinander lernen und einander gegenseitig Feedback geben.

Selbststudium und -training

Wenn klar ist, dass die SPs das Problem verstanden haben und auch die Art, wie es darzustellen ist, müssen sie das Fallmaterial nach Hause nehmen, um dort für sich selbst zu üben.

Der SP-Trainer muss Übungsvorschläge liefern, die für den Fall spezifisch sind und den Fähigkeiten der SPs entsprechen.

Sinnvolle Übungsstrategien sind:

- die Geschichte in ein Format oder eine Reihenfolge umschreiben, die ihnen liegt
- die Geschichte ohne Notizen laut vor sich hersagen, resp. mit sich selber reden wenn man allein ist, z.B. auf einem Spaziergang (Sicherheitsvorkehrungen nicht vergessen!)
- beim Umziehen im Schlaf- oder Badezimmer kurz die Darstellung eines Symptoms üben
- in diversen kleinen Dosen lernen statt alles kurz vor der nächsten Trainingssession in sich hineinzustopfen

zweite Trainingssession

Zu einer vorher bestimmten Zeit kehren die SPs für eine formale Durchsicht des Falles zurück. Der ursprüngliche Trainer sollte anwesend sein, beobachten und jedem individuell Feedback geben.

Um dies zu ermöglichen, muss ein anderer SP-Trainer oder Examinator die Rolle des Kandidaten übernehmen. Die Übungsrunden müssen nicht lang sein, aber die SPs sollten ein Gefühl für die Prüfungslänge ihrer Station bekommen. Allerdings sollten die Übungsrunden mehr vom Inhalt als vom Timing geprägt sein. Jeder SP muss die Chance erhalten, auf die ganze Auswahl von Fragen zu antworten und auf die körperlichen Un-

tersuchungen zu reagieren, denen sie möglicherweise begegnen. Eine Übungsrunde für eine 10-Minuten-Station kann also 15 Minuten dauern. SPs werden nie mit allen Eventualitäten einer Begegnung konfrontiert werden können. Sie müssen aber eine Ahnung haben, wie sie auch auf unerwartete Fragen oder Techniken reagieren können.

Im Laufe des Trainings ist es nötig, über die Kandidaten zu sprechen.

Diese sind angespannt und die meisten nehmen wahrscheinlich erstmals an einem OSCE teil. Die SPs können sie an die Aufgabe heranzuführen und ihnen helfen, in ihrer Rolle zu bleiben. Wenn Kandidaten richtig in das Szenario „einsteigen“, können sie wahrscheinlich ihre Fertigkeiten gut unter Beweis stellen. Wenn ein SP eine Kandidatenfrage nicht versteht, muss er nachfragen, so wie er es in einer Klinik tun würde. Es wäre aber unangebracht, ungeduldig oder rüde zu werden.

Wenn mehr als ein SP für ein und dieselbe Rolle vorbereitet worden sind, sollten bei der Hauptprobe, alle anwesend sein, denn dies ist der entscheidende Schritt bei der Standardisierung der Falldarstellung.

Hauptprobe

Der Kandidat soll diesmal z.B. von einem Mitglied der Prüfungskommission verkörpert werden. Der Fall soll 1:1 so durchgespielt werden wie an der Prüfung. Das heißt, dass die richtigen Einrichtungsgegenstände und Untersuchungsutensilien vorhanden sein müssen, dass der SP richtig ausgestattet ist (Kleidung, Frisur usw.), dass der „Kandidat“ auf ein akustisches Signal den Raum betritt, die Aufgabe nach den Kandidateninstruktionen erfüllt und nach dem Schlussignal den Raum wieder verlässt.

Es ist zu empfehlen, dass der „Kandidat“ bei einem ersten Durchlauf nur die Kandidateninstruktion kennt und noch nicht die Checkliste, da dies sein Verhalten sonst zu sehr steuern und verschiedene Durchgänge zu ähnlich machen kann. In weiteren Durchgängen darf der „Kandidat“ den SP mit verschiedenem, auch unangenehmem Verhalten auf die Probe stellen. Er sollte sich z.B. kalt, schroff geben oder nervös, desorganisiert. Er sollte variieren zwischen verschiedenen Frageformen (offene, geschlossene, Mehrfach-Fragen) und dabei prüfen, wie gut der SP checklistenrelevante Informationen zurückhalten kann, bis er spezifisch danach gefragt wird. Er sollte längere Pausen entstehen lassen und darf auch versuchen, den SP durch überraschende Aktionen aus seiner Rolle zu bringen.

Wenn zur Station eine körperliche Untersuchung gehört, muss der „Kandidat“ gewährleisten, dass der SP alle Befunde zeigt. Er sollte verschiedene Untersuchungstechniken anwenden, um den SP darauf vorzubereiten, sollte ihn während einer Untersuchung mit einem Gespräch ablenken und auch unübliche oder falsche Techniken anwenden, um zu schauen, ob er in seiner Rolle bleibt.

Andere SPs, der Trainer und jemand von der Prüfungskommission schauen zu und geben Feedback, möglichst mit konkreten Beispielen.

Konstruktives Feedback

Sehr gut war, wie Sie die Stimmung im richtigen Moment wechselten. Sie schüttelten die Hand bei der Begrüßung zu fest für eine unterwürfige Person. ...

Um die Authentizität der Darstellung zu überprüfen, ist die Anwesenheit eines Facharztes wünschbar, der mit dem Fallproblem gut vertraut ist, z.B. der Hauptautor der Station.

7.5 Wie wird ein Circuit angelegt?

Grundprinzipien für die Circuitplanung

Sobald bekannt ist, ob alle Stationen in einem Circuit angelegt oder auf zwei aufgeteilt werden sollen, kann nach geeigneten Prüfungsräumlichkeiten Ausschau gehalten werden. Gut bewährt haben sich Polikliniken mit Kojen, da diese räumlich beieinander liegen und für etliche Prüfungsaufgaben schon eine geeignete Grundausstattung aufweisen. Allerdings bedeutet dies in der Regel, dass die Prüfung an einem Wochenende stattfinden muss.

Langstationen (von z.B. 14 Minuten) können in einem Circuit mit Kurzstationen (von 7 Minuten) gemischt werden. Allerdings müssen die Langstationen dann doppelt geführt werden, da nach 7 resp. 8 Minuten der nächste Kandidat von der vorangehenden Kurzstation vor der Tür steht.

Sind grössere Kandidatenzahlen zu prüfen, sollten Circuits mit Langstationen möglichst doppelt geführt werden. Damit kann auch besser gewährleistet werden, dass nicht die Kandidaten der vierten oder fünften Schicht Informationen über einzelne Stationen mitbekommen, die ihre Leistung verbessern können.

Dann ist auch an die notwendigen Pausen, sowohl für Kandidaten wie für SPs und Examinatoren zu denken.

Sofern ein Circuit aus Stationen einheitlicher Länge besteht, kann er einfach nach ca. 1¹/₂-2 Stunden einmal für eine längere Pause unterbrochen werden. Dies ist nicht möglich, wenn er sowohl Kurz- wie Langstationen enthält. In diesem Fall sind für die Kandidaten im Circuit Ruhestationen einzuplanen (vgl. Anhang 5, S.177). Damit auch SPs und Examinatoren ab und zu eine längere Pause haben (Trinken, Snack, WC), kann man z.B. an zwei Stationen des Circuits keinen Kandidaten starten lassen. Der Zeit- und Pausenplan muss dann aber allen genau bekannt sein und stimmen.

erforderliche Räume

Für jeden Circuit ist ein Plan zu zeichnen, welche Station in welchem Raum stattfindet. Nicht zu vergessen sind Räume für die Registration und Information der Kandidaten, der Examinatoren, der SPs und des Hilfspersonals, für die Verpflegung sowie Umkleieräume für die SPs. Bei zwei Kandidatenschichten muss darauf geachtet werden, dass sich die beiden Gruppen beim Übergang nicht begegnen.

Einrichtung und Materialien

Für jede Station muss eine Checkliste der nötigen Einrichtungs- und Untersuchungsgegenstände vorliegen (z.B. Untersuchungsfläche, Kissen, 3 Stühle, Blutdruckmessgerät, Stethoskop). Es ist empfehlenswert, die Einrichtung in einer Skizze festzuhalten.

Die kleineren Utensilien sind eine Quelle möglicher Probleme. Wenn eine Station z.B. für die Beurteilung einer körperlichen Untersuchung mit Stethoskop und Reflexhammer ausgerüstet wird, packt plötzlich eine Kandidat während der laufenden Prüfung diese

Utensilien unabsichtlich in seine Tasche. Der nächste Kandidat kann die Aufgabe nicht lösen und schon droht der Circuit auseinanderzufallen. Abhilfe: Alle Kandidaten erhalten vor der Prüfung die Anweisung, diese Untersuchungshilfsmittel selber mitzunehmen, oder eine Hilfsperson kontrolliert bei jedem Postenwechsel, ob die Station noch vollständig ausgerüstet ist.

Zur Einrichtung gehört auch eine klare Beschilderung des Circuits und der einzelnen Stationen sowie die Kandidateninstruktionen, welche möglichst neben der Tür aufzuhängen sind, damit ein Kandidat diese lesen kann, auch wenn die Türe geöffnet wird.

Der Stationenwechsel muss absolut koordiniert erfolgen. Es ist deshalb ein zentral gesteuertes System erforderlich, welches die Wechsel akustisch anzeigt. Besonders bei kurzen Stationen kann es für die Kandidaten hilfreich sein, wenn eine halbe Minute vor dem Ende ein Warnton gegeben wird.

akustisches Zeichensystem

Für alle an der Durchführung Beteiligten sind detaillierte Zeitpläne ihres Einsatzes zu erstellen und Checklisten, was sie bei sich haben müssen resp. was im Raum, für den sie verantwortlich sind, vorhanden sein muss.

Checklisten, Zeitpläne

Die Einrichtung des Circuits sollte am Vorabend der Prüfung bereit sein und vom Durchführungsverantwortlichen kontrolliert werden. Bevor am Morgen der Circuit gestartet wird, ist anhand der Checklisten noch einmal eine Kontrolle durchzuführen. Dazu gehört auch die Überprüfung des akustischen Systems. Dies mag schon fast zwanghaft erscheinen, aber der Teufel steckt beim OSCE im Detail. Eine kleine Panne kann einen Circuit zum Erliegen bringen.

Kontrolle

Es ist deshalb auch sinnvoll, „trouble shooters“ zur Verfügung zu haben, welche eine eventuelle Panne möglichst schnell beheben können. Vor allem bei längeren Circuits gibt es ab und zu auch Kandidaten, die dem Stress nicht gewachsen sind und „ausrasten“ oder nicht mehr weiter wissen. Auch dafür ist es gut, eine Betreuungsperson vorzusehen.

Retter in der Not

In Anhang 5 ist als Beispiel ein Circuit des „Examen de certification du Collège des médecins de famille du Québec“ aufgezeichnet.

7.6 Wie wird ein OSCE ausgewertet?

Auswertungsziele Die aus der Prüfungsauswertung gewonnenen Daten sollen primär einen Entscheid ermöglichen, welche Kandidaten über ungenügende praktische Kompetenzen verfügen, um als Facharzt zertifiziert werden zu können. Sie sollen aber zudem ermöglichen, die Qualität der Prüfung zu evaluieren, Prüfungsstationen oder auch einzelne Checklistenitems zu identifizieren, welche keine Trennschärfe aufweisen und in dieser Form nicht wiederverwendet werden sollten.

elektronische Datenerfassung Diese Prüfungsevaluation bedingt, dass nicht nur die Summenwerte der Stationen, sondern die Werte aller einzelnen Items für eine statistische Verarbeitung mittels Computer erfasst werden. Dafür ist es sinnvoll, wenn immer möglich mit elektronisch lesbaren Check- und Beurteilungslisten zu arbeiten. Dies eliminiert die Fehlermöglichkeiten bei einer manuellen Dateneingabe und ist auch ökonomischer.

inhaltsorientierte Standardsetzung Wie in Kapitel 8 begründet, sollte die Standardsetzung inhaltsorientiert erfolgen. Dafür kommt das Angoff-Verfahren in Frage (vgl. Abb. 8.2, S.189), das bereits vor dem Prüfungstermin durchgeführt werden kann oder Grenzgruppen- oder Kontrastgruppen-Verfahren, welche das Vorliegen der ausgefüllten Checklisten erfordern und damit erst nach gelauener Prüfung durchführbar sind. Mit all diesen Verfahren wird die Bestehensgrenze für jede einzelne Station bestimmt.

Entscheid über Kompensationsmöglichkeiten Ein wichtiger Entscheid betrifft die Frage, ob ungenügende Leistungen an einzelnen Stationen durch sehr gute Leistungen an anderen kompensiert werden können oder nicht. Wird dies verneint, ist zusätzlich zur Standardsetzung für die einzelnen Stationen eine Entscheidung darüber erforderlich, wie viele Stationen minimal bestanden werden müssen, damit das gesamte OSCE als bestanden gewertet werden kann. Eine besondere Überlegung erfordert die Bewertung von Doppelstationen. Mit Sicherheit sind sie thematisch voneinander abhängig und wahrscheinlich ist die Leistung an der Fragestation auch von der Leistung an der vorangehenden Prozedurstation abhängig. Es dürfte daher sinnvoll sein, Doppelstationen als eine Station zu werten.

