

# Die Umsetzung der Leitlinien für Fakultäts-interne Leistungsnachweise am Lehrbereich Allgemeinmedizin der Ludwig-Maximilians-Universität (LMU) München: "MC 2.0" mit dem IMSm – die Prüfung im Wandel

## Zusammenfassung

Benotete Prüfungen sind Voraussetzung für die Zulassung zum zweiten Abschnitt der Ärztlichen Prüfung (Staatsexamen) in der Medizin. Daraus lässt sich die Notwendigkeit ableiten, qualitativ hochwertige Prüfungen zu konzipieren, die einen Leistungsvergleich unter Absolventen einer Fakultät und darüber hinaus auch interfakultär erlauben. Hierbei sind Kernqualitätsmerkmale Objektivität, Validität und Reliabilität zu beachten. Die im Leitlinienkatalog der GMA genannten Kriterien sollen die Qualität der Prüfungen sicherstellen. Das Prüfungskonzept des Lehrbereichs Allgemeinmedizin an der LMU erreichte bei Betrachtung der MC-Klausur 2008 nur 14 von 48 möglichen Kriterien. Ein fest eingeplanter Review-Prozess, stetige Weiterbildung des Prüfungsverantwortlichen und die Einführung der Prüfungsverwaltungssoftware IMSm waren Kernpunkte der Veränderung. Heute liegen die Anzahl der erreichten GMA-Kriterien bei 30. Besonders die Einführung der Prüfungsverwaltungssoftware IMSm wurde als zukunftsweisend betrachtet. Die eingebauten Review-Elemente des Systems erleichtern die Etablierung dieser Prozesse und erhöhen die Qualität der Fragen (Item-Analyse, Erkennung von Cueing und weiteren Problemen im Review-Prozess). Die aktuellen Verbesserungsmaßnahmen sind im Sinne einer Qualitätssicherung der Prüfungen als sehr positiv zu bewerten. Es erscheint erstrebenswert, den Kontakt zu Prüfungsverantwortlichen anderer Fakultäten zu suchen, um durch Vernetzung der gegenseitigen Fragepools zu profitieren.

**Schlüsselwörter:** Prüfungen, Allgemeinmedizin, Item Management System

## Hintergrund

Mit der Überarbeitung der Ärztlichen Approbationsordnung (ÄAppO) von 2002 [1] standen die medizinischen Fakultäten vor der Aufgabe benotete Prüfungen zu etablieren, um Leistungen der Studierenden zu bewerten. Die benoteten Leistungsnachweise sind Voraussetzung für die Zulassung zum Staatsexamen und werden seither im Zeugnis abgedruckt. Sie können bei Bewerbungsverfahren und der Vergabe von Stipendien eine wichtige Rolle spielen.

Studien haben gezeigt, dass die Vorbereitung, Strukturierung, Durchführung und Auswertung von Prüfungen an medizinischen Fakultäten in Deutschland häufig nicht optimal ist. Hier besteht weiterhin Aufholbedarf, wenn auch mittlerweile Bestrebungen zur Prüfungsverbesserung messbar sind [2], [3]. Am Beispiel der Allgemeinmedizin an der LMU München wollen wir den aktuellen Stand der Prüfungsentwicklung darstellen und evaluieren.

## Einleitung

Mit Hinblick auf die gewonnene Relevanz der fakultätsinternen Leistungsnachweise müssen Prüfungsformen und -modalitäten gefunden werden, die aussagekräftige und vergleichbare Bewertungen der Leistungen der Studierenden zulassen.

Um die Erstellung solcher Prüfungen zu erleichtern, hat das Didaktikzentrum der Eidgenössischen Technischen Hochschule Zürich (ETHZ) 2003 erstmals den „Leitfaden für das Planen, Durchführen und Auswerten von Prüfungen(...)“ [4] veröffentlicht. Er stellt eine Art Checkliste dar und beachtet dabei auch testtheoretische Gütekriterien, die Anhalt für die Qualität einer Prüfung sind. Neben den Hauptgütekriterien Objektivität, Validität und Reliabilität werden auch Vergleichbarkeit, Bezugssystem/Normierung und Ökonomie als Qualitätsmerkmale formuliert.

In Anlehnung an den Züricher Leitfaden hat der Ausschuss Prüfungen der Gesellschaft für Medizinische

Niklas Boeder<sup>1</sup>

Matthias Holzer<sup>2</sup>

Jörg Schelling<sup>1</sup>

1 LMU München, Medizinische Klinik und Poliklinik IV, Bereich Allgemeinmedizin, München, Deutschland

2 Klinikum der LMU München, Lehrstuhl für Didaktik und Ausbildungsforschung in der Medizin, München, Deutschland

Ausbildung (GMA) und das Kompetenzzentrum Prüfungen Baden-Württemberg im Jahr 2008 „Leitlinien für Fakultäts-interne Leistungsnachweise während des Medizinstudiums“ veröffentlicht [5]. Darin benennen die Autoren Empfehlungen für die Gestaltung von Prüfungen, um den anspruchsvollen Spagat zwischen der knappen personellen und technischen Infrastruktur im universitären Prüfungswesen und der Forderung nach objektiven und zuverlässigen Prüfungen bewältigen zu können. Der Katalog umfasst 48 Kriterien und ist in die folgenden sieben Bereiche gegliedert:

1. Allgemeine strukturelle Vorbedingungen
2. Prüfungskonzeption und -bewertung
3. Organisatorische Vorbereitungen zur Durchführung der Prüfung
4. Durchführung der Prüfung
5. Auswertung und Dokumentation
6. Rückmeldung an die Studierenden
7. Prüfungsnachbereitung

Die Kriterien sollen sicher stellen, dass die Prüfungen organisatorisch gut geplant, inhaltlich relevant gestaltet und nachhaltig konzipiert werden. So sind zum Beispiel die Auswahl der passenden Prüfungsform und Verrechnung von Teilnoten vorab zu klärende Punkte. Insgesamt wird großer Wert auf die Benennung eines Prüfungsverantwortlichen, die stetige Weiterbildung der Prüfungsauctoren sowie den ökonomischen Einsatz von technischem Equipment bei der Auswertung und im Review-Prozess gelegt.

