

Effekte und Nachhaltigkeit von Trainingsworkshops für den mündlich-praktischen Teil des M2-Examens

Zusammenfassung

Zielsetzung: Die vielfältigen Limitationen mündlich-praktischer Prüfungen können bekanntermaßen durch spezifische Trainings günstig beeinflusst werden. Die vorliegende Studie analysiert daher anhand eines Fragebogens die in Ulm durchgeführten Trainings für Staatsexamensprüfer, deren Nachhaltigkeit und mögliche Unterschiede zwischen den Teilnehmer-Subgruppen.

Methode: Alle 367 Teilnehmer der bisherigen Ulmer Prüfertrainings (2007 – 20012) wurden per e-Mail angeschrieben. 63 Personen beantworteten die Umfrage, die insgesamt 28 Items zu demografischen Daten und zu Effektivität und Nachhaltigkeit der Trainings enthielt.

Ergebnisse: Es ließen sich 6 wesentliche Trainingseffekte nachweisen (Trainingseffekte, die von mindestens zwei Dritteln der Befragungsteilnehmer auf einer Skala von 1 = „trifft zu“ bis 6 = „trifft nicht zu“ mit den Skalenstufen 1 oder 2 beantwortet wurden; kumulierte Prozentzahl in Klammern):

1. Bewusster Umgang mit Stärken und Schwächen mündlicher Prüfungen (71%)
2. Kenntnis von Faktoren mit Einfluss auf die Reliabilität mündlich-praktischer Prüfungen (76%)
3. Kenntnis von Faktoren mit Einfluss auf die Validität mündlich-praktischer Prüfungen (75%)
4. Erhöhung der Sicherheit bei der Aufgabenkonstruktion (68%)
5. Erhöhung der Sicherheit in Bezug auf Prüfungsregularien (75%)
6. Umsetzung des Konzepts des „strukturierten mündlichen Prüfens“ (Festlegung von Themenbereichen, Vorformulierung von Aufgaben und Erwartungshorizonten sowie Bewertungskriterien) (86%)

Die Antworten der Teilnehmer, deren Training länger als 2 Jahre zurücklag, unterschieden sich nicht signifikant von den Antwortmittelwerten der übrigen Teilnehmer, was für die Nachhaltigkeit der Trainingseffekte spricht.

Teilnehmer ohne relevante Vorerfahrung als Prüfer profitierten signifikant stärker von den Trainings, insbesondere in den Bereichen Stressreduktion, Sicherheit in der Notenfindung und kritischere Notengebung.

Schlussfolgerung: Die Studie liefert Hinweise für nachhaltig positive Effekte der M2-Prüfertrainings, die für die weitere Durchführung der Trainings sprechen, insbesondere als Trainingsangebot für unerfahrene Prüfer. Die Ausweitung der Trainings erscheint daher auch für Prüfer in den M1-Prüfungen sinnvoll.

Schlüsselwörter: Lehrforschung, Staatsexamen Medizin, mündlich-praktische Prüfungen, Prüfertrainings

Hintergrund

In der Humanmedizin sind mündlich-praktische Prüfungen fester Bestandteil der Staatsexamina [2]. Daran wird auch in der aktuell anstehenden Änderung der Approbationsordnung festgehalten [http://www.bundesrat.de/cln_116/nn_8538/DE/presse/pm/2012/069-2012.html], [12].

Die durch Multiple-Choice-Fragen erreichbare hohe Reliabilität und gute rechtliche Überprüfbarkeit der schriftlichen Examina [10], [13] ist bei mündlich-praktischen Prüfungsteilen allerdings eingeschränkt [14], [17]. Unzureichende Reliabilität führt zu mangelnder Validität; diese beide Größen sind aber wesentliche Gütekriterien für Prüfungen [10], [16].

Wolfgang Öchsner^{1,2}

Sandra Geiler³

Markus Huber-Lang⁴

1 Universitätsklinikum Ulm, Abteilung Kardioanästhesiologie, Ulm, Deutschland

2 Universität Ulm, Medizinische Fakultät, Studiendekanat, Ulm, Deutschland

3 Universität Ulm, Medizinische Fakultät, Evaluation und Qualitätssicherung Lehre, Ulm, Deutschland

4 Universitätsklinikum Ulm, Abteilung für Unfallchirurgie, Hand-, Plastische- und Wiederherstellungschirurgie, Ulm, Deutschland

Trotzdem tragen die Noten aus beiden mündlich-praktischen Examensprüfungen (M1 und M2) hälftig zur Gesamtnote für die Ärztliche Prüfung bei; der Anteil der mündlich-praktischen M2-Prüfung beträgt immerhin 33% [2].

In Baden-Württemberg wurde 2007 das „Kompetenznetz Lehre in der Medizin“ als Verbund aller fünf Fakultäten des Landes gegründet [4]. In diesem „Kompetenznetz Lehre“ wurde für den mündlich-praktischen Teil des M2-Examens ein Grundkonzept für Prüferschulungen im Umfang von 8 Unterrichtseinheiten entwickelt. Dieses Konzept bildet auch die Grundlage für die M2-Prüfertrainings der Medizinischen Fakultät Ulm. Ziele und Inhalte sind dabei folgende:

- Sicherer Umgang mit den einschlägigen formalen Bestimmungen
- Beurteilung von Stellenwert, Stärken und Schwächen mündlich-praktischer Prüfungen
- Beherrschen des „strukturierten mündlichen Prüfens“ (Festlegung von Themenbereichen, Vorformulierung von Aufgaben und Erwartungshorizonten sowie Bewertungskriterien)
- Kenntnis von und Umgang mit positiven und negativen Einflussgrößen auf Reliabilität und Validität mündlich-praktischer Prüfungen

Die Trainings beinhalten seminaristische (Experten-Inputs mit interaktiven Elementen, Eigen- und Gruppenarbeit) und praktische Anteile (Prüfungssimulationen mit Peer- und Expertenfeedback).

Zielsetzung

Dass die Limitationen mündlicher Prüfungen durch kriterienorientierte Auswahl der Prüfer (beispielsweise Fachkompetenz mit im Prüfungsteam einheitlich akzeptierten Denkweisen und Handlungsstrategien, spezifische Prüfungskompetenz in Planung und Durchführung mündlicher Prüfungen, konstruktive Teamfähigkeit im kleinen Prüfungsteam), sowie durch Training und kontinuierliches Monitoring der Prüfer günstig beeinflusst werden können, ist aus der Literatur bekannt [14], [17], [18]. Die in vorliegender Studie verfolgte Fragestellung war daher, welche positiven Effekte sich durch die seit 2007 an der Medizinischen Fakultät Ulm durchgeführten M2-Prüfertrainings denn tatsächlich erreichen lassen.