Gewichtung der Stationen Wenn eine gegenseitige Kompensationsmöglichkeit als sinnvoll erachtet wird, können im Prinzip einfach die für das Bestehen der einzelnen Stationen erforderlichen Minimalpunktzahlen zusammengezählt werden. Dies ergibt die Bestehensgrenze für die Gesamtprüfung. Problematisch wird dies dann, wenn Stationen deutlich unterschiedliche Maximalpunktzahlen aufweisen und damit verbunden wahrscheinlich auch un-

terschiedliche Punktzahlen für das Bestehen. Bei einer einfachen Punkte-
summierung wirken in diesem Fall die Stationen mit unterschiedlichem
Gewicht auf die globale Bestehensgrenze ein. Dies kann verhindert wer-
den, indem mit Prozentwerten gerechnet wird, wie in folgendem Bei-
spiel anhand der Kombination zweier Stationen mit ungleicher Itemzahl
gezeigt wird.

	Station 1	Station 2	Total
Maximalpunktzahl:	40	20	60
Bestehensgrenze absolut:	24	15	39
Bestehensgrenze in %:	60	75	67.5
			(entspricht 40.5 Punkten)

Dieses Prozedere kann auch angewandt werden, um gezielt gewissen Sta-
tionen ein grösseres Gewicht zu geben als anderen (z.B. Lang- vs. Kurz-
stationen, Prozedur- vs. Fragestationen, essentielle vs. weniger wichtige
Prozeduren).

Es ist natürlich auch möglich, eine kombinierte Bestehensanforderung
zu stellen: z.B. „Es müssen mindestens 16 der 22 Stationen bestanden
und in der Gesamtprüfung mindestens 350 Punkte erzielt werden“.

Soll im Interesse der Qualitätssicherung die Zahl der Kandidaten, wel-
che trotz ungenügender Kompetenz bestehen (falsche Positive), möglichst
klein gehalten werden, kann die Bestehensgrenze in der Gesamtprüfung
noch um den Standardmessfehler (se) oder gar um ein 5 %-Vertrauens-
intervall (1.96 se) erhöht werden. Dies bedingt allerdings, dass die
Prüfungsreliabilität ermittelt wird.

Sicherheitsmargen

René Krebs

Literatur

- ¹ van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58-76
- ² van der Vleuten CPM, Norman GR, de Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ* 1991;25:110-18
- ³ Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-97
- ⁴ Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skills via an innovative "bench station" examination. *Am J Surg* 1997;173:226-30
- ⁵ Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ* 1964;39:802-5
- ⁶ Barrows HS. Simulated (standardized) patients and other human simulations: A comprehensive guide to their training and use in teaching and evaluation. Chapel Hill, NC: Health Science Consortium, 1987
- ⁷ Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1994;23:23-30

Schema für die strukturierte Erfassung der Angaben zu einem Fall mit Standardisierten Patienten (Anamnese und körperliche Untersuchung)

I. Anamnese zum jetzigen Leiden / Konsultationsgrund

Lokalisation	Verlauf/Entwicklung
Art	Häufigkeit
Ausmass/Schweregrad	Umstände/Auslöser
Beginn	verstärkende resp. erleichternde Faktoren
Dauer	damit verbundene Symptome

II. Persönliche Anamnese

Gesundheitszustand allgemein	Allergien
frühere Krankheiten	Impfstatus
frühere Unfälle	frühere Hospitalisationen (inkl. Zweck resp. durchgeführte Operation)

III. Familienanamnese

	<u>Eltern</u>	<u>Geschwister</u>	<u>andere nahe Verwandte</u>
Alter			
Gesundheitszustand			
Todesursache			

IV. Psychosozialanamnese

Zivilstand	Wohnsituation (wo, mit wem)
Rauchen	andere wichtige Lebensstilelemente
Genussmittel	familiäre Belastungselemente
Drogen	spezielle Arbeitsplatzbelastungen/-risiken
Bewegung	Sexualanamnestisches

V. Medikamentenanamnese

ärztlich verordnete laufende Medikation
evtl. Einnahme pharmakologisch wirksamer Substanzen auf eigene Initiative

VI. Systemanamnese

Allgemeinzustand	gynäkologische Anamnese
endokrinologische Auffälligkeiten	Organe des Bewegungsapparats
Haut	Blut
Kopf/Ohren/Nase/Hals	Neurologisches
Mammae	Psychiatrisches
Atmung	Schlaf
Herz und Kreislauf	Antrieb, Energie
Verdauungsorgane	Appetit
Geschlechts- und Harnorgane	allgemeines Wohlbefinden

Befunde der körperlichen Untersuchung

Gesamteindruck	Herz-/Kreislaufsystem
Vitalzeichen/Bewusstsein	Gastrointestinaltrakt
Orientierung/Merkfähigkeit/geistige Verfassung	urogenitales System
Allgemeinzustand/endokrinolog. Auffälligkeiten	Bewegungsapparat
Haut	periphere Gefässe
Kopf/Ohren/Nase/Hals	Neurostatus
Mammae	hämatologische Auffälligkeiten
Atmung	

Das Formular muss so gestaltet werden, dass bei jedem Punkt genügend Platz vorhanden ist, um die entsprechende Information für den geplanten Fall einzutragen. Je nach hauptsächlichem Prüfungsziel können weitere Punkte wichtig werden. Geht es z.B. primär um eine Evaluation der Arzt/Patienten-Kommunikation, müssten die Inhalte und Ziele des Gesprächs skizziert werden.

Globale Beurteilungsskala zur Einschätzung der technischen Fertigkeiten in einem chirurgischen OSCE in Kombination mit aufgabenspezifischen Checklisten

Bitte umkreisen Sie in jeder Kategorie die Zahl, welche der Leistung des Kandidaten entspricht.					
Respekt für Gewebe:	1 übte häufig unnötig Kraft auf das Gewebe aus oder verursachte Beschädigungen durch unangemessene Instrumenten-Verwendung	2	3 behandelte das Gewebe vorsichtig, verursachte aber gelegentlich eine unbeabsichtigte Beschädigung	4	5 behandelte das Gewebe durchwegs angemessen mit minimaler Beschädigung
Zeit und Bewegung:	1 viele unnötigen Bewegungen	2	3 effiziente Arbeitsabläufe aber einige unnötige Bewegungen	4	5 klare Bewegungsökonomie und maximale Effizienz
Instrumentenhandhabung:	1 machte wiederholt versuchsweise oder ungeschickte Bewegungen durch unangemessene Instrumentenanwendung	2	3 kompetente Verwendung der Instrumente, wirkte aber gelegentlich steif oder ungeschickt	4	5 flüssige Bewegungen mit Instrumenten ohne Ungeschicklichkeiten
Instrumentenkenntnis:	1 verlangte häufig falsches Instrument oder verwendete unangemessenes Instrument	2	3 kannte die Namen der meisten Instrumente und benützte das angemessene	4	5 offensichtlich vertraut mit Instrumenten und ihren Namen
Operationsfluss:	1 unterbrach häufig die Operationstätigkeit und schien unsicher über die nächste Bewegung	2	3 zeigte eine gewisse Vorausplanung mit vernünftigem Vorschreiten der Prozedur	4	5 plante offensichtlich den Verlauf der Operation mit mühelosem Fluss von einer Bewegung zur nächsten
Einsatz der Assistenten:	1 platzierte Assistenten durchwegs schlecht oder unterliess es, sie einzusetzen	2	3 setzte die Assistenten die meiste Zeit angemessen ein	4	5 setzte Assistenten strategisch durchwegs optimal vorteilhaft ein
Kenntnis der spezifischen Prozedur:	1 mangelhafte Kenntnis, brauchte bei den meisten Schritten spezifische Anweisungen	2	3 kannte alle wichtigen Schritte der Operation	4	5 zeigte Vertrautheit mit allen Aspekten der Operation
Der Kandidat soll bei dieser Aufgabe insgesamt: <input type="checkbox"/> nicht bestehen <input type="checkbox"/> bestehen					

Globale Beurteilungsskala zur Gestaltung der Arzt/Patient-Beziehung (Kommunikation/Interaktion)

Bitte umkreisen Sie in jeder Kategorie die Zahl, welche der Leistung des Kandidaten entspricht.						
			ungenügend		genügend	
Beginn des Interviews	0 kühl, gehetzt, abrupt, unbeholfen, unpersönlich	1 geringe Beachtung des Patienten	2 neutral, beachtet Patienten	3	4 recht ent- spannt, beachtet Patienten aufmerksam	5 aufmerksam, warm, ent- spannt, umgänglich
Fähigkeit zuzuhören	0 unterbricht Patienten, überhört dessen Antwor- ten	1 nimmt Ant- worten des Patienten kaum zur Kenntnis	2 einigermassen	3 aufmerksam	4 beachtet Antworten des Patienten	5 geht systema- tisch auf die Mitteilungen des Patienten ein
Patienten- beziehung	0 herablassend, gönnerrhaft, verletzend	1 nur minimale Höflichkeit	2 respektvoll,	3 höflich	4 warm, respektvoll, höflich	5 warm, mitfühlend, tröstend
Befragungsfertigkeit	0 unbeholfen, ausschliess- licher Ge- brauch ge- schlossener oder suggesti- ver Fragen, Fachjargon	1 etwas unbe- holfen, unan- gebrachte Wortwahl, minimaler Gebrauch offener Fragen	2 mässig entspannt,	3 angemesse- ne Sprache, Gebrauch ver- schiedenartiger Fragenarten	4 entspannt, klare Fragen, angebrachte Verwendung offener und geschlossener Fragen	5 sichere und kompetente Befragung
Patienten- information u. -beratung	0 kein oder ungeeigneter Versuch, auf die Punkte einzugehen	1 unbeholfene und/oder unvollständige Versuche, auf die Punkte einzugehen	2 einigermassen	3 entspannt, Versuche, auf die Punkte einzugehen	4 entspannt, be- spricht die The- men, achtet einigermassen darauf, ob der Patient versteht	5 sichere und kompetente Gesprächsfüh- rung, dem Patienten angemessen
Organisation	0 sprunghaft, völlig unsyste- matisch	1 minimale Organisation	2 logischer Fluss	3	4 logischer Fluss, recht zweck- gerichtet	5 zielgerichtete, ganzheitliche Begegnungs- gestaltung
Abschluss	0 abrupt	1 zeigt Schluss der Befragung an	2 versucht	3 Abschluss	4 klarer Ab- schluss	5 überlegter, organisierter Abschluss

Unterlagen für eine Anamnese-Doppelstation modifiziertes Beispiel aus der kanadischen Qualifikationsprüfung für Ärzte

Kurzbeschreibung der Station

Typ:	Doppelstation
Kompetenzbereich:	Anamnese
Fachgebiet:	Innere Medizin (Hämatologie)
Problem:	Müdigkeit, jüngste Blutuntersuchung zeigt Anämie
Prüfungsziel/e:	Anamneserhebung im Rahmen einer Anämieabklärung
Unterlagen:	1. Anweisung an den Kandidaten 2. Informationen für den SP 3. Checkliste für den die Anamneserhebung beobachtenden Examinator 4. Beurteilungsliste für nachfolgende Befragung 5. Erforderliche Einrichtung und Utensilien

A. Anweisung an den Kandidaten

Herbert Meister, 61-jährig, hat vor drei Tagen Ihren Kollegen aufgesucht wegen Müdigkeit. Nun kommt er zu Ihnen in die Praxis. Ihr Kollege hat eine Blutuntersuchung veranlasst und ging dann in die Ferien. Seine Praxisassistentin hat die Resultate übermittelt:

Hämoglobin	85 g/l
MCV	75 fl
Leukozyten	4500, Blutbilddifferenzierung normal
Thrombozyten	500'000

Sie haben 5 Minuten Zeit, um eine fokussierte und relevante Anamnese zu erheben.
In der nachfolgenden Station werden Sie über dieses Problem Fragen zu beantworten haben.

B. Informationen für den SP

Konsultationsgrund: Patient fragt: „Die Praxishilfe von Dr. Bieber hat gesagt, ich solle bei Ihnen vorbeikommen, weil Sie nun das Resultat meiner Blutuntersuchung hätten. Weiss man jetzt, warum ich mich die ganze Zeit müde fühle?“

Vorgeschichte: Herr Meister hat sich in den vergangenen 2 Monaten zunehmend müde gefühlt. Er ist noch 50 % berufstätig, wurde von seinem Geschäft zur Hälfte pensioniert. Er war bislang noch ziemlich aktiv. Aber sowohl die Freizeit- als auch die beruflichen Aktivitäten hat er wegen seiner Müdigkeit beschneiden müssen. Seit den letzten drei Wochen fühlt er sich so erschöpft, dass er nicht mehr ohne Mittagsschlafchen auskommen kann und dieses verlängert sich gerne in den Nachmittag hinein. Er bemerkte eine Kurzatmigkeit bei Anstrengung, z.B. beim Treppensteigen. Einmal im vergangenen Monat sei sein Stuhlgang schwarz gewesen.

