

Visions and reality: the idea of competence-oriented assessment for German medical students is not yet realised in licensing examinations

Abstract

Objective: Competence orientation, often based on the CanMEDS model, has become an important goal for modern curricula in medical education. The National Competence Based Catalogue of Learning Objectives for Undergraduate Medical Education (NKLM) has been adopted in Germany. However, it is currently unknown whether the vision of competence orientation has also reached the licensing examination procedures.

Methods: Therefore, a prospective, descriptive, single-centre, exemplary study design was applied to evaluate 4051 questions/tasks (from 28 examiners at 7 two-day licensing oral-practical exams) for undergraduate medical students at the University of Ulm. The oral and practical questions/tasks as well as the real bedside assessment were assigned to specific competence roles (NKLM section I), categories (NKLM section II) and taxonomy levels of learning domains.

Results: Numerous questions/tasks were set per candidate (day 1/2: $70 \pm 24 / 86 \pm 19$ questions) in the licensing oral-practical exam. Competence roles beyond the “medical expert” were scarcely considered. Furthermore, practical and communication skills at the bedside were hardly addressed (less than 3/15 min). Strikingly, there was a significant predominance of questions with a low-level taxonomy.

Conclusions: The data indicate a misalignment of competence-oriented frameworks and the “real world” licensing practical-oral medical exam, which needs improvement in both evaluation and education processes.

Keywords: medical competence, high-stakes exam, alignment, roles, NKLM

Markus Huber-Lang¹
Annette Palmer¹
Claudia Grab²
Anja Boeckers³
Tobias Maria Boeckers³
Wolfgang Oechsner⁴

1 University Hospital of Ulm, Medical School, Institute for Clinical- and Experimental Trauma-Immunology, Ulm, Germany

2 University of Ulm, Medical Faculty, Dean's Office, Ulm, Germany

3 University of Ulm, Institute of Anatomy and Cell Biology, Ulm, Germany

4 University Hospital of Ulm, Department of Cardiac Anaesthesiology, Ulm, Germany

1. Introduction

1.1. Background

The classical concept of learning-objective orientated education, which over many decades has represented the conceptual basis for higher education, has been increasingly broadened by a much more complex concept of competence based education. Since their first publication in the 1990s, the 7 CanMed roles (medical expert, communicator, collaborator, manager, health advocate, scholar, and professional) have served worldwide as the “archetype” for medical competence orientation [1], [2], [3]. With reference to the CanMEDS framework, the National Competence-Based Learning Objectives for Undergraduate Medical Education (NKLM) has been recently introduced throughout Germany as a framework for competence-based medical education on the basis of the national licence order [4], [5].

To assess the grade of clinical competence finally attained by medical students, oral-practical examinations have been used as a major instrument for decades [6], [7], [8], [9]. In general, classical oral-practical exams remain subject to serious criticism based not only on low-rated validity and reliability [6], [10], [11], but also on difficulties in addressing the whole spectrum of “medical competences”. In contrast, positive examples for well-structured and competence oriented oral examinations are only occasionally described [12].

Currently, there are no direct observation data of the licensing oral-practical examinations in Germany available in regard to the degree of their competence orientation and to the taxonomy level of the questions and tasks used to assess the students.

1.2. The German Licensing Exam

Though recently changed, at the time period of our study the licensing exam for medicine in Germany consisted of

two sections. Both of them, section I after 2 years and section II after 6 years, contained a written and an oral / oral-practical part. The change recently performed associated the written part of the exam section II to an earlier study phase (after 5 years) and declared the oral-practical part as exam section III after 6 years. However, the actual process of the oral-practical part, lasting two days, has not been affected by this change. The concrete details of the examination process as described in the following article are, within the framework of the German National License Order of Germany, regulated by the responsible State Examination Office of Baden-Württemberg.

During the two days of the oral-practical examination part, the examiners are given an ample scope within the regulatory framework in regard to the content of the questions and tasks they plan to set the candidates [13], [14]. On day 1, after a 3-h time slot for each candidate to investigate and document a real patient's case at the bedside, the candidate (one of up to four students) has approximately 15 minutes to present the patient's case to the three examiners, to demonstrate examination techniques (on the examiners' demand) and to answer patient-oriented questions (see Figure 1). Off-bedside, an interrogation of each candidate follows (7-11 minutes per examiner). On day 2, each candidate is interrogated by each examiner for 11-15 minutes. The examination is performed by the same four examiners, who are permitted to ask questions or to set tasks from any discipline.

Each student is finally scored on a scale of 1–5, having successfully passed the oral-practical part of the exam if a score ≤ 4 is achieved.

Having successfully passed all examination parts, the candidates may apply for the full licence to practise medicine (approbation) in Germany and in the European Union, with the licence being valid for an unlimited time. In the present prospective, exemplarily performed single-centre study, we investigated to which degree medical competences other than the traditional “medical expert” are represented in the German licensing oral-practical exam. Furthermore, we explored the taxonomy levels (in approximation to Bloom's taxonomy of learning domains [15]) of the questions and tasks addressed during this important examination.

2. Methods

2.1. Study design

The present analyses were based on a prospective, observational-descriptive, single-centred, exemplary study design. Approval was obtained from the local independent Ethics Committee of the University of Ulm (17122013) and from the dean's office. A pilot study with observation of two licensing oral-practical exams tested the feasibility and practicability of the observation tools. The experience of the pilot lead to the hypotheses mentioned above. The main study then included the observation of days 1 and

2 of $n=7$ exams (during the winter semester 2013/14 and summer semester 2014). In accordance with the nationwide medico-legal framework (AeAppO 2002/2012), one study observer (not involved as an examiner) was present per oral-practical exam. The overall two observers met the following criteria: they were members of the faculty teaching body, have been actively involved in academic teaching and evaluation processes for more than 10 years, functioned as trainers in examiner training for more than 5 years, completed a master of medical education (MME) study-course, have been acquainted with NKLM criteria, and have been members of the university's task force for research in medical education. Two observers were intensively pre-trained based on the developed observation tools. The inter-rater reliability (Cohen's Kappa) between the two evaluators was calculated based on the analysis of 50 questions/tasks from real examinations according to both the assessment of the taxonomy levels (in approximation to the criteria described by Bloom 1956 [15]) and the attribution to various competence roles and content categories/subjects [http://www.nkml.de accessed:13 Oct 2015]. A combined inter-rater reliability of 0.78 was considered sufficient for the study. However, if any question could not be classified onsite, the exact wording of the question was documented and later clarified between the two observers to reach a consensus (which occurred in 20 cases).

2.2. Observation and Evaluation Instruments

2.2.1. Bedside exam document

For data acquisition during the bedside exam, one document has been designed to record the measured time as well as the number of questions/tasks and their assignment to the respective taxonomy level.

2.2.2. Off-bedside exam document day 1 and 2

For data acquisition during the off-bedside examination, an one-page document has been developed, which allows categorisation of each individual question/task (of each examiner) for different categories/subjects and taxonomy levels, inclusive of open ended questions for the observer's comments. All the described documents were tested for feasibility and practicability in a pilot study (during two full exams), and consequently slightly adjusted.

2.3. Chronometry of patient-contact during the oral-practical bedside exam

To assess the utilisation of the examination time for direct patient-examinee-communication, -contact, -interaction, and -practical-skill demonstration, the respective time was recorded in seconds using a conventional stopwatch and documented. The *post hoc* analysis then added the timing in relation to the total bedside examination time.

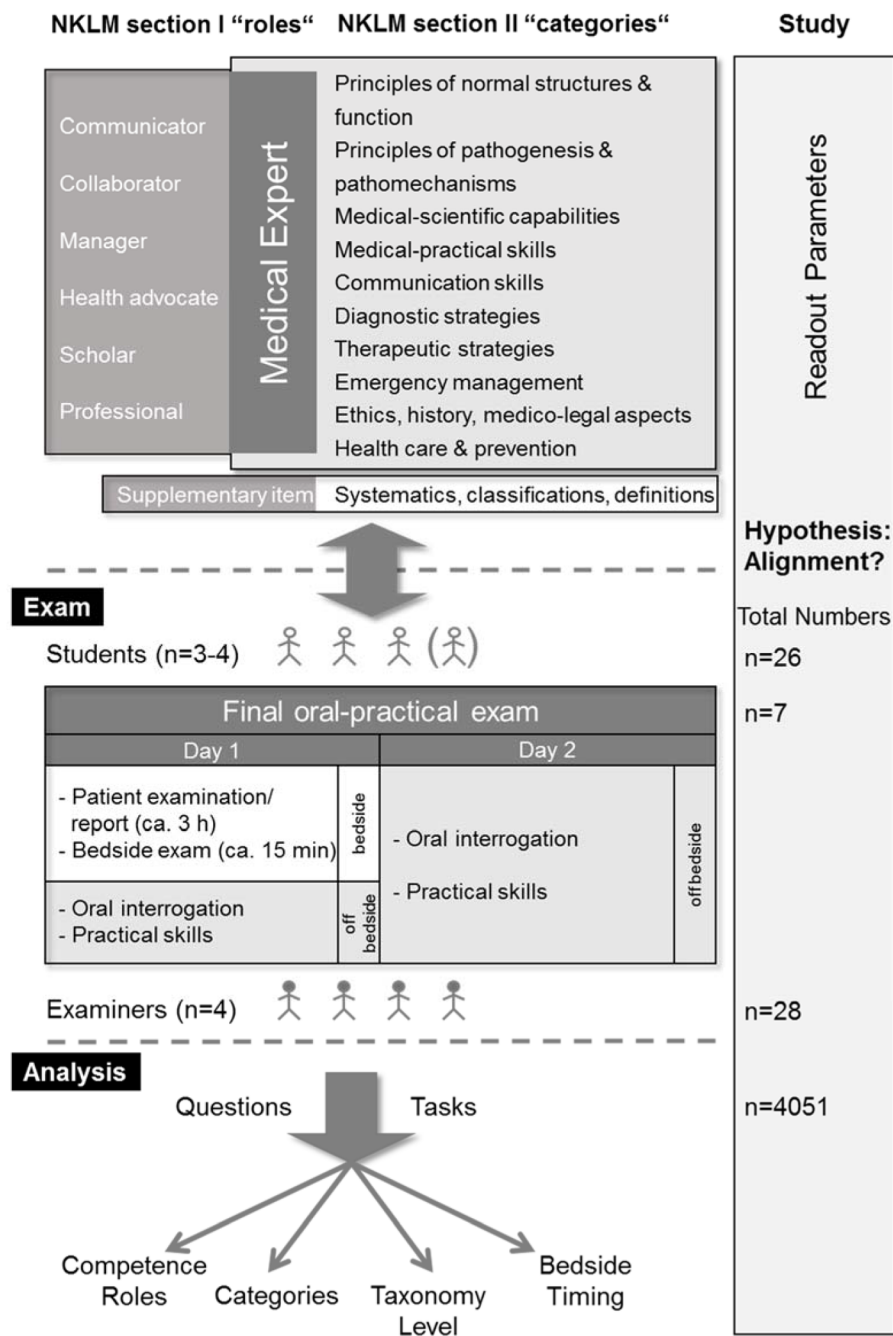
Competence Orientation: Roles & Categories of NKLM sections I and II

Figure 1: Competence Orientation: Roles & Categories of NKLM sections I and II

2.4. Assessment of the licensing oral-practical exam regarding competence-oriented issues

After each exam the data were transferred by an independent person to a digital database and – after complete data gathering – analysed regarding the following read-out parameters:

- Number of questions/tasks (per examiner, examinee, time slot)
- Taxonomy level of questions/tasks. In approximation to the Bloom's classification [15], a simplified three-levelled attribution was performed for the medical questions and tasks, with "knowledge" (in the sense of pure reproduction of factual knowledge) as the first and lowest taxonomy level, "analysis/interpretation" as the second, intermediate level, and "problem solving" as the third level. The pilot study confirmed the feasibility of this simplified 3-step categorisation during the onsite situation of the examination.
- Assignment, if suitable, of questions/tasks to one of the competence roles apart from the central "Medical Expert" role (in accordance to CanMEDS and NKLM): Communicator, Collaborator, Manager, Health Advocate, Scholar, Professional.