Zudem galt es im Hinblick auf die Nachhaltigkeit der Trainings zu untersuchen, inwieweit positive Effekte auch nach einem längeren Zeitraum noch nachweisbar sind.

Methode

Alle Ulmer Prüfertrainings werden seit 2007 kontinuierlich unter Leitung desselben Trainers standardisiert durchgeführt; der Trainer ist Arzt und verfügt über einen Masterabschluss in der Medizinididaktik (Master of Medical Education).

Bis zum Zeitpunkt der Umfrage hatten 367 Personen an den freiwilligen Trainings teilgenommen. Alle Teilnehmer wurden per E-Mail gebeten, an einer quantitativen Befragung teilzunehmen, die mit dem Umfragetool EvaSys Version 5.0 durchgeführt wurde. Aufgrund der im Universitätsbetrieb üblichen Fluktuation konnten erwartungsgemäß einige der angeschriebenen Personen nicht mehr erreicht werden. Der Rücklauf betrug $n=63$ Personen, davon 32 Personen, die in den akademischen Lehrkrankenhäusern der Medizinischen Fakultät als Prüfer für den mündlichen Teil des M2-Staatsexamens eingesetzt werden, und 31 Personen aus dem Universitätsklinikum Ulm. Die Stichprobe setzt sich sowohl aus Medizinern aus dem operativen Bereich ($n=28$), als auch aus Medizinern nicht-operativer Fachdisziplinen ($n=35$) zusammen.

Die Umfrage umfasste insgesamt 28 Items zu demografischen Daten und zur Effektivität der Trainings, unter Berücksichtigung der im Jahr 2012 in einem Positionspapier der Gesellschaft für Medizinische Ausbildung in Anlehnung an Griffith et al. [1], [6] vorgestellten Kriterien zur Überprüfung der Wirksamkeit medizindidaktischer Maßnahmen [3]. Die Bewertung der Effektivität der Trainings erfolgte über Skalen zur Selbsteinschätzung des eigenen Kompetenzgewinns. Insbesondere waren die Items konzipiert, um auf Seiten der Trainingsteilnehmer Veränderungen von Einstellungen, von Wissen und Fertigkeiten zu erfragen, sowie die Bereitschaft, diese Veränderungen in künftige mündlich-praktische Prüfungen einzubringen.

Alle Daten wurden anonymisiert erfasst und ausgewertet. Es wurden sowohl Lage- und Streuungsparameter berechnet als auch bei angenommener Normalverteilung t-Tests für unabhängige Stichproben durchgeführt. Eine Signifikanz wurde bei $p<0.05$ angenommen.

Ergebnisse

Rücklauf

An der Befragung nahmen 63 von insgesamt 367 angeschriebenen Trainingsteilnehmern teil.

Die Häufigkeit des Einsatzes als M2-Prüfer liegt bei den Personen, die an der Studie teilnahmen ($n=63$), mehrheitlich bei mindestens 2-4 Prüfungstagen pro Jahr.

Wesentliche Trainingseffekte der M2-Trainings

Als wesentliche Trainingseffekte wurden Trainingseffekte definiert, die von mindestens zwei Dritteln der Teilnehmer mit den Skalenstufen 1 oder 2 auf der 6-stufigen Antwortskala (von 1=„trifft zu“ bis 6=„trifft nicht zu“) bewertet wurden.

Die Befragung ergab sechs solcher Effekte (in Klammern die kumulierte Antworthäufigkeit der Skalenstufen 1 und 2, siehe Tabelle 1):

1. Das M2-Training hat dazu beigetragen, dass ich mit den Stärken und Schwächen mündlicher Prüfungen

Tabelle 1: Deskriptive Statistik der gefundenen wesentlichen Trainingseffekte (M2T=M2-Prüfertraining; SD=Standardabweichung)

	Das M2T hat dazu beigetragen, dass ich mit den Stärken und Schwächen mündlicher Prüfungen (im Vergleich zu schriftlichen Prüfungen) bewusster umgehen kann.	Durch das M2T wurden mir Faktoren bewusst, die die Reliabilität (formale Verlässlichkeit: „kommen alle Prüfer zum gleichen Ergebnis?“) mündlicher praktischer Prüfungen beeinflussen.	Durch das M2T wurden mir Faktoren bewusst, die die Validität (Gültigkeit: „Erfasst meine Prüfung die Kompetenz, die ich abprüfen will?“) mündlich-praktischer Prüfungen beeinflussen.	Das M2T hat meine Sicherheit in Bezug auf die Aufgabenkonstruktion in mündlich-praktischen Examenprüfungen erhöht.	Das M2T hat meine Sicherheit in Bezug auf Prüfungsregularien erhöht.	In meinen M2 Prüfungen strebe ich an, das Konzept des „strukturierten mündlichen Prüfens“ umzusetzen (Festlegung von Themenbereichen, Vorformulierung von Aufgaben und Erwartungshorizonten sowie Bewertungskriterien)
n	62	62	63	63	63	62
Mittelwert	3,6	2,3	2,3	2,3	2,1	1,8
Median	3,0	2,0	2,0	2,0	2,0	2,0
SD	1,7	1,3	1,3	1,3	1,1	1,1
Antwortoption 1 und 2	71%	76%	75%	68%	75%	86%

(im Vergleich zu schriftlichen Prüfungen) bewusster umgehen kann (71%)

2. Durch das M2-Training wurden mir Faktoren bewusst, die die Reliabilität (formale Verlässlichkeit: „Kommen alle Prüfer zum gleichen Ergebnis?“) mündlich-praktischer Prüfungen beeinflussen (76%)
3. Durch das M2-Training wurden mir Faktoren bewusst, die die Validität (Gültigkeit: „Erfasst meine Prüfung die Kompetenz, die ich abprüfen will?“) mündlich-praktischer Prüfungen beeinflussen (75%)
4. Das M2-Training hat meine Sicherheit in Bezug auf die Aufgabenkonstruktion in mündlich-praktischen Examensprüfungen erhöht (68%)
5. Das M2-Training hat meine Sicherheit in Bezug auf Prüfungsregularien erhöht (75%)
6. In meinen M2-Prüfungen strebe ich an, das Konzept des „strukturierten mündlichen Prüfens“ umzusetzen (Festlegung von Themenbereichen, Vorformulierung von Aufgaben und Erwartungshorizonten sowie Bewertungskriterien) (86%)

Unterhalb der 2/3 Zustimmung (bei 47%) lag der Bereich „Kompetenzzuwachs als Prüfungsvorsitzender“.