Mit seiner vorherigen Gesundheit sei er sehr zufrieden ausser den Magengeschwüren vor ca. 22 Jahren und den gelegentlichen Knieschmerzen seit ungefähr einem Jahr. Regelmässige Medikamente hat er nicht, aber wenn die Knie weh tun, behilft er sich mit Aspirin 500 bis 1000 mg. Manchmal ist das nur zwei- bis dreimal im Monat, manchmal bis zu zweimal in der Woche. Seinerzeit habe man ihn angewiesen, wegen des Dünndarmgeschwürs möglichst eine reizlose Diät einzuhalten und verteilt zu essen mit kleinen Snacks zwischen den Mahlzeiten. Damit sei es gut gekommen, bloss vor 10 Jahren habe er nochmals Verdauungsbeschwerden gehabt. Der Arzt habe ihm dann für zwei Monate Tagamet gegeben. Was den Urin anbelangt, so sei der Strahl seit einem Jahr nicht mehr so kräftig und es störe ihn zudem, dass er nicht mehr ohne weiteres einsetze.

Herr Meister ist verheiratet, bewohnt mit seiner Frau eine Eigentumswohnung. Er hat mit Rauchen aufgehört, als er vor 10 Jahren Verdauungsbeschwerden hatte. Zuvor hatte er 30 Jahre lang geraucht, gewohnheitsmässig ein Dutzend Zigaretten pro Tag. Alkohol nimmt er nur gelegentlich zu sich, in Form eines Glases Wein zum Abendessen

Der Vater ist 84-, die Mutter 88-jährig verstorben, Ursache unbekannt. Er hat nur ein Geschwister, einen Bruder, der ein Jahr jünger ist. Bei diesem hat man vor 5 Jahren einen Diabetes festgestellt, welcher mit Diät unter Kontrolle gehalten werden kann.

Verhalten, Stimmung, Benehmen des Patienten: Er soll ein wenig in sich zusammengesackt im Sprechstundenzimmer sitzen. Er ist dankbar, dass er in die Sprechstunde kommen kann. Er ist besorgt, was die Ursache seiner Müdigkeit sein könnte und möchte dringend wissen, was ihm fehlt. Redet mit freundlicher, aber flach und müde tönender Stimme.

Er soll über seinen Stuhlgang, seine Kurzatmigkeit und die Aspirineinnahme nur Bescheid geben, wenn er direkt danach gefragt wird.

Fragen, die der Patient stellen muss: Keine

Fragen, die der Patient stellen kann: Wird es jemals wieder besser? Was stimmt nicht? Was kann es denn sein, was mich dermassen müde macht?

C. Checkliste für den die Anamneseerhebung beobachtenden Examinator

Thema	Punkte
Erfragen der Umstände der Müdigkeit:	
- Dauer (2 Monate)	2
- Entwicklung (allmählich zunehmend)	2
- Auswirkungen im Alltag (Treppen schwierig, benötigt Mittagsschlaf, kann nicht mehr soviel tun)	4
Erfragen assoziierter Symptome:	
- Dyspnöe (ja, bei Anstrengung)	2
- Thoraxschmerz (nein)	2
Medikamentenanamnese:	
- bestehende Medikationen (nicht regelmässig)	4
- fragt spezifisch nach NSAID's/Medikamente gegen Gelenkbeschwerden (benutzt Acetylsalicylsäure wegen Knieschmerzen)	4
- vergangene Medikationen (Tagamet 2 Monate, vor 10 Jahren)	2
Frühere Krankheiten: (Ulkus vor 22 Jahren)	2
Gewohnheitenanamnese:	
- Alkohol (Wein zu Mahlzeit)	2
- Rauchen (vor 10 Jahren aufgehört)	2
Systemanamnese:	
- Abdominalschmerz/Verdauungsstörung (nein)	2
- Stuhlfarbe (normal ausser einmal schwarz vor einem Monat)	2
- Stuhlkaliber (normal)	2
- Obstipation vor kurzer Zeit (nein)	2
- Appetitveränderung (nein)	2
- Gewichtsveränderung (nein)	2
- blaue Flecken oder andere Blutungszeichen vor kurzer Zeit (nein)	2
- vor kurzer Zeit Fieber, Nachtschweiss, Infektion (nein)	2

D. Beurteilungsliste für nachfolgende Befragung**Fragen und Antworten:****Punkte****1. Welches sind bei diesem Patienten die drei anamnestischen Schlüsselbefunde, welche zur Diagnose führen?**

- vergangener Monat schwarzer Stuhl 5
- in Vorgeschichte/Krankheitsanamnese Ulkus bzw. Verdauungsstörung 3
- allmähliche Symptomentwicklung 2

Maximalpunktzahl 10**2. Nennen Sie drei differentialdiagnostische Ursachen für seine Anämie.**

- Gastrointestinaltraktmalignom 2
- Peptische(s) Ulkus/Ulzera 2
- Gastritis (mit Acetylsalicylsäure oder Alkohol in Zusammenhang stehend) 2
- Anderes Malignom 1
- Ösophagusblutung (infolge Ösophagitis oder Speiseröhrenkrebs) 1

Maximalpunktzahl 6**Nochmals die Laborresultate des Patienten, die Sie in der vorangehenden Station bereits gesehen haben:**

Hämoglobin	85 g/l
MCV	75 fl
Leukozyten	4500, Blutbilddifferenzierung normal
Thrombozyten	500'000

3. Welche zusätzlichen Laboruntersuchungen würden Sie durchführen, um die Ursache der Anämie dieses Patienten abzuklären?

- Serumeisen 1
- Totale Eisenbindungskapazität oder Transferrin 1
- Serumferritin 2
- Hämocult-Test 1
- B₁₂-Serumspiegel 0
- Rotes Blutbild oder Serumfolat 0

Maximalpunktzahl 2**4. Welche zwei weiteren Untersuchungen helfen mit der grössten Wahrscheinlichkeit das Problem dieses Patienten zu diagnostizieren?**

- Gastroskopie / Endoskopie des oberen Gastrointestinaltraktes 4
- Kolonoskopie / Sigmoidoskopie / Endoskopie des unteren Gastrointestinaltraktes 3
- Serienaufnahmen des oberen Gastrointestinaltraktes 2
- Barium-Kontrastmitteleinlauf 2
- Dünndarmpassage 1
- Hämocult-Testung 0

Maximumpunktzahl 7**E. Erforderliche Einrichtung und Utensilien**

Prozedurstation: Stühle für Patienten, Kandidaten, Examinator. Fragestation: Tisch, Stuhl, Bleistift, Gummi

<p>Kopfweh* (Langstation; doppelt geführt) <i>Anamnese</i> <i>Untersuchung</i> <i>Diagnose</i> <i>Prozedere</i></p>	<p>Rektale Blutung* <i>Anamnese</i> <i>Diagnose</i> <i>Prozedere</i></p>	<p>Posttraumatische Bursitis* <i>Punktion</i> <i>Information Pat.</i></p>	<p>Atemnot <i>Anamnese</i></p>	
<p>Screening für Prostata- Karzinom <i>Beantworten der</i> <i>Fragen des SP</i></p>	<p>Beispiel eines Circuit</p> <p>Circuit des "Examen de certification du Collège des médecins de famille du Québec". Der Circuit wurde an einem Samstagnachmittag 1998 als erster von zwei Prüfungsteilen durchgeführt. Er enthielt 15 Kurzstationen und 3 Ruhestationen à 7 Minuten sowie eine doppelt geführte Langstation à 15 Minuten. Wechsel von Station zu Station: 1 Minute. Gesamtdauer: 2 Std. 40 Min. Am nächsten Tag fand ein zweiter Circuit mit 14 Stationen statt. Insgesamt wies die Prüfung 30 Stationen auf, 25 waren mit SPs und Beobachtern "ausgerüstet", drei mit Fragebogen und eine mit Röntgenbildern. Eine Station war als Rollenspiel konzipiert (Telefonanruf, Beobachter/SP am anderen Ende).</p>		<p>Atemnot (Fragestation) <i>4 Kurzantwortfragen</i> <i>zu vorangehendem</i> <i>Fall</i></p>	
<p>Hundebiss* <i>Anamnese</i> <i>Management</i> <i>Empfehlungen</i></p>			<p>Ruhestation</p>	
<p>Orale Anti- koagulation* <i>Erklärungen</i> <i>Risiken</i> <i>Nebenwirkungen</i></p>			<p>Visusprüfung* <i>Untersuchung</i></p>	
<p>13-mt. Kind mit Fieber (Tel. der Mutter) <i>Anamnese</i> <i>Diagnose</i> <i>Prozedere</i></p>			<p>Gesichtsfeld- prüfung* <i>Untersuchung</i></p>	
<p>Ruhestation</p>			<p>Schulter- schmerzen* <i>Anamnese</i> <i>Diagnose</i> <i>Prozedere</i></p>	
<p>Röntgenbild (Fragestation) <i>3 Kurzantwort-</i> <i>fragen</i></p>	<p>Husten (13-j. SP)* <i>Anamnese</i> <i>Diagnose</i> <i>erklären</i></p>	<p>Gedächtnisverlust* <i>Anamnese</i> <i>Mentalstatus</i> <i>Diagnose erklären</i> <i>Prozedere</i></p>	<p>Ruhestation</p>	<p>Mutter mit Kind, Schulprobleme* <i>Anamnese</i> <i>3 Fragen</i> <i>des Beobachters</i></p>

* Station mit Standardisiertem Patienten (SPs)

8

Wer besteht
die Prüfung?

INHALT

- 8.1 Warum ist die Standardsetzung wichtig? - 183
- 8.2 Wie funktionieren die verschiedenen Verfahren? - 186
- 8.3 Wie wird geplant? - 192

Beim Festlegen der Anforderungen, die erfüllt werden müssen, um ein Examen gerade noch zu bestehen, sind verschiedene Gesichtspunkte zu berücksichtigen. An erster Stelle stehen fachliche Kriterien: Facharztprüfungen sollten gewährleisten, dass Ärzte tatsächlich über die Kompetenzen verfügen, welche notwendig sind, um ihren Beruf auszuüben. Dann müssen gesellschaftspolitische Überlegungen stattfinden: die Bestehensgrenze bestimmt darüber, wie viele Ärzte im Land praktizieren. Auch standespolitische Kriterien sind einzubeziehen: die Facharztprüfung selektioniert, welche Kandidaten als Mitglieder des Standes aufgenommen werden und welche nicht. Schliesslich sind wissenschaftliche Kriterien zu berücksichtigen: welche Verfahren zum Festlegen der Bestehensgrenze sind anerkannt?

Die Tragweite der Entscheidung „bestanden“ oder „durchgefallen“ ist enorm. Ob die Leistung eines Kandidaten genügend ist oder nicht, muss deshalb gültig, zuverlässig und professionell beurteilt werden. Die Bestehensgrenze muss objektiv, sachlich gerechtfertigt und begründbar festgesetzt werden.

Dieses Kapitel handelt davon, wie die Anforderungsgrenzen für das Bestehen von Examen wissenschaftlich fundiert festgelegt werden können und was zu beachten ist, wenn Facharztprüfungen geplant werden.

In diesem Kapitel erfahren Sie,

- warum der Prozess der Standardsetzung so wichtig ist, und welche Kreise darauf Einfluss nehmen wollen
- wie sich normorientierte und inhaltsorientierte Methoden unterscheiden, und warum letztere den ersteren unbedingt vorzuziehen sind
- wie die Bestehensgrenze, d.h. die Anzahl richtig zu beantwortender Fragen, in zwei konkreten Beispielen bestimmt wird
- worauf Sie achten müssen, wenn Sie eine Standardsetzung planen

8.1 Warum ist die Standardsetzung wichtig?

Bevor wir uns mit den möglichen Verfahren zum Festlegen der Bestehensgrenze befassen, möchten wir einige wichtige allgemeine Aspekte zum Thema beleuchten¹. Der Zweck einer summativen Prüfung besteht in der Regel darin zu entscheiden, ob die Leistung eines Kandidaten genügt, um ihm eine Qualifikation zuzusprechen. Überwindet ein Prüfling eine bestimmte Hürde, hat er das Examen bestanden und ist für eine Tätigkeit qualifiziert. Der Beschluss über die Höhe dieser Hürde ist also ein wesentliches und brisantes Element eines Examens.

Wer besteht und wer fällt durch?

Die Entscheidungsfindung, wie hoch eine Messlatte angesetzt werden soll, ist gerade bei qualitativ hochstehenden Examen oft komplex und schwierig². Die Gesellschaft, der Berufsstand, die Wirtschaft und die Kandidaten machen ihre Ansprüche geltend. Die unterschiedlichen Interessen können unter Umständen miteinander kollidieren. Schlussendlich muss aber eine Entscheidung getroffen werden, wo die Grenze zwischen „kompetent“ und „inkompetent“ zu liegen kommt (vgl. Abb. 8.1).

Verschiedene Gruppen wollen mitreden.

