Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin

Zusammenfassung

Die Praxis der Leistungserfassung bei Studierenden der Human-, Zahn- und Tiermedizin an Hochschulen und Universitäten im deutschsprachigen Raum hat in der letzten Dekade bedeutende Änderungen erfahren. Die Betonung der praktischen Anforderungen an die ärztliche Tätigkeit in der Ausbildung weg von einer oft theoriedominierten Lehre, die wissenschaftliche Auseinandersetzung mit den Grundlagen der Vermittlung von ärztlichem Wissen und Fertigkeiten sowie geänderte gesetzliche Rahmenbedingungen erfordern einen stetigen Anpassungsprozess von Lehre und der Art und Weise, Prüfungen im Medizinstudium durchzuführen. Um hier Qualitätsstandards zu etablieren, wurden im Jahr 2008 von der Gesellschaft für medizinische Ausbildung Empfehlungen zur Durchführung fakultätsinterner Prüfungen verabschiedet, die nunmehr einer Aktualisierung unterzogen wurden und gemeinsam vom Ausschuss Prüfungen der GMA mit dem Medizinischen Fakultätentag (MFT) als Empfehlungen für die Durchführung qualitativ hochwertiger Prüfungen verabschiedet wurden.

Schlüsselwörter: Empfehlungen, Prüfungen

Jana Jünger[1]
Ingo Just[2]

1 Für den GMA-Ausschuss Prüfungen, Leiterin, Heidelberg, Deutschland

2 Für die Arbeitsgruppe Prüfungen des Medizinischen Fakultätentags, Hannover, Deutschland

Einleitung

Diese Empfehlungen für fakultätsinterne Prüfungen sind an alle Mitarbeiterinnen und Mitarbeiter[1] der human-, zahn- und tiermedizinischen Fakultäten in Deutschland, Österreich und der Schweiz gerichtet, die mit der Planung, Durchführung und Auswertung von fakultätsinternen Prüfungen betraut sind, also Dozentinnen und Dozenten, Studiendekanate und aufgrund der engen Verzahnung von Prüfungen und Lehre auch Curriculumsentwickler und Lehrverantwortliche. Die Empfehlungen beinhalten Qualitätsstandards, die u. a. für eine objektive, zuverlässige, valide – und damit justitiable – Prüfung Voraussetzung sind. In Form einer Checkliste geschrieben, sollen die Empfehlungen als praktisches Arbeitsinstrument zur Organisation von Prüfungen dienen.

Hintergrund

Im Jahr 2008 legte der GMA-Ausschuss Prüfungen gemeinsam mit dem Kompetenzzentrum für Prüfungen in der Medizin Baden-Württemberg „Leitlinien für fakultätsinterne Leistungsnachweise in der Medizin“ vor [1]. Diese sollten dabei helfen, konsentierte Qualitätsstandards für die durch die Neufassung der Ärztlichen Approbationsordnung des Jahres 2002 erforderlichen Prüfungen an den medizinischen Fakultäten in Deutschland zu etablieren, die den international anerkannten Ansprüchen an qualitativ hochwertige Verfahren der Leistungserfassung genügen (z. B. [2], [3], [4]). Ihre Bedeutung zeigt sich an verschiedenen Publikationen zu Prüfungsformaten und der Qualität universitärer Prüfungen, die auf dem Hintergrund dieser Empfehlungen entstanden [5], [6], [7].

Die wesentliche Bedeutung von Leistungsrückmeldungen und Leistungserfassungen und ihrer lernsteuernden Wirkung für die medizinische Ausbildung und die daraus folgende Notwendigkeit einer systematischen Einbindung des Prüfungsgeschehens in das Curriculum („constructive alignment“, „programmatic assessment“ [8], [9], [10], [11], [12]) sind allgemein anerkannt, ihre praktische Umsetzung ist vielfach jedoch noch defizitär. Dies gilt insbesondere bei Lehrinhalten, die über die traditionell vorherrschende Vermittlung von medizinischem Expertenwissen hinausgehen, wie sie etwa im CanMEDS-Rollenmodell der ärztlichen Tätigkeiten beschrieben sind. Auf der Grundlage dieses Rollenmodells werden im Schweizer Lernzielkatalog und dem sich in Entwicklung befindenden „Nationalen kompetenzbasierten Lernzielkatalogs Medizin“ (NKLM) [13] in Deutschland die erforderlichen Qualifikationen und Kompetenzen der ärztlichen Ausbildung definiert.

Diese Entwicklungen in den Anforderungen an die ärztliche Ausbildung müssen sich somit auch in den Verfahren zur Leistungserfassung spiegeln, neue Prüfungsformate und -methoden zur Erfassung der für die Ausübung des

ärztlichen Berufs erforderlichen Qualifikationen und Kompetenzen müssen entwickelt und eingesetzt werden. Für die Praxis der Prüfungen in der medizinischen Ausbildung bedeutet dies, dass häufiger verschiedene Prüfungsformen kombiniert werden, formative Prüfungen im Vergleich zu summativen Prüfungen einen breiteren Raum einnehmen [12], sowie kriteriumsorientierte Bewertungen einen höheren Stellenwert aufweisen sollten. Dem soll die vorliegende Aktualisierung der Empfehlungen von 2008 Rechnung tragen. Insbesondere ist dabei sicherzustellen, dass an innovative Prüfungsformen die gleichen Qualitätsansprüche bezüglich Messzuverlässigkeit und Aussagekraft gestellt werden wie an traditionelle Prüfungsmethoden.

Schwerpunkt der Empfehlungen sind nach wie vor Prüfungen, die zur Erlangung von Leistungsnachweisen an den medizinischen Fakultäten erbracht werden müssen. Solche „summativen oder bilanzierenden Beurteilungen bezwecken die abschließende Ermittlung eines Lernstands“ [14]. Die formalen – insbesondere gesetzlichen – Anforderungen an rein formative Prüfungen sind i. A. deutlich geringer, für die inhaltliche Qualität der Aufgaben gelten jedoch die gleichen Ansprüche wie bei summativen Prüfungen.

Den Verfassern dieser Empfehlungen ist bewusst, dass eine vollständige Umsetzung die medizinischen Fakultäten vor erhebliche organisatorische und personelle Probleme stellt, die nur mittel- oder sogar längerfristig zu bewältigen sind. Dennoch zeigen Beispiele an den medizinischen Fakultäten, dass sämtliche Punkte der Empfehlungen erfüllbar sind. Die Fakultäten sind deshalb aufgerufen, in einem kontinuierlichen Prozess die Qualität ihrer Leistungserfassungen und -bewertungen zu verbessern. Um dies zu unterstützen, ist vorgesehen, durch den Ausschuss Prüfungen der GMA beispielgebende praktische Ansätze zu Umsetzungen der Anforderungen der vorliegenden Empfehlungen zu veröffentlichen.

Aktualisierung der Empfehlungen

Auf Grund der oben erwähnten Entwicklungen wurde vom Ausschuss Prüfungen der GMA im Jahr 2012 eine Aktualisierung der Empfehlungen aus dem Jahr 2008 beschlossen. Im Rahmen der „International Conference in competency-based Assessment“ in Heidelberg am 04.07.2012 wurden erste Verbesserungsvorschläge (vgl. [15]) entworfen und in einer weiteren Sitzung am 27.09.2012 bei der Jahrestagung der GMA in Aachen gemeinsam mit der Arbeitsgruppe Prüfungen des MFT die Themenbereiche 1-4 der Empfehlungen (Allgemeine strukturelle Vorbedingungen, Prüfungskonzeption und -bewertung, organisatorische Vorbereitungen zur Prüfungsdurchführung, Durchführung der Prüfung) eingehend diskutiert und Verbesserungen erarbeitet. Eine weitere Diskussion sowie die Behandlung der Themenbereiche 5-7 (Auswertung und Dokumentation, Rückmeldung an die Studierenden, Prüfungsnachbereitung) erfolgten auf der Sitzung des Ausschusses Prüfungen der GMA und der AG Prüfungen des MFT während der GMA-Tagung am 26.9.2013 in Graz. Nach der Einarbeitung der dort beschlossenen Veränderungen wurde die Aktualisierung in einem Umlaufverfahren weiter ergänzt. Im Januar 2014 erfolgte eine externe juristische Prüfung[2] dieser Version durch eine auf Prüfungsrecht spezialisierte Kanzlei in Hannover.

Die daraus erwachsenen Änderungen wurden Anfang Februar 2014 eingearbeitet und am 11.2.2014 in einer Sitzung des Ausschusses Prüfungen der GMA diskutiert. Noch offene Punkte wurden auf dieser Sitzung geklärt und eingearbeitet. Die Empfehlungen wurden sowohl in der Arbeitsgruppe Lehre des Medizinischen Fakultätentags (MFT) und dem Vorstand der GMA im Mai 2014 vorgestellt und verabschiedet. Sowohl MFT und GMA unterstützen die Empfehlungen, die Leitliniencharakter haben.

Erläuterung zur neuen Version der Empfehlungen

Die erste Version der Empfehlungen [1] bestand aus den als Checkliste formulierten Einzelpunkten und zugehörigen nummerierten Erläuterungen. Um die Lesbarkeit zu erleichtern, sind in dieser Version die einzelnen Punkte der Empfehlungen mit entsprechenden Erläuterungen als fortlaufender Text formuliert, eine zusätzliche Checkliste befindet sich im Anhang. Die in der Checkliste aufgelisteten Einzelkriterien sind im folgenden Text jeweils kursiv gesetzt (siehe Anhang).

1. Allgemeine strukturelle Vorbedingungen: Inhaltliche und formale Voraussetzungen

Die strukturellen Vorbedingungen umfassen Kriterien, die die curriculare Einbindung der Lehrveranstaltung(en), auf die sich die Prüfungen beziehen, gewährleisten sollen, formale Anforderungen zur Information der Studierenden und Regularien sowie Qualifizierung der Prüfungsverantwortlichen. Sie beziehen sich damit nicht auf Vorbereitung oder Durchführung einer konkreten Prüfung sondern betreffen die Rahmenbedingungen, die für qualitativ hochwertiges Prüfen erforderlich sind.

1.1. Gesamtprüfungsprogramm

Ein Gesamtprüfungsprogramm, in dem Anzahl, Umfang, Inhalt, zeitlicher Ablauf und Format der im Medizinstudium durchzuführenden summativen wie formativen Einzelprüfungen aufeinander abgestimmt sind, liegt allen Studierenden und Lehrenden vor.

Die an der Fakultät bzw. im Studiengang Medizin/Zahnmedizin/Tiermedizin verwendbaren Prüfungsformen sollten in den entsprechenden formalen Regelungen (Studienordnung, Prüfungsordnung oder in geeigneten Ausführungsbestimmungen) aufgeführt und hinsichtlich ihrer Durchführung und Bewertung festgelegt sein. Dabei

ist Sorge zu tragen, dass die Bestimmungen hinreichend Raum für die Etablierung innovativer Prüfungsformen bieten.

Es ist darauf zu achten, dass die Prüfungsinhalte mit adäquaten Prüfungsformen, die sowohl Methoden zur Leistungserfassung theoretischer Kenntnisse wie auch praktischer Fertigkeiten beinhalten, abgeprüft werden („Triangulation": Leistungserfassung auf Basis unterschiedlicher Quellen, zu unterschiedlichen Zeitpunkten, unter unterschiedlichen Bedingungen, durch verschiedene Personen und mit unterschiedlichen Methoden [16], [17]). Z. B. können theoretische Kenntnisse mit schriftlichen Klausuren, praktische Prüfungsinhalte mit objektiv-strukturierten praktischen/klinischen Prüfungen (OSPE/OSCE) angemessen erfasst werden. Die Prüfungsformen sollten den jeweiligen Qualitätsanforderungen an Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) genügen. Basiert die Leistungsbewertung auf verschiedenen Prüfungsteilen, so bezieht sich die Anforderung an die Messzuverlässigkeit auf die Gesamtprüfung, nicht notwendigerweise auf die einzelnen Teile (siehe auch Erläuterung zu 2.6).

Um die Studierenden bestmöglich auf ihre spätere ärztliche Berufsausübung vorzubereiten, sind in die curricularen Lernziele Kompetenzen aufzunehmen, die wesentlich über medizinisches Expertenwissen und fachlichen Fertigkeiten hinausgehen. Damit ist es aber auch erforderlich, geeignete Prüfungsformen und Strukturen zu entwickeln, die eine angemessene, zuverlässige und praktikable Leistungserfassung dieser Kompetenzen ermöglichen. Dies bedingt den Einsatz neuer Prüfungsformen, insbesondere arbeitsplatzbasierter Prüfungsformen, wie z. B. DOPS, Encounter-Cards oder 360°-Assessment, zur Erfassung von kommunikativen Kompetenzen, professionellem Handeln, Managementfähigkeiten. Besondere Beachtung erfordert die Qualitätssicherung dieser Prüfungsformate, so ist etwa im Vorfeld eine ausreichende Schulung der Prüfer sicherzustellen oder bei der Analyse der Prüfungsergebnisse zur Kontrolle der bei arbeitsplatzbasierten Leistungserfassungen geringeren Standardisierbarkeit der Prüfungssituation der Einsatz angemessener Auswerteverfahren (z. B. Generalisierbarkeitstheorie) vorzusehen.