- Assignment of the questions/tasks to various content categories in approximation to the NKLM section II (see Figure 1). The pilot study revealed numerous questions addressing “systematics, classifications, and definitions”; therefore, we added this category as an additional read-out parameter.

2.5. Statistics:

The present analyses are based on a prospective, observational-descriptive, single-centred, exemplary study design. Results are presented as the mean $\pm$ standard error of the mean (SEM). A one-way analysis of variance (ANOVA) followed by the Student-Newman-Keuls test as a post hoc test for multiple comparisons was performed. A comparison of read-outs between exam days 1 and 2 was performed by the Student-t test. A value of $p < 0.05$ was considered statistically significant.

3. Results

3.1. A considerable number of questions and tasks are demanded in the licensing oral-practical medical examination

In present study of seven licensing oral-practical medical exams (each of 2-day duration) with $n=28$ examiners (4 in each exam) and $n=26$ examinees (3-4 in each exam), a total of $n=4051$ raised questions/tasks were analysed. During the first day, $n=70 \pm 24$ questions/tasks were set per candidate and during the second day $n=86 \pm 19$ questions/tasks were allotted per candidate (data not shown).

3.2. Minimal address of practical and communication skills by examiners at the bedside during the licensing medical exam

The bedside presentation of the patient, including the main findings and the demonstration of specific practical and communication skills on demand, is scheduled (according to the State Examination Office of Baden-Württemberg) for a time period of regularly 15 minutes. The scheduled total bedside exam time was perfectly adhered to by the examiners (see Figure 2, Point A). However, during this bedside time, $>80\%$ of exam time was used to ask and to answer theoretical questions, $<20\%$ of the exam time was used to demonstrate hands-on practical skills and virtually no communication skills were explicitly addressed. The student was challenged with $>17 \pm 9$ (theoretical) questions at the bedside (i.e. >1 question per minute). Furthermore, the analysis of the taxonomy level of these questions revealed that 13 ± 2 targeted a low taxonomy level (i.e. the “knowledge” level; see Figure 2, Point B), 3 ± 0.1 (18%) bedside questions considered the analysis or interpretation level and one question was based on the level of problem solving.

3.3. Competence roles beyond the “medical expert” are scarcely reflected in the licensing oral-practical medical exam

The NKLM section I framework, derived from the CanMEDS Physician Competence Framework [1], is to some extent based on the seven roles as a prerequisite for all physicians for good and reliable patient care: Medical Expert, Communicator, Collaborator, Manager, Health Advocate, Scholar, and Professional. The “medical expert”, considered as the classical and traditional role, and standing “in the centre” of all roles, certainly plays the main role in teaching, learning, and assessing medicine. Therefore, we were mainly interested in learning to what degree the six roles, other than the “medical expert”, are represented as examination subjects. The present analysis revealed that on both examination days a mean of <1 (maximal 2) question representing competences of a role beyond the medical expert was put to each candidate (see Figure 3), and that virtually no question addressed the role of the professional or scholar.

3.4. The licensing oral-practical exam focuses on very “traditional” subjects and categories

Professional expertise for the competent physician comprises a range of items and subjects from various categories, including knowledge, skills, and attitudes, which section II of the NKLM attempts to define (see Figure 1). Each question or task of the two day examination was classified in one of the 11 content categories of the observation tool. The main subjects were associated (in progressional order) with questions to the “classical” or “traditional” line of systematics/classifications/definitions, pathomechanisms, therapy, and finally diagnostics (see Figure 4). Only a minority of the questions/tasks were aimed at emergency management issues (together 5 ± 4.5 questions on both days of the exam) and almost none at medico-legal and ethical aspects or health care and prevention. In addition, a mean of <1 question per candidate was raised both regarding scientific issues and in the context of medical conversation/counselling techniques (communication skills). The assessment of clinical and practical skills, although permitted by the legal examination framework, played only a minor role.

3.5. Predominance of questions with a low-level taxonomy

All the questions raised were categorised by the observers within three taxonomy levels, namely “knowledge”, “analysis/interpretation”, and “problem solving”. The majority of the questions (70% on both days) were categorised as low-level, pure “knowledge” questions. Less than 15% of the tasks were evaluated as mid-level tasks requiring some analytical and/or interpretative competence. Similarly, the level of “problem solving” was addressed in $<10\%$ on day 1 and $<20\%$ on day 2. Overall,

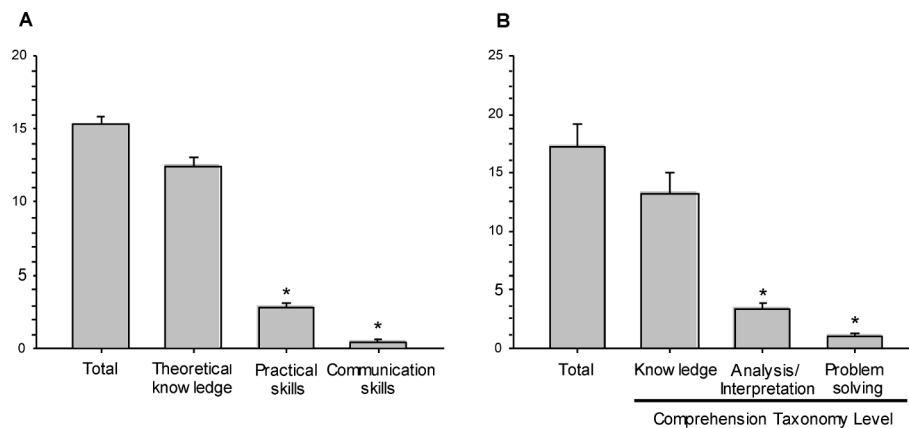


Figure 2: Bedside practical-oral exam

A Time dedicated to theoretical knowledge, practical or communication skills;

* $p < 0.05$ vs. theoretical knowledge. **B** Taxonomy level of questions; Number of questions during the exam addressing three taxonomy levels: knowledge, analysis/interpretation and problem solving. * $p < 0.05$ vs. knowledge; data are presented as mean $\pm$ SEM; (One-Way-Anova, $n=26$ students/ $n=28$ examiners)

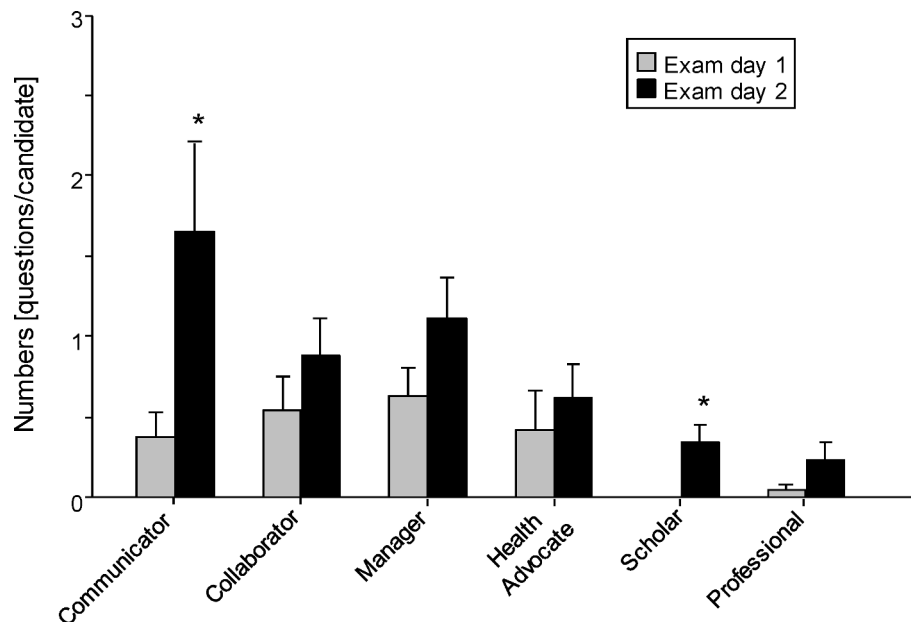


Figure 3: Off bedside practical-oral exam: Addressed roles

Number of questions per candidate addressing the different roles on both days of the exam; data are presented as mean $\pm$ SEM; * $p < 0.05$ vs. number of questions at day one (t-test).

the taxonomy level of each examiner's questions per candidate did not significantly differ between the first and second days of the licensing exam (see Figure 5). A more detailed analysis of the taxonomy levels in relation to the mainly examined expertise subjects or categories (>3 questions/subject) shows that the examination was dominated by "traditional" categories and by low-level (knowledge) questions (see Figure 6). Even the category "therapeutic strategies" failed to induce questions of higher taxonomy levels such as analytical, interpretative and problem solving competences.

4. Discussion

Based on an etymological analysis, the word "examine" originates from the Latin "examinare" meaning "accurately

weighing" [16]. To exactly "weigh" the performance profile of a candidate in a high-stakes medical exam, an assessment of multiple medical and personal competencies is required [6], [17], [18], [19]. Theoretically, the German licensing examination offers the chance to "weigh the candidates' performance" by a written 2-day multiple-choice-test and a 2-day oral-practical examination, each format aiming towards different professional requirements and competences [20].

However, the present study indicates a clear gap between the demanded competence-orientation in medical education (that is generally rated as highly relevant by clinical teachers and examiners) [21], [22] and the real-life situation during the oral-practical part of the investigated licensing medical exam (which tends to represent an assessment of pure knowledge within a single competence role). In concrete terms, the physician's roles beyond the

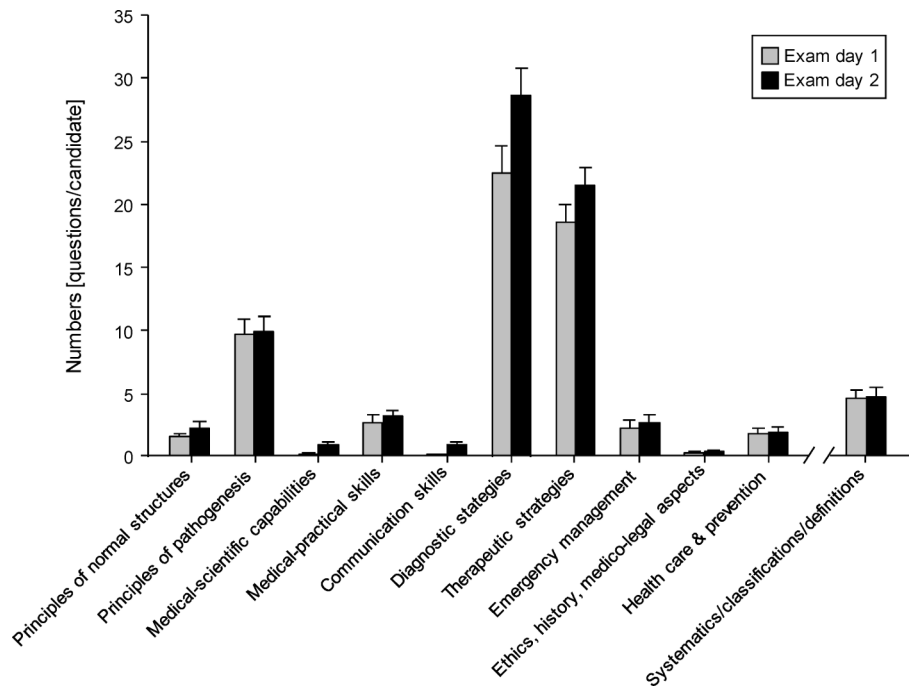


Figure 4: Off bedside practical-oral exam: Addressed topics

Number of questions regarding the different categories per candidate on both days of the exam; data are presented as mean $\pm$ SEM;

No statistically significant differences ($p < 0.05$) between the number of questions of day one and two (t-test).