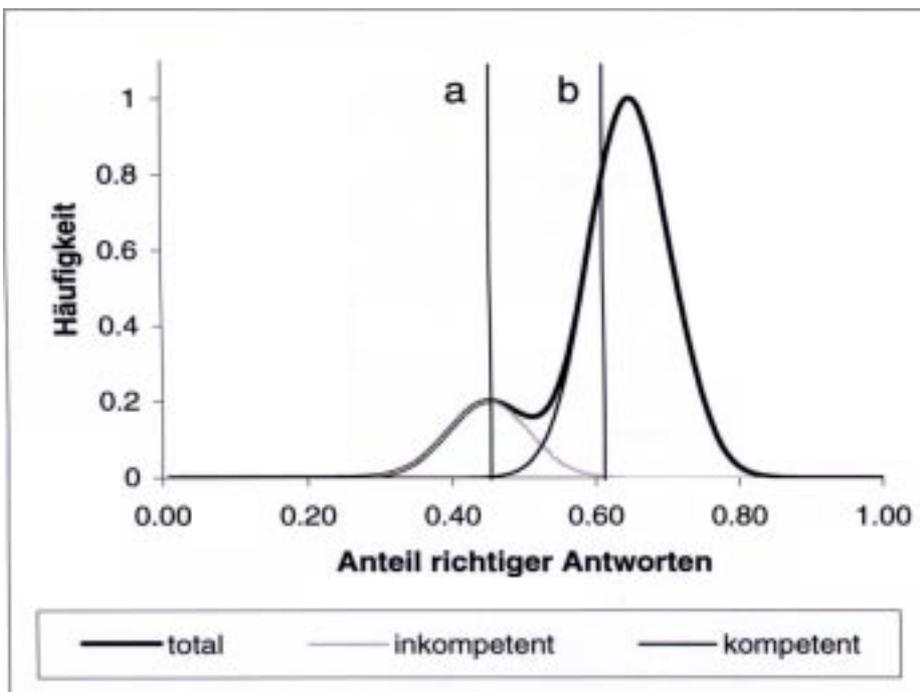


Abb. 8.1 Hypothetische Leistungsverteilung einer Kompetenzprüfung mit 90 Prozent kompetenten und 10 Prozent inkompetenten Kandidaten bei einer Streuung von 0.1. Wenn die Bestehensgrenze bei a festgelegt wird, bestehen zwar alle kompetenten, aber auch mehr als die Hälfte der inkompetenten Kandidaten. Kommt die Grenze bei b zu liegen, fallen alle inkompetenten durch, mit ihnen aber auch ein grosser Teil der kompetenten.

Die Gesellschaft will Sicherheit.

Eine entscheidende Rolle beim Festlegen der Bestehensgrenze spielt der Einfluss der Gesellschaft. Nehmen wir als Beispiel die Autofahrprüfung: Aus Sicherheitsgründen toleriert die Öffentlichkeit hier nur wenige, geringfügige Fehler. Wer bestehen will, darf die Fehlerquote von beispielsweise fünf Prozent nicht überschreiten. Es handelt sich bei der Höhe der Hürde also um eine Norm (im soziologischen Sinne von „Regel“, „Gesetz“), die primär gesellschaftspolitisch bestimmt wird. Auch im Zusammenhang mit der Facharztprüfung meldet die Allgemeinheit ihre berechtigten Interessen an. Die Leute wollen eine genügend grosse Anzahl sehr gut qualifizierter Ärzte und eine entsprechend sichere Trennung von kompetenten und inkompetenten Kandidaten. In der Schweiz fordert das Krankenversicherungsgesetz³ beispielsweise, dass nur qualifizierte Ärzte der Krankenkasse Rechnung stellen können. Die Facharztprüfung muss also ein wesentliches Element der Qualitätssicherung im Gesundheitssystem darstellen (vgl. Einleitung).

Berufsaussichten führen zu standespolitischen Normen.

Auch für die Standespolitik hat die Facharztprüfung und insbesondere die Höhe der Messlatte grosses Gewicht. Das Examen entscheidet über die Mitgliedschaft in der „Zunft“. Seine Bedeutung nimmt im Zuge der Internationalisierung und einer europa- oder weltweiten Anerkennung stetig zu. Betrachten wir aber zuerst wieder ein nicht-medizinisches Beispiel, die Pilotenprüfung: Pilot ist noch immer ein Traumberuf, und es gibt wesentlich mehr Bewerber als Stellen. Also wird eine Fluggesellschaft bestrebt sein, nur die besten Kandidaten zuzulassen, und zwar etwa so viele, wie beschäftigt werden können. Die Bestehensgrenze wird im Zusammenhang mit den Berufsperspektiven festgelegt. Solche berufsbezogene oder standespolitische Normen können auch in der Medizin – z.B. in chirurgischen Spezialdisziplinen – sinnvoll sein.

Vor rund 20 Jahren herrschte in der Schweiz ein akuter Mangel an Krankenschwestern. In der Folge hat man bei entsprechenden Prüfungen bewusst eine möglichst niedrige Bestehensgrenze definiert und es wurde auch die Motivation gewertet. Inhaltlich wurde aber trotzdem ein Minimum an Kompetenzen verlangt, damit eindeutig Inkompetente ausgeschlossen werden konnten.

Für die Kandidaten ist die Bestehensgrenze von existentieller Bedeutung.

Im Weiteren müssen beim Festlegen der Bestehensgrenze die Interessen der Kandidaten berücksichtigt werden. Bei der Facharztprüfung handelt es sich aus ihrer Sicht um ein entscheidendes, existentielles Examen, das am Ende einer langen akademischen Aus- und Weiterbildung steht. Die Höhe der Hürde kann für einen Kandidaten schicksalshafte Bedeutung

haben. Ohne den Titel darf er seinen erlernten Beruf nicht ausüben. Ein negatives Resultat kann für den Prüfling eine existentielle Krise bedeuten. Bei der Festlegung der Bestehensgrenze ist also eine Diskussion über die Problematik der definitiv Erfolglosen und das Risiko einer falsch-negativen Entscheidung einzubeziehen.

Schliesslich prägen ökonomische Überlegungen den Entscheidungsfindungsprozess. Wie wir noch sehen werden, sind qualitativ gute Methoden zum Festlegen der Bestehensgrenzen nicht billig. Solange nur nationale Lösungen machbar sind, müssen auch die manchmal bescheidenen personellen und finanziellen Möglichkeiten der Fachgesellschaften berücksichtigt werden.

Die erwähnten Aspekte zeigen, dass es nicht nur ein gültiges, einmal festgelegtes Verfahren gibt. Das Festsetzen der Bestehensgrenze ist ein vielseitig beeinflusster Prozess, in dem sorgfältig ein den Rahmen- und Prüfungsbedingungen angepasstes, möglichst auch wissenschaftlich abgestütztes Verfahren zu etablieren ist. Wesentlich ist aber nicht nur die Glaubwürdigkeit der verlangten Anforderungen, sondern auch die Glaubwürdigkeit des gesamten Examens⁴.

Nicht alle Fachgesellschaften können sich das Beste leisten.

Die Standardsetzung ist ein Prozess.

8.2 Wie funktionieren die verschiedenen Verfahren?

*normorientiert,
inhaltsorientiert oder
ein Kompromiss*

Grundsätzlich unterscheiden wir in Anlehnung an führende Experten der Standardsetzung in der Medizin^{5,6,7} drei Verfahrensgruppen: Normorientierte Verfahren orientieren sich an der Leistungsverteilung einer Referenzbevölkerung. Sie bestimmen anhand von Vergleichsdaten oder Bezugsgrößen über „bestanden“ oder „durchgefallen“. Sie nehmen Bezug auf gesellschaftliche, standespolitische oder wirtschaftliche Zustände oder Interessen. Mit der Höhe der Hürde wird ein Soll-Zustand angepeilt, es soll eine Norm erfüllt werden. Es wird vor dem Examen z.B. eine Quote von reüssierenden Kandidaten festgelegt. Anders funktionieren inhalts- oder kriteriumsorientierte Verfahren. Die Standardsetzung basiert hier auf dem Fragen- oder Prüfungsinhalt und den geforderten Kompetenzen. Experten beurteilen die Prüfungsfragen im Idealfall einzeln und bestimmen die zu erfüllenden Kriterien und Anforderungen. Unabhängig von äusseren Umständen besteht, wer die vor der Prüfung festgelegten Ansprüche erfüllt. So definierte Bestehensgrenzen werden auch als absolute Standards bezeichnet. Wir verwenden im Folgenden aber ausschliesslich den Begriff inhaltsorientiert. Als dritte Verfahrensgruppe zur Standardsetzung gibt es verschiedene Kompromissmethoden.

Wie funktionieren normorientierte Verfahren?

Normorientierte Verfahren vergleichen Kandidaten mit einer Referenzbevölkerung. Es gibt zahlreiche Varianten und verschiedene Möglichkeiten, zu Vergleichsdaten und Bezugsgrößen zu kommen: Man kann ein Examen mit früheren Prüfungen vergleichen (vielleicht hat sich ein mittlerer Prozentsatz richtiger Antworten als Bestehensanforderung bewährt) oder mit Examen anderer Ausbildungsgänge (Berufsprüfungen haben eine Erfolgsquote von beispielsweise 70 Prozent) oder mit Referenzgruppen und Daten aus verschiedenen Ländern (in der EU gibt es vielleicht eine bestimmte Anzahl Handchirurgen auf eine Million Einwohner).

*Beispiel 1:
Die besten 50
Kandidaten bestehen.*

Diese Methode wird oft angewendet, wenn die Zahl der zu Selektionierenden klar begrenzt ist. Unabhängig davon, wie schwierig eine Prüfung ist und wie kompetent die Kandidaten sind, bestehen immer gleichviele. Damit wird in Kauf genommen, dass Qualifizierte zurückgewiesen werden. Andererseits wird nicht gewährleistet, dass alle erfolgreichen Kandidaten tatsächlich über die nötigen Qualifikationen verfügen.

*Beispiel 2:
Die besten 85 Pro-
zent bestehen.*

Prozentuale Quoten werden bei Medizinalprüfungen häufig angewendet. Es wird davon ausgegangen, dass sich die Leistungen einer geprüften Kandidatengruppe bezüglich Mittelwert und Verteilungsform nicht von denen der theoretischen Gesamtpopulation aller Kandidaten unterscheiden. Die Kandidaten werden nicht dahingehend beurteilt, ob sie genug wissen, sie werden nur mit andern Prüflingen verglichen. Die Erfolgsaussichten sind damit von der jeweiligen Gruppenleistung abhängig.

Oft wird ein tradierter Prozentsatz richtiger Fragen als Standard gesetzt (z.B. als langjähriger Mittelwert von anderen Prüfungen übernommen). Vielleicht war dieser absolute Standard früher einmal auf den Inhalt abgestützt. Prüfungen weisen aber selten konstante Schwierigkeiten auf und je kürzer sie sind, umso grösser sind die Schwierigkeitsschwankungen. 60 Prozent richtige Antworten sind bei einer leichten Prüfung eine niedrige Hürde. Umgekehrt sind 60 Prozent in einer schweren Prüfung vielleicht ein zu hoher Grenzwert.

*Beispiel 3:
60 Prozent der
Fragen richtig*

Beim Vergleich mit einer Referenzgruppe, ebenfalls ein normorientiertes Verfahren, beantwortet neben den Prüflingen auch eine repräsentative und bereits qualifizierte Gruppe die Prüfungsfragen. Gemäss deren Leistungen wird die Bestehensgrenze definiert. Die Ergebnisse der zu prüfenden Kandidaten werden mit den Leistungen der qualifizierten Gruppe verglichen, wobei die Festlegung der Grenzpunkte (eine Standardabweichung unter dem Mittelwert) willkürlich bleibt. Diese Methode wird beispielsweise in der Schweizerischen Gesellschaft für Rheumatologie angewendet. Probleme kann es geben, wenn nicht die ganze Gesellschaft hinter dem Entscheid steht, überhaupt eine Prüfung durchzuführen. Dann ist es schwierig, eine adäquate und anerkannte Referenzgruppe zu identifizieren. Vermutlich noch schwieriger ist es, die Referenzgruppe zu motivieren, ein solches Examen zu absolvieren.

*Beispiel 4:
Vergleich mit der
Leistung einer
Referenzgruppe*

Eine normorientierte Festlegung der Bestehensgrenze mag auf den ersten Blick durchaus plausibel erscheinen. Oft verbergen sich dahinter aber tradierte Werte und Vorstellungen oder die Scheu vor dem Aufwand, sich umfassend mit dem Prüfungsinhalt auseinanderzusetzen. Normorientierte Verfahren setzen den Standard darum eher willkürlich. Sie lassen sich weder wissenschaftlich begründen noch halten sie ohne spezifische gesetzliche Bestimmungen einer juristischen Argumentation stand. Werden sie trotzdem angewendet, ist es unerlässlich zu belegen, woraus der Standard abgeleitet wurde; die Vorgehensweise sollte im Reglement festgehalten werden.

*Normorientierte
Methoden sind
„willkürlich“.*

Wie funktionieren inhaltsorientierte Verfahren?

Inhaltsorientierte Methoden bestimmen die Bestehensgrenze aufgrund fachlich-inhaltlicher Kriterien^{8,9,10}. Banal ist die Standardsetzung dann, wenn eine Prüfung aus einer einzigen Frage und nur einer möglichen Antwort besteht. In diesem Fall besteht logischerweise, wer die richtige Antwort gibt. Schwieriger wird es, wenn auf eine Frage eine differenzierte Antwort erwartet wird. Nun muss genau definiert werden, welche Antworten inhaltlich korrekt sind und den Anforderungen genügen und welche nicht. Konsequenterweise braucht es für jede Prüfungsfrage eine solche Entscheidung. In den vergangenen Jahren sind dazu verschiedene Methoden entwickelt worden, die heute etabliert sind (Angoff¹¹, Ebel¹², Nedelsky¹³ usw.¹⁴).