Die Leistungsmessung bei nichtfachspezifischen Lernzielen ist logistisch oft nicht im Rahmen der einzelnen Fachprüfungen durchzuführen. Hier sind andere Prüfungsstrukturen denkbar, bei denen Prüfungsbestandteile einzelner Prüfungen fachübergreifend und analog zu einem Portfolio zusammengestellt und beurteilt werden. So könnten z. B. die Kommunikationsstationen in OSCEs verschiedener Fächer für eine Bewertung der Kompetenz als „Kommunikator" zusammengefasst werden. Dieses Portfolio könnte auch die Erfassung von kritischen Ereignissen (etwa zur Beurteilung professionellen Verhaltens) umfassen.

1. 2. Lernzielkatalog

Für jede in der Studienordnung definierte Unterrichtseinheit (z. B. Fach, Modul, Kurs, Seminar, Querschnittbereich) im vorklinischen und klinischen Abschnitt des Studiums liegt ein vollständiger schriftlicher Lernzielkatalog vor.

Aus der Gesamtheit der Lernzielkataloge der Unterrichtseinheiten muss entnommen werden können, welche Lernziele bei Vorliegen eines Gesamtlernzielkatalogs in welchen Veranstaltungen vermittelt werden.

1.3. Information der Studierenden bzgl. Lernzielkatalog

Die Studierenden werden vor jeder Unterrichtseinheit/jedem Modul über die spezifischen Lern- und Prüfungsziele zeitgerecht informiert.

1.4. Adäquate Prüfungsformate

Die in den Lernzielen formulierten Kenntnisse, Fähigkeiten und Haltungen werden mit adäquaten Prüfungsformaten geprüft. Insbesondere sind Verfahren einzusetzen, die geeignet sind, ärztliche Entscheidungs- und Handlungskompetenzen sowie Fertigkeiten der ärztlichen Gesprächsführung zu erfassen (s. 1.1).

Neben schriftlichen Prüfungsformaten (als Multiple-Choice-Prüfung oder mit offenen Fragen), die vornehmlich der Prüfung theoretischen Wissens dienen, sind zur Leistungserfassung praktischer Fertigkeiten in medizinischen Studiengängen OSCEs etabliert. Zur Erfassung anderer Kompetenzbereiche ärztlichen Handelns sind darüber hinaus weitere Prüfungsformen erforderlich, mit denen zuverlässige arbeitsplatzbezogene Leistungserfassungen möglich sind. Hierzu gehören z. B. miniCEX, 360°-Assessments, Encounter-Cards, Direct Observation of Practical Skills (DOPS).

1.5. Schriftliche Regelungen für Prüfungsvorbereitung und Prüfungsablauf

Für die nachfolgend aufgeführten Bereiche sollten schriftliche Regelungen vorhanden sein.

1. *Teilnahmevoraussetzungen*
2. *Festsetzung von Prüfungsterminen (incl. Wiederholungstermine) und formaler Prüfungsablauf.*
   Bei jeder Prüfung sollten klare Regularien für den formalen Prüfungsablauf standardmäßig eingehalten werden. Diese Regularien sollten schriftlich niedergelegt sein und folgende Aspekte enthalten:
   – Form und Terminvorgaben für Prüfungsankündigung
   – Form und Terminvorgaben für die Anmeldung der Studierenden zur Prüfung, ggf. automatische Anmeldung zur Prüfung durch Einteilung zum Modul
   – Zahl und Qualifikation der Prüfer (z. B. Facharzt, Habilitation usw.)
   – Dauer der Prüfung
   – Prüfungseinführungen (z. B. eigener Termin zur Einweisung für computerbasierte Prüfungen)
   – Ansagen zu Beginn der Prüfung
   – Bei der Prüfung erlaubte Hilfsmittel
   – Mitnahme von Prüfungsunterlagen
   – Umgang bei verspätetem Erscheinen zur Prüfung
   – Rücktritt und Versäumnis von Prüfungen

– Vorgehen bei Täuschungsversuchen
– Regelungen zum Prüfungsabbruch

3. *Regelungen zu den im Studiengang einsetzbaren Prüfungsformen (s. 1.1)*
4. *Festlegung räumlicher und zeitlicher Voraussetzungen und Bedingungen für die Prüfungsdurchführung (s. 3.3)*
5. *Bewertungsskalen, Bestehensgrenzen, Anwendung einer Gleitklausel*[3] *(s. 2.5, 2.8)*
6. *Bewertung bei fehlerhaft gestellten Aufgaben (s. 5.2)*
7. *Gewichtung von Teilprüfungen (s. 3.1)*
8. *Kompensationsmöglichkeiten und Nachteilsausgleich bei Prüfungen (s. 1.6)*
9. *Teilnahmebedingungen und Verfahren für Nach- und Wiederholungsprüfungen (s. 1.6)*
10. *Bekanntmachung und Einsichtnahme in Prüfungsergebnisse (s. 6.2)*
11. *Regelungen bei Einsprüchen gegen Bewertung und Prüfungsaufgaben (s. 5.2, 6.3)*
12. *Umgang mit Verletzungen der Durchführungsbedingungen und außergewöhnlichen Störungen der Prüfungsdurchführung sowie Regelungen für dadurch erforderliche Prüfungswiederholungen (s. 4.3)*
13. *Veröffentlichung von Aufgaben (s. 6.5)*
14. *Dokumentation der Prüfung und der Prüfungsergebnisse (s. 5.5)*

1.6. Kompensation von Prüfungsleistungen, Nach- und Wiederholungsprüfungen

1. *Können Leistungsnachweise oder Teile von Leistungsnachweisen seitens Studierender nicht oder nur unter nicht zumutbaren Bedingungen erbracht werden, die in der Art und Form der Prüfungsdurchführung begründet sind, sollte grundsätzlich geklärt sein, unter welchen Bedingungen Prüfungsleistungen kompensiert werden können.*
   Dies betrifft z. B. Studierende mit körperlichen Beeinträchtigungen, bei denen u. U. der Behindertenbeauftragte hinzugezogen werden sollte, oder Studierende mit eingeschränkten Kenntnissen der deutschen Sprache, die nicht regulär im Studiengang eingeschrieben sind (Studierende in internationalen Studienaustauschprogrammen, z. B. Erasmus).
2. *In den maßgeblichen rechtlichen Bestimmungen (Studienordnung, Prüfungsordnung) sind die Bedingungen für die Durchführung und Teilnahme an Nach- und Wiederholungsprüfungen festzulegen. Ebenfalls geregelt sein muss, ob und inwieweit notenverbessernde Prüfungen durchgeführt werden.*
   Das Prüfungsformat für Wiederholungs- und Nachprüfungen sollte mit dem Format der Erstprüfung übereinstimmen, z. B. sollte keine schriftliche oder mündliche Nachprüfung bei nicht bestandenem OSCE durchgeführt werden. Ebenfalls sollte bei Nichtbestehen einer schriftlichen Prüfung keine mündliche Nachprüfung erfolgen[4].
   Bei eigenständigen Wiederholungsprüfungen (also Prüfungen, in denen mehrheitlich Kandidaten geprüft werden, die mindestens einmal nicht bestanden haben), ist u. U. eine Modifikation der Gleitklausel zu empfehlen (s. auch 2.5).

1.7. Prüfungsverantwortliche

1. *In jedem Fach ist mindestens ein Prüfungsverantwortlicher nebst Stellvertreter benannt, dessen Verantwortlichkeiten klar definiert sind. (Verantwortungsbereiche: z. B. Blueprint, Fragenerstellung, Durchführung, Korrektur, Prä- und Postreview, Auswertung, Rückmeldung an Curriculumsentwickler).*
2. *Die Prüfungsverantwortlichen haben an Weiterbildungsmaßnahmen zum Thema Prüfungen teilgenommen.*
   Jeder Prüfungsverantwortliche für einen Lehrbereich (Fach, Modul, Block etc.) sollte eine zertifizierte Weiterbildung zum Thema Prüfungen aufweisen können.

2. Prüfungskonzeption und bewertung

Die folgenden Empfehlungen beziehen sich auf die Vorbereitung konkreter Prüfungen. Sie betreffen die curriculare Anbindung der Prüfungsinhalte und Maßnahmen zur Qualitätssicherung von Aufgaben und Gesamtprüfung (Reliabilität und Validität) sowie die ökonomische und für die Studierenden transparente Durchführung.

2.1. Abstimmung der Prüfungen mit Gesamtprüfungsprogramm

Die Einzelprüfungen sind mit dem Gesamtprüfungsplan des Studiengangs abgestimmt. Diese Abstimmung betrifft sowohl summative als auch formative Leistungsrückmeldungen.

2.2. Validität

Jeder Einzelprüfung liegt ein schriftliches Gesamtkonzept („Blueprint“) zugrunde, das die fachspezifischen Prüfungsinhalte repräsentativ abbildet.
Der Blueprint dient der Sicherung der inhaltlichen Validität der Prüfung. Diese wird gewährleistet

1. durch die Repräsentativität der Aufgaben für den abzuprüfenden Bereich und
2. die Vermeidung von für diesen irrelevanten Prüfungsinhalten („konstuktirrelevante Varianz“).

Die Validität ist das Kriterium für die Testgüte. Sie ist ein Maß dafür, ob die bei der Messung erzeugten Daten wie beabsichtigt die zu messende Größe, also die Kenntnisse oder Fertigkeiten in dem durch die Prüfung abzudeckenden Fachgebiet o. ä., repräsentieren: Misst der Test das Merkmal, das er messen soll[5]?
Nach der Auswertung der Prüfungen können weitere Quellen der Validität untersucht werden:

- stellen sich die Testergebnisse plausibel dar?
- Gibt es eine hohe Korrelation zwischen dem Test und anderen Tests, die das gleiche Konstrukt messen sol-

len (z. B. MC-Klausur Innere Medizin und Anteil Innere Medizin in Staatsprüfungen)?

2.3. Einbindung von Fachvertretern

Bei der Zusammenstellung der Prüfungen sind Vertreter aus allen beteiligten Lehrgebieten beteiligt.

2.4. Begutachtung der Prüfungsaufgaben (Prä-Review) und Prüfung der inhaltlichen Validität

1. *Vor der Durchführung einer Prüfung findet eine standardisierte, inhaltliche und formale Bewertung (Prä-Review) der Prüfungsaufgaben statt.*
   Bei Prüfungsformaten, in denen nur eingeschränkte Möglichkeiten zu ihrer Standardisierung bestehen (z. B. arbeitsplatzbasierte Prüfungen), ist festzulegen, wie unterschiedliche Rahmenbedingungen und Schwierigkeitsgrade berücksichtigt werden (z. B. detailliertes Standard-Setting).
   Bei der Erstellung von Prüfungen sollten im Hinblick auf die Validität die folgenden Punkte in einer Gesamtsicht beachtet werden:
   – Ist jede einzelne Aufgabe qualitativ hochwertig? Hier ist insbesondere wichtig, dass nur die zu messende Fertigkeit/Fähigkeit (z. B. Wissen in einem bestimmten Fachgebiet) zur richtigen Beantwortung führt und nicht andere Fähigkeiten (z. B. Sprachkenntnisse) zur Lösung der Aufgabe erforderlich sind.
   – Sind die Inhalte allgemeingültig/evidenzbasiert und z.B. keine lokale Lehrmeinung?
   – Deckt sich der Inhalt der Prüfung mit der Lehre/den Lernzielen?
   – Handelt es sich um Wissen, das beim gegenwärtigen Ausbildungsstand erwartet werden kann und nicht z.B. um Inhalte, die in einen späteren Abschnitt des Studiums oder der medizinischen Weiterbildung gehören?
   – Ist der Inhalt des zu prüfenden Stoffgebiets mit seinen Subgebieten verhältnismäßig angemessen und umfassend im Test vertreten? Um dies zu gewährleisten müssen angemessene Methoden der Fragenzusammenstellung standardmäßig gewählt werden (s. 2.2, „Blueprint“).
   – Wurden die Prüfungsaufgaben sowie die ganze Prüfung einem sorgfältigen Review-Prozess unterzogen?
   – Ist das theoretische Rahmenwerk argumentativ nachvollziehbar?
   – Erscheint der Test den Prüflingen plausibel? (Akzeptanz)
2. *Am Review nehmen mindestens zwei Fachvertreter und ein Vertreter eines anderen Faches teil.*
3. *Das Ergebnis der Begutachtung muss dokumentiert werden.*

2.5. Bestehensgrenzen

1. *Vor der Durchführung einer Prüfung wird die Bestehensgrenze durch ein interdisziplinäres Expertengremiums nach inhaltlichen Kriterien (z. B. mittels eines Standard-Setting-Verfahrens) oder anhand eines formalen Kriteriums (z. B. 60%-Regel) schriftlich festgelegt.*
   Bestehensgrenzen sollten möglichst anhand inhaltlicher Kriterien entsprechend einer kriteriumsorientierten Leistungsmessung festgesetzt werden (vgl. z. B. Verfahren des Standardsetzens beim OSCE). Bei MC-Fragen sollten mindestens formale Kriterien (z. B. 60%-Regel) eingesetzt werden.
2. *Eine Regelung zur Anwendung einer Gleitklausel ist schriftlich festgelegt.*
   Eine Gleitklauselregelung ist i. A. bei Prüfungen mit Multiple-Choice-Aufgaben erforderlich. Im Studiengang sollte durch eine einheitliche Regelung klargestellt sein, bei welchen Prüfungsformen und in welcher Weise eine Gleitklausel einheitlich zur Anwendung kommt. Es ist festzulegen, wie Prüfungen mit gemischten Formaten (z.B. Multiple-Choice und offene Fragen) zu behandeln sind.
   Ergänzend zu einer kriteriumsorientierten Bestehensgrenze sollten auch bei anderen Prüfungsformen entsprechende Regelungen zur Kompensation unangemessen schwieriger Prüfungen getroffen werden, die den Studierenden rechtzeitig bekannt gegeben werden müssen.
   Wir empfehlen für Prüfungen mit MC-Fragen zur Vereinfachung eine modifizierte Gleitklausel, die die durchschnittliche Prüfungsleistung aller Teilnehmer, die zum ersten Mal in direktem Anschluss an den Kurs an der Prüfung teilnehmen, (ohne Beschränkung auf Studierende in der Regelstudienzeit o. Ä.) berücksichtigt. Für Nachhol- und Wiederholungsprüfungen mit einem erheblichen Anteil an Teilnehmern, die die Prüfung nicht zum ersten Mal ablegen, sind geeignete Regelungen zu treffen.
3. *Rundungen der Bestehens- und Notengrenzen sind verbindlich festzulegen.*
   Ergibt sich z. B. bei 99 Aufgaben und einer Bestehensgrenze von 60% der maximalen Punktzahl die Bestehensgrenze von 59,4 Punkten, so wird empfohlen, diese auf 60 Punkte aufzurunden, falls bei den Prüfungsaufgaben nur ganze Punkte vergeben werden. Werden auch halbe Punkte vergeben, so wäre die Bestehensgrenze entsprechend auf 59,5 Punkte zu setzen (nach der deutschen ÄAppO muss die Mindestprozentzahl zum Bestehen erreicht oder überschritten sein, d. h. es finden in keinem Fall Abrundungen statt).