Die Beschreibung des Review-Prozesses in den Leitlinien schließt zum Beispiel die kritische Durchsicht der Prüfung, am besten zusammen mit fachfremden Vertretern, vor dem Prüfungstermin, aber auch eine Qualitätssicherung mit teststatistischen Analysen nach der Prüfungsdurchführung – vor allem mit Hinblick auf zukünftige Prüfungen – ein (siehe Abbildung 1).

#### 7. Prüfungsnachbereitung

	Erläut.	Erlie-digt	Anmerkungen
7.1	Zur Qualitätssicherung und -verbesserung künftiger Prüfungen findet eine schriftlich dokumentierte Nachbewertung (Post-Review) der Prüfung statt, an der die Prüfungsbeauftragten teilnehmen. Anhand inhaltlicher Kriterien und teststatistischer Auswertungsergebnisse (z. B. Schwierigkeiten, Trennschärfen, Reliabilität) werden in dieser Nachbewertung Verbesserungsvorschläge für Prüfungsaufgaben und Prüfungszusammenstellung erarbeitet.	<input type="checkbox"/>	
7.2	Die Prüfungsergebnisse, deren Auswertung sowie die Ergebnisse des Post-Review-Prozesses sollen zeitnah mindestens einmal im Semester an die Fragenautoren, die Curriculumentwickler und Fachvertreter weitergegeben werden.	<input type="checkbox"/>	

**Abbildung 1: GMA-Leitlinien - Abschnitt 7: Prüfungsnachbereitung als Beispiel für die Gliederung des Kriterienkatalogs [5]**

Um den Stand der Prüfungsentwicklung an den medizinischen Fakultäten seit der Novellierung der ÄAppO zu erheben wurde bereits 2007 vom deutschen Fakultätentag eine Umfrage in Auftrag gegeben. Die medizinischen Fakultäten gaben im Rahmen dieser Umfrage Informationen zur zentralen Organisation, Prüfungsplanung und -durchführung, zu den Prüfungsformen, zur Durchführung eines Aufgabenreviews als interne Qualitätssicherung und zur statistischen Auswertung an. Eine erste Veröffent-

lichung der Ergebnisse dieser Umfrage wurde von Herrn Professor Resch bereits auf dem Fakultätentag 2008 publiziert [2]. Die Erhebung zeigte, dass nur weniger als die Hälfte der Fakultäten zentrale Einrichtungen für die Erstellung oder Auswertung von Prüfungen nutzten, obwohl der Großteil der Prüfungen als Multiple-Choice-Klausur konzipiert war. Die von Möltner et al im Jahr 2010 veröffentlichte genauere Analyse [3] derselben Umfrage zeigte, dass Multiple-Choice-Klausuren die häufigste Prüfungsform war und nur wenige Universitäten praktische Prüfungen, so genannte OSCE („objektive strukturierte klinische Prüfungen“), einsetzten. Immerhin knapp 40% der Universitäten nutzten nach den Studienergebnissen strukturierte Verfahren zur Beurteilung der Fragenqualität und eine testanalytische Auswertung. Die Autoren kommen zu dem Schluss, dass zwar vermehrt Anstrengungen zur Verbesserung der Prüfungskultur betrieben werden, diese aber sehr heterogen ausfallen.

## Allgemeinmedizin an der LMU München

Seit Einführung des MeCuM-Konzeptes („Medizinisches Curriculum München“) an der Medizinischen Fakultät München ist der Lehrbereich Allgemeinmedizin maßgeblich an der Gestaltung von zwei Prüfungen [6], [[https://e-learning.mecum-online.de/Informationambu\\_Allg\\_L6\\_L7.PDF](https://e-learning.mecum-online.de/Informationambu_Allg_L6_L7.PDF)] und vielen weiteren Unterrichtsveranstaltungen im Rahmen der Ausbildung angehender Ärzte beteiligt.

Im klinischen Studienjahr wird eine Vorlesungsreihe über die ambulante Patientenversorgung in Verbindung mit Seminaren und Hospitationen in Lehrpraxen angeboten [[https://e-learning.mecum-online.de/Informationambu\\_Allg\\_L6\\_L7.PDF](https://e-learning.mecum-online.de/Informationambu_Allg_L6_L7.PDF)]. Der Lernerfolg dieser Kursreihe wird am Ende des dritten Studienjahres mit einer aus 40 Fragen bestehenden Multiple-Choice-Klausur überprüft und wird Grundlage für die Betrachtung der Prüfungsentwicklung in diesem Artikel sein.

Über viele Jahre hinweg waren die Vorlesungsverantwortlichen – ein Dozent pro Thema – in der Pflicht Fragen mit passenden Antwortmöglichkeiten für die Klausur zu konzipieren. Die Prüfung bestand dann aus jeweils einer Frage pro Vorlesungsthema. Die Fragenauswahl und die zur Verhinderung von Täuschungsversuchen erforderliche Versionierung der Klausuren musste manuell durchgeführt werden. Die Auswertung der Antwortbögen erfolgte mit hohem personellen Ressourceneinsatz. Im Laufe der Zeit entstand aus den Einsendungen der Dozenten eine Fragensammlung, welche in Form einer unübersichtlichen Computerdatei gepflegt wurde. Eine teststatistische Analyse oder ein organisierter Review-Prozess war nicht eingeplant.

Nutzt man die Kriterien der von der GMA veröffentlichten Leitlinie [5], um die Prüfung zu beschreiben, erfüllte der Lehrbereich mit der bisherigen Prüfungsdurchführung lediglich 14 von 48 möglichen Kriterien. Die Frage nach der Beurteilung der Kriterienerfüllung erschließt sich entweder aus der eindeutigen Formulierung der Kriterien oder aber aus den ausführlichen Erläuterungstexten.

Aufgrund der neu gewonnenen Relevanz der Leistungsnachweise und den damit verbundenen qualitativen Ansprüchen leitete der Lehrbereich die Notwendigkeit ab, das Prüfungskonzept zu überdenken.

## Methoden

### Maßnahmen zur Verbesserung des Prüfungskonzeptes

#### Einführung des IMSm

Besonders unter dem Gesichtspunkt eines ressourcensparenden Konzeptes wurde intensiv an der Etablierung der Prüfungssoftware ItemManagementSystem für Medizin (IMSm) gearbeitet [7]. Bereits seit 2006 wurde es an den Pilotuniversitäten Berlin, Heidelberg und München eingesetzt. Im Sommer 2008 fand es das erste Mal Anwendung im Lehrbereich Allgemeinmedizin.