Die Methode nach Angoff

Im Folgenden werden als Beispiel eines inhaltsorientierten Verfahrens die wesentlichen Schritte der Methode nach Angoff¹ kurz beschrieben. Inzwischen wurde diese Methode von verschiedenen Autoren modifiziert^{15,16}, so dass heute für verschiedene Ausgangssituationen erprobte, angepasste Angoff-Verfahren vorliegen.

1. Schritt: Bestimmen der Expertenkommission

Zunächst gilt es, eine Expertenkommission zusammenzustellen¹⁷, die innerhalb des Fachgebietes und der Ärzteschaft anerkannt ist. Sie sollte von den zuständigen Organisationen ein klares Mandat erhalten und unabhängig sein.

2. Schritt: Die Experten einigen sich über den gerade noch akzeptablen Facharzt.

Als erste wesentliche Aufgabe definiert das Expertenteam eine Grenzfall-(Borderline-)Gruppe, bzw. einen Grenzfall-Kandidaten. Das Team einigt sich, welche Kompetenzen potentielle Fachärzte aufweisen sollten, die weder klar qualifiziert noch klar unqualifiziert sind¹⁸. Die Experten beschreiben also, was sie von jemandem minimal erwarten, dem eine Fachkompetenz zugesprochen werden soll. Sie legen auch fest, wann es gerade nicht mehr reicht. Die Kommission entwickelt damit ein gemeinsames Bild eines gerade noch akzeptablen Facharztes. Dieser Schritt ist sehr wichtig, um produktiv weiterarbeiten zu können. Mit diesem Hintergrund gehen die Experten daran, die einzelnen Fragen zu beurteilen.

3. Schritt: Wie schwierig sind die einzelnen Fragen?

Jeder Experte wägt die Schwierigkeit der ersten Frage ab, vor allem hinsichtlich Relevanz und Komplexität. Schliesslich schätzt jedes Mitglied die Wahrscheinlichkeit, dass ein Borderline-Kandidat, wie er in der vorherigen Phase definiert wurde, die Frage richtig beantwortet. Der Schätzwert sollte nicht tiefer liegen als die Ratewahrscheinlichkeit (z.B. nicht tiefer als 20 Prozent für eine Frage mit fünf Wahlantworten).

4. Schritt: Konsens in der Expertengruppe

Die Experten geben ihre Schätzungen bekannt, grosse Diskrepanzen werden diskutiert. Mit einem vorher festgelegten Prozedere wird ein Konsens gesucht. Schliesslich wird der Durchschnittswert der Kommission ermittelt. Ein tiefer Wert deutet darauf, dass das Item schwierig ist (wenige werden die Frage korrekt beantworten). Eine hohe Bewertung zeigt, dass das Item leicht ist (viele werden die Frage richtig beantworten).

5. Schritt: Der Mittelwert über alle Fragen ergibt den Standard.

In gleicher Weise erfolgt die Beurteilung der weiteren Prüfungsfragen. Am Schluss wird die Durchschnittsbewertung der Experten über alle Fragen berechnet, es resultiert eine einzige Prozentzahl (z.B. 63 Prozent). Diese bildet die Bestehensgrenze. Kandidaten, welche 63 oder mehr Prozent der Prüfungsfragen richtig beantworten, bestehen das Examen.

Das Angoff-Verfahren lässt sich nicht nur bei Prüfungen mit schriftlichen Fragen anwenden. Wie das Beispiel in Abb. 8.2 zeigt, funktioniert es ebenso bei der Standardsetzung für praktische Prüfungen mit Kriterienlisten, selbst wenn die Kriterien unterschiedlich gewichtet sind. Es setzt lediglich voraus, dass die Fragen resp. Aufgaben und die Bewertungskriterien für alle Kandidaten gleich sind und dass die Bewertung quantitativ erfolgt.

*Angoff-Verfahren
auch bei mündlichen
und praktischen
Prüfungen*

Checklisten-Items	Gewicht	Angoff-Wert
Der Kandidat		
1. fragt nach Art des Schmerzes	3 x	0.9 = 2.7
2. Begleitsymptomen	3 x	0.8 = 2.4
3. früheren Episoden	3 x	0.7 = 2.1
4. Zusammenhang mit Anstrengung	3 x	0.9 = 2.7
5. Blutdruckanamnese	3 x	0.6 = 1.8
6. Rauchen	3 x	0.8 = 2.4
7. Herzkrankheiten in der Familie	3 x	0.9 = 2.7
8. Cholesterin	3 x	0.3 = 0.9
9. Medikamenten	3 x	0.4 = 1.2
10. gibt wahrscheinliche Diagnose	3 x	0.8 = 2.4
11. erklärt, dass Hospitalisation am sichersten ist	3 x	0.1 = 0.3
12. fragt nach Diabetes in der Familie	2 x	0.5 = 1.0
13. erklärt, dass gastro-intest. und kardiale Schmerzen ähnlich sind	2 x	0.5 = 1.0
14. fragt nach Ernährung	1 x	0.4 = 0.4
möglicher Maximalwert	38	Bestehensgrenze 24

Abb. 8.2 Anamnese-Checkliste (Angoff-Methode mit gewichteten Items) einer praktischen Prüfung mit Standardisiertem Patienten. Fall: instabile Angina pectoris

Grenzgruppen- und Kontrastgruppenverfahren

Neben dem Angoff-Verfahren werden eine Reihe anderer inhaltsorientierter Methoden angewendet. Vor allem bei strukturierten Prüfungen von praktischen Fertigkeiten, welche die Leistung der Kandidaten mittels Checklisten oder Skalen beurteilen, haben sich Grenz- und Kontrastgruppenverfahren als nützlich erwiesen. Beim Grenzgruppenverfahren richtet sich das Augenmerk auf die Gruppe der Kandidaten, die sich gerade an der Grenze zwischen kompetenter und inkompetenter Leistung befinden. Das Kontrastgruppenverfahren stellt die klar kompetenten Kan-

-didaten ihren klar inkompetenten Kollegen gegenüber. Die detaillierte Beschreibung dieser beiden Methoden mit ihren Vor- und Nachteilen kann in den entsprechenden Originalarbeiten nachgelesen werden^{19,20,21}.

Welches Verfahren eignet sich bei kleinen Kandidatengruppen?

*Methoden nach
Angoff, Ebel oder
Nedelsky*

Bei kleinen Gruppen (<30) kommen nur inhaltsorientierte Verfahren in Frage, die einen absoluten Standard liefern. Auch hier muss eine Expertenkommission gebildet und entsprechend trainiert werden. Je nach gewählter Prüfungsmethode kommen (evtl. modifizierte) Methoden nach Angoff, Ebel oder Nedelsky in Frage.

minimale Kompetenzen eindeutig beschreiben

Je kleiner die Kandidatengruppe, umso wichtiger ist, dass ein klares Konzept über die Kompetenzen des Facharztes besteht und dass möglichst eindeutig beschrieben wird, welche minimalen Kompetenzen von einem Grenzfallkandidaten erwartet werden. So wird es der Kommission möglich, für jede Aufgabe (Prüfungsfrage, Fertigkeit) oder jeden Aufgabenkomplex (Fall, Thema) und für die gesamte Prüfung die Bestehensanforderungen festzulegen (z.B. wie viele Punkte bei der Bearbeitung eines Falles erreicht und wieviele Fälle „richtig“ gelöst werden müssen). Dies sollte vor der Prüfung, also unabhängig von der gezeigten Leistung der Kandidaten, erfolgen.

*möglichst hohe
Standardisierung*

Je weniger die Prüfungsaufgaben standardisiert sind (z. B. praktische Prüfung am Krankenbett mit zufällig vorhandenen Patienten oder mündliche Prüfung anhand ad hoc ausgewählter Patientenfälle), umso unspezifischer müssen die Beurteilungskriterien formuliert werden (vgl. Kapitel 5 und 6). Aber auch bei solchen globalen Beurteilungsskalen lassen sich kriteriumsorientierte Bestehensanforderungen bestimmen. Grundsätzlich ist auch bei kleinen Gruppen zu empfehlen, mit zuvor bekannten, standardisierten Aufgaben oder Patienten zu prüfen. So fällt es leichter, eindeutig überprüfbare Bestehensgrenzen festzulegen und im Rekursfall den Entscheidungsprozess klar nachzuvollziehen.

Kompromissmodelle und Mischformen – die Hofstee-Methode

In der Literatur sind verschiedene Verfahren beschrieben, die Kompromisse zwischen norm- und inhaltsorientierten Methoden darstellen oder die beiden Verfahren mischen²². Wir beschränken uns an dieser Stelle auf eine anerkannte Kompromissmethode, die Hofstee-Methode²³. Die Hofstee-Methode geht davon aus, dass es selten einen einzig korrekten Stan-

dard gibt wie bei der einfachen 1-Frage-Prüfung. Das Verfahren erfasst deshalb einen gewissen Ermessensspielraum oder Toleranzbereich. Auf mathematisch-grafischem Weg wird ein Kompromiss zwischen den inhaltlichen Anforderungen und der erwarteten bzw. erwünschten Durchfallrate gefunden.

Das Expertenkomitee bestimmt insgesamt vier Werte, zwei beziehen sich auf den Prüfungsinhalt und zwei auf die Durchfallrate. Jedes Mitglied definiert je einen akzeptablen tiefsten und höchsten Prozentsatz richtiger Antworten, der erreicht werden muss, um das Examen gerade noch zu bestehen (minimale und maximale Bestehensgrenze). Dann legt jeder Experte je einen tiefsten und höchsten akzeptablen Prozentsatz nicht-erfolgreicher Kandidaten fest (minimale und maximale Durchfallrate). In diese Entscheidung fließen standespolitische, gesellschaftliche und ökonomische Aspekte ein. Aus den vier Werten eines jeden Experten werden die vier Mittel- oder Medianwerte des Komitees gebildet.

Wie funktioniert die Methode konkret?

In ein Koordinatensystem mit den Achsen „Durchfallrate in Prozent“ und „Richtige Punkte in Prozent“ wird mit den Prüfungsergebnissen die Durchfallrate als Funktion der Bestehensgrenze eingezeichnet. In das Koordinatensystem kommen dann auch die vier Expertenwerte zu liegen, die ein Rechteck bilden. In dieses Rechteck wird nun die Diagonale von links oben nach rechts unten gezogen. Der Schnittpunkt der Diagonalen mit der kumulativen Kurve gibt die Bestehensgrenze (vgl. Abb. 8.3).