2.6. Reliabilität der Prüfung

Bei summativen Prüfungen ist eine Reliabilität von mindestens 0,8 für den Leistungsnachweis zu erwarten.
Soweit methodisch möglich, wird empfohlen, Leistungsnachweise eines Fachs auf Basis mehrerer Teilprüfungen zu erstellen (s. Erläuterung zu 1.1). In diesem Fall ist das Kriterium der Mindestreliabilität von 0,8 auf die Gesamtbewertung anzuwenden, nicht notwendigerweise auf die einzelnen Teilprüfungen. Ein Beispiel hierfür wäre, wenn in einem Fach sowohl eine schriftliche Prüfung für das theoretische Wissen wie auch eine OSCE-Prüfung für die praktischen Fertigkeiten abzulegen sind. Hier kann sowohl

bei Klausur wie bei OSCE die Reliabilität der beiden einzelnen Prüfungen jeweils 0,8 unterschreiten, die Reliabilität der zusammengesetzten Prüfung kann aber merklich höher sein. Zur Bestimmung der Reliabilität zusammengesetzter Leistungsnachweise sei auf die einschlägige Literatur verwiesen (z. B. [18]).

Dabei ist zu beachten, dass Teilprüfungen, die nicht durch andere Prüfungsleistungen kompensiert werden können, eine hinreichende Messzuverlässigkeit der Entscheidung bestanden/nicht bestanden aufweisen, um zu verhindern, dass einem Studierenden auf Grund einer einzelnen wenig zuverlässigen Teilprüfung der Leistungsschein verweigert wird. Beispiele hierfür sind etwa „K.O.-Stationen" in einem OSCE oder Teilprüfungen bei fächerübergreifenden Leistungsnachweisen, die jede für sich bestanden werden müssen.

Um in einzelnen Prüfungen eine Reliabilität von mindestens 0,8 zu erreichen, sind in der Regel bei MC-Klausuren wenigstens 40 qualitativ hochwertige Fragen erforderlich, bei OSCE wenigstens 12 Stationen. Diese Angaben können nur als grobe Anhaltspunkte dienen, in Abhängigkeit von Prüfungsziel, Aufgabenqualität und zu prüfender Studierendenkohorte sind erhebliche Schwankungsbreiten möglich, weshalb zur Abschätzung der zu erwartenden Reliabilität die Kennwerte entsprechender früherer Prüfungen des Faches herangezogen werden sollten.

Insbesondere „kleine" Fächer stehen vor dem Problem, dass auf Grund des Prüfungsumfangs nur schwer eine Reliabilität von wenigstens 0,8 erreicht werden kann. Eine Lösung im Bereich der Humanmedizin in Deutschland bieten die sog. „fächerübergreifenden Leistungsnachweise", bei denen mehrere Fächer, die ihre Unterrichtseinheiten in zeitlicher Nähe durchführen, eine gemeinsame Prüfung durchführen und eine Gesamtnote bilden können. Sind keine fächerübergreifenden Leistungsnachweise möglich, sollte zumindest durch eine intensive Qualitätssicherung eine möglichst hohe Validität durch Repräsentativität der Aufgaben für den Lehrstoff und der Vermeidung von Aufgaben, die lernzielfremde Kenntnisse oder Fertigkeiten prüfen (konstruktirrelevante Varianz), gesichert werden.

2.7. Ressourcenaufwand

Die geplante Prüfung ist ressourcensparend konzipiert.

Hierunter sind Möglichkeiten einer Einsparung von Ressourcen bei der Konzeption, Durchführung und Auswertung der Prüfungen zu verstehen. Dazu gehören z. B. Einlesen der Antwortbögen durch Belegleser, adäquate Anzahl der Aufsichtspersonen, Verwendung eines fakultäts-/studiengangsübergreifenden Prüfungspools, Einsatz computerbasierter Durchführung, standardisierte teststatistische Auswertung (z. B. zentral in der Fakultät), Verwendung der Mindestanzahl von Prüfern (z. B. beim OSCE einer pro Station bei zentraler Aufsicht ausreichend), Wahl ressourcensparender Prüfungsformate und Aufgabenformate (offene Fragen auf das Notwendige beschränken).

2.8. Bewertung der Aufgaben

1. *Die zu verwendenden Bewertungsskalen (Noten, Punkte) von Prüfungen sollten für den Studiengang einheitlich und verbindlich sein.*
2. *Die richtigen Antworten, der Erwartungshorizont, die Korrekturrichtlinien und Bewertungsmodus sind vor der Durchführung der Prüfung schriftlich festgelegt.*
   Die richtigen Antworten und der Erwartungshorizont liegen dem Prüfer schriftlich vor. Die schriftliche Korrekturanleitung für eine Klausur ist eindeutig (z. B. zur Vergabe halber Punkte oder zur Korrektur offener Fragen). Empfehlung: Jeweils derselbe Prüfer sollte die Antworten aller Studenten einer offenen Frage korrigieren.
   Der Bewertungsmodus bei einem OSCE ist eindeutig festgelegt. Für jede OSCE-Station/-aufgabe ist eindeutig festgelegt, wie viele Punkte anhand einer Checkliste oder auf Basis einer globalen Beurteilung („Global Rating") der Fertigkeit/Fähigkeit vergeben werden. Für mündliche Prüfungen gilt Entsprechendes.
3. *Die Anzahl der Punkte für jede einzelne Frage/Aufgabe ist vor Prüfungsbeginn festgelegt.*
   Bei schriftlichen Prüfungen ist bei nicht einheitlicher Gewichtung der Aufgaben die jeweils zu erzielende Punktzahl in der Klausur anzugeben. Es ist zu beachten, dass bei MC-Aufgaben, die nicht vom Einfachauswahltyp sind (z.B. „Mehrfach richtig/falsch"), erbrachtes Teilwissen ebenfalls angemessen zu berücksichtigen ist.

2.9. Bewertung von Teilprüfungen

1. *Setzen sich die im Zeugnis aufzuführenden Noten aus mehreren Teilprüfungen zusammen, sollten die Bewertungen der Teilprüfungen auf einer hinreichend differenzierten Bewertungsskala vorgenommen werden.*
   Notenskalen, wie etwa das deutsche System der Vergabe von 4 Notenstufen bei bestandener Prüfung, bilden die Prüfungsleistungen nur grob ab. Werden wenig abgestufte Noten von Teilprüfungen zur Bildung einer Gesamtnote zusammengefasst, können durch die Mittelung Verzerrungen der Beurteilung der Gesamtleistung entstehen.
2. *Die Rundung der Noten sollte eindeutig festgelegt werden*.
   Rundungen auf ganze Zahlen, wie sie bei der Bildung der durch die deutsche ÄAppO für das Zeugnis geforderten Notenstufen 1, 2, 3 und 4 notwendig sind, sollten immer in Richtung der nächstliegenden ganzen Zahl durchgeführt werden. Bei gleichem Abstand (Dezimalstellen 0,500) ist zu Gunsten der Studierenden zu runden, so ist etwa 1,500 auf die ganze Note 1, hingegen ist 1,501 auf die ganze Note 2 zu runden. Es wird empfohlen, Teilbewertungen, die zu einer Gesamtnotenbildung verwendet werden, auf einer Skala mit wenigstens drei Kommastellen durchzuführen. Die Verwendung von drei Kommastellen ist im Normalfall hinreichend genau, um Verzerrungen durch

iterierte Rundungen zu vermeiden (wie etwa, dass im deutschen System aus 2,54 durch die Rundung auf eine Stelle 2,5 und durch eine nachfolgende Rundung in Richtung der besseren Note eine 2 wird).
Die Bewertungsskala sollte eine geforderte Gleichabständigkeit von Notengrenzen berücksichtigen. Ist z. B. bei schriftliche Prüfungen vorgegeben, dass ab 60% bis zu 70% der erreichten Punktzahl die Note 4, ab 70% bis zu 80% die Note 3, ab 80% bis 90% die Note 2 und ab 90% eine 1 zu vergeben ist, so müsste eine notenäquivalente Dezimalskala von 0,5 bis 4,5 reichen. Damit wird das Intervall von 80-90% der erreichten Punktzahl (Note 2) auf die Noten von 2,5 bis 1,5 und 90-100% auf ein gleich großes Intervall von 1,5 bis 0.5 abgebildet. Nur so ist eine einfache lineare Umrechnung der Punktwerte in Dezimalnoten möglich.

3. Organisatorische Vorbereitungen zur Prüfungsdurchführung

Neben der inhaltlichen Vorbereitung der Prüfung bedarf es verschiedener organisatorischer und logistischer Vorarbeiten, um einen formal korrekten Ablauf der Prüfung zu gewährleisten.

3.1. Bekanntgabe von Prüfungsterminen und -formate

Die Prüfungstermine und -formate werden den Studierenden zu Beginn einer Unterrichtseinheit bekannt gegeben.
Setzt sich die Gesamtnote eines Fachs oder Querschnittsbereichs aus mehreren Teilprüfungen zusammen, so sind die Gewichtungen bekannt zu geben. Bei Änderungen der Zusammensetzung der Gesamtnote oder Gewichtung der Teilprüfungen sind klare Übergangsregelungen für Studierende zu formulieren, die Prüfungen wiederholen müssen.

3.2. Prüfungsanmeldung

Für jede Prüfung ist eine schriftliche oder Online-Anmeldung durch die Studierenden erforderlich. Die Anmeldung zu Lehrveranstaltung und Prüfung kann gemeinsam erfolgen. Unter Umständen kann bei curricular feststehenden Prüfungen eine eigenständige Anmeldung nicht erforderlich sein.
Es sollte geregelt sein, ob Studierende bei Nichtbestehen einer Prüfung für die nächstmögliche Wiederholung automatisch angemeldet sind oder ob eine gesonderte Anmeldung erforderlich ist
Summative Prüfungen sind als Abschluss einer Unterrichtseinheit zu sehen und sollten sich direkt auf das vorangegangene Curriculum beziehen. Deshalb ist es empfehlenswert, die Prüfung(en) oder letzte Teilprüfung verpflichtend für alle Studierenden zeitnah nach Abschluss der Unterrichtseinheit durchzuführen.

3.3. Räumlichkeiten und Personal für Prüfungen

1. *Zur Durchführung der Prüfung ist gewährleistet, dass ausreichend Räume zur Verfügung stehen und diese für alle Kandidaten vergleichbare Bedingungen bieten.*
2. *Zur Durchführung der Prüfung steht ausreichend geschultes Personal zur Verfügung (Prüfer, Aufsichtspersonen, Korrektoren zur Bewertung offener Fragen usw.).*

3.4. Schulung von Prüfern und Rückmeldung an Prüfer

1. *Die Prüfer und Korrektoren sind hinsichtlich einheitlicher Bewertungskriterien vor der Prüfung geschult.*
   Es soll ein gemeinsames Training derjenigen, die die Prüflinge bewerten, zur Erhöhung der Interrater-Reliabilität durchgeführt werden. Dies ist insbesondere bei parallelen Prüfungsparcours eines OSCE, bei mündlichen Prüfungen oder bei schriftlichen Prüfungen mit offenen Fragen notwendig.
   Für Prüfungen, in denen der Prüfer mit dem Prüfling direkt in Kontakt tritt, sind insbesondere Schulungen mit videographierten Prüfungen sinnvoll.
2. *Prüfer sind hinsichtlich Rückmeldung und Erläuterung der abgeprüften Leistungen und ihrer Bewertung an die Studierenden („Feedback“) geschult. Dies gilt insbesondere bei allen formativen Prüfungen.*
   Die Schulungsmaßnahmen sind an die speziellen Erfordernisse des Prüfungsformats anzupassen, neben einer eingehenderen Erstschulung sind Auffrischungsschulungen durchzuführen. Die Wirksamkeit der Schulungsmaßnahmen ist zu überprüfen (z. B. durch standardisierte Studierende).
3. *Prüfer erhalten Rückmeldung über ihre Prüfungsleistung.*
   Bei Prüfungen, in denen Prüfereinflüsse bei der Bewertung zu berücksichtigen sind, erfolgt eine Rückmeldung an die Prüfer (s. 5.1). So ist z. B. bei mündlichen oder mündlich-praktischen Prüfungen eine Rückmeldung hinsichtlich Strenge oder der Ausnutzung der Bewertungsskalen zu geben. U. U. ist vor dem Einsatz bei der nächsten Prüfung eine Nachschulung von Prüfern durchzuführen.