Nachhaltigkeit der Trainingseffekte

Ein wünschenswertes Merkmal didaktischer Trainings besteht in der Nachhaltigkeit der erzielten Effekte. Hofer et al. [7] haben Effekte didaktischer Trainings als nachhaltig definiert, wenn sie nach Verlauf von 1 Jahr noch entsprechend nachweisbar waren. Da Prüfertätigkeiten im Vergleich zu den von Hofer et al. untersuchten Lehrtätigkeiten weniger kontinuierlich ausgeübt werden, haben wir die zeitliche Marke bei 2 Jahren gesetzt; von Nachhaltigkeit haben wir gesprochen, wenn die erzielten Effekte auch 2 Jahre nach Abschluss des M2-Trainings noch nachweisbar waren. Deshalb haben wir mittels eines t-Tests für unabhängige Stichproben überprüft, inwieweit sich die Antworten der Teilnehmer, deren M2-Training länger als 2 Jahre (24 Monate und länger) zurückliegt (n=28), in Bezug auf die o.g. wesentlichen Trainingseffekte von den Antworten der Teilnehmer unterscheiden, deren Teilnahme am M2-Training weniger als 2 Jahre (1-23 Monate) zurückliegt (n=35). Hierbei ergeben sich bei keinem der 6 wesentlichen Trainingseffekte signifikante Unterschiede zwischen den beiden Gruppen.

Mit der methodischen Einschränkung, dass die Erhebung nicht zu zwei Messzeitpunkten erfolgte, bedeutet das, dass die Trainingseffekte als beständig über die Zeit und damit als nachhaltig bewertet werden können.

Analyse der Subgruppen

Um herauszufinden, ob sich unterschiedliche Subgruppen von Prüfern in ihrem Antwortverhalten unterscheiden, überprüften wir mittels eines t-Tests für unabhängige Stichproben eventuell vorhandene Mittelwertunterschiede folgender Teilnehmergruppen auf ihre Signifikanz: Teilnehmer mit relevanter Vorerfahrung als M2-Prüfer versus Prüfer-Neulinge, Teilnehmer aus den akademischen

Lehrkrankenhäusern versus Teilnehmer aus dem Universitätsklinikum Ulm.

Teilnehmer mit relevanter Vorerfahrung als M2-Prüfer versus Prüfer-Neulinge

Als „relevante Vorerfahrung“ in mündlich-praktischen M2-Examensprüfungen wurde eine Vorerfahrung von mindestens 2 Jahren definiert, entsprechend einem Erfahrungsschatz von mindestens 2 - 4 kompletten mündlichen M2-Prüfungen bzw. 4-8 Prüfungstagen. In 2 Fragebögen war dieses Item nicht beantwortet. Für die Subgruppe der Teilnehmer, die noch keine relevante Vorerfahrung als Prüfer mitbrachte (n=29) war die Reduktion des persönlichen Stresslevels als Prüfer in den mündlich-praktischen M2-Prüfungen nach den Trainings signifikant stärker ausgeprägt als in der Subgruppe der Teilnehmer mit einer Vorerfahrung von mindestens 2 Jahren (n=32). (Mittelwert Subgruppe ohne relevante Vorerfahrung: $3,1 \pm 1,6$; Mittelwert Subgruppe mit relevanter Vorerfahrung: $3,9 \pm 1,6$; $t_{(59)} = 2,1$; $p = 0,04$; mittlere Differenz 0,8). Weiterhin war in der Subgruppe ohne relevante Vorerfahrung die subjektiv wahrgenommene Sicherheit in der Notengebung signifikant stärker ausgeprägt als in der Subgruppe mit mindestens 2-jähriger Vorerfahrung. (Mittelwert Subgruppe ohne relevante Vorerfahrung: $2,2 \pm 0,9$; Mittelwert Subgruppe mit relevanter Vorerfahrung: $2,9 \pm 1,4$; $t_{(55)} = 2,1$; $p = 0,04$; mittlere Differenz 0,7). Die Subgruppe ohne entsprechende Vorerfahrung gab außerdem einen signifikant stärkeren Einfluss des M2-Trainings in Bezug auf kritischere Notengebung an. (Mittelwert Subgruppe ohne relevante Vorerfahrung: $2,5 \pm 1,3$; Mittelwert Subgruppe mit relevanter Vorerfahrung: $3,3 \pm 1,4$; $t_{(59)} = 2,1$; $p = 0,04$; mittlere Differenz 0,8) Zusammenfassend kann festgestellt werden, dass für die Ulmer Prüfer-Neulinge zusätzlich zu den oben genannten 6 hauptsächlichen Trainingseffekten signifikante positive Effekte der Prüfertrainings in den Bereichen Stressreduktion, Sicherheit in der Notengebung und kritischere Notengebung bestehen.

Teilnehmer aus den akademischen Lehrkrankenhäusern versus Teilnehmer aus dem Universitätsklinikum Ulm

Da am mündlich-praktischen Teil der Examensprüfungen sowohl Prüfer aus dem Universitätsklinikum Ulm teilnehmen als auch Prüfer aus den assoziierten Akademischen Lehrkrankenhäusern, untersuchten wir, ob sich bei diesen beiden Subgruppen signifikante Unterschiede in der Beantwortung der Umfrage ergaben. Es zeigte sich, dass in keinem der abgefragten Items ein wesentlicher Unterschied zwischen den Teilnehmern aus den Akademischen Lehrkrankenhäusern (n=32) und den Teilnehmern aus dem Universitätsklinikum (n=31) bestand.