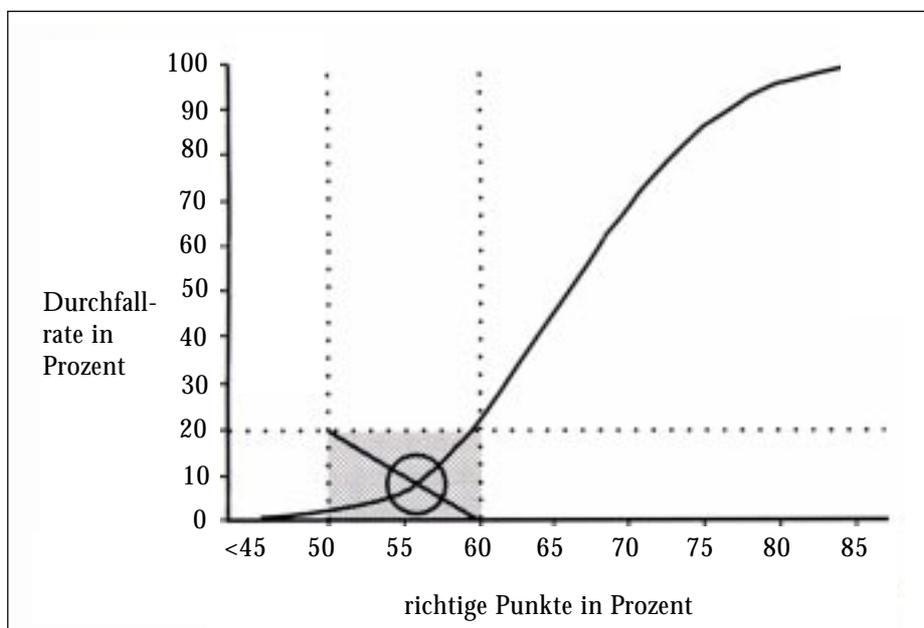


Abb. 8.3 Festlegen der Bestehensgrenze mit der Hofstee-Methode

8.3 Wie wird geplant?

- wenn immer möglich inhaltsorientierte Methoden wählen* Bei der Facharztprüfung handelt es sich um ein entscheidendes, existentielles Examen. Es verlangt ein entsprechend sorgfältiges und abgesichertes Prozedere für das Festlegen der Bestehensgrenze. Aus diesem Grund sind im Prinzip nur inhaltsorientierte Verfahren zu empfehlen. Diese gewährleisten am ehesten, dass einerseits die Anforderungen für den zukünftigen Facharzt klar definiert werden (vgl. Kapitel 1) und dass andererseits die bestehenden Kandidaten tatsächlich über die notwendigen Kompetenzen verfügen. Die Expertenarbeit in der oben beschriebenen Weise garantiert zudem ein wissenschaftlich fundiertes Vorgehen⁴.
- Standardsetzung von Anfang an einbauen* Der Prozess der Standardsetzung sollte beim Aufbau von Prüfungen von Anfang an mitgeplant, sukzessiv aufgebaut und verbessert werden. Es ist nötig, Prüfungsform und Standardsetzung aufeinander abzustimmen.
- Bestehensgrenze konkret definieren* Die Prüfungsleistung sollte möglichst quantitativ erfasst werden. Die Bestehensgrenze wird am besten als konkrete Punktzahl oder als Prozentsatz richtiger Antworten angegeben (vgl. Abb. 8.2). Diesbezügliche Probleme können bei ungenügend strukturierten praktischen und mündlichen Prüfungen auftreten. Wird die Leistung qualitativ beschrieben, müssen bei der Standardsetzung die geforderten Qualitätsmerkmale definiert und die Anzahl der zu erfüllenden Merkmale bestimmt werden.
- Standardsetzung für jeden Prüfungsteil* Besteht eine Facharztprüfung aus zwei oder mehr Teilprüfungen, die auf unterschiedliche Kompetenzschwerpunkte ausgerichtet sind, sollte die Bestehensgrenze für jeden Teil einzeln festgelegt werden. Qualifizierende Prüfungen dürfen nicht zulassen, dass ungenügende Leistungen in einem Prüfungsteil durch gute Leistungen in andern Teilen kompensiert werden können. Gut bewährt haben sich sequentielle Modelle, bei denen Kandidaten z.B. zuerst eine schriftliche Wissensprüfung bestehen müssen, bevor sie sich zur mündlich-praktischen Prüfung anmelden können.
- Experten gezielt einsetzen, ...* Für ein inhaltsorientiertes Standardsetzungs-Verfahren ist eine Gruppe von 7 - 15 Experten erforderlich²⁴. Es ist sinnvoll, die Aufgaben möglichst klar zu trennen. Die Experten sollten nur für das Fachliche zuständig sein. Administrative, organisatorische und übrige Aufgaben sollten geübtem Hilfspersonal übergeben werden.
- ...sie mit einem klaren Mandat versehen und...* Die Experten sollten die Fachgesellschaft und deren wissenschaftliches Spektrum repräsentieren und ein klares Mandat haben. Sie müssen – wenn immer möglich – unabhängig von der Herstellung und Durchführung der Prüfung arbeiten können.

Die Erfahrung zeigt, dass sich Expertengruppen mit inhaltsorientierten Methoden am Anfang oft schwer tun. Dies ist insbesondere der Fall, wenn die Experten nicht ausreichend vorbereitet werden. Unterschiedliche persönliche Gewichtungen sind dann oft schwer unter einen Hut zu bringen. Hier ist ein kollegiales Konsensfindungsprozedere gefordert, damit der gesamte Prozess nicht in Frage gestellt wird. Eine gründliche Vorbereitung (Training) ermöglicht ein effizienteres Arbeiten und ein zuverlässigeres Resultat²⁵.

...sie gut vorbereiten

Der genaue Aufwand für die Standardsetzung lässt sich schwer im Voraus abschätzen. Er hängt wesentlich von der Form und Organisation der Prüfung ab. Es kommt u.a. darauf an, wie klar die Prüfungsanforderungen beschrieben (vgl. Kapitel 1) und wie gut sie in der Fachgesellschaft und bei den Experten akzeptiert sind. Ferner ist die Prüfungslänge relevant. Bei einer hochstrukturierten Prüfung von drei bis vier Stunden Dauer kann als Richtgrösse von zwei Expertensitzungen zu rund vier Stunden ausgegangen werden.

*mindestens zwei
Sitzungen zu je vier
Stunden einplanen*

Sowohl qualitativ gute Prüfungen als auch eine fundierte Standardsetzung aufzubauen und durchzuführen bringt eine grosse personelle und finanzielle Belastung mit sich. Dies gilt besonders für kleine Gesellschaften. Diese sollten sich deshalb von erfahrenen Fachkräften und Instanzen beraten lassen und - wenn immer möglich - eine Kooperation anstreben. Es gelten für sie aber die gleichen Grundsätze.

*soviel wie möglich
zusammenarbeiten*

Standespolitische oder gesellschaftliche Umstände – beispielsweise eine Über- oder Unterversorgung mit Ärzten – sowie Engpässe oder Fehlentwicklungen in der Aus- und Weiterbildung können ausnahmsweise ein normorientiertes Verfahren zur Standardsetzung notwendig machen. Dazu braucht es allerdings eine entsprechende gesetzliche und/oder standespolitische Grundlage. Die Entscheidung sollte transparent begründet werden.

*rein normorientierte
Verfahren nur als
begründete Ausnahme*

Kompromissmethoden sind eher gerechtfertigt als rein normorientierte Verfahren. Weil sie aber die Möglichkeit bieten, einer unangenehmen, oft zeitraubenden Diskussion mindestens teilweise auszuweichen, verleiten sie dazu, dass das Prozedere zu einer Alibiübung wird. Es geht dann lediglich darum, eine für möglichst alle Beteiligten – die Kommission, die Fachgesellschaft und nicht zuletzt die Prüflinge – akzeptable Bestehensgrenze zu finden. Ökonomische Argumente allein reichen zur Rechtfertigung von Kompromissmethoden sicher nicht aus. Akzeptable Gründe sind etwa eine ungenügend strukturierte Weiterbildung, zu wenig klar definierte Anforderungen oder zu wenig repräsentative Experten.

*Kompromissmethoden
dürfen nicht zur
Alibiübung werden*

Wie mehrfach betont, sollte eine Methode gewählt werden, die vom Inhalt der Prüfung ausgeht. Das Verfahren sollte auf die jeweilige Prüfung und Situation abgestimmt sein.

Wichtige Punkte bei der Planung der Standardsetzung

- Ausgangslage, Bedeutung und gesetzliche Basis (Reglement)
- Mandat der Prüfungskommission (einschliesslich allfälliger Auflagen durch die Fachgesellschaft)
- allfällige Kombination von Prüfungsteilen
- Methode der Standardsetzung mit Begründung
- Experten für die Standardsetzung (möglichst unabhängig von der Prüfungskommission)
- Expertenvorbereitung (Training)
- erforderliche Ressourcen (evtl. Zusammenarbeit mit Prüfungsspezialisten)
- Zeitplan (abgestimmt auf den Prüfungszyklus)
- Berichterstattung über Ablauf und Resultate
- Diskussion der angewendeten Standards

Reinhard Westkämper

Literatur

- ¹ Nitko AJ. The Practical Matter of Setting Standards. *Educational Measurements: Issues and Practice* 1991:(Editorial)2
- ² Cusimano MD. Standard Setting in Medical Education. *Acad Med* 1996;71:112-20 (Review Paper)
- ³ Bundesamt f. Sozialversicherung. *Krankenversicherung*. Bern: Eidg. Drucksachen und Materialzentrale, 1997
- ⁴ Norcini JJ, Shea JA. The Credibility and Comparability of Standards. *Appl Meas Educ* 1997;10:39-59
- ⁵ Swanson DB, Dillon GF, Ross LEP. Setting content-based standards for National Board Exams; initial research for the Comprehensive Part 1 Examination. *Acad Med* 1990;65(9 suppl):17-8
- ⁶ Norcini JJ. Research on Standards for Professional Licensure and Certification Examinations. *Evaluation & the Health Professions* 1994; 17:160-77
- ⁷ Morrison Featherman C, Clauser BE, Dillon GF, Subhiyah RG. Setting Examination Standards: Learning From Experience and Meeting New Challenges. *Workshop-Handout*. 8th International Ottawa Conference, 1998
- ⁸ Berk RA. A consumer's guide to setting performance standards on criterion-referenced tests. *Rev Educ Res* 1986;56:137-72

- ⁹ Plake BS. Setting Performance Standards for Professional Licensure and Certification. *Appl Meas Educ* 1998;11:65-80
- ¹⁰ Reckase MD. Converting Boundaries Between National Assessment Governing Board Performance Categories to Points on the National Assessment of Educational Progress Score Scale: The 1996 Science NAEP Process. *Appl Meas Educ* 1998;11:9-21
- ¹¹ Angoff WA. Scales, norms and, equivalent scores. In: Thorndike RL ed. *Educational Measurement*. Washington DC: American Council on Education, 1971:514-15
- ¹² Ebel RL. *Essentials of Educational Measurement*. 3rd ed. Englewood Cliffs. NJ:Prentice-Hall. 1972:492-4
- ¹³ Nedelsky L. Absolute grading standards for objective tests. *Educ Psychol Meas* 1954;14:3-19
- ¹⁴ Livingston S, Zieky M. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service, 1982
- ¹⁵ Norcini JJ, Lipner SR, Lynn LO, Strecker AC. A comparison of three variations on a standard-setting method. *J Educ Meas* 1987;24:56-64
- ¹⁶ Skakun EN, Cook DA, Cornwall J. Do the Angoff and a Modified Angoff methods Set the Same Standards? In: Rothmen AI, Cohen R eds. *Proceedings of the 6th Ottawa Conference on Medical Education*. Toronto: University of Toronto Bookstore, 1995:232-4
- ¹⁷ Jaeger RM. Selection of Judges for Standard-Setting. *Educ Meas: Issues and Practice* 1991: 3-14
- ¹⁸ Mills CN, Melican GJ, Ahluwalia NT. Defining Minimal Competence. *Educ Meas: Issues and Practice* 1991:7-10
- ¹⁹ Clauser B, Clyman SG. A Contrasting-groups Approach to Standard Setting for Performance Assessments of Clinical Skills. *Acad Med* 1994; 69(10 suppl.):42-4
- ²⁰ Norcini J, Stillman P, Sutnick A, Regan M, Haley H, Williams R, Friedman M. Scoring and Standard Setting with Standardized Patients. *Evaluation & The Health Professions* 1993;16:322-32
- ²¹ Ross P, Clauser B, Margolis M, Orr N, Klass D. An expert-judgment approach to setting standards for a standardized-patient examination. *Acad Med* 1996;71(10suppl.):4-6
- ²² deCrujter DNM. Compromise models for establishing examination standards. *J Educ Meas* 1985;22:263-9
- ²³ Hofstee WKB. The case for compromise in educational selection and grading. In: Anderson SB, Helmick JS (eds.). *On Educational Testing*. San Francisco: Jossey Bass, 1983

- ²⁴ Norcini J, Shea J, Grosso L. The Effect of Numbers of Experts and Common Items on Cutting Score Equivalents Based on Expert Judgment. *Appl Psychol Meas* 1991;15:241-6
- ²⁵ Reid JB. Training Judges to Generate Standard-Setting Data. *Educ Meas: Issues and Practice* 1991:11-4

9

Was taugt die Prüfung?

Seit Urzeiten kann man kompetente Fachleute an der hohen Qualität ihrer Arbeit erkennen. Auch in der Medizin hat Qualitätsarbeit eine lange Tradition. Auf den ersten Blick mag daher die Forderung nach Qualitätssicherung in der Medizin im Allgemeinen und im Qualifikationsprozess der Fachärzte im Speziellen Stirnrunzeln hervorrufen. Die moderne Entwicklung in vielen Bereichen - nicht nur in der Medizin - hat aber unter bestimmten Umständen die Notwendigkeit für eine explizite Qualitätssicherung aufgezeigt. Eine systematische Qualitätssicherung ist in folgenden Situationen sinnvoll:

Das Resultat hängt von der koordinierten Zusammenarbeit vieler Fachkräfte ab. Das Resultat kann von äusseren, schwer vorauszusehenden Faktoren beeinflusst werden. Wissenschaft und Technik entwickeln sich schneller als praktische Erfahrung aufgebaut werden kann. Qualifizierte Fachleute müssen sich am Rande oder ausserhalb ihres engeren Fachgebietes betätigen.

Für die Facharztprüfung treffen mehrere, wenn nicht alle Bedingungen zu. Daher sollte ein systematisches Vorgehen bei der Planung, Durchführung und Analyse von Facharztprüfungen zur Qualitätssicherung beitragen.

In diesem Kapitel erfahren Sie,

- welche Fragen betreffend Qualitätssicherung Sie sich stellen sollten

Das Ziel der Qualitätssicherung ist es, Prüfungen zu gewährleisten, die klar zwischen kompetenten und inkompetenten Kandidaten unterscheiden, deren Urteile nicht durch Rekurse aufgehoben werden können und die eingesetzten Ressourcen nicht vergeuden. Die klassische Qualitätssicherung unterscheidet zwischen Struktur-, Prozess- und Resultatqualität. In der Praxis führt diese Unterteilung zu unnötigen Duplikationen. Daher passen wir die einzelnen Struktur-, Prozess- und Resultatelemente dem allgemeinen Ablauf einer Prüfung an.

Was soll die Qualitätssicherung leisten?

Dieses Kapitel versucht nicht, allgemein gültige Regeln für normaltypische Facharztprüfungen aufzustellen. Dazu unterscheiden sich die einzelnen Fachinhalte viel zu sehr. Lösungen, die sich bei einer grossen Anzahl von Kandidaten oder in einem bestimmten Fachbereich als ideal erweisen, wären bei einer kleinen Kandidatenzahl oder in einem anderen Fach absurd. Es erscheint uns daher nützlich, den Fachgesellschaften und Prüfungsgremien eine Reihe von sokratischen Fragen zu stellen. Das ernsthafte Bemühen, diese Fragen zu beantworten, stellt eine wirksame und anpassungsfähige Form der Qualitätssicherung dar. In diesem Sinn präsentiert der Qualitätssicherungsprozess die kollektive Erfahrung aller, die sich bisher mit der Kompetenzbeurteilung von Fachpersonen beschäftigt haben. Jede Frage weist auf eine mögliche Fehlerquelle hin, die es zu vermeiden gilt.