4. Durchführung der Prüfung

4.1. Einhaltung formaler Kriterien

Bei der Durchführung der Prüfung werden die schriftlich niedergelegten formalen Kriterien eingehalten und dokumentiert (z. B. mit Hilfe einer Checkliste zum formalen Prüfungsablauf).

4.2. Vollständigkeit der Prüfungsunterlagen

Die Vollständigkeit der Prüfungsunterlagen und des Prüfungsmaterials wird zu Beginn der Prüfung durch die Studierenden oder die Prüfungsaufsicht kontrolliert.
Eine eindeutige Zuordnung sowohl des Aufgaben- als auch des Antwortblattes zu jedem Studierenden und eine kontrollierte Abgabe ist für einen vollständigen Rückfluss aller Aufgabenblätter empfehlenswert.

4.3. Protokoll des Prüfungsverlaufs

Der Verlauf der Prüfung und dabei auftretende Probleme werden protokolliert (z. B. Nennung von Prüfungsverantwortlichen und -durchführenden, Aufsichtspersonen, spezielle Vorkommnisse, Täuschungsmanöver, Computerausfall bei computerbasierten Prüfungen).

Beispiele für die Verletzung von Durchführungsbedingungen sind:

- Lärmbeeinträchtigungen durch Baumaßnahmen während einer Klausur.
- Ungeeignete Prüfungsräume
- Mangelhafte Prüfungsmaterialien wie schlechte Kopien der Prüfungsaufgaben, fehlerhafte Fragennummerierungen.
- Ausfall von Computern bei computerbasierten Prüfungen

Studierende müssen eine Verletzung von Durchführungsbedingungen unverzüglich während oder nach der Prüfung geltend machen. Es ist nicht zulässig, zunächst das Ergebnis der Prüfung abzuwarten und sich im Falle des Nichtbestehens auf die Verletzung der Durchführungsbestimmungen zu berufen.

Bei erheblichen Beeinträchtigungen der Prüfungsdurchführung wird empfohlen, eine Wiederholungsprüfung für alle Prüfungsteilnehmer anzubieten und das bessere Prüfungsergebnis zu werten.

Geregelt werden sollte auch die Entscheidung über Störungen der Prüfung durch Prüfungsteilnehmer und deren möglichen Ausschluss, ebenso der Abbruch einer Prüfung (z. B. wegen akuter Erkrankung) und die entsprechende Dokumentation durch den Prüfungsverantwortlichen.

5. Auswertung und Dokumentation

Eine sorgfältige Auswertung der Prüfung mit Dokumentation einschließlich der statistischen Analyse ist zur inhaltlichen und insbesondere rechtlichen Absicherung erforderlich. Bei summativen Prüfungen entstehen Studierenden bei Fehlern u. U. erhebliche Nachteile, die von Mehrarbeit und Verlängerung der Studiendauer bei erforderlichen Prüfungswiederholungen, der Nichtgewährung von Stipendienleistungen bei ungerechtfertigter schlechter Bewertung bis hin zum Studienabbruch reichen können. Statistische Analyse und Dokumentation sind darüber hinaus eine wesentliche Grundlage für die Prüfungsnachbereitung (s. 7.1).

5.1. Statistische Analyse

Für alle Prüfungsformate ist eine adäquate statistische Analyse der Prüfungsergebnisse durchzuführen, die insbesondere Aufgabenschwierigkeit und -trennschärfe umfasst (Primärauswertung).

Bei Prüfungsformaten, in denen neben den Aufgaben weitere systematische Einflussfaktoren wie etwa Prüfereinflüsse existieren (z. B. OSCE), sind diese bei der Auswertung zu berücksichtigen (z. B. mit Verfahren der Generalisierbarkeitstheorie). Für Multiple-Choice-Aufgaben ist zusätzlich eine Distraktorenanalyse durchzuführen.

Ergeben sich Hinweise auf fehlerhafte oder unklare Aufgabenstellungen, so ist die Aufgabe formal und inhaltlich nachzukontrollieren.

5.2. Korrekturen der Auswertung

Nach einer evtl. notwendigen Korrektur der Aufgaben- oder Prüfungsbewertung wird eine Endauswertung der Prüfung (einschl. einer weiteren teststatistischen Analyse) durchgeführt.

Die nochmalige Überprüfung der Aufgabenstellungen nach der Prüfungsdurchführung dient der juristischen Absicherung/Rekursfestigkeit der Leistungsbewertungen. Für die Auswertung von Prüfungen wird deshalb ein zweistufiges Vorgehen empfohlen: Im ersten Schritt wird eine teststatistische Auswertung der Prüfung vorgenommen, nach der kontrolliert wird, ob einzelne Aufgaben hinsichtlich Schwierigkeit oder Trennschärfe „auffällig“ sind. Hier sind erfahrungsgemäß insbesondere sehr schwere Aufgaben (Schwierigkeiten unter 0,4) oder Aufgaben mit sehr niedriger Trennschärfe (unter 0,2) hinsichtlich ihrer inhaltlichen Korrektheit von den Prüfungsverantwortlichen zu überprüfen.

Erweisen sich dabei Aufgaben als fehlerhaft, ist eine Neuauswertung der Prüfung erforderlich. Erst im Anschluss an diese Auswertung sollten die Prüfungsergebnisse bekannt gegeben werden. Eine nochmalige Auswertung ist notwendig, wenn z. B. auf Grund begründeter studentischer Einwände weitere Korrekturen an der Aufgabenbewertung vorgenommen werden müssen (s. u).

Auch bei automatischer Auswertung wie z. B. bei computerbasierten Prüfungen ist darauf zu achten, dass alle Maßnahmen zur Qualitätssicherung vor der Bekanntgabe der Ergebnisse durchlaufen wurden. Der Prüfungsverantwortliche muss die Prüfungsergebnisse formal freigeben.

Bei der Korrektur fehlerhaft gestellten Aufgaben ist sicherzustellen, dass den Prüfungsteilnehmern hierdurch keine Nachteile entstehen. So dürfen z. B. Multiple-Choice-Aufgaben des Typs A („Eins aus Fünf“) nicht einfach aus der Wertung genommen werden, wenn mehr als eine der Antwortoptionen als zutreffend anerkannt werden muss. Stattdessen muss allen Teilnehmern, die eine der zutreffenden Antworten gegeben haben, diese Antwort anerkannt werden (man vergleiche hierzu auch die Regelungen bei den schriftlichen Staatsexamina des Instituts für medizinische und pharmazeutische Prüfungsfragen IMPP).

Bei begründeten Einsprüchen gegen Prüfungsaufgaben oder ihrer Bewertung sollen die notwendigen Korrekturen bei allen Prüfungsabsolventen durchgeführt (d. h. nicht nur bei den Beschwerdeführern) und bekannt gegeben werden. Es ist darauf zu achten, dass berechtigte Einsprüche und die daraus resultierenden Korrekturen (Verantwortlichkeiten bei Entscheidungen) dokumentiert werden.

Werden Prüfungsaufgaben als fehlerhaft erkannt, so ist ein verbindliches Vorgehen notwendig, bei dem gewährleistet ist, dass durch fehlerhafte Aufgabenstellungen Studierende nicht benachteiligt werden. Ist eine Aufgabe nicht lösbar, so kann

1. die Aufgabe aus der Wertung genommen und die maximal erreichbare Punktzahl entsprechend reduziert oder
2. die bei dieser Aufgabe ursprünglich vorgesehene erreichbare Punktzahl allen Studierenden zugebilligt werden (hier bleibt die maximal erreichbare Punktzahl unverändert).

Bei Korrekturen der Antwortmöglichkeiten nach Bekanntgabe des Ergebnisses dürfen Prüfungsbewertungen der Studierenden nicht nachträglich verschlechtert werden.

5.3. Prüfungsbericht

Es wird ein Prüfungsbericht erstellt, der die Angaben zur Bewertung und Benotung sowie die statistische Analyse der Ergebnisse umfasst. Insbesondere sind darin Veränderungen der Aufgabenbewertungen oder -gewichtungen, der als korrekt gewerteten Lösungen und nicht gewertete Aufgaben unter Angabe der für die Änderungen Verantwortlichen zu dokumentieren.

5.4. Kontrollstichproben

Es erfolgt eine stichprobenartige Kontrolle der Korrekturen und Bewertungen.

Neben einer stichprobenartigen Kontrolle der Korrekturen und Bewertungen ist eine Kontrolle der Prüfungsleistungen aller durchgefallenen Studierenden zu empfehlen. Kontrollen schriftlicher Prüfungen müssen durch unabhängige Korrektoren vorgenommen werden. Werden Klausuren mit Hilfe von Beleglesern eingelesen, so sind ebenfalls stichprobenartige Überprüfungen notwendig. Art und Umfang von Kontrollen sollten dokumentiert werden.

5.5. Dokumentation der Ergebnisse, Aufbewahrungsrichtlinien

Die Prüfungsergebnisse und Notenspiegel werden durch die Fächer oder zentral zusammengestellt und zur gesicherten Dokumentation zentral gespeichert.

Aufbewahrungsfristen für Prüfungen und Prüfungsunterlagen sind verbindlich (z. B. in der Prüfungsordnung) festzulegen. Es gibt keine einheitlichen Vorgaben für die Aufbewahrungszeiten – es gelten die entsprechenden Bestimmungen vor Ort (z. B. Landesarchivierungsordnung). Bitte informieren Sie sich bei Ihrer Rechtsabteilung. Als Anhaltspunkt können folgende Regelungen gelten: Schriftliche Prüfungen und mündliche Prüfungsprotokolle sind nach abgeschlossener Prüfung mindestens 18 Monate aufzubewahren. Bei computerbasierten Prüfungen sind die Einzelergebnisse in Form von Prüfungsprotokollen 18 Monate abzuspeichern. Die Listen über Prüfungsteilnehmer und Leistungsnachweise sind mindestens 10 Jahre in Papierform oder digital zentral aufzubewahren. Bei Einsprüchen gegen die Prüfung dürfen bis zur endgültigen Entscheidung keine Unterlagen vernichtet werden.

6. Rückmeldung an die Studierenden

Rückmeldungen an die Studierenden über ihre Prüfungsleistungen sind transparent und zeitnah zu geben. Nur so können Prüfungen als Instrument zur Lernsteuerung effektiv eingesetzt werden.

6.1. Bekanntgabe der Ergebnisse

Eine datenschutzkonforme Bekanntgabe der Prüfungsergebnisse erfolgt innerhalb eines angemessenen und vorab festgelegten Zeitraums. Dieser Zeitraum sollte 3 Wochen nicht übersteigen.

Bei der Bekanntgabe von Noten sind die datenschutzrechtlichen Bestimmungen einzuhalten. Insbesondere ist etwa ein öffentlicher Aushang der Prüfungsergebnisse mit Nennung persönlicher Daten unzulässig.

6.2. Prüfungseinsicht

Die Studierenden haben innerhalb einer angemessenen Frist die Möglichkeit zur Einsicht in ihre Prüfungsunterlagen. Die entsprechenden gesetzlichen Vorgaben sind dabei zu berücksichtigen.

Den Studierenden muss auf Nachfrage oder Antrag Einsicht in ihre eigene Prüfungsarbeit gewährt werden. Dabei ist eine angemessene Zeit zur Einsichtnahme nach Bekanntgabe des Ergebnisses zu gewährleisten. Die Möglichkeit der Einsichtnahme in die Klausur sollte den gesamten Zeitraum der Einspruchsfrist umfassen. Das Terminangebot zur Einsichtnahme muss angemessen sein. Für die Zeit der Einspruchsfrist sollte das Terminangebot öffentlich bekannt gemacht sein. Die Institution kann feste Zeiten für eine solche Einsichtnahme festsetzen. Diese müssen mit den Ankündigungen zur Prüfung veröffentlicht werden. Sollte es Studierenden aus begründetem Anlass nicht möglich sein, während dieses Termins Einsicht zu nehmen, ist die Einsichtnahme anderweitig zu ermöglichen. Die Einsichtnahme sollte unter Aufsicht erfolgen, weshalb eine Terminsetzung zur Vorbereitung einer parallelen Einsichtnahme mehrerer Teilnehmer sinnvoll ist.

6.3. Einspruchsfrist

Die Frist zum Einspruch gegen Prüfungsergebnisse muss wenigstens einen Monat ab der Bekanntgabe der Prüfungsergebnisse umfassen. Innerhalb dieses Monats sollte auch die Einsichtnahme möglich sein. Hierüber hat eine individuelle Rechtsbehelfsbelehrung zu erfolgen, die mit dem Ergebnis dem Prüfungsteilnehmer schriftlich zugestellt wird.

Diese sollte folgenden Inhalt haben:

> Sie haben an der Prüfung XY am XY teilgenommen und bestanden/nicht bestanden mit der Note XY. Rechtsbehelfsbelehrung: Gegen diesen Bescheid können Sie innerhalb von einem Monat Widerspruch beim Lehrverantwortlichen (Studiendekanat) einlegen.

Es empfehlt sich eine (automatisierte) Benachrichtigung über das Nichtbestehen mit Rechtsbehelfsbelehrung durchzuführen. Ohne Rechtsbehelfsbelehrung gilt in Deutschland eine Einspruchsfrist von einem Jahr.
Die Ablehnung eines Widerspruchs gegen die Bewertung von Aufgaben oder der Prüfungsdurchführung bedarf ebenfalls einer Rechtsbehelfsbelehrung.