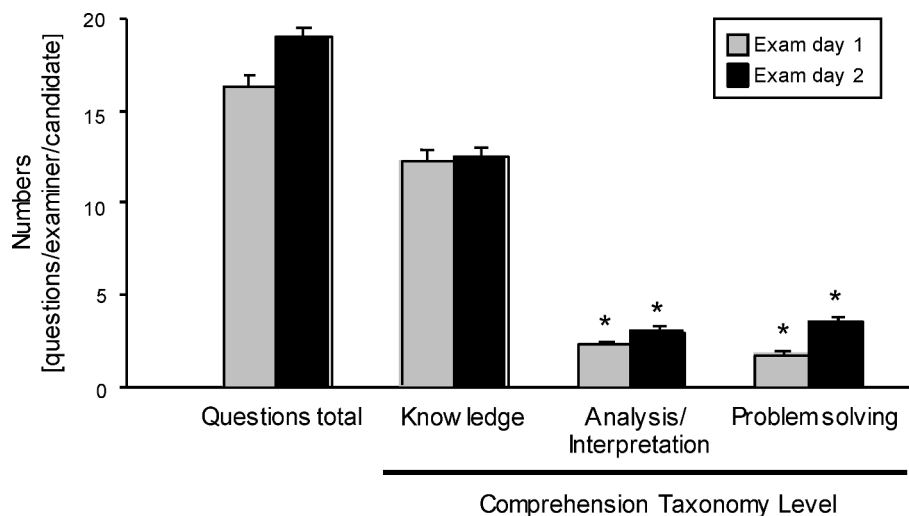


Figure 5: Off bedside practical-oral exam: Taxonomy level of questions

Number of questions per examiner and candidate attributed to three taxonomy levels: knowledge, analysis/interpretation and problem solving on both days of the exam. Data are presented as mean $\pm$ SEM; * $p < 0.05$ vs. knowledge (One-Way-Anova).

“medical expert” role were rarely addressed by the examiners, both at the bedside and in the examination room. Even the “communicator” role, which could have been addressed relatively easily during the bedside exam, was, in fact, very rarely addressed, although it is, without any doubt, of uppermost importance in a physician’s daily work. The competence of communicating with patients and their families has, to some extent, a predictive character regarding the frequency of future legal disputes between patients and physicians [23]. Furthermore, virtually no questions or tasks were related to the roles of the “professional”, “health advocate”, or “scholar”. Again, similar to the “communicator” role, the neglect of the

“scholar” role was also somewhat surprising, because all investigated examiners had – in addition to their clinical expertise – a profound academic/scientific background and all of them were considered to be scientifically active, such that the “scholar” role was supposed to play a significant role in their daily academic work. Additionally, from the perspective of the students, scientific education appears important as reflected by a generally positive attitude towards science and scientific methodology [24]. The omission of scientific/scholar aspects by the examiners may also be considered “alarming”, because mastering scientific algorithms reflects unique academic competence with a high impact in medical decision

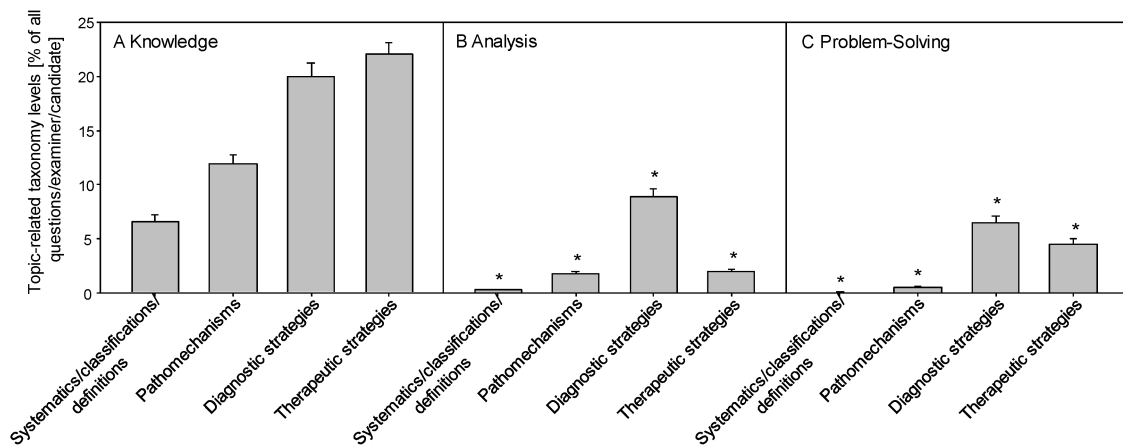


Figure 6: Off bedside practical-oral exam: Topic specific taxonomy levels

Percentage of questions regarding the different categories per examiner and candidate on both days of the exam assessed by three taxonomy levels:
(A) knowledge, (B) analysis/interpretation and (C) problem solving. Data are presented as mean $\pm$ SEM; * $p < 0.05$ vs. taxonomy level "knowledge" (One-Way-Anova)

making, self-criticism, and scientific evidence-based case-reflection that may even improve patient safety [25]. In this context it is noteworthy that the NKLM does not pre-define a specific weighting of the different roles and competences. This is certainly an important issue for future discussions among medical educators.

However, even within the singular role of the "medical expert" that has been mainly addressed by the examiners, there is surely more than pure knowledge that could be assessed: Miller's pyramid proposes "Knows" – "Knows how" – "Shows how" – "Does" as different taxonomic levels, with the higher levels perfectly suited for oral-practical exams. Additionally, within the framework of the current licensing exam, an ideal opportunity is foreseen to make use of higher taxonomic levels: the bedside exam of examination day 1, which is obligatory throughout Germany and which is (by the State Examination Office of Baden-Württemberg) recommended to take at least 15 minutes per candidate. During this time slot, one would expect the clinical experts to evaluate the candidates' competence with respect to patient interview and physical examination. However, our analysis of the bedside exam revealed a "question storm" occurring at the bedside, mostly questions of a low-level taxonomy that could have easily been asked in the later phase of the oral exam offside the patient's bed.

During the second exam day, no patients are involved, but the legal framework nevertheless provides multiple degrees of freedom for the examiners to assess the medical competences of the candidates in a multi-perspective approach, be it with the help of 3-D models or manikins and by imagined scenarios of medical dilemmas. However, the results of our analysis do not differ from the findings of the first day: roles besides the "medical expert" and subjects apart from the classical items of "diagnosis and therapy" were almost completely omitted.

The most likely explanation for this phenomenon is that the examiners' awareness and ability of assessing these roles and subjects are not yet sufficiently developed.

Both deficiencies could potentially be remedied when in future in the training of examiners greater focus is consciously placed on the importance of these roles for the overall medical competence and to provide training for appropriate examination techniques [9], [18], [26].

Nevertheless, examiner trainings (offered at many faculties) are naturally attended by predominantly novice examiners. Already active and experienced examiners are far less attainable for specific examiner trainings. At this point, all measures for faculty and human resources development should therefore be exploited in teaching and examination alike, to constantly make teachers and examiners at the faculties aware of the importance and the added value of competence orientation, and to apply the appropriate tools to implement this [19], [20], [27], [28], [29].

In this regard, the German License Order (Aerztliche Approbationsordnung) unfortunately at present even places certain constraints on further developments. Although it allows creative scope for competence orientation in teaching, the licence order makes no binding requirements in the direction of competence-oriented testing. In addition, more specific competence based exam formats for the licensing exams are currently not explicitly planned – such as an OSCE, which in the meantime has become established in the clinical intermediate exam and in the Swiss Federal Licensing Examination, or of the SOEs (Structured Oral Examinations with predefined clinical scenarios) described by Jefferies et al. [2], [12], [30], [31], [32], [33].

To our knowledge, the present study is the first onsite real-time analysis of the licensing oral-practical medical exam in Germany. However, there are some drawbacks and limitations of which to be aware. The data were obtained at only one centre and therefore translation to other nationwide exam sites may only be acceptable if further

centres are included. Although an acceptable inter-rater ratio was achieved, further observers or recording of exams and post hoc blinded analysis would enhance the validity and reliability of the data. However, in the present setting, the given medico-legal framework prevented us including more observers or to tape the exam for analytic purposes. Furthermore, live-coding of the exam also lead to restriction on explicit aspects of the examination questions with deliberate negligence of all implicit aspects. It should be emphasized, that although the idea of competence orientation was considered long before publication of the NKLM in Germany, the concrete content of the NKLM catalogue in its actual form has only recently been finalised and accepted by the authorities, and therefore requires time to be transferred to the educational curricula and various disciplines [4], [34]. At present, the NKLM has set the goals; the constructive alignment demands now subsequent development of both teaching and assessment systems. For the future of the German licensing examination system, the NKLM offers an enormous potential to push it towards a better competence orientation.

5. Conclusions

In conclusion, the present study revealed a severe misalignment between proposed competences and learning objectives (as defined in the NKLM and CanMEDS) and the learning objectives assessed in current real-life high-stakes examinations. The investigated exams followed a rather simple and traditional line, with a large number of questions on a low taxonomy level, with only minor practical parts.

Following the well-known paradigm “assessment drives learning” the multi-dimensional “competencies” aspect needs to be integrated in these oral-practical exams and addressed by the respective curricula. These changes presumably may also improve patient care and safety in the long term [35], [36].

Acknowledgements

The authors thank Stephanie Denk for graphical support. We especially thank Prof. Dr. med Jana Juenger (University of Heidelberg) for her excellent assistance in the modification of the study design. We also want to express our gratitude to Dr. Robert Blakytyn for cross reading the manuscript and correcting the language.

List of abbreviations

- ANOVA: Analysis of variance
- CanMEDS: Canadian Medical Education Directives for Specialists
- MME: Master of Medical Education

- NKLM: National Competence-based Learning objectives for undergraduate Medical education
- OSCE: Objective-Structured Clinical Examination
- SEM: Standard Error of the Mean

Ethics approval and consent to participate

Approval was obtained from the local independent Ethics Committee of the University of Ulm (17122013).

Availability of data and materials

Data are stored at the Department of Orthopaedic Trauma, Hand-, Plastic- and Reconstructive Surgery, University Hospital of Ulm, 89081 Ulm, Germany. Data are not yet shared because the extensive excel files includes additional data that are not focus of this publication.

Competing interests

The authors declare that they have no competing interests.