Globale Bewertungen der Teilnehmer zum Thema mündlich-praktische Prüfungen im M2-Examen

Die Haltung der Teilnehmer zu den mündlich-praktischen Prüfungsanteilen im M2-Examen stellt sich eindeutig dar: 96% der Teilnehmer sind dafür, mündlich-praktische Prüfungsanteile im M2-Examen beizubehalten. Allerdings votierten 93% der Teilnehmer für eine Splittung der Prüfungsanteile: der schriftliche Teil des M2-Examens sollte bereits vor dem Praktischen Jahr stattfinden, der mündlich-praktische Teil dann erst nach dem PJ. Dies entspricht den Vorgaben im Referentenentwurf zur Novellierung der Approbationsordnung, wie er im Bundesrat abgestimmt wurde [http://www.bundesrat.de/clin_228/nn_1934482/SharedDocs/Drucksachen/2010/0001-0100/96-10,templateId=raw,property=publicationFile.pdf/96-10.pdf]

Die überwiegende Anzahl der Teilnehmer (86%) sprach eine grundsätzliche Empfehlung zur Teilnahme an einem Prüfertraining zu Beginn der M2-Prüfertätigkeit aus.

Diskussion und Schlussfolgerungen

Die beiden Hauptfragestellungen der Studie nach den Effekten der Ulmer M2-Prüfertrainings und deren Nachhaltigkeit können positiv beantwortet werden: Es zeigen sich 6 wesentliche Trainingseffekte, die mit der Zielsetzung der Trainings gut übereinstimmen und die zur Qualitätsverbesserung mündlich-praktischer Prüfungen beitragen.

Für Prüfer ohne Vorerfahrung bestehen zusätzlich signifikante positive Effekte in den Bereichen Stressreduktion, Sicherheit in der Notengebung und kritischere Notengebung.

Interessant ist die Tatsache, dass sich die positiven Bewertungen der Teilnehmer auch dann nicht abschwächen, wenn das M2-Training schon mehr als zwei Jahre zurückliegt. Das spricht für die Nachhaltigkeit der Trainingseffekte, ähnlich wie es Hofer et al. für die Düsseldorfer Didaktik-Trainings nach einem Jahr Latenzzeit nachweisen konnten [7].

Deutlich geringer fiel in der Selbsteinschätzung der Benefit der Teilnehmer als Prüfungsvorsitzende aus. Da das Thema „Prüfungsvorsitz“ jedoch kein ausgesprochener Schwerpunkt des Trainings war, ist dies nicht überraschend; ein Ausbau des Trainings in diesem Bereich ist hier zukünftig zu erwägen.

Annähernd alle Teilnehmer sprachen sich dafür aus, mündlich-praktische Prüfungsanteile im M2-Examen beizubehalten, wenn auch zeitlich vom schriftlichen Teil des Examens abgekoppelt, und zu Beginn der Prüfertätigkeit ein entsprechendes M2-Prüfertraining zu absolvieren. Das betont die subjektiv wahrgenommene Wertigkeit mündlich-praktischer Prüfungen bei der Erfassung und Bewertung medizinischer Kompetenzen, in Ergänzung zu den schriftlichen Prüfungsanteilen im M2-Staatsexamen. Wenn auch die Gründe für die abgegebenen Voten nicht

abgefragt wurden, so muss doch davon ausgegangen werden, dass die Stärken dieses spezifischen Prüfungsformats, wie die Möglichkeit zur direkten Überprüfung praktischer Fertigkeiten sowie zur direkten Entwicklung von Analyse- und Problemlöse-Algorithmen, einerseits als wesentliche Bausteine der Kompetenzüberprüfung angesehen werden, andererseits Planung und Durchführung dieses Prüfungsformats als anspruchsvoll und trainingsbedürftig betrachtet werden. Letzteres könnte auch einer der Gründe für die noch immer relativ geringe Verbreitung mündlich-praktischer Prüfungen bei den fakultätsinternen Leistungsnachweisen sein [9].

Ob die Teilnehmer aus dem Universitätsklinikum oder aus einem der Akademischen Lehrkrankenhäuser kamen, spielte im Antwortverhalten keine Rolle.

In der Konsequenz ist die Weiterführung der M2-Prüfertrainings sinnvoll, insbesondere da durch die im Universitätsbetrieb übliche hohe Fluktuation auch immer wieder „unerfahrene“ Prüfer rekrutiert werden müssen. Äußerst sinnvoll erschien in diesem Zusammenhang sogar die Ausweitung des Angebots auch für die vorklinischen Prüfer der mündlichen Anteile der M1-Examensprüfung, entweder auf Landesebene im Rahmen des Kompetenznetz Lehre in der Medizin in Baden-Württemberg oder auf Fakultätsebene.

Einschränkend muss erwähnt werden, dass trotz der relativ großen Zielgruppe von über 300 Absolventen der Ulmer M2-Prüfertrainings nur etwas über 60 Personen auf unsere Online-Umfrage geantwortet haben. Dies erklärt sich einerseits aus einer mittlerweile bekannten Umfrage-Müdigkeit bei steigender Anzahl unterschiedlichster Online-Umfragen [15], andererseits aus der Tatsache, dass die Fluktuation der universitären Prüfer durch Stellenwechsel relativ hoch ist. Eine eingeschränkte Repräsentativität kann zu Verzerrungs-Effekten führen, beispielsweise durch die vorwiegende Teilnahme besonders positiv oder negativ engagierter Personen, d.h. von Personen, die entweder dem Training oder der mündlichen M2-Prüfung oder beidem überdurchschnittlich zustimmend oder ablehnend gegenüberstehen. Hier hätte eventuell eine landesweit durchgeführte Umfrage zu einer (absolut gesehen) quantitativen Steigerung geführt, allerdings auch hier wieder mit dem Risiko, vor allem überdurchschnittlich engagierte Prüfer (im positiven oder im negativen Sinn) als Teilnehmer zu gewinnen. Qualitativ hätte dies, trotz höherer Validität, außerdem die Einschränkung bedeutet, alle standort- und trainerspezifischen Besonderheiten der konkreten Trainings-Ausgestaltung zu vermischen und auf die Vorteile der sehr homogenen, standardisierten Durchführung zu verzichten.

Eine weitere Einschränkung ergibt sich aus der Tatsache, dass es sich, zumindest bei den abgefragten Trainingseffekten, jeweils um Selbsteinschätzungen der Teilnehmer und nicht um eine Interventions- oder Vergleichsstudie mit entsprechender Kontrollgruppe handelt. Die Güte von Selbsteinschätzungen wird in der Literatur in verschiedenen Studienszenarien sehr unterschiedlich beurteilt, wird aber durch Kompetenztrainings mit entsprechendem Expertenfeedback (wie in den M2-Trainings durchgeführt)

im Hinblick auf die Übereinstimmung von Selbsteinschätzung und Realität günstig beeinflusst [5], [8], [11]. Da es aufgrund des hohen Impacts der mündlich-praktischen Note auf die Gesamtnote der Ärztlichen Prüfung nicht wünschenswert sein kann, im „Ernstfall Examensprüfung“ untrainierte und/oder unerfahrene Prüfer als Ausgangs- oder Vergleichsgruppe einzusetzen, wäre ein „pre-/post-Design“ letztlich auch nur unter Simulationsbedingungen realisierbar. Zusammenfassend liefert die vorliegende Studie Hinweise, dass sich durch Prüfertrainings eine Reihe von positiven Effekten für mündlich-praktische Prüfungen erzielen lassen, die sich auch durch eine gewisse Nachhaltigkeit auszeichnen. Am stärksten profitieren dabei Prüfer zu Beginn ihrer Prüferkarriere.