6.4. Art und Umfang der Rückmeldung

Art und Umfang der Rückmeldung der Prüfungsergebnisse an die Studierenden mit dem Ziel, den Studierenden detaillierte Information zu ihrem Leistungsstand zu geben, sind festgelegt (z. B. Aufgliederung des Gesamtergebnisses nach Teilfächern o. ä.). Eine längsschnittliche Rückmeldung, die den Studierenden Informationen über ihren Leistungsstand

1. *in Bezug auf die an sie gestellten Erwartungen,*
2. *auf die anderen Prüfungsteilnehmer sowie*
3. *ihrer individuellen Leistungsentwicklung gibt,*

ist anzustreben.
Aufgrund gesetzlicher Bestimmungen ist die Anzahl an durchzuführenden summativen Prüfungen sehr hoch, was den Einsatz zusätzlicher formativer Prüfungen häufig erschwert. Es sollte daher durch die Betrachtung der einzelnen summativen Prüfungen einer Studierenden im Längsschnitt das formative Potential summativer Prüfungsleistungen genutzt werden.

6.5. Veröffentlichung von Aufgaben

Eine Veröffentlichung der Prüfungsaufgaben wird - zumindest solange kein hinreichend großer Aufgabenpool zur Verfügung steht - nicht empfohlen. Eine einheitliche Regelung und Empfehlungen diesbezüglich (z. B. Notwendigkeit eines vollständigen Rückflusses der Aufgabenblätter) sollte innerhalb einer Fakultät/eines Studiengangs angestrebt und den Studenten bekannt gemacht werden.

7. Prüfungsnachbereitung

Die Nachbereitung der Prüfung dient zunächst der Qualitätssicherung des Prüfungsgeschehens in einem Fach, indem Mängel bei Aufgaben aufgedeckt und korrigiert werden können. Weiterhin ist sie ein wichtiges Rückmeldeinstrument an die Lehrverantwortlichen, da Prüfungen darüber Auskunft geben, was die Studierenden tatsächlich gelernt haben und ob und inwieweit Änderungen im Curriculum (z. B. veränderte Schwerpunktsetzungen bei den Lehrveranstaltungen) sinnvoll oder erforderlich sind.

7.1. Nachbegutachtung der Prüfung (Post-Review)

Zur Qualitätssicherung und -verbesserung künftiger Prüfungen findet eine schriftlich dokumentierte Nachbewertung (Post-Review) der Prüfung statt, an der die Prüfungsbeauftragten teilnehmen. Anhand inhaltlicher Kriterien, teststatistischer Auswertungsergebnisse (z. B. Schwierigkeiten, Trennschärfen, Reliabilität) sowie studentischer Kommentare und Hinweise werden in dieser Nachbewertung Verbesserungsvorschläge für Prüfungsaufgaben und Prüfungszusammenstellung erarbeitet.

7.2. Rückmeldung an Autoren und Fachverantwortliche

Die Prüfungsergebnisse, deren Auswertung sowie die Ergebnisse des Post-Review-Prozesses sollen zeitnah mindestens einmal im Semester an die Fragenautoren, die Curriculumsentwickler und Fachvertreter weitergegeben werden Adäquate Konsequenzen und erforderliche Maßnahmen sollten ergriffen und dokumentiert werden.

Anmerkung

[1] Zur besseren Lesbarkeit wurde im Text z. T. auf die Nennung der weiblichen Form verzichtet, beide Geschlechter sind immer in gleichberechtigter Weise gemeint.
[2] Diese Empfehlungen haben keine rechtlich bindende oder präjudizierende Wirkung. Es gelten jeweils die entsprechenden Regelungen der für die Prüfung verantwortlichen Institutionen bzw. die Gesetzeslage
[3] Eine Gleitklausel ist eine formale Vorschrift, bei der in Abhängigkeit von den Prüfungsergebnissen der Teilnehmer bei niedrigen Gesamtergebnissen eine Korrektur der Bestehensgrenze nach unten vorgenommen wird. Hierdurch werden Prüfungen mit exorbitant hohen Durchfallquoten verhindert.
[4] Nachgewiesenermaßen besteht bei einer mündlichen Prüfung im Vergleich zur schriftlichen Prüfung eine Tendenz zur besseren Bewertung. Beispiel: Kandidat A besteht die Prüfung nicht und erhält die Möglichkeit einer mündlichen Nachprüfung. Mit hoher Wahrscheinlichkeit erhält er mindestens die Note 3. Kandidat B besteht die schriftliche Prüfung mit einer 4. Er erhält deshalb keine Möglichkeit zu einer Nachprüfung und behält im Endzeugnis die Note 4. Kandidat B hat keine Möglichkeit, die Note zu verbessern und ist gegenüber Kandidat A benachteiligt.
[5] Kritisch ist hierbei, dass aus Tests generelle Rückschlüsse über Individuen oder Gruppen gezogen werden, die auf einer sehr limitierten Anzahl von Stichproben basieren. Nur bei hoher Validität ist die Generalisierung anhand der Testergebnisse auf andere Situationen zulässig. In der klassischen Testtheorie sind Objektivität und Reliabilität Voraussetzung für eine hohe Validität.

Danksagung

Unser besonderer Dank gilt allen Mitgliedern des Ausschusses Prüfungen der GMA und der AG Prüfungen des MFT, die an der Erstellung der Empfehlungen, ihrer Diskussion und ihrer Evaluation mitgewirkt haben. Besonders hervorzuheben sind dabei

- Prof. Dr. Klaus Albegger, Graz
- Dr. Daniel Bauer, München
- Dipl. Med.-Inf. Konstantin Brass, Heidelberg
- Peter Brüstle (M.A.), Freiburg
- Dr. Corinne M. Dölling, Berlin

- PD Dr. Roman Duelli, Heidelberg
- Dr. Jan Ehlers, Hannover
- Prof. Dr. Martin Fischer (MME), München
- Dr. Volkhard Fischer, Hannover
- Prof. Dr. Johannes Forster (MME), Freiburg
- Andreas Fuhrig, Göttingen
- Prof. Dr. Matthäus Grasl (MME), Wien
- Dr. Achim Hochlehnert, Heidelberg
- Dipl.-Ing. Matthias Holzer, München
- Dr. Hans Haage, Rheinbach
- Dr. Sören Huwendiek (MME), Bern
- Prof. Dr. Jana Jünger (MME), Heidelberg
- Prof. Dr. Ingo Just, Hannover
- Dr. Roland Kabuß (MME), Hannover
- Prof. Dr. Klaus J. Klose, Marburg
- Dr. Richard Melamed, Frankfurt
- Dr. Andreas Möltner, Heidelberg
- Dr. Daniela Mohr, Tübingen
- Elisabeth Narziß (Ärztin), Mannheim
- Dr. Zineb Nouns, Berlin
- Prof. Dr. Franz Resch, Heidelberg
- Dr. Thomas Rotthoff (MME), Düsseldorf
- PD Dr. Jobst-Hendrik Schultz, Heidelberg
- Dr. Katrin Schüttpelz-Brauns, Mannheim
- Dipl.-Päd Tina Stibane, Marburg
- Markus Stieg (Arzt), Berlin
- Dipl. Wi.-Rom. Anna Vander Beken, Ulm

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Anhänge

Verfügbar unter
http://www.egms.de/en/journals/zma/2014-31/zma000926.shtml

1. Anhang.pdf (83 KB)
   Checkliste

Literatur

1. Gesellschaft für Medizinische Ausbildung; Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für fakultätsinterne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. GMS Z Med Ausbild. 2008;25(1):Doc74. Zugänglich unter/available from: http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml
2. WHO; WFME. Guidelines for Accreditation of Basic Medical Education. Geneva, Copenhagen: WHO; 2005.
3. WFME; AMSE. WFME Global Standards for Quality Improvement in Medical Education European Specifications. Copenhagen: University of Copenhagen, MEDINE Quality Assurance Task Force; 2007.
4. National Board of Medical Examiners. Test Administration Handbook. Philadelphia: National Board of Medical Examiners; 2003.
5. Jünger J, Möltner A, Lammerding-Köppel M, Rau T, Obertacke U, Biller S, Narciß E. Durchführung der universitären Prüfungen im klinischen Abschnitt des Medizinstudiums nach den Leitlinien des GMA-Ausschusses Prüfungen: Eine Bestandsaufnahme der medizinischen Fakultäten in Baden-Württemberg. GMS Z Med Ausbild. 2010;27(4):Doc57. DOI: 10.3205/zma000694
6. Reindl M, Holzer M, Fischer MR. Durchführung der Prüfungen nach den Leitlinien des GMA-Ausschusses Prüfungen: Eine Bestandsaufnahme aus Bayern. GMS Z Med Ausbild. 2010;27(4):Doc56. DOI: 10.3205/zma000693
7. Möltner A, Duelli R, Resch F, Schultz JH, Jünger J. Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten. GMS Z Med Ausbild. 2010;27(3):Doc44. DOI: 10.3205/zma000681
8. Schuwirth LW, van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
9. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. Med Teach. 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
10. Biggs JB. Enhancing teaching through constructive alignment. High Educ. 1996;32:347-364. DOI: 10.1007/BF00138871
11. Loftus S, Gerzina T. Being a Health Professional Educator. In Loftus S, Gerzina T, Higgs J, Smith M, Duffy E (Hrsg). Educating Health Professionals: Becoming a University Teacher. Rotterdam: SensePublishers; 2013. S.3-14. DOI: 10.1007/978-94-6209-353-9_1
12. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrot V, Roberts T. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-214. DOI: 10.3109/0142159X.2011.551559
13. Hahn EG, Fischer Mr. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). GMS Z Med Ausbild. 2009;26(3):Doc35. DOI: 10.3205/zma000627
14. Tremp R, Reusser K. Leistungsbeurteilung und Leistungsnachweise in Hochschule und Lehrerbildung - Trends und Diskussionsfelder. Beitr Lehrerbild. 2007;25(1):5-13.
15. Nürnberger F. Handreichung Regelungen zu Prüfungen an Medizinischen Fakultäten und Fachbereichen. Berlin: Medizinischer Fakultätentag; 2010.
16. Schuwirth LW, van der Vleuten CP. Assessing competence: Extending the appoaches to reliability. In: Hodges BD, Ligar L (Hrsg). The question of competence: reconsidering medical education in the twenty-fisrt century. Ithaca, New York: Cornell University Press; 2012. S.113-130.
17. Jackson N, Jamieson A, Khan A. Assessment in medical education and training: a practical guide. Oxford, New York: Radcliffe; 2007.
18. Lucke J. The alpha and the omega of Congeneric Test Theory: An Extension of Reliability and Internal Consistency to Heterogeneous Tests. Appl Psychol Measure. 2005;29(1):65-81. DOI: 10.1177/0146621604270882

Korrespondenzadressen:
Prof. Dr. med Jana Jünger, MME
Universitätsklinikum Heidelberg, Kompetenzzentrum für Prüfungen in der Medizin/Baden-Württemberg, Im Neuenheimer Feld 410, 69120 Heidelberg, Deutschland, Tel.: +49 (0)6221/56-8657, Fax: +49 (0)6221/56-1341
jana.juenger@med.uni-heidelberg.de
Prof. Dr. med. Ingo Just
Medizinische Hochschule Hannover, Studiendekan für Medizin und Bachelor/Masterstudiengänge, Carl-Neuberg-Straße 1, 30625 Hannover, Deutschland, Tel.: +49 (0)511/532-9014, Fax: +49 (0)511/532-2879
studiendekanat.just@mh-hannover.de

Bitte zitieren als
Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. GMS Z Med Ausbild. 2014;31(3):Doc34.
DOI: 10.3205/zma000926, URN: urn:nbn:de:0183-zma0009261

Artikel online frei zugänglich unter
http://www.egms.de/en/journals/zma/2014-31/zma000926.shtml

Eingereicht: *09.05.2014*
Überarbeitet: *01.07.2014*
Angenommen: *02.07.2014*
Veröffentlicht: *15.08.2014*

Recommendations of the German Society for Medical Education and the German Association of Medical Faculties regarding university-specific assessments during the study of human, dental and veterinary medicine

Abstract

The practice of assessing student performance in human, dental and veterinary medicine at universities in German-speaking countries has undergone significant changes in the past decade. Turning the focus to practical requirements regarding medical practice during undergraduate study away from an often theory-dominated curriculum, the academic scrutiny of the basics of teaching medical knowledge and skills, and amendments to legislation, all require ongoing adjustments to curricula and the ways in which assessments are done during undergraduate medical education. To establish quality standards, the *Gesellschaft für medizinische Ausbildung (GMA German Society for Medical Education) reached* a consensus in 2008 on recommendations for administering medical school-specific exams which have now been updated and approved by the GMA assessments committee, together with the *Medizinischer Fakultätentag (MFT German Association of Medical Faculties)*, as recommendations for the administration of high-quality assessments.

Keywords: Recommendations, assessment

Jana Jünger[1]
Ingo Just[2]

1 Representative of the GMA assessments commitee, Heidelberg, Germany

2 Representative of the working group on assessments, MFT, Hannover, Germany

Introduction

These recommendations for university-specific assessments are aimed toward all those who are employed[1] in human, dental, and veterinary medicine at universities in Germany, Austria and Switzerland, who are entrusted with the design, conduction and evaluation of school-specific exams, meaning teachers and lecturers, deans of studies, and also curricular designers and teaching coordinators due to the close interconnection between teaching and testing. These recommendations cover the quality standards requisite for objective, reliable, valid and, in turn, justifiable testing. Written in the form of a checklist, these recommendations are to serve as a practical tool for structuring and organizing exams.