Fragen statt Regeln

Allgemeine Planung und Vorbereitung

- Findet die Zusammensetzung der Prüfungsgremien Rückhalt im Plenum der Fachgesellschaft?
- Besteht ein sinnvolles Gleichgewicht zwischen Vertretern von Zentrums- und regionalen Kliniken einerseits und der Praxis andererseits?
- Entspricht die kollektive Fachkompetenz des Gremiums dem vollen Anforderungsprofil des Faches?
- Sind die individuellen Verantwortlichkeiten innerhalb des Gremiums schriftlich festgelegt?
- Genügen die vorhandenen Mittel und Arbeitskapazitäten für die anstehende Arbeit?
- Sind mögliche Interessenkonflikte offengelegt, und weiss man, wie man damit umgehen will?

- Ist das aktuelle Berufsbild des Faches entsprechend einem breiten Konsens über die Aufgabe des Facharztes in unserem Gesundheitsversorgungssystem definiert?

Wie sind die Gremien zusammengesetzt?

Ist ein Berufsbild definiert?

- Wird dieses Berufsbild regelmässig den wissenschaftlichen, technischen und gesellschaftlichen Entwicklungen angepasst?
- Sind die Schlüsselkompetenzen formuliert?*
- Sind die Komponenten der Schlüsselkompetenzen gewichtet?
 - Wird der Schlüsselkompetenzen-Katalog regelmässig dem Berufsbild sowie den wissenschaftlichen und technischen Entwicklungen angepasst?
- Am besten bestehen die Schlüsselkompetenzen in Form eines Lernzielkatalogs, der die Gesamtmenge an Wissen, Fertigkeiten und Fähigkeiten beschreibt, die für das Facharztniveau auf dem Hintergrund des Berufsbildes nötig sind.
- Ist das Prüfungsziel klar umrissen?*
- Welche Teile der Schlüsselkompetenzen müssen im Examen geprüft werden?
 - Welche Teile werden ausserhalb der Facharztprüfung qualifiziert und wie?
 - Welche wichtigen Teilkompetenzen werden überhaupt nicht überprüft und warum?
- Sind die Rahmenbedingungen klar umschrieben?*
- Wie gross ist die durchschnittliche Anzahl der Kandidaten pro Prüfung?
 - Welche finanziellen Mittel und personellen Ressourcen stehen zur Verfügung?
 - Wie wird die Prüfung finanziert?
 - Welchen Einfluss haben diese Faktoren auf die Methodenwahl und die Qualität der Prüfung?
- Welche Prüfungsformen sind vorgesehen?*
- Sind die psychometrischen Eigenschaften wie Gültigkeit und Zuverlässigkeit der vorgesehenen Prüfungsformen bekannt?
 - Wo werden diese sonst noch verwendet?
 - Wie sehen sie im Detail aus?
 - In welcher Beziehung stehen die gewählten Methoden zum Prüfungsziel und zu den Rahmenbedingungen?
 - Wenn unerprobte Prüfungsformen verwendet werden, wie werden sie methodisch abgesichert?
- Besteht für jede verwendete Prüfungsform ein Blueprint?*
- Gewichtete Inhaltsverzeichnisse (Blueprints) sind eine unerlässliche Voraussetzung für eine repräsentative Stichprobe von Problemen in der jeweiligen Prüfung. Oft sind Blueprints mit mehreren Dimensionen nötig.

- Wie sehen die „Sicherheitsdispositive“ im Revisionsprozess der Problemstellungen, in der Prüfungsdurchführung und in der Archivierung aus?

Wie wird die Geheimhaltung garantiert?

Planung und Vorbereitung (Logistik)

- Wer ist mit welchen Terminen zuständig hinsichtlich Prüfungsausschreibung, -anmeldung, Räumlichkeiten, Hilfspersonal, Fall-/Fragenpool, gedruckter Prüfungsunterlagen, Patienten-Rekrutierung?
- Genügt die Logistik den Anforderungen der gewählten Methodik?
- Wie werden die Kandidaten informiert?
- Können diese die verwendeten Prüfungsmethoden üben?

Wie sehen Terminplan und Zuständigkeiten aus?

- Wie werden die Fälle/Fragen/Problemstellungen für die Prüfung vorbereitet?
- Kommt ein für das Fach gültiger, repräsentativer Prüfungsinhalt entsprechend dem gewichteten Inhaltsverzeichnis zustande?
- Bei Mehrsprachigkeit: Wer übersetzt die Fälle? Wer kontrolliert die Übersetzung?
- Gibt es validierte Beurteilungsformulare für praktische Prüfungen?

Wie sehen Aufgabenentwicklung und Revisionsprozess aus?

- Wie werden die Examinatoren ausgewählt?
- Wie werden sie in der Anwendung der Prüfungsformen resp. Prüfungsinstrumente geschult?

Welche Qualifikationen benötigen die Examinatoren?

- Wie wird die Bestehensgrenze bei den einzelnen Prüfungsmethoden und insgesamt über alle Prüfungsformen hinweg festgelegt?
- Wie wird diese Standardsetzung durch die möglichen Konsequenzen, die eine nicht bestandene Prüfung für die Kandidaten hat, beeinflusst?

Wie wird die Bestehensgrenze bestimmt?

Durchführung

- Wie wird die Prüfung überwacht und protokolliert (inkl. Zwischenfälle)?
- Wie wird ein korrekter Prüfungsablauf garantiert?
- Sind die Examinatoren möglichst unabhängig von Mitprüfern und Kandidaten?
- Dokumentieren die einzelnen Examinatoren ihre Beurteilungen unabhängig, bevor sie mit den Mitprüfern Konsens suchen?

Analyse und Nachbereitung

- Werden die Kandidatenleistungen korrekt erhoben und ausgewertet?
- Werden die Messeigenschaften des Examens überprüft (Validität, Reliabilität und Objektivität)?
- Werden die Prüfungsprotokolle und -resultate erfolgloser Kandidaten formal unabhängig überprüft?
- Wie werden die Kandidaten über die Prüfungsergebnisse informiert, und wer ist dafür verantwortlich?
- Wie erhalten Kandidaten, Examinatoren, Prüfungsverantwortliche inhaltliche Rückmeldungen, und wer ist dafür verantwortlich?
- Findet eine Prüfungsnachlese in den verantwortlichen Gremien statt?

Hinweise aus einer Facharztprüfung auf mögliche Mängel in der Aus- und Weiterbildung oder im Fach selber sind ein ganz wichtiger Effekt von Facharztprüfungen. Diese sollen die Kompetenz der Fachärzte erhalten oder verbessern helfen.

Glossar prüfungstechnischer Fachbegriffe

Blueprint: gewichtetes Inhaltsverzeichnis der Prüfungsinhalte

Cue: „Wink“ oder „Fingerzeig“. Von einem Cue spricht man, wenn in der Art der Frage- oder Aufgabenstellung bereits ein Hinweis auf die richtige Antwort enthalten ist bzw. diese mehr hervorgehoben wird. Im Interesse der \blacktriangleright Reliabilität sollte dies vermieden werden.

Evaluation, formative vs. summative: Dienen Prüfungsergebnisse zur Orientierung über den Zwischenstand des Lernverlaufs und haben für den Geprüften keine zwingenden Konsequenzen, spricht man von formativer Evaluation. Als summative Evaluation wird dagegen eine sanktionierende Beurteilung am Schluss eines Bildungsabschnitts bezeichnet.

Face validity (Augenscheinvalidität): \blacktriangleright Validität

Formative Evaluation: \blacktriangleright Evaluation

Inhaltliche Validität: \blacktriangleright Validität

Item: Einzelheit. Im Zusammenhang mit Prüfungen einzelne Frage, Aufgabe, Beobachtungs- oder Beurteilungseinheit

Item-Analyse: Beurteilung der Messeigenschaften eines Items. Man beurteilt primär die \blacktriangleright Item-Schwierigkeit und die \blacktriangleright Item-Trennschärfe. Bei MC-Items wird zudem überprüft, ob die einzelnen falschen Antworten wunschgemäß vorwiegend schwache Kandidaten von der richtigen Antwort ablenken (sog. Distraktorenfunktion).

Item-Schwierigkeit: Verhältnis zwischen der von einer Kandidatengruppe bei einem Item erreichten zur maximal möglichen Punktzahl. Sie wird entweder als Prozentwert oder als Wahrscheinlichkeitswert angegeben.

Item-Trennschärfe: Fähigkeit eines Items, Kandidaten mit guter und schlechter Leistung in der Gesamtprüfung zu trennen. Sie wird berechnet als \blacktriangleright Korrelationskoeffizient zwischen erreichter Punktzahl in diesem Item und der Gesamtpunktzahl in der Prüfung ohne dieses Item.

Korrelationskoeffizient: statistisches Mass für den Zusammenhang zweier variabler Grössen, das Werte von -1 bis $+1$ annehmen kann. Bei Null besteht kein Zusammenhang, Eins ist der bestmögliche positive Zusammenhang, bei Minuswerten besteht ein inverser Zusammenhang.

Objektivität: Im Zusammenhang mit Prüfungen wird unter Objektivität die Unabhängigkeit der Prüfungsergebnisse von den Untersuchern verstanden. Es wird weiter differenziert zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität. Ermittelt wird die Objektivität meist als statistische Übereinstimmung zwischen verschiedenen Untersuchern. Der dabei verwendete Begriff Interrater-Reliabilität weist auf den Zusammenhang mit der \blacktriangleright Reliabilität hin.

Reliabilität: Zuverlässigkeit. Prüfungsqualitätskriterium, das danach fragt, wie genau ein Merkmal gemessen wird, gleichgültig, ob dieses Merkmal auch zu messen beansprucht wird (vgl. Validität). Der Reliabilitätskoeffizient schwankt zwischen 0 und dem Maximalwert 1. Fehlereinflüsse, welche eine Messung trüben können, sind etwa mangelnde \blacktriangleright Objektivität, Rateeinflüsse, zu kleine, nicht repräsentative Itemauswahl, Zufälligkeiten. Im Vordergrund steht heute die Reliabilität des Prüfungsinstrumentes, meist erfasst in Form des Koeffizienten alpha von Cronbach. Dieser gibt Auskunft darüber, wie stark die Ergebnisse von der spezifischen Itemauswahl abhängen, resp. wie gut eine alternative Prüfung mit einer gleichen Zahl von Items, die nach gleichen Kriterien aus dem gleichen Inhaltsbereich gezogen würden, zur gleichen Rangierung der Kandidaten führen würde.

Schwierigkeit: \blacktriangleright Item-Schwierigkeit

Summative Evaluation: \blacktriangleright Evaluation

Trennschärfe: \blacktriangleright Item-Trennschärfe

Validität: Gültigkeit. Prüfungsqualitätskriterium, das danach fragt, ob das betreffende Verfahren wirklich das misst, was beabsichtigt ist. Konkret auf die Facharztprüfungen bezogen ist es die Frage, ob die für die Berufsaufgaben erforderlichen Kompetenzen gemessen werden. Es werden folgende Aspekte unterschieden: **Inhaltvalidität** fragt danach, wie repräsentativ die ausgewählten Prüfungsitens für den Inhalt bzw. den Umfang des zu prüfenden Kompetenzbereichs oder Stoffgebiets sind. Die Beurteilung sollte durch Fachexperten erfolgen. **Kriteriumsvalidität** fragt danach, wie gut die Prüfungsergebnisse mit Leistungen ausserhalb der Prüfungssituation, z.B. im Berufsalltag übereinstimmen. Sie wird meist durch Korrelationsstudien zu klären versucht. **Konstruktvalidität** fragt danach, ob Hypothesen, die aus einer Theorie über das zu messen beabsichtigte Konstrukt (z.B. „Problemlösefähigkeit“) abgeleitet sind, durch Befunde im Zusammenhang mit den Prüfungsergebnissen gestützt werden. **Augenscheinvalidität** (Face validity), ist - wie der Name sagt - eine Scheinvalidität und damit kein Gütekriterium wie die drei vorangehenden. Eine Prüfung, die Augenscheinvalidität besitzt, erweckt beim Betrachter den Eindruck das zu erfassen, was sie zu erfassen beansprucht. Dies kann erstrebenswert sein, weil sie damit von den Kandidaten und Entscheidungsträgern besser akzeptiert wird.