References

1. Frank JR, Danoff D. The Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
2. Oechsner W, Forster J. Approbierte Ärzte – kompetente Ärzte?: Die neue Approbationsordnung für Ärzte als Grundlage für kompetenzbasierte Curricula. *GMS Z Med Ausbild*. 2005;22(1):Doc04. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000004.shtml>
3. Van den Akker M, Dornan T, Scherpbier A, Egbrink M, Snoeckx LEasing the transition: the final year of medical education at Maastricht University. *Z Evid Fortbild Qual Gesundheitsw*. 2012;106(2):92-97. DOI: 10.1016/j.zefq.2012.02.013
4. Fischer MR, Bauer D, Mohn K. Finally finished! National Competence Based Catalogues of Learning Objectives for Undergraduate Medical Education (NKLM) and Dental Education (NKLZ) ready for trial. *GMS Z Med Ausbild*. 2015;32(3):Doc35. DOI: 10.3205/zma000977
5. Hahn EG, Fischer MF. National Competence-Based Learning Objectives for Undergraduate Medical Education (NKLM) in Germany: Cooperation of the Association for Medical Education (GMA) and the Association of Medical Faculties in Germany (MFT). *GMS Z Med Ausbild*. 2009;26(3):Doc35. DOI: 10.3205/zma000627
6. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27(4):294-297. DOI: 10.1080/01421590500126437
7. Wilson GM, Lever R, Harden RM, Robertson JI, Macritchie J. Examination of clinical examiners. *Lancet*. 1969;293(7584):37-40. DOI: 10.1016/S0140-6736(69)90998-2

8. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-396. DOI: 10.1056/NEJMr054784
9. Oechsner W, Geiler S, Huber-Lang M. Effects and sustainability of trainings for the oral and practical part of the German final exam in medicine. *GMS Z Med Ausbild*. 2013;30(3):Doc36. DOI: 10.3205/zma000879
10. Van der Vleuten CP. Validity of final examinations in undergraduate medical training. *BMJ*. 2000;321(7270):1217-1219. DOI: 10.1136/bmj.321.7270.1217
11. Newstead S. Examining the examiners: Why are we so bad at assessing students? *Psych Learn Teach*. 2002;2:70-75. DOI: 10.2304/plat.2002.2.2.70
12. Jefferies A, Simmons B, Ng E, Skidmore M. Assessment of multiple physician competencies in postgraduate training: utility of the structured oral examination. *Adv Health Sci Educ Theory Pract*. 2011;16(5):569-577. DOI: 10.1007/s10459-011-9275-6
13. Pabst R. Medical education and reform initiatives in Germany. *Acad Med*. 1995;70(11):1006-1011. DOI: 10.1097/00001888-199511000-00019
14. Chenot JF. Undergraduate medical education in Germany. *Medizinstudium in Deutschland*. *GMS Ger Med Sci*. 2009;7:Doc02. DOI: 10.3205/000061
15. Bloom B. Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay; 1956.
16. Kluge F. Etymologisches Woerterbuch der Deutschen Sprache. 25th ed. Berlin: De Gruyter; 2012. DOI: 10.1515/9783110223651
17. Joughin G. Dimensions of oral assessment. *Assess Eval High Educ*. 1998;23:367-378. DOI: 10.1080/0260293980230404
18. 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
19. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach*. 2003;25(6):569-584. DOI: 10.1080/0142159032000151907
20. Stern DT, Wojtczak A, Schwarz MR. The assessment of global minimum essential requirements in medical education. *Med Teach*. 2003;25(6):589-595. DOI: 10.1080/0142159032000151295
21. Jilg S, Moeltner A, Berberat P, Fischer MR, Breckwoldt J. How do supervising clinicians of a University Hospital and associated teaching hospitals rate the relevance of the key competencies within the CanMEDS roles framework in respect to teaching in clinical clerkships? *GMS Z Med Ausbild*. 2015;32(3):Doc33. DOI: 10.3205/zma000975
22. Oechsner W, Geiler S, Huber-Lang M. The impact of a dedicated training program for oral examiners at a medical school in Germany: a survey among participants from operative and non-operative disciplines. *Patient Saf Surg*. 2013;7(1):22. DOI: 10.1186/1754-9493-7-22
23. Tamblyn R, Abrahamowicz M, Dauphinee D, Wenghofer E, Jacques A, Klass D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*. 2007;298(9):993-1001. DOI: 10.1001/jama.298.9.993
24. Pruskil S, Burgwinkel P, Georg W, Keil T, Kiessling C. Medical students' attitudes towards science and involvement in research activities: a comparative study with students from a reformed and a traditional curriculum. *Med Teach*. 2009;31(6):e254-e259. DOI: 10.1080/01421590802637925
25. Huber-Lang M, Neugebauer E. The dual role of academic surgeons as clinicians and researchers - an attempt to square the circle? *Patient Saf Surg*. 2011;5:16. DOI: 10.1186/1754-9493-5-16
26. Schickler A, Bruestle P, Biller S. The Final Oral/Practical State Examination at Freiburg Medical Faculty in 2012 - Analysis of grading to test quality assurance. *GMS Z Med Ausbild*. 2015;32(4):Doc39. DOI: 10.3205/zma000981
27. Bland CJ, Starnaman S, Wersal L, Moorhead-Rosenberg L, Zonia S, Henry R. Curricular change in medical schools: how to succeed. *Acad Med*. 2000;75(6):575-594. DOI: 10.1097/00001888-200006000-00006
28. Pruskil S, Deis N, Druener S, Kiessling C, Philipp S, Rockenbach K. Implementation of "social and communicative competencies" in medical education. The importance of curriculum, organizational and human resource development. *GMS Z Med Ausbild*. 2015;32(5):Doc50. DOI: 10.3205/zma000992
29. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach*. 2006;28(6):497-526. DOI: 10.1080/01421590600902976
30. Berendonk C, Schirlo C, Balestra G, Bonvin R, Feller S, Huber P, Jünger E, Monti M, Schanbel K, Beyeler C, Guttormsen S, Huwendiek S. The new final clinical skills examination in human medicine in Switzerland: essential steps of exam development, implementation and evaluation, and central insights from the perspective of the national Working Group. *GMS Z Med Ausbild*. 2015;32(4):Doc40. DOI: 10.3205/zma000982
31. Guttormsen S, Beyeler C, Bonvin R, Feller S, Schirlo C, Schnabel K, Schurter T, Berendonk C. The new licensing examination for human medicine: from concept to implementation. *Swiss licensing examination for human medicine*. *Swiss Med Wkly*. 2013;43:w13897.
32. Dwyer T, Glover Takahashi S, Kennedy Hynes M, Herold J, Wasserstein D, Nousiainen M, Ferguson P, Wadey V, Murnaghan ML, Leroux T, Semple J, Hodges B, Ogilvie-Harris D. How to assess communication, professionalism, collaboration and the other intrinsic CanMEDS roles in orthopedic residents: use of an objective structured clinical examination (OSCE). *Can J Surg*. 2014;57(4):230-236. DOI: 10.1503/cjs.018813
33. Nguyen LN, Tardioli K, Roberts M, Watterson J. Development and incorporation of hybrid simulation OSCE into in-training examinations to assess multiple CanMEDS competencies in urologic trainees. *Can Urol Assoc J*. 2015;9(1-2):32-36. DOI: 10.5489/cuaj.2366
34. Kadmon M, Bender MJ, Adili F, Arbab D, Heinemann MK, Hofmann HS, König S, Küper MA, Obertacke U, Rennekampff HO, Rolle U, Rücker M, Sader R, Tingart M, Tolsdorf MM, Tronnier V, Will B, Walcher F; Chirurgische Arbeitsgemeinschaft Lehre der Deutschen Gesellschaft für Chirurgie (CAL). Competency-based medical education: National Catalogue of Learning Objectives in surgery. *Chirurg*. 2013;84(4):277-285. DOI: 10.1007/s00104-012-2449-9
35. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? *Anat Sci Educ*. 2009;2(5):199-204. DOI: 10.1002/ase.102
36. Nikendei C, Kraus B, Schrauth M, Briem S, Juenger J. Ward rounds: how prepared are future doctors? *Med Teach*. 2008;30(1):88-91. DOI: 10.1080/01421590701753468

Corresponding author:

Prof. Dr. med. Markus Huber-Lang
University Hospital of Ulm, Medical School, Institute for
Clinical- and Experimental Trauma-Immunology,
Helmholtzstr. 8/1, D-89081 Ulm, Germany, Phone: +49
(0)731/500-54800, Fax: +49 (0)731/500-54818
markus.huber-lang@uniklinik-ulm.de

Please cite as

Huber-Lang M, Palmer A, Grab C, Boeckers A, Boeckers TM,
Oechsner W. Visions and reality: the idea of competence-oriented
assessment for German medical students is not yet realised in licensing
examinations. *GMS J Med Educ.* 2017;34(2):Doc25.
DOI: 10.3205/zma001102, URN: urn:nbn:de:0183-zma0011024

This article is freely available from

<http://www.egms.de/en/journals/zma/2017-34/zma001102.shtml>

Received: 2016-08-19

Revised: 2017-02-21

Accepted: 2017-03-07

Published: 2017-05-15

Copyright

©2017 Huber-Lang et al. This is an Open Access article distributed
under the terms of the Creative Commons Attribution 4.0 License. See
license information at <http://creativecommons.org/licenses/by/4.0/>.

Vision und Wirklichkeit: Die Idee einer Kompetenz orientierten Prüfung deutscher Medizinstudenten wird im mündlich-praktischen Teil des Zweiten Abschnitts der Ärztlichen Prüfung noch nicht umgesetzt

Zusammenfassung

Zielsetzung: Kompetenzorientierung basierend auf dem CanMEDS Model hat sich zu einem wichtigen Bildungsinhalt innerhalb der medizinischen Ausbildung entwickelt. Im Nationalen kompetenzbasierten Lernzielkatalog Medizin (NKLM) wurde diese als Ziel festgeschrieben. Es ist jedoch bisher nicht bekannt, inwiefern in den Prüfungen des medizinischen Staatsexamens tatsächlich kompetenzbasierend geprüft wird.

Methoden: In dieser prospektiven, deskriptiven, monozentrischen, exemplarischen Studie wurden 4051 Fragen/Aufgaben (gestellt von 28 Prüfer in n=7 zweitägigen mündlich-praktischen Prüfungen) an Medizinstudierende der Universität Ulm evaluiert. Die mündlichen und praktischen Fragen/Aufgaben wurden, ebenso wie die Prüfung am Patientenbett, auf die spezifischen Kompetenzrollen (gemäß NKLM Abschnitt I), die unterschiedlichen Kategorien (gemäß NKLM Abschnitt II) und die Taxonomielevel der unterschiedlichen Fragen/Aufgaben analysiert.

Ergebnisse: Im mündlich-praktischen Prüfungsteil wurden den Kandidaten zahlreiche Fragen/Aufgaben gestellt (Tag 1/2: 70±24/86±19 Fragen). Die Kompetenzrollen außerhalb des "Medizinischen Experten" wurden dabei selten adressiert. Zudem wurden praktische und kommunikative Fertigkeiten in der Prüfung am Patientenbett kaum geprüft (weniger als 3/15 min). Auffallend war die Dominanz an Fragen auf niedrigem taxonomischem Niveau.

Schlussfolgerungen: Die Daten zeigen ein deutliches Missverhältnis zwischen den Vorgaben zur Kompetenzorientierung und den tatsächlich geprüften Inhalten der mündlich-praktischen Prüfungen. Diese Diskrepanz sollte zukünftig mit einer verstärkten Prüfer-Schulung und -Evaluation begegnet werden.

Schlüsselwörter: Medizinische Kompetenz, Staatsexamen, Alignment, NKLM

Markus Huber-Lang¹
Annette Palmer¹
Claudia Grab²
Anja Boeckers³
Tobias Maria Boeckers³
Wolfgang Oechsner⁴

1 Universitätsklinikum Ulm, Medizinische Fakultät, Institut für klinische und experimentelle Trauma-Immunologie, Ulm, Deutschland

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

3 Universität Ulm, Institut für Anatomie und Zellbiologie, Ulm, Deutschland

4 Universitätsklinikum Ulm, Abteilung Kardioanästhesie, Ulm, Deutschland

1. Einleitung

1.1. Hintergrund

Das klassische Konzept der lernzielorientierten Ausbildung, das über Jahrzehnte hinweg die konzeptuelle Grundlage der Hochschullehre darstellte, wird zunehmend durch das bei weitem komplexere Konstrukt der kompetenzbasierten bzw. kompetenzorientierten Ausbildung ergänzt und erweitert. Für dieses komplexe Konstrukt stellen die 1990 publizierten 7 CanMEDS Rollen („medical expert, communicator, collaborator, manager, health advocate, scholar, professional“) weltweit eine geradezu archetypische Grundlage dar [1], [2], [3]. In Deutschland

wurde kürzlich der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) als Bezugsgröße für die kompetenzbasierte Mediziner Ausbildung auf dem Boden der geltenden Ärztlichen Approbationsordnung veröffentlicht, ebenfalls mit deutlichen Referenzen auf das CanMEDS-Konzept [4], [5].