Background

In 2008, the GMA assessments committee, along with the Baden Württemberg Center of Excellence for Assessment in Medicine, jointly presented the *Leitlinien für fakultätsinterne Leistungsnachweise in der Medizin (Guidelines for assessment in medical faculties)* [1]. This was to help establish agreed quality standards for exams required of medical schools in Germany by the 2002 amended version of the medical licensure act so that the internationally recognized standards for high quality methods of assessing performance are met(e.g. [2], [3], [4]). Its significance is evident in various publications that have appeared on testing formats and the quality of university assessments in response to the context behind these recommendations [5], [6], [7].

The basic importance of feedback and performance assessment along with their ability to guide learning in medical education and the resulting necessity of systematically including testing in the curriculum (constructive alignment, programmatic assessment [8], [9], [10], [11], [12]) is commonly known; however, their implementation in practice is still deficient in many cases. This applies in particular to curricular content that goes beyond the traditional and prevailing teaching of medical expertise, as it does in the CanMEDS role model, for instance. Based on this role model, the competences and skills required in medical education are defined in the Swiss Catalogue of Learning Objectives and the National Competency-based Catalogue of Learning Objectives for Undergraduate Medical Education (NKLM) [13] currently being drafted in Germany.

These developments in the requirements placed on medical education must also be reflected in the procedures for assessing performance; new testing formats and methods for evaluating the necessary skills and competences for practicing medicine must be developed and applied. In practice this means that exams during medical study will more frequently display a combination of differ-

ent testing formats, that formative tests will take on a greater presence than summative tests [12], and higher value will be placed on criteria-oriented evaluations. This revision of the recommendations from 2008 addresses these issues. In particular, it must be ensured that the same quality requirements regarding measurement reliability and validity that are placed on traditional assessment methods are also demanded of innovative testing formats.

The focus of these recommendations continues to be those assessments that must be passed at a medical school in order to receive graded credit (*Leistungsnachweis*). Such summative or accumulative evaluations aim to reflect a final determination of skill level [14]. The formal – in particular statutory – requirements placed on purely formative tests are generally much fewer; for the quality of question content, however, the same requirements are in effect as for summative tests.

The authors of these recommendations are aware that a complete change of approach confronts medical schools with substantial organizational and personnel problems which can only be dealt with over the medium or long term. Despite this, examples at medical schools demonstrate that all the points covered by these recommendations can be fulfilled. The schools are therefore called upon to improve the quality of their assessments and evaluations in an ongoing process. To provide support for this, the GMA assessments committee intends to publish practical approaches as examples for implementing these recommendations.

Revision of the Recommendations

In response to the developments mentioned above, a revision of the recommendations issued in 2008 was decided upon in 2012 by the GMA assessments committee. As part of the International Conference on Competency-based Assessment in Heidelberg on July 4, 2012, the first proposals for improvement (see [15]) were drafted. In another meeting on September 27, 2012 at the annual GMA conference in Aachen, subject areas 1-4 (general structural pre-requisites, exam design and evaluation, organizational preparation for conducting exams, administering exams) were jointly discussed in depth and compiled in cooperation with the MFT working group on assessments. A further round of discussion and focus on subject areas 5-7 (evaluation and documentation, feedback for students, post-processing) took place at the committee meeting with the MFT working group during the GMA conference in September 26, 2013 in Graz. After inclusion of the agreed changes, the revised version was supplemented further by written consent. In January 2014 an external legal review[2] of this version was undertaken by a Hanover law firm specializing in scholastic examination law. The resulting changes were included at the beginning of February 2014 and discussed on February 11, 2014 at a meeting of the GMA assessments committee. Any remaining open points were clarified at this meeting and included. The recommendations were presented to the MFT working group on teaching/curriculum and the GMA executive board in May 2014 and approved. Both the MFT and GMA support these recommendations, which have been given the character of a guideline.

Explanation of the new version of the recommendations

The first version of the recommendations [1] consisted of the individual points articulated in the form of a checklist with corresponding numbered explanations. To ease readability, the individual points of the recommendations and their explanations are formulated here as running text; an additional checklist is included in the appendix. The individual criteria from the appended checklist appear in cursive in the following text (see Attachment 1).

1. General structural pre-requisites: requirements regarding form and content

The structural pre-requisites cover criteria that should guarantee curricular inclusion of the courses and lectures upon which the exams draw, formal requirements for notifying students, as well as rules and regulations and training those responsible for the exams. They do not refer to the preparation or administration of a concrete assessment, but rather apply to the basic conditions that are needed for high-quality testing.

1.1. Comprehensive assessment program

A comprehensive assessment program, in which the number, scope, content, timeframe and format of the individual summative and formative tests to be taken during undergraduate medical study are coordinated with each other, is available to all students and teachers.

The types of assessments given at the medical school or as part of the degree programs in human, dental or veterinary medicine, along with their administration and evaluation should be listed and laid down in the relevant formal rules and regulations (*Studienordnung, Prüfungsordnung*, or in appropriate rules for implementation). Attention is to be paid that the provisions allow sufficient room for the establishment of innovative forms of assessment.

It should be noted that the exam content is tested with suitable types of assessment that not only reflect methods for assessing theoretical knowledge, but also practical skills; (Triangulation: assessment on the basis of different sources at different points in time, under different conditions, through different people and with different methods [16], [17]). For example, theoretical knowledge can be appropriately measured through written tests, practical

content with objectively structured practical/clinical exams (OSPE/OSCE). The types of assessments should fulfill the particular quality requirements for objectivity, reliability and validity. If the score is based on different exam components, the requirement regarding measurement reliability will refer to the entire exam and not only to the individual components (see explanation under 2.6).
To prepare students as well as possible for their future medical profession, learning objectives are to be included in the curricula that go substantially beyond medical expertise and technical skills. To accomplish this, it is also necessary to develop suitable types of exams and constructs that allow for appropriate, reliable and feasible assessment of these competences. This requires the use of new testing formats, in particular workplace-based exams, such as DOPS, encounter cards or 360° assessment, to assess communication skills, professional decision-making, management skills, etc. Special attention to the quality assurance of these forms of testing is required. It is important to ensure sufficient training of the examiners in advance. The use of appropriate methods of analysis (e.g. generalizability theory) must be provided for when analyzing assessment results as a control of the lower standardizability of an exam situation, as exists for workplace-based performance assessments.
Logistically, the measurement of nonsubject-specific learning objectives is often impossible within the scope of the individual subject exams. In this case, other test constructs are conceivable, in which components of other separate assessments are compiled in an interdisciplinary manner similar to a portfolio and assessed. By doing this, the communication stations in OSCE's for different subjects could be combined together for an assessment of the student's skills as communicator. This portfolio could also cover the documentation of critical events (for instance for the assessment of professional conduct).

1.2. Catalogue of learning objectives

For each curricular unit defined in the Studienordnung (e.g. subject, module, course, seminar, interdisciplinary field) in the pre-clinical and clinical phases of study there is a comprehensive written catalogue of learning objectives.
Which learning objectives are to be imparted in which courses must be evident in the learning objective catalogue as a whole, if such a catalogue exists.

1.3. Informing students about the learning objectives catalogue

The students are informed of the specific learning and assessment objectives in a timely manner prior to each curricular unit/module.

1.4. Suitable assessment formats

The knowledge, skills and attitudes defined in the learning objectives are assessed by means of suitable testing formats. In particular, procedures are to be used which are suitable for assessing skills in making medical decisions and taking medical action, as well as skills in conducting medical consultations (see 1.1).
In addition to written forms of assessment (multiple-choice or open-ended questions) which primarily serve to test theoretical knowledge, the OSCE is the type of exam established to assess practical skills taught and acquired in medical degree programs. To measure other competence areas concerning medical practice, still further testing formats are needed that make reliable, workplace-based performance assessments possible. Belonging to these types of exams are, for example, miniCEX, 360° assessments, encounter cards, and direct observation of practical skills (DOPS).

1.5. Written rules for exam preparation and assessment procedures

Written rules should exist for the following aspects and details.

1. *Pre-requisites for participation*
2. *Scheduling exam dates (including repeat sessions) and formal assessment procedures.*
   For each exam, clear rules and regulations should be followed as standard practice for the formal assessment procedure. These rules and regulations should be recorded in writing and address the following aspects:
   – Requirements regarding how and when an exam is announced
   – Requirements regarding how and when to register students for the exam. If applicable, automatic registration for the exam occurs through assignment to a module.
   – Number of examiners and their qualifications (e.g. specialist physician, post-doctorate, etc.)
   – Duration of the exam
   – Introductory sessions about the exam (e.g. individual appointments for instructions on taking computer-based assessments)
   – Announcements at the start of the exam
   – Study aids allowed during the exam
   – Rules about students keeping copies of exams afterwards
   – How to handle tardy appearances to an exam session
   – Withdrawal from or failure to attend an exam session
   – How to handle attempts at cheating
   – Rules about quitting in the middle of an exam
3. *Rules regarding the types of assessments that can be used in the degree program (see 1.1)*
4. *Definition of the pre-requisites for space and time and the conditions for conducting the assessment (see 3.3)*
5. *Rating scales, passing scores, application of a grading curve or an automatic adjustment clause*[3] *(see 2.5, 2.8)*
6. *Evaluation in the case of errors in the questions asked (see 5.2)*

7. *Weighting of component exams (see 3.1)*
8. *Compensation options and disability compensation during exams (see 1.6)*
9. *Conditions for participation and procedures for repeat and re-testing (see 1.6)*
10. *Announcement and inspection of exam results (see 6.2)*
11. *Rules regarding appeals against scores and test questions (see 5.2, 6.3)*
12. *Responses to violations of the conditions for conducting exams and extraordinary disruptions of test administration, as well as rules for any repeat testing necessary as a result (see 4.3)*
13. *Publication of questions (see 6.5)*
14. *Documentation of the assessment and its results (see 5.5)*

1.6. Compensating exam performance, retesting and repeat testing

1. *If it is impossible for students to attain graded credit or components of graded credit, or possible only under unreasonable circumstances that arise from the nature or form of exam administration or conduction, then it should be fully clarified under which conditions test performance can be compensated.*
   This applies to students with physical disabilities for whom, in certain cases, the advocate for disabled students should be involved or to students with limited German language skills who are not enrolled in the degree program as conventional students (students participating in international student exchange programs, e.g. Erasmus).
2. *The conditions for administering and sitting for repeat and re-testing are to be set down in the authoritative legal provisions (Studienordnung, Prüfungsordnung). Likewise, it must be determined if and to what extent assessments leading to grade improvement will be given.*
   The testing format for repeated and re-testing sessions should match the format of the initial assessment; for instance, no written or oral re-testing should be conducted for a failed OSCE. Likewise, in the case a written test is failed, no oral re-assessment should take place[4].
   For separate repeat assessments (meaning assessments in which mostly candidates who have failed the test at least once are tested), a modification of the automatic adjustment clause is recommended in certain cases (see also 2.5).

1.7. Persons responsible for assessments

1. *In each subject, at least one person and their deputy shall be appointed as responsible for the exam and the related tasks shall be clearly defined. (Scope of responsibility: e.g. blueprint, question generation, conduction, grading, pre- and post-review, analysis, feedback for curriculum developers).*
2. *The responsible persons must take part in professional training on the topic of assessments.*
   Each person responsible for the assessment in regard to a specific curricular area (subject, module, block, etc.) should be able to demonstrate certified training on the topic of assessments and testing.

2. Assessment design and analysis

The following recommendations refer to the preparation of concrete exams. They affect the curricular integration of test content and measures to ensure the quality of questions and overall assessment (reliability and validity), as well as test administration that is economically feasible and transparent for students.

2.1. Coordination of exams with the comprehensive assessment program

The individual exams are to be coordinated with the medical school's comprehensive assessment program. This coordination affects not only summative, but also formative performance feedback.

2.2. Validity

Each individual exam is based on a written blueprint that representatively maps out the subject-specific exam content.
The blueprint serves to ensure the validity of the assessment's content. This guarantees

1. that the questions represent the subject area being tested and
2. avoids the presence of any content irrelevant to this assessment (construct-irrelevant variance).

Validity is the criterion for test quality. It is a measure of whether or not the data gained through the measurement represent, as intended, the quantity to be measured, meaning the knowledge or skills in the subject area to be covered by the assessment: Does the test measure what it is supposed to measure[5]?
After analyzing the assessments, further sources of validity can be investigated:

- Are the exam scores plausible?
- Is there a high correlation between this exam and other exams that are meant to measure the same construct (e.g. correlation between a multiple-choice test on internal medicine and the sections on internal medicine contained in the state medical examinations)?

2.3. Inclusion of subject area representatives

Representatives from all the affected subject areas are involved in putting the exams together.