Register

A

Ablenker 72
Angoff-Verfahren 168, **188f**
Anspruchsniveau 65
Antworten 71, 73ff, 83ff, 91, 103, 105, 118, 155
Antwortschlüssel 51, **105**, 109f, 118
Anweisung **154**, 174
Anwendungsbezogenheit 67
Aufwand **16**, 49, 50ff, 54f, 57, 193

B

Beobachtungsinstrumente **132ff**
Beobachtungszeit 55
Bestehensgrenze 16, 108, 125, 168f, **183ff**, 190ff
Beurteilungskriterien 146, **151f**, 155
Bewertung 118, 121, 124, 138, 168, **188f**
Blueprint 23, **28ff**, 131, 117, 202
Borderline-Gruppe **108**

C

Checkliste 52, 56f, **132**, 166ff, 172, 175
Circuit **56**, 148, **166f**, 177
computerunterstützte Prüfungsmethoden **47**
Cue 16, 52, 84f, **211**
Cueing-Effekt **50**, 75

E

Eröffnungsfrage **119**
Examinatorentaining 52, **54**, 118

F

Face validity **211**
Fall **81**, **101ff**, 151ff
Fallspezifität **45**, 57, 131
Fangfragen **81**
Fertigkeiten 16, 28, **45**, 48, 54, 56, 189
formative Evaluation **10**, **211**
Formulierung 81ff, 102ff, 155
Fragenanalyse 50
Fragenthema 65, **67**
Fragentypen siehe Itemtypen

G

Gewichtetes Inhaltsverzeichnis siehe Blueprint
Gewichtung **28ff**, 91, 155, 168
Gültigkeit siehe Validität

H

Handlungsfragen **118f**

I

inhaltsorientiert **186ff**, 192f
Item 65ff, 80ff, 155, **211**
Itemformular **89**, 92
Itemformulierung **83ff**, 155
Itemgewichtung **155**
Itemrevision **89f**, 92
Itemthema **67**
Itemtypen **71ff**, 87f, 94f

K

KAF siehe Kurzantwortfragen
Kandidateninstruktion **154**, 167
Kandidatenzahl **47**, 50, 52, 55, 109f, 147, 166, 201
Kenntnisse **48f**, 53
Key feature siehe Schlüsselprobleme
Kommunikationsfähigkeit **53**, 160
Kompetenzen **15**, 26, 32f, 43, 48f, 51, 53, 55f, 125, 131, 184, 190, 192
Konsensgremien **15**
Konsensprozess 29, 108
kriteriumsorientiert **186**
Kurzantwortfragen 48, 51f, 83, **99ff**
Kurzfragen **119f**

L

Lernziel **26**
Lernzielkatalog 65, **202**
Lösungshinweise siehe Cue

M

MC siehe Multiple Choice
Messzuverlässigkeit siehe Reliabilität
Methodenwahl **43ff**
Multiple Choice **48ff**, **61ff**

N

normorientiert 51, **186f**, 193

O

Objektives strukturiertes klinisches Examen 56f, **143ff**
Objektivität 43f, 48, 115, 117f, 135, **211**
OSCE siehe Objektives strukturiertes klinisches Examen
OSCE-Station 56f, **148ff**, 168f, 174f

P

Patient Management Problems 47
Patienten 55f, 136, 190
Patiententrainer **146**, 156
Patiententraining siehe SP-Training
Patientenvignette **82f**
PMPs siehe Patient Management Problems
Problem- oder Fallspezifität 45, 50
Prüfungsauswertung **168**
Prüfungsinhalt 15, **25**, 28, 43, 117, 203
Prüfungsmethoden 15, **47ff**
Prüfungsqualität 47, **201ff**

R

Rahmenbedingungen **46**, 202
Referenzgruppe **186f**
Relevanz **67**, 90
Reliabilität 16, **44f**, 48, 50, 52ff, 57, 66, 79, 117, 138, **212**
Rückmeldung 204

S

SB siehe Strukturierte Beobachtung
Schlüsselkompetenz 15f, **26ff**, 202
Schlüsselkriterien **33f**, 102, 104
Schlüsselprobleme **32f**, 102, 104
Signalwirkung **45f**
SMP siehe Strukturierte mündliche Prüfung
SP siehe Standardisierte Patienten
SP-Training 152f, 156, **163ff**
Stamm 72, **80ff**
Standardisierte Patienten 135f, 147, 152f, 155, **160ff**, 174f
Standardisierung 43f, 117, 129, 190
Standardsetzung 108, 138, 168, **183ff**, 192ff, 203
Stichprobe 15, 23, **45**, 131, 202
Strukturierte Beobachtung 48, **54**, 129ff

Strukturierte mündliche Prüfung 48, **52f**, 115ff
Strukturierung **43f**, 117, 129
Summative Evaluation 10, **211**

T

Teilpunktbewertung **76**
Trennschärfe 66, 83, **212**

V

Validität 16, **43ff**, 53, 65, 117, **212**
Verständnis 53, 65, 81
Vignette 72, **81f**

W

Wahlantwort **49**, 65, 71ff, 83f, 188
Wahlantwortfragen 15, **49**

Z

Zuordnungsfragen 71, **73ff**, 94
Zwischenfragen **120**

Checklisten zur Planung und Überwachung des Prüfungszyklus

Ist eine Facharztprüfung von Grund auf neu zu entwickeln, erfordert dies eine sorgfältige Projektplanung. Die folgenden Checklisten sind nicht dafür angelegt, sondern für die Planung und Überwachung alljährlich wiederkehrender Arbeiten der Prüfungsherstellung, -durchführung und -auswertung (sog. Prüfungszyklus). Wir empfehlen dafür Checklisten mit den Rubriken Arbeitsschritte, Verantwortung, Termin.

Vor allem bei der Herstellung von Facharztprüfungen aber auch bei deren Durchführung und Auswertung sind stets etliche Personen beteiligt. Jede davon kann die ihr zugeteilten Arbeitsschritte immer erst ausführen, wenn der je vorangehende Arbeitsschritt getan ist. Die meisten dieser Personen erledigen die anfallenden Arbeiten neben einem bereits stark befrachteten Alltagsarbeitsprogramm („Milizsystem“). Zudem sind sie oft nicht nur an einer Prüfungskomponente beteiligt (z.B. sowohl an der theoretischen wie an der praktischen Prüfung).

Nebenamtliche Arbeit vieler örtlich getrennter Beteiligter ...

Für die Planung eines Prüfungszyklus bedeutet das,

- dass diese langfristig, mind. für die Zeitdauer eines Jahres erfolgen muss
- dass für die einzelnen Arbeitsschritte genügend Zeit vorgesehen werden muss
- dass zwischen der Fertigstellung der Prüfung und dem Prüfungstermin wenn möglich ein zeitliche Reserve eingeplant werden sollte

... erfordert eine langfristige, grosszügige Planung ...

Für die Herstellung, Durchführung und Auswertung einer qualitativ guten Facharztprüfung ist mindestens ein Jahreszyklus vorzusehen. Bei den theoretischen Schlussprüfungen für Ärzte (MC-Examen) in der Schweiz haben sich Zyklen von anderthalb Jahren gut bewährt, wobei selbst hier noch ab und zu Engpässe auftreten.

... mit Zyklen von 1-1½ Jahren.

Das erste der zwei folgenden Beispiele zeigt die Checkliste für eine KAF-Prüfung, deren Herstellung innerhalb eines Jahres erfolgen soll. In der Rubrik „Termin“ würden real natürlich konkrete Kalenderdaten stehen. Die Bereiche Prüfungsherstellung, Anmeldungsadministration, Prüfungsdurchführung und Prüfungsauswertung sind im Interesse einer guten Übersicht nebeneinandergestellt.

Beispiel mit Herstellungszyklus von 1 Jahr

Das zweite Beispiel zeigt einen möglichen Planungsraster für eine MC-Prüfung mit einem Herstellungszyklus von anderthalb Jahren. Es beschränkt sich auf den Arbeitsbereich Prüfungsherstellung.

Beispiel mit Herstellungszyklus von 1½ Jahren

Checkliste zur Herstellung, Durchführung und Auswertung einer Prüfung mit KAF

Instanzen: EG Expertengruppe des Faches
PK Prüfungskommission

PE Prüfungsexperte
S Prüfungssekretariat

a) Herstellung des Prüfungsmaterials

Arbeitsschritte	Verantwortung	Termin x=Prüfung 1bis12=Monate
Herstellen der Aufgaben: Schlüsselprobleme, -kriterien, Auswahl/Instruktion/Schulung der Fragenautoren	„EG Prüfungsinhalt“	x - 12
Fall-/Fragenherstellung nach Vorgaben der EG Prüfungsinhalt (in Workshops), Reinschrift/Erfassen dieser Rohfragen	„EG Autoren“/S	x - 10
inhaltliche Vorrevision der vorbereiteten Rohfragen der EG Autoren	„EG Revision“	x - 8
formal-sprachliche Vorrevision der Rohfragen, Organisation/Koordination der Revisionsitzung	PE	x - 7
Revision der Rohfragen: Aufnahme gemeinsam akzeptierter Fragen in den Fragenpool	„EG Revision“	x - 6
Rückversand der abgelehnten Rohfragen mit Begründung/ Korrekturvorschlägen an EG Autoren oder EG Prüfungsinhalt	PE	x - 6
Vortest der akzeptierten Rohfragen: Organisation (evtl. Testlauf in definitiver Prüfung)	„EG Antwortschlüssel“/PE	x - 5
Revision akzeptierter und getesteter Rohfragen (inkl. Antwortschlüssel): Bereinigung, Ablage in Fragenpool, evtl. Rückversand EG Revision	„EG Antwortschlüssel“	x - 4
Zusammenstellen der Prüfung aus Fragen des Fragenpools aufgrund des Blueprints	PE	x - 3
Evtl. Übersetzen: Organisation, Übereinstimmungskontrolle sprachlich verschiedener Versionen	PE	x - 3
Standardsetzung in der zusammengestellten Prüfung, Festlegen einer inhaltsorientierten Bestehensgrenze	„EG Standard“	x - 2
Organisation der Korrektur: Erstellen der Korrektur- und Auswertungsunterlagen, Organisation der Korrekturveranstaltung	PE/S	x - 2
Erstellen des Prüfungsheftes	S	x - 1

b) Anmeldungsadministration

Arbeitsschritte	Verantwortung	Termin x=Prüfung
Prüfungsausschreibung	PK	x - 6
Versand des Anmeldeformulars	S	x - 5
Sammeln der Anmeldungen inkl. Zahlungskontrolle	S	x - 3
Informationen zur Prüfung (auch methodisch inkl. Übungsmöglichkeiten für Kandidaten)	PK/PE/S	x - 3 oder früher
Erstellen/Erfassen der Kandidatenliste mit Personendaten	S	x - 2
Versand der Anmeldebestätigungen	S	x - 2

c) Prüfungsdurchführung

Reservation des Prüfungslokals	S	x - 12
Kontrolle des Prüfungsmaterials vor der Prüfung	S	x - 0.5
Einrichten des Prüfungslokals	S	x
Prüfungsaufsicht	PK	x
Kontrolle des Prüfungsmaterials nach der Prüfung	S	x

d) Prüfungsauswertung

Korrektur der Prüfung	„EG Korrektur“	x
erste Auswertung: Analysen, Vorschläge zur Elimination von Fragen, Rechnen der Bestehensgrenze	PE	x + 0.3
Entscheide zur definitiven Bewertung der Kandidatenleistungen	PK	x + 0.5
definitive Auswertung aufgrund der Entscheide der Prüfungskommission	PE	x + 0.6
Nachprüfen der Prüfungsergebnisse, v.a. der Leistungen von Kandidaten, die nicht bestehen	„EG Überprüfung“	x + 0.7
Feedback an Kandidaten: Bestehensentscheid, Leistung in verschiedenen Blueprint-Kategorien	PK/PE	x + 1
Analyse/Bericht: Überprüfen der Gütekriterien der Prüfung, Feedback an Expertengruppen und Prüfungsverantwortliche	PE	x + 2
Nachlese	PK	x + 3

Checkliste zur Herstellung einer MC-Prüfung mit einem Zyklus von 1½-Jahren

Instanzen: EG Expertengruppe des Faches PE Prüfungsexperte
 PK Prüfungskommission (Präs Präsident, M Mitglied) S Sekretariat

Prüfung 2001**Prüfung 2002**

Termin	Arbeitsschritte	Verantwortung	Arbeitsschritte	Verantwortung
April 00	Fragenbestellung: Anzahl, Inhalte, Typen	PK (Präs)		
Juni 00	Fragenherstellung Inhaltsbereich 1 Inhaltsbereich 2	EG Team A,B Team C,D		
Juli 00	Grobe formale Kontrolle und Koordination der inhaltlichen Vorrevision	PE,S		
Aug. 00	Inhaltlich-fachliche Vorrevision Inhaltsbereich 1 Inhaltsbereich 2	EG Team B,A Team D,C		
Sept. 00	Formal-sprachliche Vorrevision	PE		
Okt. 00	Versand der vorrevidierten Fragen an Teilnehmer der Revisionsitzung	S		
Nov. 00	Revisionsitzung	PK,PE		
Dez. 00	Erfassen der akzeptierten Fragen	S		
Jan. 01	Prüfungszusammenstellung	PK (M)		
Feb. 01	Kontrolle der zusammengestellten Prüfung Versand zur Übersetzung	PK (Präs) S		
März 01	Übersetzung Erfassung Übersetzungen Letzte Kontrolle	PK (M) S PK (M)		
April 01	Standardsetzungssitzung	PK,PE,EG	Fragenbestellung	PK (Präs)
Juni 01	Erstellen der Druckvorlagen	S	Fragenherstellung	EG
Juli 01	Druck der Prüfungshefte	S	Grobe formale Kontrolle	PE
Aug. 01	Bedrucken d. Antwortbelege u. Versand Prüfungsmaterial	S	Inhaltlich-fachliche Vorrevision	EG
Sept. 01	Prüfungstermin 2000		Formal-sprachliche Vorrev.	PE