Das IMSm entstand aus einem Gemeinschaftsprojekt der genannten Universitäten und ermöglicht den Prüfungsverantwortlichen ein hohes Maß an Standardisierung und Qualitätssicherung bei der Gestaltung von Prüfungen [7].

#### Arbeitsweise des IMSm

Über das internetbasierte Webinterface der IMSm-Software sind Prüfungsverantwortliche in der Lage Fragen zu importieren oder zu erstellen. Ein Rechte- und Rollenmanagement erlaubt die Einbindung von Mitarbeitern auf verschiedenen Ebenen.

Neben dem verbreiteten MC-Fragentyp mit Einfachauswahl (siehe Abbildung 2) steht auch der Fall-Typ bei der Fragenerstellung zur Verfügung. Beim Fall-Typ kann gewählt werden, ob die Fragen aufeinander aufbauen („Key Feature“ für elektronische Prüfungen) oder ob sie unabhängig voneinander beantwortet werden können („Sternförmige Fallvorlage“). In der aktuellen IMSm-Version können sogar OSCE-Stationen für praktische Prüfungen im System gespeichert werden. Die Möglichkeiten fakultäre Prüfungsworkflows im IMSm abzubilden sind somit weitreichend.

Eine weitere wichtige Grundidee der Software ist eine mögliche Vernetzung verschiedener Lehrbereiche oder sogar Fakultäten. Auf diese Weise sinkt der Aufwand der Prüfungserstellung an jeder einzelnen Fakultät, da qualitätsgeprüfte Fragen aus anderen Fakultäten zur Klausurerstellung verwendet werden können.

Um den Überblick über große Fragenpools zu behalten und um Fragen bei der Klausurerstellung gezielt auswählen zu können, kann den Fragen nicht nur ein bestimmtes Fach, sondern darüber hinaus auch ein Lernziel, ein Leitsymptom und weitere Metadaten zugeordnet werden. Ein Beispiel für ein Lernziel könnte in der Lernzielgruppe „Spezielle hausärztliche Versorgungsprobleme“ lauten: „Versorgung von alten Menschen, Pflegebedürftigkeit“ oder „Abwartendes Offenhalten“. Im Falle einer fachübergreifenden Frage zum Beispiel „Innere Medizin“ → „Ödeme“. Die Auswahlmöglichkeiten beziehen sich hier auf einen definierten Lernzielkatalog, der innerhalb der

Fakultät zur Zuordnung von Lernzielen zu weiteren MeCuM-Lehrveranstaltungen verwendet wird. Für die Allgemeinmedizin sind fachspezifische Metadaten zur Beschreibung der Fragen wie spezifische Lernziele, Leitsymptome und allgemeinmedizinische Handlungsweisen (AGV, abwartendes Offenhalten, Hausbesuch, uvm.) implementiert. Für das Zusammenstellen einer neuen Klausur werden Fragen aus dem Pool ausgewählt. Jederzeit können in der Planungsphase Fragen editiert oder ausgetauscht werden. Definierte Optionen wie zum Beispiel Seitenzahlen, Versionsangabe, Fakultätslogo sind beim Druck der Prüfungen fakultativ auswählbar.

#### Technische Verbesserungen bei der Auswertung

Die Umstellung der Prüfungsverwaltung von einer oder mehreren einfachen Textdateien auf eine datenbankbasierte Websoftware blieb im Rahmen der Überarbeitung der Prüfungsorganisation nicht die einzige Änderung. Das IMSm vereint neben der reinen Datenverwaltung auch die Fähigkeit die Datensätze für ein Scanner-basiertes Auswertungssystem aufzubereiten.

Die Antwortbögen der Studierenden werden dann zentral mit einem Dokumentenscanner ausgewertet, um eine fehlerarme und zeitnahe Veröffentlichung der Ergebnisse zu ermöglichen. Aus den Scandaten werden darüber hinaus die Item- und Prüfungskennwerte berechnet und an das IMSm übermittelt. Vor der Ergebnisveröffentlichung kann noch auf etwaige Schwachstellen (z.B. fehlerhafte Fragen) reagiert werden. Für zukünftige Prüfungen können die Kennwerte wie Trennschärfe, Schwierigkeit und Distraktorenhäufigkeiten zur Verbesserung von Prüfungsfragen verwendet werden.

#### Strukturelle Maßnahmen

Was die allgemeinen Vorbedingungen einer Prüfung betrifft, wird versucht die Kommunikation zwischen den Studierenden und dem Lehrbereich zu verbessern, indem u.a. ein vollständiger und für die Prüfung verbindlicher Lernzielkatalog online abrufbar ist. Des Weiteren werden Modalitäten wie Termin, Ort, Zeit, Ablauf und Möglichkeiten der Einsichtnahme schriftlich mitgeteilt.

Mit Frageautoren und im Rahmen von Kerngruppentreffen werden Prüfungsfragen und Prüfungen – inhaltlich und organisatorisch – besprochen, um nachhaltig die Qualität zu sichern und weiter zu verbessern. Der Review-Prozess im Vorfeld der Prüfungen wurde ausgebaut. Eingesendete Prüfungsfragen und Antwortmöglichkeiten werden vom Prüfungsbeauftragten nach formalen und inhaltlichen Kriterien überprüft und gegebenenfalls in der Kerngruppe der Lehrbeauftragten als Review-Team bezüglich einer notwendigen Überarbeitung diskutiert. Als weiteres Qualitätskriterium wird eine Reliabilität (Cronbachs alpha) von 0,8 angestrebt, was aber mit einer Fragenzahl von 40 Fragen nur bedingt zu realisieren ist. Damit erfüllt der Lehrbereich Allgemeinmedizin an der LMU weite Teile der GMA-Anforderungen an Review-Prozesse.