Interessenkonflikt

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

Literatur

- Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach*. 2001;23(2):164-170. DOI: 10.1080/0142150020031084
- Bundesministerium für Gesundheit. Ärztliche Approbationsordnung (ÄAppO) in der Fassung vom 27.6.2002. *Bundesgesetzbl*. 2002;l:2405.
- Fabry G, Lammerding-Köppel M, Hofer M, Ochsendorf F, Schirlo C, Breckwoldt J. Hochschuldidaktische Qualifizierung in der Medizin IV: Messung von Wirksamkeit und Erfolg medizindidaktischer Qualifizierungsangebote: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg. *GMS Z Med Ausbild*. 2010;27(4):Doc62. DOI: 10.3205/zma000699
- Fegert JM, Obertacke U, Resch F, Hilzenbecher M. Die Qualität der Lehre nicht dem Zufall überlassen. *Dtsch Arztebl*. 2009;106(7):A290-A291.
- Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*. 1991;66(12):762-769. DOI: 10.1097/00001888-199112000-00012
- Griffith III CH, Georgesen JC, Wilson JF. Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. *Acad Med*. 2000;75(10 Suppl):S62-S64. DOI: 10.1097/00001888-200010001-00020
- Hofer M, Jansen M, Soboll S. Effektive Didaktiktrainings für Dozenten der Medizin. *GMS Z Med Ausbild*. 2005;22(1):Doc07. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000007.shtml>
- Jünger J, Schellberg D, Nikendei C. Subjektive Kompetenzeinschätzung von Studierenden und ihre Leistung im OSCE. *GMS Z Med Ausbild*. 2006;23(3):Doc51. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000270.shtml>
- 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
- Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild*. 2006;23(3):Doc53. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000272.shtml>
- Nagler M, Feller S, Beyeler C. Retrospektive Anpassung der Selbsteinschätzung ärztlicher Kompetenzen – Beachtenswert bei der Evaluation praktischer Weiterbildungskurse. *GMS Z Med Ausbild*. 2012;29(3):Doc45. DOI: 10.3205/zma000815
- Richter-Kuhlmann EA. Medizinstudium: Hammerexamen soll bald Geschichte sein. *Dtsch Arztebl*. 2011;108(40):A-2061/B-1757/C-1741.
- Schulze J, Drolshagen S. Format und Durchführung schriftlicher Prüfungen. *GMS Z Med Ausbild*. 2006;23(3):Doc44. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000263.shtml>
- Seyfarth M, Reincke M, Seyfarth J, Ring J, Fischer MR. Neue ärztliche Approbationsordnung und Notengebung beim Zweiten Staatsexamen. Eine Untersuchung an zwei bayerischen medizinischen Fakultäten. *Dtsch Arztebl Int*. 2010;107(28-29):500-504.
- Sheehan K. E-mail Survey Response Rates: A Review. *JCMC*. 2001;6(2). Zugänglich unter/available from: <http://jcmc.indiana.edu/vol6/issue2/sheehan.html>
- Van der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ*. 1996;1(1):41-47. DOI: 10.1007/BF00596229
- Wakeford R, Southgate L, Wass V. Improving oral examinations: selecting, training and monitoring examiners for the MRCGP. *BMJ*. 1995;311(7010):931-935. DOI: 10.1136/bmj.311.7010.931
- Wass V, Wakeford R, Neighbour R, Vleuten CV. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ*. 2003;37(2):126-131. DOI: 10.1046/j.1365-2923.2003.01417.x

Korrespondenzadresse:

Dr. med. Wolfgang Öchsner, MME (Bern)
 Universitätsklinikum Ulm, Abteilung
 Kardioanästhesiologie, Albert-Einstein-Allee 23, 89081
 Ulm, Deutschland, Tel.: +49 (0)731/500-55408
wolfgang.oechsner@uniklinik-ulm.de

Bitte zitieren als

Öchsner W, Geiler S, Huber-Lang M. Effekte und Nachhaltigkeit von Trainingsworkshops für den mündlich-praktischen Teil des M2-Examens. *GMS Z Med Ausbild*. 2013;30(3):Doc36. DOI: 10.3205/zma000879, URN: urn:nbn:de:0183-zma0008795

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2013-30/zma000879.shtml>

Eingereicht: 27.07.2012

Überarbeitet: 13.01.2013

Angenommen: 02.05.2013

Veröffentlicht: 15.08.2013

Copyright

©2013 Öchsner 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.

Effects and Sustainability of Trainings for the Oral and Practical Part of the German Final Exam in Medicine

Abstract

Study Goals: It is known that the manifold limitations of oral and practical examinations can be improved by specific training. With the help of an online survey, our present study analyzes the effects that can be achieved by the training conducted at the University of Ulm for examiners in the final medical examination, the long-lasting impact of the training, and differences among participant subgroups.

Method: All 367 participants in the training at Ulm (2007- 2012) were contacted via email. Sixty-three persons responded to the survey that included 28 items concerning demographic data, effectiveness, and sustainability.

Results: Six main effects of the training were identified (meaning effects rated with a grade of 1 or 2 on a 6-point scale by two thirds of the participants, with 1="applicable" and 6="not applicable"; cumulated percentage of answers of 1 or 2 in parentheses):

1. Conscious handling of strengths and weaknesses of oral examinations (71%),
2. Knowledge of factors contributing to the reliability of oral/practical examinations (76%),
3. Knowledge of factors contributing to the validity of oral/practical examinations (75%),
4. Improvement of competence in task construction (68%),
5. Improvement of competence in respect to examination formalities (75%),
6. Implementation of the concept of "structured oral examinations" (a priori planning of examination subjects, tasks, levels of expectation and grading criteria) (86%).