Als Gradmesser für die erworbene Kompetenz medizinischer Absolventen kommen seit langem und mit hoher Gewichtung mündlich-praktische Prüfungen zum Einsatz [6], [7], [8], [9]. Allerdings wurden traditionelle mündlich-praktische Prüfungen häufig auch ernsthaft kritisiert, zum einen wegen ihrer oft zu geringen Reliabilität und Validität [6], [10], [11], zum anderen oft auch wegen ihres in Bezug auf die Bandbreite medizinischer Kompetenzen zu engen Prüfungsspektrums. Positive Gegenbeispiele im Sinne gut strukturierter und breit kompetenzorientierter

mündlicher Prüfungen wurden deutlich seltener beschrieben [12].

Zum gegenwärtigen Zeitpunkt gibt es außerhalb der vorliegenden Studie gemäß unseres Kenntnisstands keine Daten aus der direkten Beobachtung mündlich-praktischer medizinischer Staatsexamina, die auf den Grad der Kompetenzorientierung und auf das Taxonomieniveau der dort gestellten Fragen und Aufgaben abzielen.

1.2. Das Medizinische Staatsexamen in Deutschland

Zum Zeitpunkt der Durchführung unserer Studie war die Ärztliche Prüfung (Staatsexamen) in Deutschland gemäß geltender Approbationsordnung in zwei Abschnitte unterteilt (Erster Abschnitt der Ärztlichen Prüfung, nach 4 Fachsemestern, und Zweiter Abschnitt der Ärztlichen Prüfung, nach 10 Fachsemestern). Beide Abschnitte bestanden wiederum aus je einem schriftlichen und einem mündlich-praktischen Anteil, die regelhaft in direkter Folge zusammenhingen. In der heute, also zum Publikationszeitpunkt geltenden, veränderten Approbationsordnung wurden der schriftliche und der mündlich-praktische Teil des zweiten Abschnitts zu jeweils eigenständigen Abschnitten der Ärztlichen Prüfung erklärt und zeitlich getrennt. Die im folgenden beschriebenen Details beziehen sich auf die Verfahrensweisen, wie sie in der Ärztlichen Approbationsordnung und vom Landesprüfungsamt Baden-Württemberg für Medizin und Pharmazie zum Zeitpunkt der Durchführung der Studie entweder vorgeschrieben oder zumindest empfohlen waren; grundlegende Änderungen im Ablauf des hier betrachteten mündlich-praktischen Abschnitts haben sich durch die veränderte Approbationsordnung nicht ergeben.

Innerhalb der wie genannt vorgegebenen Rahmenbedingungen waren die geforderten vier bis fünf (in der Studie regelhaft vier) Examensprüfer relativ frei in Bezug auf die fachlich-inhaltliche, konkrete Ausgestaltung der gestellten Fragen und Aufgaben während der beiden Prüfungstage [13], [14]. Am ersten Prüfungstag hatten die (bis zu vier; in der Studie regelhaft drei bis vier) Kandidaten innerhalb von drei Stunden jeweils einen realen Patienten zu untersuchen, zu befragen, seine Krankheitsgeschichte zu dokumentieren und einen Fallbericht anzufertigen. Im Anschluss war vom Landesprüfungsamt ein Zeitraum von 15-20 Minuten vorgesehen, in dem der Patient vorzustellen war, und konkrete bettseitige Fertigkeiten (Untersuchungstechniken, Anamnesetechniken, etc.) sowie gezielte patientenbezogene Fragen abgeprüft werden konnten (siehe Abbildung 1). Außerhalb des Krankenzimmers, im Prüfungsraum, wurde dann jeder Kandidat erneut geprüft (7-11 Minuten pro Prüfer nach Maßgabe des zuständigen Landesprüfungsamts). Am zweiten Prüfungstag wurde jeder Kandidat von derselben Prüfungskommission erneut für 11-15 Minuten mündlich geprüft. Fachübergreifende Fragen und Aufgaben waren zugelassen und erwünscht. Zum Bestehen des mündlich-praktischen Examensteils war mindestens die Note 4 („ausreichend“) zu erreichen.

Mit dem Bestehen der Ärztlichen Prüfung insgesamt konnte (und kann weiterhin) seitens der Kandidaten die Approbation beantragt werden, die in Deutschland und in der Europäischen Union zeitlich unbegrenzt Gültigkeit besitzt.

In der vorliegenden prospektiven, exemplarisch an einer Fakultät durchgeführten Studie wurde untersucht, inwieweit ärztliche Kompetenzen jenseits des traditionellen „Medizinischen Fachexperten“ im mündlich-praktischen Teil des abschließenden deutschen Staatsexamens geprüft werden. Darüberhinaus wurden (in Anlehnung an die von Bloom formulierten Taxonomiestufen, [15]) die Taxonomie-Niveaus der in diesem bedeutsamen Examensteil gestellten Prüfungsaufgaben untersucht.

2. Methoden

2.1. Studienaufbau

Die vorliegende Untersuchung basierte auf einer exemplarischen, prospektiven, deskriptiven Beobachtungsstudie an der Universität Ulm. Die Studie wurde durch die unabhängige lokale Ethikkommission (17122013) und das Studiendekanat der Universität Ulm genehmigt. Eine Pilotstudie mit der Beobachtung von zwei mündlich-praktischen Staatsexamina überprüfte die Machbarkeit und Durchführbarkeit der Beobachtungsinstrumente. Die Erfahrungen der Pilotierung führten zu den oben genannten Hypothesen. Die Hauptstudie beinhaltete dann die Beobachtung von Tag 1 und 2 von n=7 Examina (während des Wintersemesters 2013/14 und Sommersemesters 2014). In Übereinstimmung mit dem deutschlandweit gültigen Approbationsordnung (ÄAppO 2002/2012) war ein Beobachter pro mündlich-praktischem Examen (der selbst nicht als Prüfer beteiligt war) anwesend. Die insgesamt zwei Beobachter erfüllten folgende Kriterien: sie waren Mitglieder des Lehrkörpers der Medizinischen Fakultät, waren seit mehr als 10 Jahre aktiv in der akademischen Lehre und im Prüfungswesen tätig, fungierten seit mehr als 5 Jahre als Trainer von Prüfertrainings, hatten erfolgreich eine Weiterbildung zum Master of Medical Education (MME) absolviert, waren mit den NKLM Kriterien vertraut und waren Mitglieder der Arbeitsgemeinschaft für Lehrforschung an der Universität Ulm. Die zwei Beobachter haben sich mit den entwickelten Beobachtungsinstrumenten vorab intensiv beschäftigt. Die Inter-Rater Reliabilität (Cohen's Kappa) zwischen den zwei Beobachtern wurde anhand einer Analyse von 50 Fragen/Aufgaben von realen Examenssituationen bezüglich der Einschätzung des Taxonomieniveaus (in Annäherung an die durch Bloom 1956 beschriebenen Kriterien [15]) und bezüglich der Zuordnung zu den verschiedenen Kompetenzrollen und fachlich-inhaltlichen Kategorien [<http://www.nklm.de>, Version vom 13. Okt. 2015] ermittelt. Die gefundene Inter-Rater Reliabilität von 0,78 wurde als ausreichend für die Studie angesehen. Im Falle dass eine Frage nicht direkt während der Beobachtung klassifiziert werden konnte, wurde die Fragestellung wörtlich

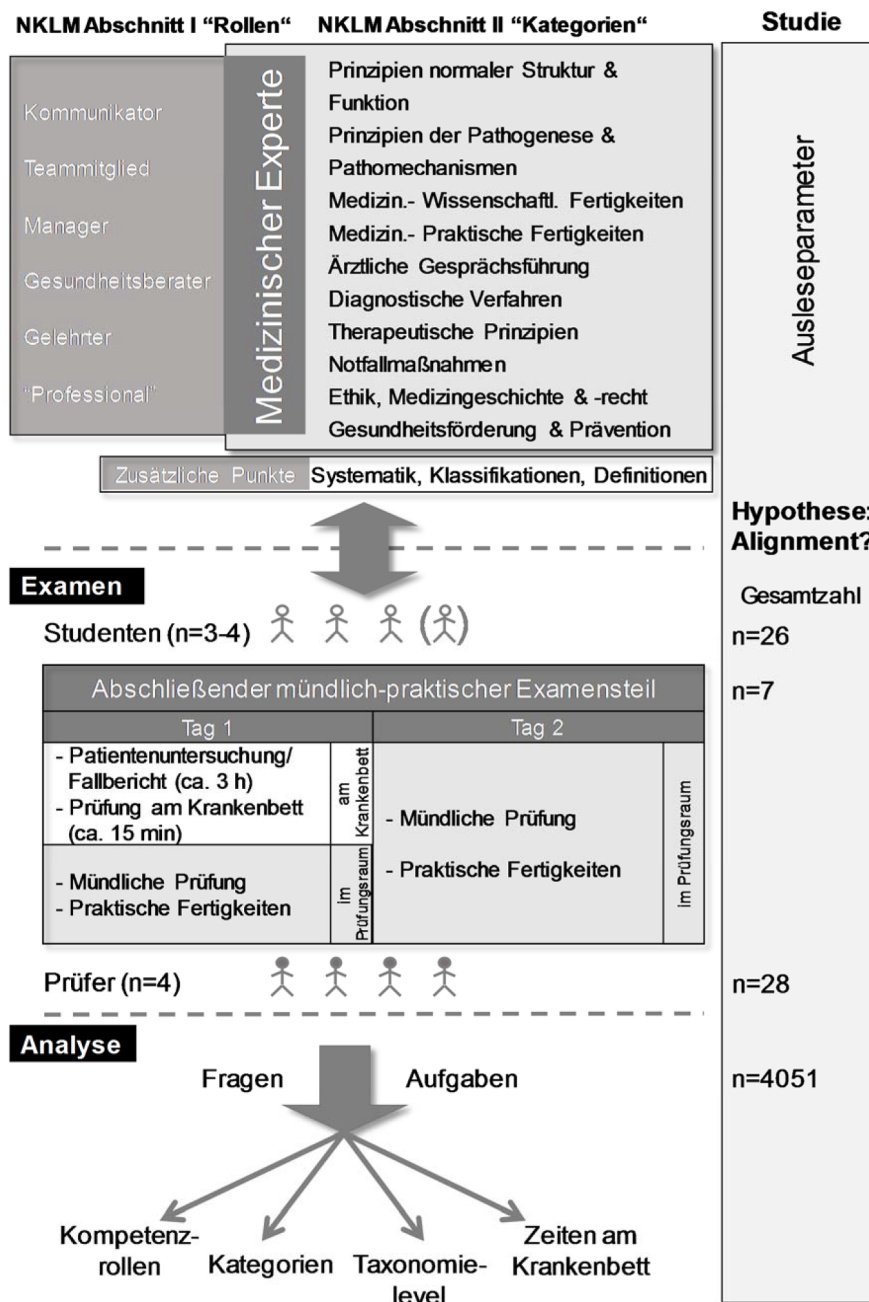
Kompetenzorientierung: Rollen & Kategorien der NKLM Abschnitte I & II

Abbildung 1: Kompetenzorientierung – Rollen und Kategorien in den NKLM-Abschnitten I und II, Prüfungsablauf und Studiendesign

dokumentiert und post hoc zwischen den zwei Beobachtern im Konsens geklärt (in 20 Fällen).