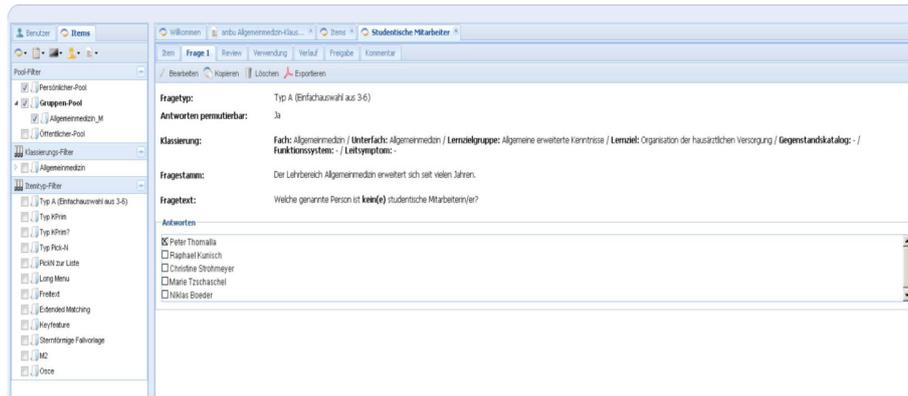


Abbildung 2: Beispiel für die Ansicht einer neu erstellten Frage vom Typ Multiple Choice

## Ergebnisse

### Re-Evaluation mit Hilfe der GMA-Leitlinien

Die Bemühungen der Prüfungsverbesserung schlagen sich in einem deutlichen Anstieg der erfüllten Leitlinienempfehlungen der GMA nieder. Bei aktueller Berechnung werden 30 von 48 möglichen Kriterien im Vergleich zu 14 Kriterien vor der Verwendung des IMSm erfüllt (siehe Abbildung 3).

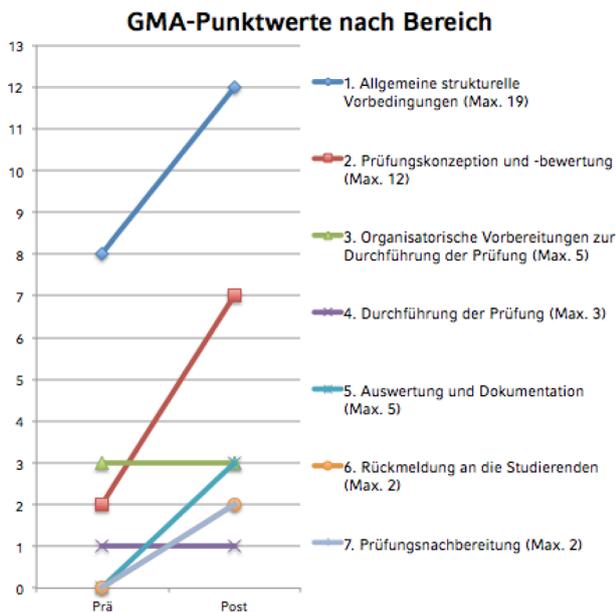


Abbildung 3: Zahl der erfüllten Leitlinienkriterien nach Abschnitt (vor und nach der Einführung des IMSm)

Die Steigerung im Bereich „Allgemeine strukturelle Vorbedingungen“ ergibt sich im Wesentlichen aus folgenden Punkten: Einführung eines vollständigen schriftlichen Lernzielkatalogs, formale Verbesserungen im Prüfungsablauf (z.B. eindeutige zugeordnete Lösungsbögen, klare Einteilung in Gruppen, feste Sitzordnung), Einführung eines einheitlichen Verfahrens zur Bewertung bei fehlerhaft gestellten Aufgaben und die Möglichkeit zur strukturierten Einsichtnahme in die Prüfungsergebnisse.

Im Bereich „Prüfungskonzeption und -bewertung“ konnten wir die deutlichste Verbesserung erreichen. Dies betrifft die folgenden Punkte: Einführung einer standardisierten, inhaltlichen und formalen Bewertung der Fragen (Prä-Review), Dokumentation der Review-Ergebnisse im IMS, Ressourceneinsparung durch Delegation von Aufgaben an die zentrale Prüfungsstelle und Einführung eines verbindlichen Bewertungsschemas. Die Reliabilität (Cronbachs Alpha) der Gesamtprüfungen mit 40 MC-Fragen liegt bei etwa 0,65 und wurde durch die eingeleiteten Maßnahmen nicht signifikant verändert. In Anbetracht der vorgegebenen Fragenzahl betrachten wir diesen Wert als ausreichend.

Abbildung 3 zeigt neben dem generell positiven Trend der Entwicklung in den Bereichen „Organisatorische Vorbereitungen zur Durchführung der Prüfung“ und „Durchführung der Prüfung“ keine messbare Verbesserung. Hier sind jeweils zwei Kriterien nicht erfüllt.

Im Bereich „Auswertung und Dokumentation“ haben wir uns von null auf drei Kriterien stark verbessert. Dies bezieht sich auf die teststatistische Auswertung vor der Veröffentlichung der Noten zur Identifikation von problematischen Fragen. Zudem wird jetzt ein Prüfungsbericht erstellt, in dem Veränderungen der Aufgabenbewertungen oder -gewichtungen, der als korrekt gewerteten Lösungen und nicht gewerteten Aufgaben dokumentiert sind.

Die „Rückmeldung an die Studierenden“ hat sich durch die Bekanntgabe der Prüfungsergebnisse innerhalb eines vorab festgelegten Zeitraums von zwei Wochen verbessert. Zusätzlich haben die Studierenden innerhalb einer angemessenen Frist von einem Monat die Möglichkeit zur Einsicht in ihre Prüfungsunterlagen.

In der „Prüfungsnachbereitung“ werden jetzt beide Kriterien erfüllt: es findet eine schriftlich dokumentierte Nachbereitungssitzung (Post-Review) der Prüfung mit den Prüfungsbeauftragten statt und die Prüfungsergebnisse werden zeitnah an die Fragenautoren weitergegeben.

## Diskussion

Die Novellierung der ÄAppO 2002 hat weitreichende Relevanz für die Durchführung der hochschulinternen Prüfungen. Leistungsnachweise werden seither benotet und

können eine Rolle bei der Vergabe von Arbeitsplätzen oder zum Beispiel Stipendien spielen. Daraus lässt sich die Notwendigkeit ableiten, qualitativ hochwertige Prüfungen zu konzipieren, die einen Leistungsvergleich unter Absolventen einer Fakultät und darüber hinaus auch einen interfakultären Vergleich erlauben.