The responses of participants trained more than two years ago were not significantly different from the answers given by recently trained persons. This is an argument for the sustainability of the training effects. Furthermore, participants without relevant prior experience in oral/practical examinations profited significantly more from the trainings, especially in the areas of stress reduction, confidence in grading, and competence in critical discrimination of grading.

Conclusion: The positive and sustained effects of the examiner training argue for continuing the training program, especially for inexperienced examiners. Expansion of the successful training program to include the first medical exam should be considered.

Keywords: Medical Education, Final Exam, Oral Examinations, Examiner Training

Background

In the study of human medicine, oral and practical examinations are an integral part of state examinations [2]. This will continue to be adhered in the current pending changes to the licensing regulations [http://www.bundesrat.de/cln_116/nn_8538/DE/presse/pm/

2012/069-2012.html] [12]. The high reliability and good legal verifiability of the written exams through the use of multiple-choice questions (10, 13) is, however, limited in terms of oral and practical exams [14], [17]. Insufficient reliability leads to a lack of validity, even though these two factors are essential criteria for ensuring the good quality of examinations [10], [16]. Despite this, the grades for both oral exams (M1 and M2) comprise half of the overall grade for the medical exam-

Wolfgang Öchsner^{1,2}

Sandra Geiler³

Markus Huber-Lang⁴

1 University Hospital Ulm, Department for Cardiac Anaesthesiology, Ulm, Germany

2 University of Ulm, Medical Faculty, Office of the Dean of Education, Ulm, Germany

3 University of Ulm, Department for Evaluation and Quality Management in Medical Education, Ulm, Germany

4 University Hospital Ulm, Clinic for Trauma Surgery, Hand Surgery, Plastic Surgery and Reconstructive Surgery, Ulm, Germany

ination; the percentage assigned to the M2 oral exam is after all 33% [2].

In Baden-Württemberg the Kompetenznetz Lehre in der Medizin (Competency Network for Teaching Medicine) was founded in 2007 as an alliance of all five state university medical schools [4]. In this Competency Network a basic concept encompassing eight instructional units was developed to train examiners for the oral part of the M2 exam. This concept also forms the basis for the M2 examiner training program at the medical school in Ulm. The aims and content are as follows:

- Confident handling of the relevant formal provisions,
- Evaluation of the significance as well as the strengths and weaknesses of oral and practical exams,
- Mastery of “structured oral examinations” (predetermination of topics, tasks, levels of expectation and grading criteria),
- Awareness and handling of positive and negative factors influencing the reliability and validity of oral and practical exams.

The training sessions took the form of seminars (expert input with interactive elements, individual and group work), along with offering practical elements (test simulations with peer and expert feedback).

Study Goals

It is known from the literature that the limitations of oral examinations can be favorably influenced not only by criteria-oriented selection of the examiners (e.g. subject knowledge coupled with uniformly accepted methods of thought and strategy among co-examiners, specific competence in designing and conducting oral examinations, and constructive team skills in small groups of co-examiners), but also through training and continually monitoring the examiners [14], [17], [18]. Thus, the issue investigated in this study was to identify which positive effects are actually achieved by the M2 examiner training that has been conducted since 2007 at the medical school in Ulm. In terms of the sustainability of the training, a second aim was to analyze whether the positive effects could still be determined after a longer period of time.

Methods

Since 2007, all of the examiner training sessions in Ulm have been held under the leadership of the same trainer; this trainer is a physician and holds a master’s degree in Medical Education.

Up until the time of the survey, 367 people had participated in the voluntary training program. All participants were requested by email to take part in a quantitative survey, which was administered using EvaSys version 5.0. Due to the normal fluctuation seen at universities, it was expected that some of the former participants could no longer be reached. The response rate was n=63 per-

sons, of which 32 are deployed as examiners for the oral part of the M2 state examination at the teaching hospitals of the medical school, and 31 persons from the Ulm University Clinic. The random sample was comprised of both physicians working in the operative field (n=28) and those from non-operative disciplines (n=35).

The survey covered a total of 28 items on demographic information and the effectiveness of the training, taking the criteria introduced in a 2012 position paper by the Gesellschaft für Medizinische Ausbildung based on Griffith et al. [1], [6] for verifying the effectiveness of educational measures in medicine [3]. The evaluation of the training’s effectiveness was done using a scale for self-assessment of the acquired competency. The items were specifically designed to elicit information not only on changes in the participants’ attitudes, knowledge and skills, but also on their willingness to include these changes in future oral and practical examinations.

All data was collected and analyzed anonymously. Location and dispersion parameters were calculated, and presuming normal distribution, t tests were performed for independent random samples. A significance of $p < 0.05$ was assumed.

Results

Rate of response

Sixty-three of a total of 367 participants contacted responded to the survey. The frequency of being deployed as an M2 examiner was at least 2-4 examination days per year for the majority of those who took part in the survey (n=63).

Basic effects of M2 training

A basic effect of training was defined as an effect which was rated by at least two-thirds of the participants with a 1 or 2 on a six-point scale (with 1=applicable and 6=not applicable). The survey yielded six such effects; the cumulated frequency of 1 or 2 as the response is given in parentheses (see table 1):

1. The M2 training helped me to be able to deal more consciously with the strengths and weaknesses of oral examinations (compared to written exams). (71%)
2. As a result of the M2 training, I have become aware of factors that influence the reliability of oral/practical examinations (formal reliability: “Do all examiners come to the same conclusions?”). (76%)
3. As a result of the M2 training, I have become aware of factors that influence the validity of oral/practical examinations (validity: “Does my exam cover the skills I wish to test for?”). (75%)
4. The M2 training has increased my confidence in regard to designing tasks for oral/practical testing. (68%)
5. The M2 training has increased my confidence in respect to examination formalities. (75%)

Table 1: Descriptive statistics on the basic training effects detected (M2T=M2 examiner training; SD=standard deviation)

	The M2T helped me to be able to deal more consciously with the strengths and weaknesses of oral exams (compared to written exams).	As a result of M2T, I have become aware of factors that influence the reliability of oral/practical examinations (formal reliability: "Do all examiners come to the same conclusions?").	As a result of the M2T, I have become aware of factors that influence the validity of oral/practical exams (validity: "Does my exam cover the skills I wish to test for?").	The M2T has increased my confidence in regard to designing tasks for oral and practical testing.	The M2T has increased my confidence in respect to examination formalities.	When conducting M2 exams, I strive to implement the concept of "structured oral examinations" (predefinition of topics, tasks, levels of expectation and grading criteria).
n	62	62	63	63	63	62
Mean	3.6	2.3	2.3	2.3	2.1	1.8
Median	3.0	2.0	2.0	2.0	2.0	2.0
SD	1.7	1.3	1.3	1.3	1.1	1.1
Rating of 1 or 2	71%	76%	75%	68%	75%	86%

6. When conducting M2 examinations, I strive to implement the concept of “structured oral examinations” (prior definition of topics, tasks, levels of expectation and grading criteria). (86%)

Falling below the two-thirds majority (at 47%) was the aspect concerning increased confidence in the role of examination committee chairman.