2.4. Pre-review of the test questions and analyzing content validity

1. *Prior to administering an exam, a standardized analysis is carried out regarding the content and form of the test questions (pre-review).*

In respect to testing formats, for which only limited standardization options exist (e.g. workplace-based exams), it must be determined how different conditions and degrees of difficulty are to be taken into account (e.g. detailed standard setting).
When creating exams, the following aspects should be considered overall in regard to validity:
– Is each question of a high quality? It is especially important that only the skill/ability being tested (e.g. knowledge of a specific subject area) is necessary to arrive at the correct answer and not other skills (e.g. language skills).
– Is the content generally valid/evidence-based and does not, for instance, represent local doctrine?
– Does the exam's content correspond with the curriculum/learning objectives?
– Does the content involve knowledge that can be expected in terms of the current level of education and does not involve content that, for example, belongs to a later phase of study or advanced medical training?
– Is the content of the material to be tested represented appropriately and extensively with its sub-areas? To ensure this, suitable methods for compiling questions must be selected as standards (see 2.2 Blueprint).
– Have the test questions and the entire exam been subjected to a thorough review process?
– Is the theoretical framework based on sound and comprehensive reasoning?
– Does the exam appear credible to the candidates? (Acceptance)

2. *At least two representatives from the subject area and one from another discipline take part in the review.*
3. *The results of the review must be documented.*

2.5. Passing scores

1. *Prior to administering an exam, the lowest possible passing score will be set down in writing by an interdisciplinary board of experts and determined according to content-related criteria (e.g. by means of a standard setting procedure) or a formal criterion (e.g. 60% rule).*
   Passing scores should be determined to the extent possible using content-related criteria according to a criteria-oriented assessment scale (as an example, see standard-setting methods for OSCE's). For multiple-choice questions, formal criteria (e.g. the 60% rule) should be applied at the very least.
2. *A rule for applying an automatic adjustment clause is set down in writing.*
   A rule for automatic adjustments to the grading curve is generally necessary for exams with multiple-choice questions. In the degree program, a uniform rule should clearly state for which types of tests and in what manner automatic adjustment will be universally applied. It must be determined how exams of mixed formats are to be treated (e.g. multiple-choice and open-ended questions).
   In addition to a criteria-oriented passing score, appropriate rules to compensate for unreasonably difficult exams should also be made for other types of assessments and these must be communicated to the students in a timely manner.
   As a simplification for exams with multiple-choice questions, we recommend a modified automatic adjustment clause that takes into account the average grade of all candidates sitting the test for the first time directly following the course (without restricting this to traditional, full-time students, etc.). Appropriate rules need to be defined for re-testing and repeat tests where a substantial proportion of the participants are not taking the test for the first time.
3. *The procedure for rounding the lowest passing score and borderline point totals must be definitively set down in writing.*
   If the lowest passing score for a test with 99 questions and a minimum percent of 60% is 59.4 points, then rounding the passing score up to 60 points is recommended if only full points are given for the test questions. If half points are assigned, the passing score would then be set at 59.5 points (according to the German medical licensure act (ÄAppO) the minimum percentage to pass must be achieved or exceeded, meaning that in no case are scores to be rounded downward).

2.6. Assessment reliability

For summative tests, a reliability of at least 0.8 is to be expected for the achievement of graded credit (Leistungsnachweis).
We recommend that graded credits for a subject are based on multiple component exams to the extent that is methodically possible (see explanation under 1.1). Here, the criterion of a minimum reliability of 0.8 is to be applied to the overall assessment and not necessarily to the individual component exams. An example for this would be if in a subject a student must take a written exam on theoretical knowledge and undergo an OSCE of practical skills. It is possible that, for the exam and the OSCE, the reliability of each individual assessment is lower than 0.8, but the reliability of the two combined together can be distinctly higher. To determine reliability of combined graded credits, we refer to the relevant literature (e.g. [18]).
It must be noted that component exams which cannot be compensated for by other assessment scores must possess sufficient measurement reliability regarding the decision to pass or fail students, in order to avoid students being denied credit on their academic transcripts due to one deficiently reliable component exam. Examples of this are knock-out stations in an OSCE or components for interdisciplinary graded credits that must be passed separately.
So that an assessment fulfills the minimum reliability of 0.8 as an individual exam, as a general rule at least 40

high quality questions are necessary for a multiple-choice test and at least 12 stations for an OSCE. This information can only serve as an approximate reference. Depending on the test objective, quality of questions, and the student cohort being assessed, considerable fluctuations are possible which is why the statistical values of corresponding past exams on the subject should be drawn upon to estimate the expected reliability.

"Exotic" subjects, in particular, are confronted by the problem that a minimum reliability of 0.8 can only be attained with difficulty due to the scope of the exam. A solution offered in human medicine in Germany is the concept of interdisciplinary graded credits (fächerübergreifende Leistungsnachweise) allowing the combination of multiple subjects which are covered at about the same time in the curriculum into one exam. These are then represented by one overall score. If no interdisciplinary graded credits are possible, then an in-depth quality assurance program should ensure the highest possible validity resulting from the representativeness of the questions in terms of the curricular material and the avoidance of questions that test for knowledge or skills not included in the learning objectives (construct-irrelevant variance).

2.7. Use of resources

The scheduled exam is conceived in such a way that it conserves resources.

Under this heading, the possibilities for conserving resources in the development, administration and evaluation of exams are covered. Belonging to this are feeding the answer sheets into scanners, adequate numbers of test monitors, use of school/degree program test question pools, use of computer-based administration, standardized test-statistical analysis (e.g. centrally in the medical school), deployment of a minimum number of examiners (e.g. one per station for an OSCE is sufficient when monitored centrally), selection of resource-saving testing formats and question types (limit open-ended question to what is necessary).

2.8. Evaluating the answers

1. *The rating scale to be applied (grades, points) to assessments should be uniform and binding for the degree program.*
2. *The correct answers, the expectations, the grading guidelines, and mode of analysis must be determined in writing before the exam is administered.*
   The correct answers and expectations are available to the examiner in writing. The written instructions for grading an exam are clear (e.g. regarding the assignment of partial points or evaluating open-ended questions). Recommendation: the same examiner should rate all student responses to a particular open-ended question.
   The mode of assessment for an OSCE is clearly defined. For each OSCE station or question, the number of points assigned is clear based on a checklist or global rating of skill/ability. The same applies for oral exams.
3. *The number of points for each individual question/task is determined before the start of the exam.*
   For written exams with non-uniformly weighted questions, the number of possible points for each question must be indicated on the exam. It must be noted that for multiple-choice questions which are not of the single-choice type (e.g. more than one true/false answer) the demonstration of partial knowledge is to be taken into consideration.

2.9. Evaluation of component exams

1. *If the graded credits are composed of more than one component, the evaluations of the individual components should be done using a sufficiently differentiated rating scale.*
   Grading scales, such as the German system of applying a four-point grading scale to successfully passed exams, only roughly reflect actual test performance. If poorly nuanced grades from component exams are compiled to yield an overall grade, distortions in the assessment of the overall performance can arise as a result of any averaging.
2. *The procedure for rounding the grades must be clearly defined.*
   Rounding to whole numbers, as for the four-point grading scale required by the German medical licensure act (ÄAppO) to indicate proficiency levels on the officially recognized certificate, should always be in the direction of the nearest whole number. In the case of equal distance (decimal places 0.500), rounding should be to the advantage of the student, meaning that 1.500 is rounded to the grade of 1, while 1.501 is rounded to the grade of 2.
   It is recommended that partial evaluations used to compute an overall score are done on a scale with at least three decimal places. The use of three decimal places is sufficiently precise in normal cases to avoid distortion through repeated rounding (as occurs in the German system when 2.54 is rounded one decimal place to 2.5 and then, through repeated rounding, results in the better grade of 2).
   The rating scale should take a required equal distance between grade categories into account. For instance, if it is required of written exams that 60% to 70% of the possible points yields the grade of 4, and 70% to 80% the grade of 3, 80% to 90% the grade of 2, and 90% and above the grade of 1, then a grade-equivalent decimal scale of 0.5 to 4.5 must also suffice. As a result, the interval of 80-90% of the possible points (grade of 2) reflects a same-sized interval in the grades of 2.5 to 1.5, and 90-100%, the same-sized interval of 1.5 to 0.5. A simple linear conversion of point scores into decimal values is only possible in this manner.

3. Organizational preparation for conducting exams

Along with preparing exam content, various organizational and logistical preparations are called for to ensure a proper course of events during the assessment.

3.1. Announcing exam dates and formats

Exam dates and formats are announced to students at the beginning of a curricular unit.

If the overall grade for a subject or interdisciplinary area is the product of multiple component exams, then the weighted value of each exam is to be announced. In the case of changes in the make-up of the overall grade or in the weighting of the components, clear transitional rules must be drawn up for students who are required to repeat the exams.

3.2. Registering for exams

For each assessment, written or online registration is required of students. Registering for a course and an exam can be done at the same time. Under certain circumstances, it is possible that active registration is not required for exams which are mandatorily part of the curriculum.

It should be determined in advance if students who fail an assessment are automatically registered for the next possible repetition or if separate registration is required. Summative tests are to be viewed as the conclusion of a curricular unit and should refer to the curricular material just covered. Therefore, it is recommended that the assessment(s) or final component exam be mandatorily administered to all students shortly after completing the curricular unit.

3.3. Rooms and personnel for conducting exams

1. *To administer the exam it is ensured that sufficient rooms are available and that these pose comparable conditions and environments for all candidates.*
2. *Sufficiently trained personnel are available to administer the exam (examiners, monitors, graders for open-ended questions, etc.).*

3.4. Training and feedback for examiners

1. *Prior to administering the exam, the examiners and graders have received training regarding uniform grading criteria.*
   Joint training of all who evaluate the candidates should be conducted to increase inter-rater reliability. This is especially necessary for simultaneously conducted exams during an OSCE, oral exams, and written exams with open-ended responses.
   For assessments where the examiner comes in direct contact with the candidate, training sessions with video recordings of exams are particularly helpful.
2. *Examiners have received training regarding giving feedback to students and explaining the tested material and its evaluation. This applies in particular to all formative tests.*
   The training sessions need to be adjusted to meet the specific requirements of the testing format; in addition to more detailed initial training sessions, refresher courses must be conducted. The effectiveness of the training must be verified (e.g. through simulated students).
3. *Examiners receive feedback on their own performance giving the exam.*
   In the case of assessments where the influence of the examiner must be taken into consideration in the evaluation, feedback is to be given to the examiner (see 5.1). This means that for oral or oral practical exams, feedback is to be given regarding strictness or utilization of the rating scales. In certain cases, prior to the next assessment, examiners must undergo repeat training.

4. Conducting exams

4.1. Observance of formal criteria

When administering the exam, the formal criteria defined in writing are adhered to and documented (e.g. using a checklist for the formal assessment procedure).

4.2. Completeness of exams

The completeness of the exams and materials are double-checked by the students or the test monitors prior to starting the exam.

A clear assignment of both the question and answer sheets to each student and a monitored return of the same are recommended so that all sheets are returned at the end of the exam.

4.3. Documenting the course of an exam

The course of the assessment and any arising issues or problems are documented (e.g. recording the name of the persons responsible for the exam and for administering it, the monitors, specific events, incidents of cheating, and any computer problems in the case of computer-based exams).

Examples of violations to the conditions for administering the exam include:

- Noise and disturbance through construction work during an exam
- Rooms unsuitable for testing
- Deficient test material or poor copies of test questions, errors in the numbering of questions
- Computer failure during computer-based exams

Students must assert immediately during or after the exam that a violation of proper administration has occurred. It is not permissible to wait for the exam scores and then, in the case of failure, claim that proper administration of the exam did not take place.

In the case substantial problems arise during the administration of an exam, it is recommended that a repeat

session be offered for all candidates and the better of the two results be counted.
There should also be rules set down for reaching decisions about disruptions caused by test-takers and their possible exclusion, as well as the discontinuation of an exam (e.g. due to acute illness) and the corresponding documentation by the responsible person.

5. Analysis and documentation

A thorough, documented analysis of the assessment, including statistical analysis, is required to ensure the exam's content validity and legality. Errors in summative tests can cause considerable disadvantages to students which can range from increased study load and lengthened study time as a result of required repeat tests to the cancellation of scholarships as a result of unjustifiably low scores and dropping out of the degree program. In addition, statistical analyses and documentation are basic to the post-review of assessments (see 7.1).

5.1. Statistical Analysis

For all testing formats, an appropriate statistical analysis of the exam results is to be performed that covers, in particular, question difficulty and discrimination (primary analysis).
For testing formats in which, in addition to the questions, other systematic influencing factors exist, such as examiner influences (e.g. OSCE), these are to be taken into consideration in the analysis (e.g. methods of the generalizability theory). For multiple-choice questions an additional distractor analysis must be performed.
Should there be indications of erroneous or unclear questions, then any such questions need to be double-checked in respect to form and content.

5.2. Corrections of the analysis

After any needed corrections to the evaluation of the questions or the exam, a final analysis of the exam shall take place (including further test-statistical analysis).
The second review of the questions after administering the exam serves to solidify the legal conformity/non-appealability of the exam scores. For this reason, a two-step procedure is recommended for analyzing assessments. The first step is test-statistical analysis of the exam, after which there a check is conducted to see if any of the questions are conspicuous in terms of difficulty or discrimination. According to current experience, very difficult questions (difficulty under 0.4) and questions with very low discrimination (below 0.2) are to be checked in terms of content accuracy by the responsible persons.
If the questions are determined to be erroneous, re-analysis of the exam is required. Only after performing the new analysis, should the exam scores be announced. Re-analysis is necessary if, for instance, additional corrections in how questions are graded must take place in response to student appeals (see below).
Even in the case of machine grading, such as for computer-based exams, attention must be paid that all measures to ensure quality have been followed prior to announcing the scores. The person responsible for the assessment must formally release the results.
When correcting erroneously asked questions, it must be ensured that no disadvantages to the candidates arise as a result. For instance, type-A multiple-choice questions (one of five) are not simply dropped from the evaluation if more than one of the possible responses must be recognized as correct. Instead, all candidates who gave one of the correct responses must be given credit for it (see also the rules for the written state examinations issued by the Institut für medizinische und pharmazeutische Prüfungsfragen [IMPP]).
In the case of justifiable objections to the test questions or their evaluation, the necessary corrections must be undertaken for all those who completed the exam, meaning not just for the student filing the appeal, and made public. Attention must be paid that justifiable objections and the resulting corrections are documented (e.g. scope of responsibility of those making decisions).
If test questions are acknowledged as problematic, a legally binding approach is needed which guarantees that no disadvantages to students are caused by deficient test questions. If a question cannot be solved, then

1. the question can be excluded from the valuation and the maximum number of possible points is reduced accordingly, or
2. the total possible number of points allotted for this question is credited to all students (in this case the total number of possible points remains unchanged).