Um die Erstellung solcher qualitativ hochwertiger Prüfungen zu erleichtern, hat erstmals das Didaktikzentrum der ETHZ und wenige Jahre später der Ausschuss Prüfungen der GMA in Zusammenarbeit mit dem Kompetenzzentrum Prüfungen Baden-Württemberg Qualitätsmerkmale und Leitlinien für den Prozess der Prüfungskonzeption veröffentlicht. Als Kernqualitätsmerkmale gelten Objektivität, Validität und Reliabilität. Besonderer Wert wird im Weiteren auf transparente Kommunikation (Prüfungsankündigung, Bestehensregelung, uvm.), Review-Prozesse inklusive der Einbindung fachfremder Vertreter, sowie auf die teststatistische Analyse gelegt. Die im Leitlinienkatalog der GMA genannten 48 Kriterien sollen die Qualität der Prüfung und dadurch die Nachhaltigkeit des Fragenpools für zukünftige Prüfungen sicherstellen.

Noch 2008 erfüllte das Prüfungskonzept des Lehrbereichs nur 14 der 48 Kriterien. Bei den notwendigen lehrbereichsinternen Diskussionen über die möglichen Verbesserungen orientierten wir uns stark an den GMA-Leitlinien. Grundlegende Änderungen innerhalb des Lehrbereichs waren die Motivation zur stetigen Weiterbildung der Prüfungsverantwortlichen, eine verbesserte Ausnutzung der fakultätsinternen Ressourcen (z.B. Belegleser und zentraler Druck) und die Einführung der Prüfungsverwaltungssoftware. Wir erachten besonders die Etablierung eines geregelten Review-Prozesses, sowohl vor als auch nach der Prüfung, als Kernpunkt.

Heute erfüllt der Lehrbereich damit insgesamt 30 von 48 möglichen Kriterien (Abbildung 3).

In den Bereichen „Allgemeine strukturelle Vorbedingungen“ und „Prüfungskonzeption und -bewertung“ war uns besonders die Einführung eines Lernzielkataloges ein Anliegen. Der Lernzielkatalog ist Grundlage für die Erstellung der Prüfungsfragen und hilft den Fragenautoren bei der Formulierung von relevanten Fragen. Er dient auch den Studierenden als Erwartungshorizont zur Prüfungsvorbereitung. Zusätzlich sind die Vorlesungsbeiträge und Seminare der Dozenten am Lehrzielkatalog ausgerichtet und über das Studierendenportal herunterladbar.

Zu einem hohen Maß wird die Einführung der Prüfungsverwaltungssoftware IMSm im Lehrbereich als zukunftsweisend betrachtet. Die Möglichkeit der technisch-unterstützten Prüfungsauswertung (Dokumentenscanner und entsprechende Software) brachte enorme Vorteile. Bei nur einer 40%-Stelle im Lehrbereich konnte der personelle Ressourceneinsatz bei der Korrektur der Klausur gesenkt werden und zum anderen besteht damit die Möglichkeit die Klausurergebnisse unmittelbar für teststatistische Analysen zu verwenden.

Ein weiterer Vorteil ist die Online-Funktionalität der IMSm-Administrationsoberfläche, denn so kann die Verwaltung des Fragenpools ortsunabhängig gestaltet werden und zugleich auf mehrere Mitarbeiter verteilt werden. Die

eingebauten Review-Elemente des Systems erleichterten die Etablierung der Prozesse und erhöhen die Qualität der Fragen (z.B. Item-Analyse, Erkennung von Cueing). In den Bereichen „Organisatorische Vorbereitungen zur Durchführung der Prüfung“ und „Durchführung der Prüfung“ zeigte sich keine messbare Verbesserung. Der Grund liegt in der Gestaltung der Prüfungsdurchführung, denn das Dekanat meldet zum einen alle in die Kurse eingeschriebenen Studierenden automatisch zur Prüfung an und zum anderen erfolgt die Auswertung der Multiple-Choice-Prüfung mittlerweile durch einen Belegleser und nicht mehr manuell.

Bezüglich der Durchführung der Prüfung wollen wir in Zukunft auftretende Fehler oder Täuschungsmanöver erfassen und zentral dokumentieren. Eine Checkliste für den formalen Ablauf der Prüfung soll erarbeitet werden. Es erfolgt bisher aus Ressourcengründen keine stichprobenartige Kontrolle der Korrekturen und Bewertungen. Dies ist bei maschineller Auswertung und manueller Problembearbeitung auch nur beschränkt sinnvoll.

In den Bereichen „Rückmeldung an die Studierenden“ und „Prüfungsnachbereitung“ sind keine Verbesserungen mehr möglich, da inzwischen alle Kriterien erfüllt sind.

Die aktuellen Verbesserungsmaßnahmen sind hinsichtlich der Qualitätssicherung der Prüfung im Lehrbereich Allgemeinmedizin an der LMU München als sehr positiv zu bewerten. Zu bedenken ist allerdings, dass die Berechnung der erfüllten Leitlinienkriterien nicht alle Aspekte der Prüfungsqualität (wie beispielsweise die Studentenzufriedenheit) erfassen kann. Der positive Trend kann aber als Indiz dafür angesehen werden, dass sich die Prüfungssituation verbessert hat.

Der Stichprobenumfang der Prüfungsevaluation durch die Studierenden war leider zu gering, als dass eine statistische Auswertung aussagekräftig gewesen wäre. Für die Zukunft wird hier eine detaillierte Untersuchung angestrebt, um Schwachstellen aufzudecken.

Im Weiteren können wir uns die Einbindung fachfremder Vertreter in die Fragenreviews vorstellen. Es erscheint zudem erstrebenswert den Kontakt zu Prüfungsverantwortlichen anderer Fakultäten zu suchen, um durch Vernetzung der gegenseitigen Fragenpools zu profitieren.

## Interessenkonflikt

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

## Literatur

1. Bundesministerium für Gesundheit. Approbationsordnung für Ärzte vom 27. Juni 2002. Bundesgesetzbl. 2002;1:2405-2435.
2. Resch F. Universitäre Prüfungen im Licht der neuen ÄAppO. Tagungsband des MFT. Berlin: Medizinischer Fakultätentag; 2008.