Sustainability of the training effects

A desirable characteristic of didactic training is the sustainability of the achieved effects. Hofer et al. [7] have defined the effects of didactic training as sustained if they are still measurable after one year. Since the activities as examiner are exercised much less regularly than the teaching activities studied by Hofer et al., the authors here have chosen a longer timeframe of two years; we have defined sustainability to be when the attained effects can still be detected two years after the completion of the M2 training. Applying a t test for independent random samples, we therefore analyzed the extent to which the responses of the participants whose M2 training took place over two years ago (24 months or longer, n=28) differed from the responses of participants whose M2 training was received less than 2 years ago (1-23 months; n=35). No significant differences arose between the two groups in terms of the six basic training effects. With the methodological limitation that the collection of data did not take place at two points in time, this means that the training effects can be assessed as existing during this period of time and therefore viewed as sustained.

Analysis of the subgroups

In order to discern whether various subgroups of examiners differed in their responses, we performed a t test for independent random samples to analyze the significance of any potential difference of means in the following participant subgroups: participants with relevant prior experience as M2 examiner versus new examiners, and participants from the university teaching hospitals versus participants from the Ulm University Clinic.

Participants with relevant prior experience as M2 examiners versus new examiners

“Relevant prior experience” was defined as experience in administering M2 oral exams that amounted to at least two years and corresponded to a minimum of two to four complete M2 oral examinations or four to eight days of testing. On two of the surveys this item was left unanswered. For the subgroup of participants who did not possess any relevant prior experience as examiner (n=29), the reduction of personal stress levels as examiner in the M2 oral exams was significantly more pronounced than in the subgroup of participants with at least two years of previous experience (n=32). (Mean value for the subgroup without relevant prior experience: 3.1±1.6; mean value for the subgroup with relevant prior experi-

ence: 3.9±1.6; $t_{(59)}=2.1$; $p=0.04$; mean difference 0.8). Moreover, the subjectively perceived level of confidence in assigning grades was significantly more pronounced in the subgroup without relevant prior experience than in the subgroup with a minimum of two years previous experience. (Mean value for the subgroup without relevant prior experience: 2.2±0.9; mean value for the subgroup with relevant prior experience: 2.9±1.4; $t_{(55)}=2.1$; $p=0.04$; mean difference 0.7). In addition, the subgroup without prior experience indicated a significantly stronger influence of the M2 training sessions in regard to critical discrimination in grading. (Mean value for the subgroup without relevant prior experience: 2.5±1.3; mean value for subgroup with relevant prior experience: 3.3±1.4; $t_{(59)}=2.1$; $p=0.04$; mean difference 0.8). From this it can be concluded that in addition to the six main training effects identified above, significant positive effects of the examiner training exist in the areas of stress reduction, confidence in grading, and critical discrimination of grade assignment for the new examiners in Ulm.

Participants from the university teaching hospitals versus participants from the Ulm University Clinic

Since examiners from both the Ulm University Clinic and the affiliated university teaching hospitals participate in administering the oral part of the medical exams, we investigated whether or not any significant differences between these two groups arose in terms of the responses given on the survey. It was discovered that none of the surveyed items showed a significant level of difference between the participants from the academic teaching hospitals (n=32) and the participants from the Ulm University Clinic (n=31).

Global assessment of M2 oral exams by the participants

The attitude of the participants toward the oral sections of the M2 exams is clear: 96% of participants are in favor of retaining oral components in the M2 examination. However, 93% of the participants voted to split the examination parts, with the written section of the M2 exam taking place before the fifth year of medical study (“Praktisches Jahr”) and the oral section afterward. This corresponds to the contents of the legislative draft to amend the licensing regulations as it was agreed upon by the Federal Council [http://www.bundesrat.de/cln_228/nn_1934482/SharedDocs/Drucksachen/2010/0001-0100/96-10,templateId=raw,property=publicationFile.pdf/96-10.pdf].

The overwhelming majority of the participants (86%) recommended participation in the examiner training upon taking up the responsibility of an M2 examiner.

Discussion and Conclusions

The two main questions addressed by this study concerning the effects of the M2 examiner training in Ulm and its sustainability can be answered positively: six basic effects were detected that comply closely with the training goals and contribute to improvement in the quality of oral examinations.

For examiners without relevant prior experience there are additional significant positive effects regarding stress reduction, confidence in grading, and critical discrimination of grade assignment.

Of interest is the fact that the positive responses of the participants do not taper off even if the M2 training occurred more than two years ago. This speaks for the sustainability of the training effects, similar to Hofer et al. finding sustainability in the didactic training conducted in Düsseldorf after a latent period of one year [7].

The benefit to the participants as examination committee chairman was rated much lower in the self-assessments. This is not surprising since the topic of “chief examiner” was not a defined focus of the training. Expansion of the training program to include this aspect should be considered in future. Almost all participants were in favor of retaining oral exam components as part of the M2 examinations, even if these were to be administered at a different time point than the written part of the examination. Another point of near consensus concerned participation in M2 examiner training when first taking up responsibilities as examiner. This emphasizes the subjectively perceived value of oral exams in the measuring and assessment of medical skills as a supplement to the written sections of the M2 state examinations. Even though the participants were not questioned as to why they cast their votes this way, it still must be assumed that not only the strengths of this specific testing format (such as the possibility to directly test practical skills, as well as directly develop analytical and algorithmic problem-solving) are viewed as basic elements for testing competence, but also the design and conduction of this kind of testing are considered as demanding and requiring training. This final aspect could also be one of the reasons for the still relatively low prevalence of oral exams among in the school-specific graded assessments [9].

Whether the participants came from the University Clinic or one of the teaching hospitals had no influence on how the survey questions were answered. Consequently, the continuation of the M2 examiner training program is worthwhile, particularly since the normally high turnover in staff at universities often makes it necessary to recruit “inexperienced” examiners. In response to this, expanding the training program to also include preclinical examiners for the oral parts of the M1 examination, either on the state level within the scope of the Competency Network for Teaching Medicine in Baden-Württemberg or at the medical school level, would appear to be the most sensible approach.