2.2. Beobachtungs- und Evaluationsinstrumente

2.2.1. Prüfungsdokument am Patientenbett:

Für die Datenerhebung des Exams am Patientenbett wurde ein Dokument erstellt, das sowohl die gemessene Zeit als auch die Anzahl der Fragen/Aufgaben und deren Zuordnung zu den jeweiligen Taxonomielevels festhielt.

2.2.2. Dokumentation für die Prüfung im Prüfungsraum:

Für die Datenerhebung des Exams fern des Patientenbettes (im Prüfungsraum) wurde ein einseitiges Dokument entwickelt, das die Kategorisierung jeder einzelnen Frage/Aufgabe (jedes Prüfers) für die unterschiedlichen fachlich-inhaltlichen Kategorien und für das Taxonomie-niveau erlaubte, einschließlich offener Kommentare durch den Beobachter. All die beschriebenen Dokumente wurden auf ihre Machbarkeit und Durchführbarkeit in einer Pilotstudie getestet (i.R. von 2 vollständigen Examen) und anschließend geringfügig angepasst.

2.3. Zeitmessung des Patientenkontakts während des mündlich-praktischen Exams am Patientenbett

Um die Ausnutzung der Prüfungszeit für die unmittelbare Patient-Prüfling-Kommunikation, -Kontakt, -Interaktion und -Demonstration praktischer Fähigkeiten einzuschätzen, wurde die entsprechende Zeit in Sekunden mit einer handelsüblichen Stoppuhr gemessen und notiert. Die *post hoc* Analyse setzte dann die ermittelte Zeit in das Verhältnis zur gesamten Prüfungszeit am Patientenbett.

2.4. Einschätzung des mündlich-praktischen Staatsexamens bezüglich der Kompetenzorientierung

Nach jedem Examen wurden die Daten durch eine unabhängige Person in eine digitale Datenbasis übertragen und nach vollständiger Datenerhebung bzgl. folgender Parameter analysiert:

- Anzahl der Fragen/Aufgaben (pro Prüfer, Prüfling, Zeiteinheit)
- Taxonomielevel der Fragen/Aufgaben. In Anlehnung an die Bloom Einteilung [15] erfolgte eine vereinfachte dreistufige Zuordnung der medizinischen Fragen/Aufgaben; eine erste und niedrigste Taxonomiestufe beinhaltete „Wissen“ (im Sinne einer reinen Wiedergabe von Faktenwissen), eine zweite mittlere Taxonomiestufe adressierte „Analyse und Interpretation“, und eine dritte Stufe zielte auf die „Problemlösung“. Die Pilotstudie bestätigte die Machbarkeit dieser vereinfachten, 3-stufigen Zuordnung in Echtzeit während des Exams.
- Zuordnung der Fragen/Aufgaben -falls möglich- zu einer Kompetenzrolle neben der zentralen Rolle als „Medizinexperte“ (in Übereinstimmung mit den CanMEDS und NKLM Definitionen): Kommunikator, Mitglied eines Teams, Verantwortungsträger/Manager, Gesundheitsberater und -fürsprecher, Gelehrter, Professionell Handelnder.
- Einteilung der Fragen/Aufgaben bzgl. der fachlich-inhaltlichen Kategorien in Anlehnung an den II. Abschnitt des NKLM (siehe Abbildung 1). Die Pilotierung wies zahlreiche Fragen zur „Systematik, Klassifikationen und Definitionen“ auf; daher haben wir diese Kategorie als einen zusätzlichen Ausleseparameter eingeschlossen.

2.5. Statistische Analyse:

Die vorliegende Analyse war auf einer exemplarischen, prospektiven, deskriptiven, monozentrischen Beobachtungsstudie aufgebaut. Die Ergebnisse werden als Mittelwert $\pm$ Standardfehler (SEM) dargestellt. Es kam eine einfaktorielle ANOVA Analyse (one-way analysis of variance) mit anschließender *post hoc* Analyse mittels dem Student-Newmann-Keuls Test zur Anwendung. Ein Vergleich der Ausleseparameter zwischen Tag 1 und 2 erfolgte

durch den Student-t Test. Ein p-Wert von $p < 0,05$ wurde als statistisch signifikant angenommen.

3. Ergebnisse

3.1. Im mündlich-praktischen Staatsexamen werden die Kandidaten mit einer beträchtlichen Anzahl an Prüfungsfragen und Aufgaben konfrontiert

In vorliegender Studie wurden 7 je 2-tägige mündlich-praktische Staatsexamina mit insgesamt 28 Prüfern (vier pro Examen) und 26 Kandidaten (drei bis vier pro Examen) untersucht. Es wurden dabei insgesamt 4051 Aufgaben- und Fragenstellungen analysiert. Am ersten Tag wurden jedem Kandidaten $n=70 \pm 24$ Aufgaben gestellt, am zweiten Tag $n=86 \pm 19$ Aufgaben (Daten nicht grafisch dargestellt).

3.2. Praktische und kommunikative Fertigkeiten werden auch während des Prüfungsteils am Patientenbett kaum adressiert.

Der gemäß des zuständigen Landesprüfungsamt für die Vorstellung des Exams-Patienten vorgesehene Zeitkorridor von 15-20 Minuten pro Kandidat wurde von den Prüfern perfekt eingehalten (siehe Abbildung 2, Punkt A). In diesem Zeitkorridor können wichtige Befunde Prüfungsgegenstand sein, ebenso wie spezifische praktische und kommunikative Kompetenzen des Kandidaten. Allerdings wurde mehr als 80% der bettseitig verfügbaren Prüfungszeit genutzt, um theoretische Fragen zu stellen, dementsprechend wenig Zeit ($< 20\%$) blieb für die Überprüfung praktischer Fertigkeiten am Krankenbett. Kommunikative Fertigkeiten wurden so gut wie nicht explizit adressiert. Jeder Kandidat wurde mit durchschnittlich mehr als 17 ± 9 theoretischen Fragen am Krankenbett konfrontiert (d.h. > 1 Frage/Minute). Die weitere Analyse der am Krankenbett gestellten Fragen zeigte, dass 13 ± 2 dieser Fragen sich auf dem taxonomisch relativ niedrigen Niveau der Reproduktion reinen Faktenwissens bewegten (siehe Abbildung 2, Punkt B). Analyse- oder Interpretationsfertigkeiten wurden nur in $3 \pm 0,1$ (18%) der Aufgaben adressiert, das Level der Problemlösung wurde noch seltener von den Kandidaten gefordert.

3.3. Kompetenzrollen jenseits des „Medizinischen Experten“ werden im mündlich-praktischen Staatsexamen kaum adressiert

In direktem Bezug auf das CanMEDS-Konzept [1] werden im Abschnitt I des NKLM 7 für die ärztliche Tätigkeit wichtige Kompetenzrollen definiert: Medizinischer Experte, Gelehrter, Kommunikator, Mitglied eines Teams, Ge-

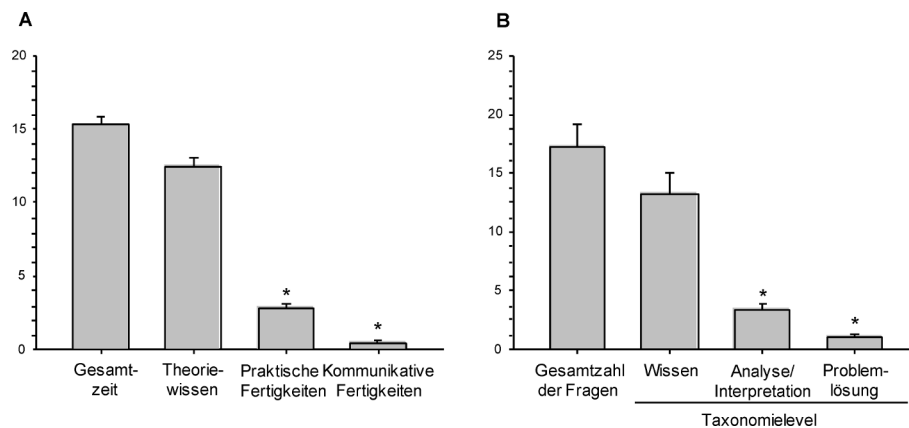


Abbildung 2: Prüfung am Patientenbett

A Aufgewendete Prüfungszeit für Theoriewissen, praktische oder kommunikative Fertigkeiten; * $p < 0,05$ gegenüber Prüfungszeit für Theoriewissen

B Taxonomielevel; Zahl der Prüfungsfragen auf den Taxonomielevels (reproduzierendes) Wissen, Analyse/Interpretation, Problemlösung; * $p < 0,05$ gegenüber (reproduzierendem) Wissen; Daten sind Mittelwerte + SEM (One-Way-Anova, $n=26$ Kandidaten / $n=28$ Prüfer.

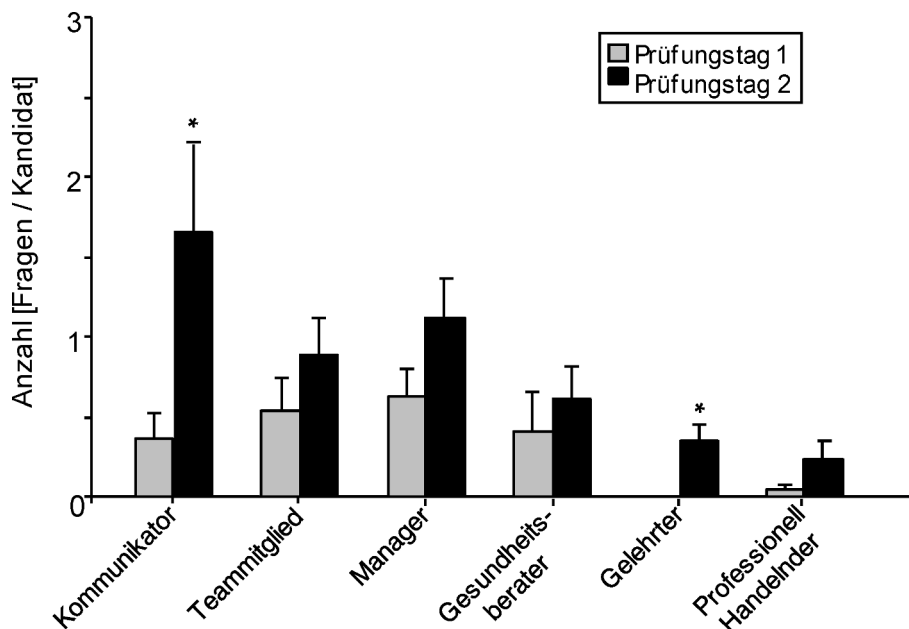


Abbildung 3: Prüfung im Prüfungsraum (außerhalb Patientenzimmer)

Zahl der Prüfungsfragen/aufgaben pro Kandidat mit Bezug auf die jeweilige Rolle an den beiden Prüfungstagen; Daten sind Mittelwerte $\pm$ SEM; * $p < 0,05$ Anzahl Fragen mit Bezug auf die Rolle an Tag 2 gegenüber Tag 1 (t Test).

sundheitsberater und -försprecher, Verantwortungsträger und Manager, Professionell Handelnder. Der traditionellen und klassischen Rolle des „Medizinischen Experten“ kommt in Lehre und Prüfung, damit im Lernprozess insgesamt, sicherlich eine zentrale Bedeutung und Wertigkeit zu. Deshalb waren wir vornehmlich interessiert zu erfahren, in welchem Maß die übrigen 6 Rollen in der mündlich-praktischen Staatsprüfung als Prüfungsgegenstände repräsentiert waren. Die Analyse zeigte, dass an beiden Prüfungstagen im Schnitt höchstens 1-2 Fragestellungen pro Kandidat auf diese übrigen Rollen entfielen (siehe Abbildung 3). Die Rollen „Professionell Handelnder“ oder „Gelehrter“ wurden so gut wie nicht adressiert.