3. 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
4. Eugster B, Lutz L. Leitfaden für das Planen, Durchführen und Auswerten von Prüfungen an der ETHZ. Zürich: Didaktikzentrum Eidgenössische Technische Hochschule Zürich; 2004.
5. Gesellschaft für Medizinische Ausbildung, Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für Fakultäts-interne 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>
6. Schelling J, Boeder N, Schelling U, Oberprieler G. Evaluation des "Blockpraktikums Allgemeinmedizin". Überblick, Auswertung und Rückschlüsse an der LMU. Z Allg Med. 2010;86:461-465.
7. Brass K, Hochlehnert A, Jünger J, Fischer MR, Holzer M. Stadiumbegleitende Prüfungen mit einem System: ItemManagementSystem für die Medizin. GMS Z Med. 2008;25(1):Doc37. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000521.shtml>

**Korrespondenzadresse:**

Dr. Jörg Schelling  
LMU München, Medizinische Klinik und Poliklinik IV,  
Bereich Allgemeinmedizin, Pettenkoferstraße 8 a, 80336  
München, Deutschland  
[joerg.schelling@med.uni-muenchen.de](mailto:joerg.schelling@med.uni-muenchen.de)

**Bitte zitieren als**

Boeder N, Holzer M, Schelling J. Die Umsetzung der Leitlinien für Fakultäts-interne Leistungsnachweise am Lehrbereich Allgemeinmedizin der Ludwig-Maximilians-Universität (LMU) München: "MC 2.0" mit dem IMSm – die Prüfung im Wandel. GMS Z Med Ausbild. 2012;29(3):Doc42.  
DOI: 10.3205/zma000812, URN: urn:nbn:de:0183-zma0008123

**Artikel online frei zugänglich unter**

<http://www.egms.de/en/journals/zma/2012-29/zma000812.shtml>

**Eingereicht:** 22.06.2011

**Überarbeitet:** 08.12.2011

**Angenommen:** 21.03.2012

**Veröffentlicht:** 15.05.2012

**Copyright**

©2012 Boeder et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.

# Realisation of the guidelines for faculty-internal exams at the Department of General Medicine at the University of Munich: Pushing medical exams one step ahead with IMSm

## Abstract

Graded exams are prerequisites for the admission to the medical state examination. Accordingly the exams must be of good quality in order to allow benchmarking with the faculty and between different universities. Criteria for good quality need to be considered - namely objectivity, validity and reliability. The guidelines for the processing of exams published by the GMA are supposed to help maintaining those criteria. In 2008 the Department of General Medicine at the University of Munich fulfils only 14 of 18 items. A review process, appropriate training of the staff and the introduction of the IMSm software were the main changes that helped to improve the 'GMA-score' to 30 fulfilled items. We see the introduction of the IMSm system as our biggest challenge ahead. IMSm helps to streamline the necessary workflow and improves their quality (e.g. by the detection of cueing, item analysis). Overall, we evaluate the steps to improve the exam process as very positive. We plan to engage co-workers outside the department to assist in the various review processes in the future. Furthermore we think it might be of value to get into contact with other departments and faculties to benefit from each other's question pools.

**Keywords:** Tests, General Medicine, Item Management System

Niklas Boeder<sup>1</sup>

Matthias Holzer<sup>2</sup>

Jörg Schelling<sup>1</sup>

1 LMU München, Medizinische Klinik und Poliklinik IV, Bereich Allgemeinmedizin, München, Deutschland

2 Klinikum der LMU München, Lehrstuhl für Didaktik und Ausbildungsforschung in der Medizin, München, Deutschland

## Background

With the implementation of a new licensing regulation for doctors in 2002 [1] the faculties have to organise graded exams to benchmark students' course achievements. The course achievements are prerequisites for the admission to the medical state examination and are being printed in the certificate since the new licensing requirements in 2002. Student grades can be important in the process of applications and scholarships. Recent studies showed that preparation, structuring, accomplishment and correction of exams in Germany are often not ideal. Space for improvement exists and even first attempts at improving the process were noted [2], [3]. We want to demonstrate the exam processing in the Department of General Medicine and evaluate its progress.

## Introduction

Given the new importance of faculty internal exams, the faculties had to find exam types and modalities to ensure significant and comparable grades.

It was in 2003 when the Didaktikzentrum der Eidgenössischen Technischen Hochschule Zürich (ETHZ) first published the „Leitfaden für das Planen, Durchführen und Auswertung von Prüfungen (...)“ [4]. This guideline was meant to be a helpful checklist and includes criteria of good quality which aim to be indicators for the quality of an exam. Those main criteria are objectivity, reliability and validity but the guideline also includes factors such as comparability, scaling and economy.

On the basis of the Swiss' guideline, the GMA (Gesellschaft für Medizinische Ausbildung) and the Kompetenzzentrum Prüfungen Baden-Württemberg published the „Leitlinien für Fakultäts-interne Leistungsnachweise während des Medizinstudiums“ in 2008 [5]. The authors formulate suggestions for the design of exams to balance the resources with respect to the amount of people that are required and the technical help that exists for the preparation of the exam in comparison with the high prerequisites. The catalogue has 48 items that are grouped into seven categories:

1. General structural prerequisites
2. Exam design
3. Organisational preparations for the realisation of the exam
4. Realisation of the exam
5. Correction of the exam and documentation

6. Publishing the results
7. Review process

The items are presented to help organising exams effectively and relevantly. For example, the appropriate exam design and a consensus on how to correct and deal with grades is a topic to be discussed before the exam. The catalogue focuses on the designation of a person responsible for the exam, appropriate training and an economical use of technical instruments for the correction and the review process.

The review process encompasses a careful revision of the exam, including possible collaboration with faculty outside the department, before and after the actual exam takes place. It would include the calculation of the criteria for good quality (see figure 1).

7. Exam process				
		Comment	Completed	Note
7.1	To ensure and improve the quality of future exams a post exam review process is taking place. Co-workers from outside of the department assist. Analytical calculations are used as a basis for future exam processing.		<input type="checkbox"/>	
7.2	Exam results, the analytical calculations and the conclusions drawn from the post exam review process are published once per semester.		<input type="checkbox"/>	

Figure 1: GMA guidelines, Chapter 7: Review process [5]

To evaluate the progress and changes that have been made in the field of exams since the implementation of a new licensing regulation for doctors, the Deutsche Fakultätentag initiated a survey in 2007. The German faculties were asked to describe the organisation, exam processing and different review processes at their university. Professor Resch published the first results of this survey at the Fakultätentag in 2008 [2]. One of the main results was that, despite the fact that most of the exams were designed as multiple choice-exams, less than half of the faculties used existing central infrastructures for the preparation and correction. A more detailed analysis of the data was published by Möltner et al two years later [3]. Only 40% of the universities used organised review processes. The authors pointed out that efforts that have been made in the field of exams were measurable but that the results have been very heterogeneous.