It must be mentioned as a limitation that despite the relatively large target group of over 300 M2 training program

graduates, only a little over 60 responded to our online survey. This can be explained in part by what has been identified as “survey exhaustion” resulting from an increasing number of various online questionnaires [15], and also by the fact that the fluctuation among university examiners is relatively high due to job changes. A limited representativity can lead to distortion, for instance through the participation of persons with predominantly positive or negative impressions, meaning those who agree or disagree with either the training or the oral M2 examination, or both of them, more strongly than the average. In respect to this, a survey conducted at the national level may have led to a quantitative increase (when considered absolutely); however, even in this case the risk is present of primarily acquiring overly-engaged examiners (in a positive or negative sense). Qualitatively, this would have, despite high validity, also indicated the limitations of mixing all unique aspects of the concrete training programs specific to each location and trainer, as well as waiving the advantages of very homogenous, standardized study implementation.

A further limitation results from the fact that, at least in terms of the training effects surveyed, the focus is on self-assessment by the participants and not on conducting an intervention or comparative study with the corresponding control groups. Although the quality of self-evaluations in various study scenarios is assessed very differently in the literature, it is favorably influenced by skills training paired with expert feedback (such as in the M2 training sessions here) in terms of conformity between self-evaluation and reality [5], [8], [11].

Due to the high impact of the grade for the oral exam in respect to the overall grade on the medical examination, it cannot be seen as desirable to deploy untrained and/or inexperienced examiners in experimental or control groups in a real case of test taking. As a result, it would only be possible to realize a pre-post study design under simulated conditions.

In conclusion, the present study provides indications that a series of positive effects for oral examinations can be achieved through training the examiners and that these effects demonstrate a certain level of sustainability. Those who profit most strongly are the examiners who are just beginning their activities as examiners.

Competing interests

The Authors declare that they have no competing interests.

References

1. Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach*. 2001;23(2):164-170. DOI: 10.1080/0142150020031084

2. Bundesministerium für Gesundheit. Ärztliche Approbationsordnung (ÄAppO) in der Fassung vom 27.6.2002. Bundesgesetzbl. 2002;I:2405.
3. Fabry G, Lammerding-Köppel M, Hofer M, Ochsendorf F, Schirlo C, Breckwoldt J. Hochschuldidaktische Qualifizierung in der Medizin IV: Messung von Wirksamkeit und Erfolg medizindidaktischer Qualifizierungsangebote: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg. *GMS Z Med Ausbild.* 2010;27(4):Doc62. DOI: 10.3205/zma000699
4. Fegert JM, Obertacke U, Resch F, Hilzenbecher M. Die Qualität der Lehre nicht dem Zufall überlassen. *Dtsch Arztebl.* 2009;106(7):A290-A291.
5. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med.* 1991;66(12):762-769. DOI: 10.1097/00001888-199112000-00012
6. Griffith III CH, Georgesen JC, Wilson JF. Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. *Acad Med.* 2000;75(10 Suppl):S62-S64. DOI: 10.1097/00001888-200010001-00020
7. Hofer M, Jansen M, Soboll S. Effektive Didaktiktrainings für Dozenten der Medizin. *GMS Z Med Ausbild.* 2005;22(1):Doc07. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000007.shtml>
8. Jünger J, Schellberg D, Nikendei C. Subjektive Kompetenzeinschätzung von Studierenden und ihre Leistung im OSCE. *GMS Z Med Ausbild.* 2006;23(3):Doc51. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000270.shtml>
9. 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
10. Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild.* 2006;23(3):Doc53. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000272.shtml>
11. Nagler M, Feller S, Beyeler C. Retrospektive Anpassung der Selbsteinschätzung ärztlicher Kompetenzen – Beachtenswert bei der Evaluation praktischer Weiterbildungskurse. *GMS Z Med Ausbild.* 2012;29(3):Doc45. DOI: 10.3205/zma000815
12. Richter-Kuhlmann EA. Medizinstudium: Hammerexamen soll bald Geschichte sein. *Dtsch Arztebl.* 2011;108(40):A-2061/B-1757/C-1741.
13. Schulze J, Drolshagen S. Format und Durchführung schriftlicher Prüfungen. *GMS Z Med Ausbild.* 2006;23(3):Doc44. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000263.shtml>
14. Seyfarth M, Reincke M, Seyfarth J, Ring J, Fischer MR. Neue ärztliche Approbationsordnung und Notengebung beim Zweiten Staatsexamen. Eine Untersuchung an zwei bayerischen medizinischen Fakultäten. *Dtsch Arztebl Int.* 2010;107(28-29):500-504.
15. Sheehan K. E-mail Survey Response Rates: A Review. *JCMC.* 2001;6(2). Zugänglich unter/available from: <http://jcmc.indiana.edu/vol6/issue2/sheehan.html>
16. Van der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ.* 1996;1(1):41-47. DOI: 10.1007/BF00596229
17. Wakeford R, Southgate L, Wass V. Improving oral examinations: selecting, training and monitoring examiners for the MRCGP. *BMJ.* 1995;311(7010):931-935. DOI: 10.1136/bmj.311.7010.931
18. Wass V, Wakeford R, Neighbour R, Vleuten CV. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ.* 2003;37(2):126-131. DOI: 10.1046/j.1365-2923.2003.01417.x

Corresponding author:

Dr. med. Wolfgang Öchsner, MME (Bern)
University Hospital Ulm, Department for Cardiac
Anaesthesiology, Albert-Einstein-Allee 23, 89081 Ulm,
Germany, Tel.: +49 (0)731/500-55408
wolfgang.oechsner@uniklinik-ulm.de

Please cite as

Öchsner W, Geiler S, Huber-Lang M. Effekte und Nachhaltigkeit von Trainingsworkshops für den mündlich-praktischen Teil des M2-Examens. *GMS Z Med Ausbild.* 2013;30(3):Doc36. DOI: 10.3205/zma000879, URN: urn:nbn:de:0183-zma000879

This article is freely available from

<http://www.egms.de/en/journals/zma/2013-30/zma000879.shtml>

Received: 2012-07-27

Revised: 2013-01-13

Accepted: 2013-05-02

Published: 2013-08-15

Copyright

©2013 Öchsner 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.