In the case of corrections of the possible responses to a test question that occur after announcement of the exam results, the candidates' scores may not be subsequently lowered.

5.3. Assessment reports

An assessment report regarding the exam is generated covering information on evaluation and grading, along with the statistical analysis of the scores. In particular, any changes to the value or weighting of questions, the answers evaluated as correct, and unevaluated questions must be documented along with the name of the person responsible for the changes.

5.4. Random checks

A random check is carried out on the corrections and evaluations.
Along with a random check of the corrections and evaluations, a check of the performance of all failed students is recommended. Inspection of written exams must be undertaken by impartial graders. If tests are read with the help of scanners, then random checks are also necessary. The nature and scope of these checks should be documented.

5.5. Documentation of the results, guidelines on archiving

The exam scores and performance records are compiled centrally, or by the subject departments, and saved centrally to ensure documentation.

The lengths of time for keeping exams and test documents are to be bindingly set down (e.g. in the exam regulations *[Prüfungsordnung]*). There are no uniform requirements concerning the length of time: the relevant valid provisions at the local level apply (e.g. state regulations on archiving *[Landesarchivierungsordnung]*). Please seek advice from your legal department in regard to this aspect. As a point of reference, the following rule can apply: written exams and records of oral exams are to be kept for at least 18 months after completion of the assessment. For computer-based exams, the individual scores are to be saved for 18 months in the form of test records. The lists of candidates and graded credits are to be centrally kept for at least ten years as hardcopy or digitally. In the case of appeals against the assessment, no documents may be destroyed until the final decision has been reached.

6. Feedback for students

Feedback for students regarding their performance on exams must be given in a timely and transparent manner. This is the only way assessments can be effectively used as an instrument to guide learning.

6.1. Announcement of scores

Announcement of the scores in a manner compliant with data privacy law occurs within an appropriate amount of time that has been defined in advance. This time period must not exceed three weeks.

When announcing exam scores, the provisions under data privacy law must be observed. In particular, it is impermissible to publicly post test results with personal information.

6.2. Inspection of assessment documents

Students have the option of inspecting their exams within an appropriate period of time. The relevant statutory provisions are to be taken into account in respect to this.

Students must be granted access to their own exams upon request or application. An appropriate period of time should be allotted for inspecting exams after announcement of the scores. The option to view the completed test should be possible throughout this entire time period. The dates and times for inspection must be reasonable and should be announced for the period in which any appeals may be submitted. The educational institution can determine fixed times for such inspections. These must be made known at the same time the exam is announced. Should students not be able to review the documents during this time for a justified reason, inspection of the documents should be made possible in another way. The inspection should take place under supervision, which is why it makes sense to set a date for preparing simultaneous inspections by more than one test-taker.

6.3. Deadline for appeals

The deadline to appeal the exam score must be at least a month starting from the announcement of the results. The possibility to view exam documents should also be possible within this month-long period. Information about these rights must be individually communicated in writing and sent to the candidate with the exam result.

This information should contain the following:

> You have sat for the XY exam on (date) and have passed/failed with the grade of XY.
>
> Information on right to appeal: You may file an appeal against this notification with the Dean of Studies within a time period of one month.

Sending out (automatically generated) notifications regarding failure of an exam, along with instructions on filing an appeal, is recommended. If there are no instructions on submitting an objection or appeal, then the time period allotted for this in Germany will be one year.

Rejections of appeals against the evaluation of questions or objections concerning test administration must also contain instructions about legal recourse.

6.4. Nature and scope of feedback

The nature and scope of the feedback for students regarding assessment results are defined with the goal of giving students detailed information on their proficiency levels (e.g. breaking the overall score down according to sub-disciplines, etc.). Longitudinal feedback is to be aimed for that gives students information on their proficiency level

1. *in relation to the requirements placed on them,*
2. *in relation to the other candidates, and*
3. *their own individual educational development.*

As a result of statutory requirements, the number of summative tests to be conducted is very high, which frequently makes the administration of additional formative tests difficult. For this reason, the formative potential of the summative test results should be utilized by considering the individual summative tests of a student over the long-term.

6.5. Publishing test questions

Publication of the test questions is not recommended – as long as no sufficiently large question pool exists. Uniform rules and recommendations on this (e.g. the necessary collecting of all sheets of paper with test questions) are to be striven for by the medical school or degree program and these are to be communicated to the students.

7. Post-processing assessments

Following up on an assessment initially serves to ensure the quality of a subject exam by allowing deficiencies in

questions to be discovered and corrected. Moreover, it is an important feedback instrument for teaching coordinators, since assessments provide information on what the students have actually learned, as well as if and to what extent changes to the curriculum are needed and would be meaningful (e.g. a change in focus during class sessions).

7.1. Post-Review

To assure and improve the quality of future exams, a written and documented post-review of the assessment will take place, in which the persons responsible for the exam participate. Using content-based criteria, results of test-statistical analysis (e.g. difficulty, discrimination, reliability) and student comments and suggestions, recommendation for improvements to test questions and exam structure will be compiled in the post-review.

7.2. Feedback for authors and subject representatives

The assessment results, their analysis and the results of the post-review process need to be forwarded in a timely manner, once each semester, to the authors of the questions, curriculum developers, and the responsible subject representatives. Appropriate consequences should be drawn and necessary measures implemented and documented.

Notes

[1] To facilitate the readability of the German version, the feminine grammatical form does not additionally appear in the text; the meaning includes both genders equally in all cases.

[2] These recommendations have no legally binding or precedential effect. The relevant statutory provisions and regulations of the educational institution responsible for the assessment apply in each individual case.

[3] An adjustment clause is a formal rule that allows the minimum passing grade to be lowered if the results of the candidates are overall low. By adjusting the grading curve, assessments with an exorbitantly high number of failures are prevented.

[4] It has been proven that there is a tendency to give better evaluations for oral assessments than for written ones. Example: candidate A does not pass the test and receives the opportunity to be re-examined orally. There is a high probability that he will be rated with a 3 at least. Candidate B passes the written assessment with a 4 and is not given the opportunity to be re-examined, leaving him with the grade of 4 on his official academic transcript. Candidate B has no opportunity to improve his grade and is thus disadvantaged in relation to Candidate A.

[5] It is critical here that generalized conclusions are drawn from the assessments about individuals or groups which are based on a very limited number of random checks. Only in the case of a high validity is the applicability of these generalizations to other situations permissible using the assessment results. In traditional test theory, objectivity and reliability are pre-requisites for high validity.

Acknowledgements

We wish to extend our gratitude to all members of the GMA assessments committee and the MFT working group on assessment who took part in drawing up these recommendations and the related discussions and evaluations. We wish to particularly thank:

- Prof. Dr. Klaus Albegger, Graz
- Dr. Daniel Bauer, Munich
- Dipl. Med.-Inf. Konstantin Brass, Heidelberg
- Peter Brüstle (M.A.), Freiburg
- Dr. Corinne M. Dölling, Berlin
- PD Dr. Roman Duelli, Heidelberg
- Dr. Jan Ehlers, Hanover
- Prof. Dr. Martin Fischer (MME), Munich
- Dr. Volkhard Fischer, Hanover
- Prof. Dr. Johannes Forster (MME), Freiburg
- Andreas Fuhrig, Göttingen
- Prof. Dr. Matthäus Grasl (MME), Vienna
- Dr. Achim Hochlehnert, Heidelberg
- Dipl.-Ing. Matthias Holzer, Munich
- Dr. Hans Haage, Rheinbach
- Dr. Sören Huwendiek (MME), Bern
- Prof. Dr. Jana Jünger (MME), Heidelberg
- Prof. Dr. Ingo Just, Hanover
- Dr. Roland Kabuß (MME), Hanover
- Prof. Dr. Klaus J. Klose, Marburg
- Dr. Richard Melamed, Frankfurt
- Dr. Andreas Möltner, Heidelberg
- Dr. Daniela Mohr, Tübingen
- Elisabeth Narziß (physician), Mannheim
- Dr. Zineb Nouns, Berlin
- Prof. Dr. Franz Resch, Heidelberg
- Dr. Thomas Rotthoff (MME), Düsseldorf
- PD Dr. Jobst-Hendrik Schultz, Heidelberg
- Dr. Katrin Schüttpelz-Brauns, Mannheim
- Dipl.-Päd Tina Stibane, Marburg
- Markus Stieg (physician), Berlin
- Dipl. Wi.-Rom. Anna Vander Beken, Ulm

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from
http://www.egms.de/en/journals/zma/2014-31/zma000926.shtml

1. Attachment.pdf (78 KB)
   Checklist

References

1. Gesellschaft für Medizinische Ausbildung; Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für fakultätsinterne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. GMS Z Med Ausbild. 2008;25(1):Doc74. Zugänglich unter/available from: http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml
2. WHO; WFME. Guidelines for Accreditation of Basic Medical Education. Geneva, Copenhagen: WHO; 2005.
3. WFME; AMSE. WFME Global Standards for Quality Improvement in Medical Education European Specifications. Copenhagen: University of Copenhagen, MEDINE Quality Assurance Task Force; 2007.
4. National Board of Medical Examiners. Test Administration Handbook. Philadelphia: National Board of Medical Examiners; 2003.
5. Jünger J, Möltner A, Lammerding-Köppel M, Rau T, Obertacke U, Biller S, Narciß E. Durchführung der universitären Prüfungen im klinischen Abschnitt des Medizinstudiums nach den Leitlinien des GMA-Ausschusses Prüfungen: Eine Bestandsaufnahme der medizinischen Fakultäten in Baden-Württemberg. GMS Z Med Ausbild. 2010;27(4):Doc57. DOI: 10.3205/zma000694
6. Reindl M, Holzer M, Fischer MR. Durchführung der Prüfungen nach den Leitlinien des GMA-Ausschusses Prüfungen: Eine Bestandsaufnahme aus Bayern. GMS Z Med Ausbild. 2010;27(4):Doc56. DOI: 10.3205/zma000693
7. Möltner A, Duelli R, Resch F, Schultz JH, Jünger J. Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten. GMS Z Med Ausbild. 2010;27(3):Doc44. DOI: 10.3205/zma000681
8. Schuwirth LW, van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
9. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. Med Teach. 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
10. Biggs JB. Enhancing teaching through constructive alignment. High Educ. 1996;32:347-364. DOI: 10.1007/BF00138871
11. Loftus S, Gerzina T. Being a Health Professional Educator. In Loftus S, Gerzina T, Higgs J, Smith M, Duffy E (Hrsg). Educating Health Professionals: Becoming a University Teacher. Rotterdam: SensePublishers; 2013. S.3-14. DOI: 10.1007/978-94-6209-353-9_1
12. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrot V, Roberts T. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-214. DOI: 10.3109/0142159X.2011.551559
13. Hahn EG, Fischer Mr. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). GMS Z Med Ausbild. 2009;26(3):Doc35. DOI: 10.3205/zma000627
14. Tremp R, Reusser K. Leistungsbeurteilung und Leistungsnachweise in Hochschule und Lehrerbildung - Trends und Diskussionsfelder. Beitr Lehrerbild. 2007;25(1):5-13.
15. Nürnberger F. Handreichung Regelungen zu Prüfungen an Medizinischen Fakultäten und Fachbereichen. Berlin: Medizinischer Fakultätentag; 2010.
16. Schuwirth LW, van der Vleuten CP. Assessing competence: Extending the appoaches to reliability. In: Hodges BD, Ligar L (Hrsg). The question of competence: reconsidering medical education in the twenty-fisrt century. Ithaca, New York: Cornell University Press; 2012. S.113-130.
17. Jackson N, Jamieson A, Khan A. Assessment in medical education and training: a practical guide. Oxford, New York: Radcliffe; 2007.
18. Lucke J. The alpha and the omega of Congeneric Test Theory: An Extension of Reliability and Internal Consistency to Heterogeneous Tests. Appl Psychol Measure. 2005;29(1):65-81. DOI: 10.1177/0146621604270882

Corresponding authors:

Prof. Dr. med Jana Jünger, MME
Universityhospital Heidelberg, Center of Excellence for Assessment in Medicine/Baden-Württemberg, Im Neuenheimer Feld 410, D-69120 Heidelberg, Germany, Phone: +49 (0)6221/56-8657, Fax: +49 (0)6221/56-1341
jana.juenger@med.uni-heidelberg.de

Prof. Dr. med. Ingo Just
Medizinische Hochschule Hannover, Study dean for medicine, Carl-Neuberg-Straße 1, 30625 Hannover, Germany, Phone: +49 (0)511/532-9014, Fax: +49 (0)511/532-2879
studiendekanat.just@mh-hannover.de

Please cite as

Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. GMS Z Med Ausbild. 2014;31(3):Doc34. DOI: 10.3205/zma000926, URN: urn:nbn:de:0183-zma0009261

This article is freely available from

http://www.egms.de/en/journals/zma/2014-31/zma000926.shtml

Received: *2014-05-09*
Revised: *2014-07-01*
Accepted: *2014-07-02*
Published: *2014-08-15*