3.4. Die mündlich-praktische Prüfung fokussiert auf eher „traditionelle“ Prüfungsgegenstände

In Anlehnung an den Abschnitt II des NKLM (siehe Abbildung 1), in dem unterschiedliche thematische Kategorien für die Ausbildung des kompetenten Arzts beschrieben sind, wurde für die vorliegende Studie ein Beobachtungskatalog aus 11 inhaltlich-thematischen Kategorien entwickelt. Jede in der Prüfung gestellte Aufgabe wurde mittels Strichliste einer dieser Kategorien zugeordnet. Die meisten Prüfungsfragen ließen sich eher „klassischen“ bzw. „traditionellen“ Kategorien zuordnen, nämlich (in quantitativ aufsteigender Reihenfolge) den Beobachtungskategorien „Systematik/Klassifikation/Definitionen“, „Pathomechanismen“, „Therapie“ und „Diagnosestellung“ (siehe

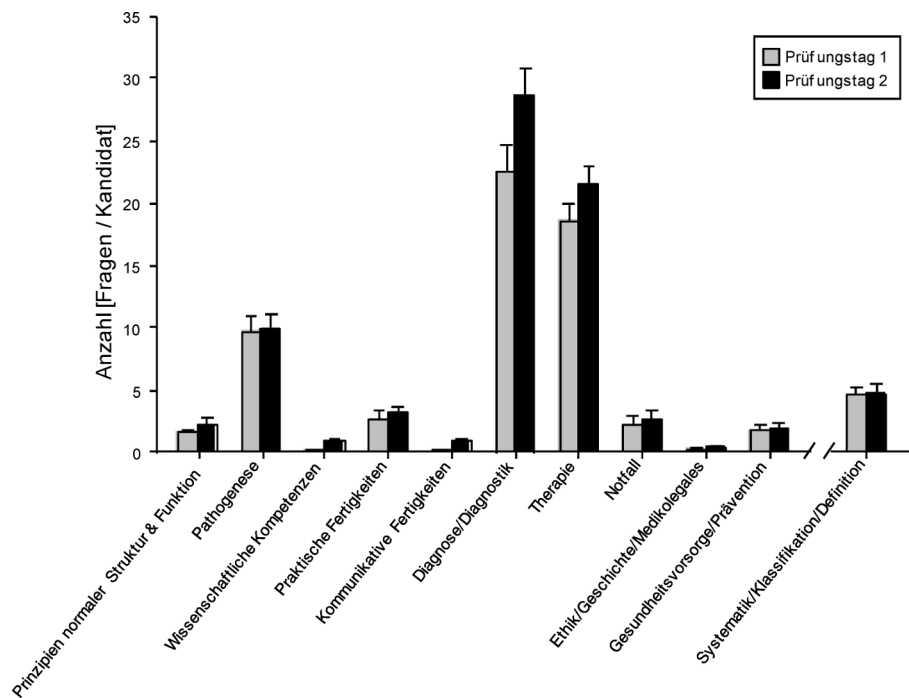


Abbildung 4: Prüfung im Prüfungsraum (außerhalb Patientenzimmer)

Zahl der Prüfungsfragen/aufgaben pro Kandidat mit Bezug auf die jeweilige Kategorie an den beiden Prüfungstagen; Daten sind Mittelwerte $\pm$ SEM; es gab keine signifikanten Unterschiede ($p < 0,05$) der Anzahl Fragen mit Bezug auf die jeweilige Kategorie an Tag 2 gegenüber Tag 1 (t-Test).

Abbildung 4). Nur eine Minderheit der gestellten Prüfungsaufgaben zielte auf Notfallversorgung, und praktisch keine zielte auf medico-legale, ethische oder präventionsbezogene Aspekte oder auf Aspekte der Gesundheitsförderung in der Gesellschaft. Ebenfalls so gut wie nicht vertreten waren Fragen zu Wissenschaftlichkeit/Forschung oder zu Ärztlicher Gesprächsführung. Klinisch-praktische Aufgaben („skills“), z.B. anhand von Übungsmodellen, wurden nur in geringem Maß abgeprüft, obwohl die gesetzlich-formalen Rahmenbedingungen dies an beiden Prüfungstagen ermöglicht hätten.

3.5. Die gestellten Prüfungsfragen bewegen sich auf relativ niedrigem Taxonomie-Level

Zusätzlich zu den oben genannten inhaltlich-thematischen Beobachtungskategorien wurden alle gestellten Prüfungsaufgaben von den Beobachtern „live“ je einer von drei relativ rasch zuordenbaren Taxonomie-Levels zugewiesen: „Wissen/Reproduktion“, „Analyse/Interpretation“, „Problemlösung“. Mehrheitlich (70% an beiden Prüfungstagen) wurden die gestellten Fragen von den Beobachtern der relativ niedrigen Taxonomiestufe der reinen Wissensfragen zugeordnet. Weniger als 15% der Fragen bewegten sich auf dem Level von Analyse/Interpretation, auch das Level „Problemlösung“ wurde am jeweils ersten Prüfungstag in weniger als 10%, am zweiten Prüfungstag in weniger als 20% der Prüfungsfragen adressiert. Insgesamt gab es unter taxonomischen Aspekten keine signifikanten Differenzen zwischen den beiden Prüfungstagen (siehe Abbildung 5).

Abbildung 6 verdeutlicht, dass die beobachteten Prüfungssituationen geprägt waren von traditionellen inhaltlichen Kategorien, die auf relative niedriger Taxonomiestufe abgeprüft wurden. Selbst in der Kategorie Therapie/therapeutische Strategien wurde nicht vermehrt auf höhere taxonomische Levels wie Analyse, Interpretation oder Problemlösung gesetzt.

4. Diskussion

Ethymologisch betrachtet, leitet sich das Wort Examen vom lateinischen „examinare“ ab, was soviel wie „wiegen, abwägen, messen“ meint [16]. Um das Kompetenzprofil eines Kandidaten in einer medizinischen Abschlussprüfung möglichst genau zu messen, bedarf es der Überprüfung multipler medizinischer und persönlicher Kompetenzen [6], [17], [18], [19]. Die Ärztliche Approbationsordnung bietet mit der Kombination aus 2-tägigem MC-Test und 2-tägigem mündlich-praktischen Examen grundsätzlich diese Chance, denn jedes dieser Formate zielt bei kluger Nutzung der Formate auf unterschiedliche berufsrelevante Kompetenzen [20].

Die vorliegende Studie zeigt allerdings eine noch deutlich wahrnehmbare Diskrepanz zwischen der gewünschten, von klinischen Lehrern wie von Prüfern gleichermaßen als relevant eingestuften Kompetenzorientierung [21], [22] und der realen Prüfungssituation in der mündlich-praktischen Abschlussprüfung, in der noch immer der Trend zum Abprüfen von reinem Reproduktionswissen im Rahmen des medizinischen Fachexperten besteht. Weder unmittelbar am Patientenbett noch im Prüfungs-

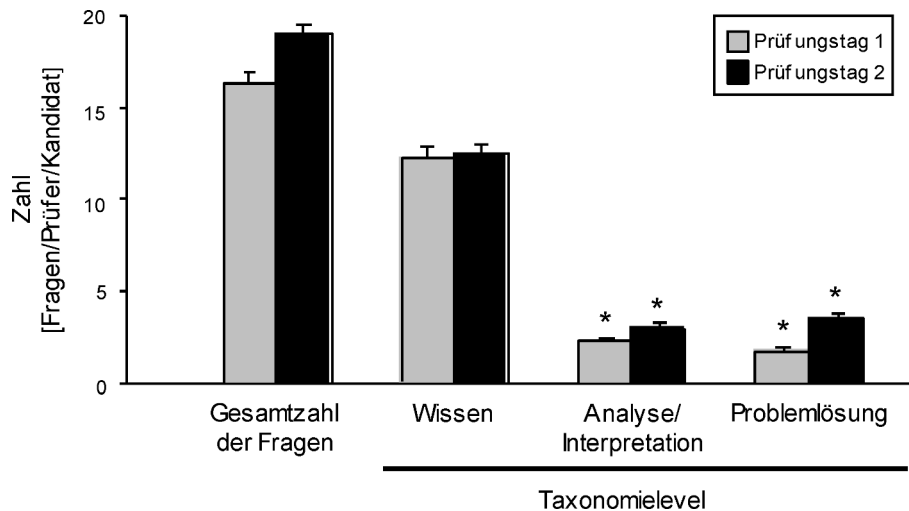


Abbildung 5: Prüfung im Prüfungsraum (außerhalb Patientenzimmer) Taxonomielevel; Zahl der Prüfungsfragen auf den Taxonomielevels (reproduzierendes) Wissen, Analyse/Interpretation, Problemlösung an den beiden Prüfungstagen; Daten sind Mittelwerte $\pm$ SEM; * $p < 0,05$ gegenüber (reproduzierenden) Wissensfragen (One-Way-Anova).

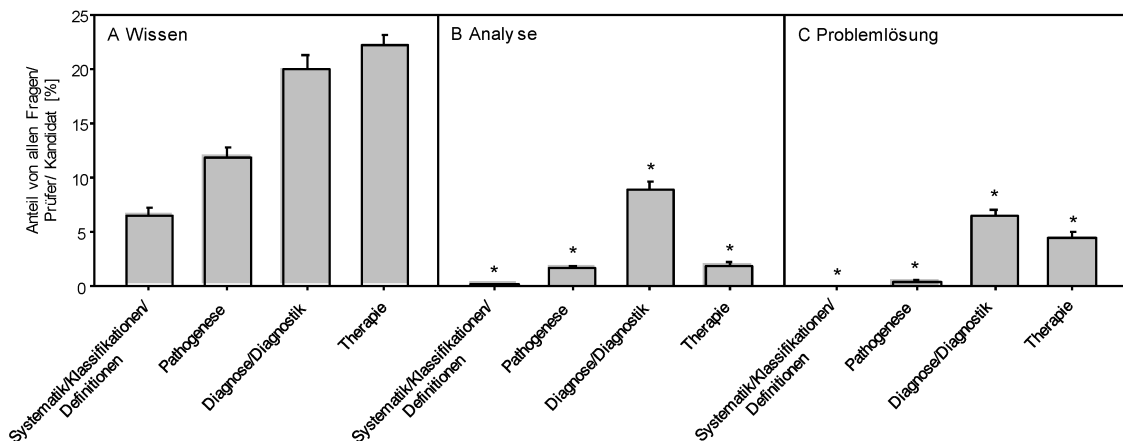


Abbildung 6: Prüfung im Prüfungsraum (außerhalb Patientenzimmer) Prozentsatz der Fragen im Hinblick auf die unterschiedlichen Kategorien pro Prüfer und Kandidat an den beiden Prüfungstagen und auf die drei Taxonomielevels A (reproduzierendes) Faktenwissen, B Analyse/Interpretation, C Problemlösung. Daten sind Mittelwerte $\pm$ SEM; * $p < 0,05$ gegenüber (reproduzierenden) Wissensfragen (One-Way-Anova).