## General medicine at the University of Munich

Since the implementation of "MeCuM", the name of the new curriculum for medical students in Munich, the Department of General Medicine is involved in medical education with two exams [6], [[https://e-learning.mecum-online.de/Informationambu\\_Allg\\_L6\\_L7.PDF](https://e-learning.mecum-online.de/Informationambu_Allg_L6_L7.PDF)] and a variety of classes.

During the clinical part of the medical studies the department offers classes and lectures covering the clinic outpatient [[https://e-learning.mecum-online.de/Informationambu\\_Allg\\_L6\\_L7.PDF](https://e-learning.mecum-online.de/Informationambu_Allg_L6_L7.PDF)]. The course achievements are benchmarked with the help of a multiple choice (MC) exam consisting of 40 questions taking place at the end of

year four. This exam will be evaluated for further consideration.

For many years each lecturer was responsible for the creation of multiple choice (MC) questions covering their topic. The final exam consists of only one question per lecture. Automatic versioning or export for printing was not available. The correction of the exams was done by a large number of assistants in order to be able to publish the results in an adequate time. In the course of time a confusing and unordered file record originated. A statistical analysis or a review process was never planned.

Using the items published by the GMA [5] to describe the exam, the department fulfils only 14 of 48 possible items. Items were fulfilled if they matched the criteria that were described in detail.

On the basis of the new relevance the faculty exams gained, the Department of General Medicine drew the conclusion that an evaluation of the actual exam concept was necessary.

## Methods

### Steps to improve the exam concept

#### Introduction of IMSm

To make the process more efficient, the department worked hard on the introduction of a software to accompany the whole process of exams called ItemManagement-System für Medizin (IMS<sub>m</sub>) [7]. The software had already been tested by universities in Berlin, Heidelberg and Munich before and was used by the Department of General Medicine for the first time in 2008.

The software was developed as a joint project by the universities mentioned above and insures standardisation and quality management to a high degree [7].

#### How IMS<sub>m</sub> works

Users can log into a web interface that gives full control over all steps of the exam. A well-thought-out user management process allows the integration of co-workers at each of the steps individually. IMS<sub>m</sub> facilitates the co-operation between departments or faculties by giving the parties the possibility to exchange their pool of questions. The system supports not only multiple choice questions (see figure 2) but also cases that can consist of multiple questions that can be linked. The current version even supports OSCE exams and thereby can be used in almost every case.

The users can tag the questions to certain categories, e.g. educational objectives, in order to help not losing the overview over a growing database. The educational objectives are chosen to be general but the software supports department specific objectives as well.

To generate a new exam the user only has to choose questions out of the database. At all steps of the creation questions can be edited or exchanged. The software also allows the user to integrate a faculty logo, versioning or page numbering, for example.

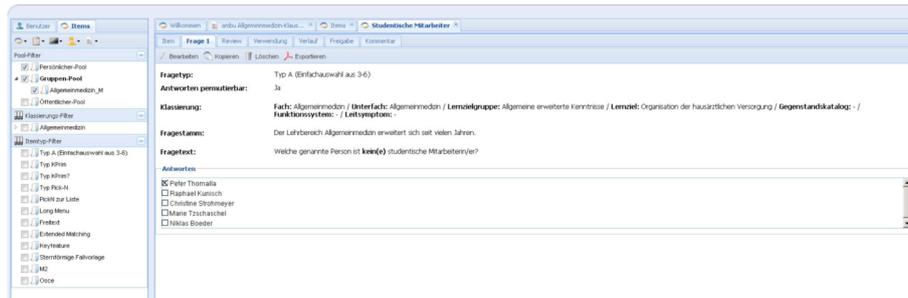


Figure 2: Creation of a new multiple choice question

### Technical improvements

Introducing IMSm as a tool to facilitate the exam preparation thereby moving the file-based maintenance to a database-based web interface was not the only change. IMSm supports the ability to read the answer sheets after scanning.

The answer sheets are scanned centrally to ensure a fast and accurate processing and publishing of the results. Also the scanned data is being used by the software to calculate statistical values. This is important for further exams as values like the discriminatory power can be used to improve the quality.

### Structural steps

The department tries to improve communication with the students by publishing the exact date, place and time the exam will take place as well as a written, detailed catalogue of topics that will be covered by the exam. To ensure quality the co-workers responsible for the questions meet regularly to talk about the questions and their possible answers. Questions must fulfil criteria regarding their content and form. This happens not only before but also after the exam. The department aims to reach a reliability (Cronbachs alpha) of 0.8 which because of the small number of questions is not always easy.

## Results

The steps that have been undertaken to improve the exam concept of the Department of General Medicine at the University of Munich lead to a significant increase with respect to the fulfilment of the criteria enunciated by the GMA (see figure 3).

The number of fulfilled items in the category “general structural prerequisites” mainly evolves from the realisation of the following: introduction of a catalogue of educational objectives, formal improvements of the general process (e.g. structured seating plan and grouping), introduction of a system to cope with wrong questions and announcing regularly dates for inspection of the exams by students.

In the category “exam design” we achieved the best improvement by: introduction of a pre-review process and its results, improvement of the economical use of existing infrastructures by delegating tasks to the central examining section and the introduction of a binding correction scheme. The overall reliability (Cronbachs alpha) was calculated to be 0.65 and did not change very much.

Figure 3 shows the positive trend towards improvements in the categories “organisational preparations for the realisation of the exam” and “realisation of the exam”. In each of these categories two items remain unimproved. In the category “correction of the exam and documentation” we have improved on three items. Those items cover the analytical calculation before results are published. In addition a protocol is created covering any change in the process of correction.

With the introduced software and the technical help we were able to speed up the exam correction process and are now able to publish the results within at least two weeks. As mentioned before, the students now have the chance to inspect their exams afterwards.

Within the category “review process” both items are fulfilled: there is a documented session reviewing the exam afterwards. Almost immediately any relevant topic will be forwarded to the person that handed in the particular question.

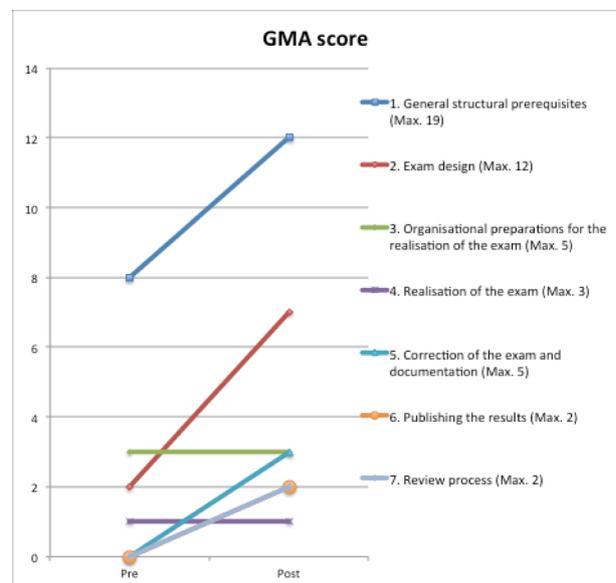


Figure 3: Fulfilled items before and after the introduction of IMSm

## Discussion

The implementation of a new licensing regulation for doctors in 2002 has great impact on the relevance of exams within the faculty. Since 2002 the faculties have

to organise graded exams to benchmark students' course achievements. The grades can be important when it comes to the admission of work places for scholarships. Accordingly the exams must be of good quality in order to allow benchmarking with the faculty and between different universities.

To assist in the creation of higher quality exams the Didaktikzentrum der ETHZ followed by the Ausschuss Prüfungen der GMA and the Kompetenzzentrum Prüfungen Bayern-Württemberg, published guidelines and quality criteria for the whole process. The main criteria for good quality were objectivity, reliability and validity. Effort was put into transparent communication, review process including the integration of external department reviews and analytical calculations. The published guideline consists of 48 items the aim of which is to improve the quality and design of exams effectively and relevantly including current and future exams.

During internal discussions about the possible steps to improve the exam process of our department we oriented ourselves strongly on the GMA guidelines. Main changes were the declaration of one co-worker who is responsible for the entire process and educational objectives. We expect the new review processes will guarantee quality. Today, the department fulfils 30 out of the 48 possible items ('GMA-score', see figure 3).

In the category "general structural prerequisites" and "exam design" we focused on the implementation of a catalogue of educational objectives. This catalogue can be seen as a basis for the creation of questions, but is also helpful for the students as the lectures are geared to it.

We see the introduction of the IMSm system as our biggest challenge ahead. The possibility to use technical help, such as reading the answer sheets automatically, takes a burden of the small department that mainly consists of one 40%-full time position. The technical help allows automatic correction and calculation of analytic values almost instantaneously. IMSm helps to streamline the necessary workflow as it allows the integration of co-workers at each individual step by a well-thought-through user management, especially in the review process (e.g. detection of cueing).

There were no measureable improvements in the categories "organisational preparations for the realisation of the exam" and "realisation of the exam". This is largely the result of the faculty that registers every single student for the exam automatically. Only if a student decides not to take part, he or she must contact the faculty.

In the future we will advise those people watching the students during the exam to document those who try to cheat and the department will work on a general procedure on how to deal with those cases.

Overall, we evaluate the steps to improve the exam process as very positive. It has to be noted that the items being used to calculate the 'GMA-score' do not cover all aspects of an exam, e.g. they do not cover the question of how satisfied the students were with the process. A survey about this question was not yet performed. The

very positive trends toward a better system are an encouraging sign that the process has improved.

We plan to engage co-workers outside the department to assist in the various review processes in the future. Furthermore we think it might be of value to get into contact with other departments and faculties to benefit from each other's question pools.

## Competing interests

The authors declare that they have no competing interests.

## References

1. Bundesministerium für Gesundheit. Approbationsordnung für Ärzte vom 27. Juni 2002. Bundesgesetzbl. 2002;1:2405-2435.
2. Resch F. Universitäre Prüfungen im Licht der neuen ÄAppO. Tagungsband des MFT. Berlin: Medizinischer Fakultätentag; 2008.
3. 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
4. Eugster B, Lutz L. Leitfaden für das Planen, Durchführen und Auswerten von Prüfungen an der ETHZ. Zürich: Didaktikzentrum Eidgenössische Technische Hochschule Zürich; 2004.
5. Gesellschaft für Medizinische Ausbildung, Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für Fakultäts-interne 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>
6. Schelling J, Boeder N, Schelling U, Oberprieler G. Evaluation des "Blockpraktikums Allgemeinmedizin". Überblick, Auswertung und Rückschlüsse an der LMU. *Z Allg Med.* 2010;86:461-465.
7. Brass K, Hochlehner A, Jünger J, Fischer MR, Holzer M. Studienbegleitende Prüfungen mit einem System: ItemManagementSystem für die Medizin. *GMS Z Med.* 2008;25(1):Doc37. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000521.shtml>

### Corresponding author:

Dr. Jörg Schelling

LMU München, Medizinische Klinik und Poliklinik IV, Bereich Allgemeinmedizin, Pettenkoferstraße 8 a, 80336 München, Deutschland  
joerg.schelling@med.uni-muenchen.de

### Please cite as

Boeder N, Holzer M, Schelling J. Die Umsetzung der Leitlinien für Fakultäts-interne Leistungsnachweise am Lehrbereich Allgemeinmedizin der Ludwig-Maximilians-Universität (LMU) München: "MC 2.0" mit dem IMSm – die Prüfung im Wandel. *GMS Z Med Ausbild.* 2012;29(3):Doc42. DOI: 10.3205/zma000812, URN: urn:nbn:de:0183-zma0008123

**This article is freely available from**

<http://www.egms.de/en/journals/zma/2012-29/zma000812.shtml>

**Received:** 2011-06-22

**Revised:** 2011-12-08

**Accepted:** 2012-03-21

**Published:** 2012-05-15

**Copyright**

©2012 Boeder et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share – to copy, distribute and transmit the work, provided the original author and source are credited.