raum wurden Rollen jenseits des Medizinischen Experten in nennenswertem Umfang in die Prüfungsinhalte mit einbezogen. Selbst die Rolle des Kommunikators, die verhältnismäßig leicht in den Prüfungsteil am Patientenbett zu integrieren wäre, war kein relevanter Prüfungsschwerpunkt, trotz ihrer Bedeutung für den ärztlichen Alltag, die sogar medikolegale Aspekte umfasst [23]. Prüfungsinhalte mit klarem Bezug zu den Rollen des „professionell Handelnden“, des „Gesundheitsberaters und -fürsprechers“ oder des „Gelehrten“, wie sie im NKLM formuliert sind, waren so gut wie nicht vorhanden. Insbesondere das Fehlen von Prüfungsinhalten der letztgenannten Rolle, des „Gelehrten“, wirkt überraschend, denn die Prüfer waren sämtlich neben ihrer klinischen Tätigkeit auch wissenschaftlich aktiv, so dass die Rolle des Wissenschaftlers durchaus einen signifikanten Raum in ihrem Berufsalltag einnimmt. In gewissem Sinn ist das Fehlen solcher Prüfungsinhalte sogar alarmierend, denn das Beherrschen wissenschaftlicher Kompetenzen stellt eine herausragende akademische Fähigkeit dar. Diese Fähig-

keit wird studierendenseitig grundsätzlich positiv bewertet [24], und hat hohen Impact auf unterschiedliche Aspekte ärztlichen Handelns, wie beispielsweise klinische Entscheidungsfindung, selbstkritisches und fallbezogenes Reflektieren, evidenzbasiertes Arbeiten, sogar auf Aspekte verbesserter Patientensicherheit [25]. Für künftige Examensprüfungen wird sicherlich zu diskutieren sein, in welchem Umfang bzw. in welcher Gewichtung und in welcher Form die unterschiedlichen im NKLM definierten Rollen in die Prüfungsaufgaben einfließen können oder sollen. Aber auch innerhalb der in den untersuchten Examina hauptsächlich adressierten Rolle des „Medizinischen Experten“ kann bzw. könnte sicherlich mehr als nur rein reproduzierendes Faktenwissen geprüft werden. Mündlich-praktische Prüfungsteile bieten sich für die Überprüfung der höheren taxonomischen Levels der Miller-Pyramide prinzipiell an (Knows – Knows how – Shows how – Does). Gerade die in den aktuellen Examensregularen für den Tag 1 vorgeschriebene Prüfungszeit am

Krankenbett bietet dafür den perfekten Rahmen. Das baden-württembergische Landesprüfungsamt empfiehlt, hierfür 15-20 Minuten pro Kandidat vorzusehen. In diesem Zeitkorridor könnten praktische (z.B. Untersuchungs-) oder auch anamnestiche Kompetenzen überprüft werden, anstelle des in vorliegender Studie beobachteten theoretisch geprägten „Fragensturms“ mit Betonung reproduzierenden Faktenwissens. Theoriefragen hätten ausreichend Raum in den Prüfungsabschnitten, die sich außerhalb des Patientenzimmers abspielen. Aber auch in den Prüfungsabschnitten ohne realen Patientenkontakt bieten die aktuell geltenden Examensvorgaben den Prüfern genügend Spielraum, um die ärztliche Kompetenz der Kandidaten unter verschiedenen Perspektiven und Aspekten auszuloten, sei es beispielsweise mit Hilfe von Übungsmodellen oder von Bildsequenzen medizinischer Problemsituationen. In unserer Analyse waren die Resultate an beiden Prüfungstagen praktisch gleich: Rollen jenseits des „Medizinischen Experten“ und Inhalte jenseits der klassischen Kategorien „Diagnose und Therapie“ hatten als Prüfungsgegenstände so gut wie keine Bedeutung. Die wahrscheinlichste Erklärung für dieses Phänomen dürfte darin liegen, dass das Bewusstsein für die Prüfungsrelevanz der unterschiedlichen Rollen und Inhaltskategorien (wie sie der NKLM als Lernziel definiert) sowie das entsprechende Prüfungs-Know-How seitens der Examensprüfer noch nicht genügend ausgeprägt ist. Diesem Defizit könnte künftig abgeholfen werden, indem z.B. in Prüfertrainings ein deutlicher Fokus auf die Wertigkeit und auf die Wichtigkeit einer umfassenden Kompetenzüberprüfung sowie auf das Trainieren der entsprechenden Prüferfertigkeiten gelegt wird [9], [18], [26]. Allerdings erreichen viele dieser Trainingsangebote vor allem diejenigen Kollegen und Kolleginnen, die als Novizen in die Examensprüfungen einsteigen; bereits aktive und erfahrene Prüfer sind oft deutlich schwerer zu erreichen. Deshalb müssen zur realen Umsetzung der Idee der Kompetenzorientierung in der medizinischen Ausbildung (d.h. in der Lehre und in den Prüfungen) und zur Bewusstmachung des Mehrwerts dieser Idee gezielt weitere Maßnahmen der Fakultäts- und Personalentwicklung ausgeschöpft werden [19], [20], [27], [28], [29]. Die aktuell gültige Ärztliche Approbationsordnung erlaubt und unterstützt zwar in einem gewissen Rahmen die kreative Ausgestaltung des mündlich-praktischen Examensteils im Hinblick auf kompetenzorientiertes Prüfen, stellt aber gleichzeitig auch gewisse Hürden auf. So sind bestimmte, deutlich stärker kompetenzorientierte Prüfungsformate wie z.B. OSCE (Objective Structured Clinical Examinations) im mündlich-praktischen Staatsexamen in Deutschland bisher nicht ausdrücklich vorgesehen, im Gegensatz zu den Regelungen in der Schweiz oder auch zu den Gepflogenheiten in vielen deutschen innerfakultären Leistungsnachweisen; auch an verbindlichere Vorgaben für mehrdimensionale strukturierte mündliche Prüfungen könnte hier künftig gedacht werden [2], [12], [30], [31], [32], [33].

Unserer Kenntnis nach stellt die vorliegende Studie die erste direkte „Live“-Beobachtung und Analyse des

mündlich-praktischen Staatsexamensabschnitts in Deutschland dar (zum Zeitpunkt der Studiendurchführung „mündlich-praktischer Teil des Zweiten Abschnitts der Ärztlichen Prüfung“). Einige Limitationen der Studie sind jedoch zu beachten. Die Daten wurden nur an einem Zentrum erhoben; für eine abgesicherte Übertragbarkeit der Ergebnisse müssten die Examensprüfungen weiterer Studienorte analysiert werden. Es wurde eine akzeptable Interrater-Reliabilität erreicht; weitere aus statistischer Sicht wünschenswerte Maßnahmen (mehrere Untersucher/Beobachter pro Prüfung, audiovisuelle Aufnahme der Prüfungen mit anschließender Analyse) mussten aufgrund formaler juristischer Erwägungen unterbleiben. Auch ging der Charakter der „Live“ Beobachtung als solcher mit der Einschränkung einher, dass sich die Beobachter in ihrer Echtzeitanalyse auf explizit geäußerte Aspekte der Prüfungsaufgaben beschränken mussten; implizit enthaltene Aspekte konnten keine Berücksichtigung finden. Außerdem muss betont werden, dass die Idee kompetenzorientierter Mediziner Ausbildung zwar deutlich weiter zurückreicht als nur bis zur Publikation des NKLM, dass aber dennoch der NKLM als solcher erst vor sehr kurzer Zeit veröffentlicht wurde [4], [34]. Insofern muss den Fakultäten, den Dozenten und den Prüfern Zeit gegeben werden, um die Kompetenzorientierung in den Curricula und in den Fachbereichen zu etablieren. Der NKLM hat Ziele formuliert; das Konzept des „constructive alignment“ würde nun deren Umsetzung sowohl in der Lehre als auch in den Prüfungen erfordern. Für die Zukunft der medizinischen Staatsexamina in Deutschland birgt der NKLM jedenfalls ein enormes Entwicklungspotential in Richtung auf Kompetenzorientierung im zentralen Prüfungswesen.

5. Schlussfolgerung

Zusammenfassend kann festgestellt werden, dass die vorliegende Studie eine deutliche Lücke zwischen den Kompetenzen und Lernzielen, wie sie etwa im NKLM formuliert sind, und den im mündlich-praktischen Staatsexamensteil nach dem Praktischen Jahr realiter überprüften Lernzielen aufdeckt. Die untersuchten Examensprüfungen folgten einer eher einfachen und traditionellen Linie, mit einer hohen Zahl an Fragen auf relativ niedrigem Taxonomielevel, und nur einem geringen Anteil praktischer Aufgaben. Gemäß des Paradigmas von „assessment drives learning“ muss der multidimensionale Kompetenzbegriff nicht nur in den Curricula, sondern auch in den Prüfungen eine wahrnehmbar wichtige Rolle spielen. Möglicherweise können derartige Weiterentwicklungen langfristig einen wirksamen Beitrag zu weiter verbesserter Patientenversorgung und Patientensicherheit leisten [35], [36].

Danksagung

Die Autoren bedanken sich bei Stephanie Denk für ihre Mithilfe bei den grafischen Darstellungen. Ganz besonderer Dank gilt Prof. Dr. med. Jana Jünger für ihre Beiträge zur verbesserten Konzeption des Studiendesigns, und Dr. Robert Blakytyn für das fachlich kompetente Korrekturlesen der englischsprachigen Texte.

Liste der verwendeten Abkürzungen

- ANOVA: Analysis of variance
- CanMEDS: Canadian Medical Education Directives for Specialists
- MME: Master of Medical Education
- NKLM: Nationale Kompetenzbasierte Lernzielkatalog Medizin
- OSCE: Objective-Structured Clinical Examination
- SEM: Standard-Fehler (Standard error of the mean)

Ethikantrag

Die unabhängige Ethikkommission der Universität Ulm hat dem Vorhaben zugestimmt (17122013).

Datenverfügbarkeit

Die Daten sind in der Klinik für Unfallchirurgie, Hand , Plastische und Wiederherstellungschirurgie des Universitätsklinikums Ulm gespeichert. Ein Datensharing findet nicht statt, da die sehr umfangreichen Excel Files auch Daten inkludieren, die außerhalb des Fokus dieser Publikation liegen.

Interessenkonflikt

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

Literatur

1. Frank JR, Danoff D. The Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
2. Oechsner W, Forster J. Approbierte Ärzte – kompetente Ärzte?: Die neue Approbationsordnung für Ärzte als Grundlage für kompetenzbasierte Curricula. *GMS Z Med Ausbild*. 2005;22(1):Doc04. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000004.shtml>
3. Van den Akker M, Dornan T, Scherpbier A, Egbrink M, Snoeckx LEasing the transition: the final year of medical education at Maastricht University. *Z Evid Fortbild Qual Gesundhwes*. 2012;106(2):92-97. DOI: 10.1016/j.zefq.2012.02.013
4. Fischer MR, Bauer D, Mohn K. Finally finished! National Competence Based Catalogues of Learning Objectives for Undergraduate Medical Education (NKLM) and Dental Education (NKLZ) ready for trial. *GMS Z Med Ausbild*. 2015;32(3):Doc35. DOI: 10.3205/zma000977
5. Hahn EG, Fischer MF. National Competence-Based Learning Objectives for Undergraduate Medical Education (NKLM) in Germany: Cooperation of the Association for Medical Education (GMA) and the Association of Medical Faculties in Germany (MFT). *GMS Z Med Ausbild*. 2009;26(3):Doc35. DOI: 10.3205/zma000627
6. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27(4):294-297. DOI: 10.1080/01421590500126437
7. Wilson GM, Lever R, Harden RMcg, Robertson JI, Macritchie J. Examination of clinical examiners. *Lancet*. 1969;293(7584):37-40. DOI: 10.1016/S0140-6736(69)90998-2
8. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-396. DOI: 10.1056/NEJMr054784
9. Oechsner W, Geiler S, Huber-Lang M. Effects and sustainability of trainings for the oral and practical part of the German final exam in medicine. *GMS Z Med Ausbild*. 2013;30(3):Doc36. DOI: 10.3205/zma000879
10. Van der Vleuten CP. Validity of final examinations in undergraduate medical training. *BMJ*. 2000;321(7270):1217-1219. DOI: 10.1136/bmj.321.7270.1217
11. Newstead S. Examining the examiners: Why are we so bad at assessing students? *Psych Learn Teach*. 2002;2:70-75. DOI: 10.2304/plat.2002.2.2.70
12. Jefferies A, Simmons B, Ng E, Skidmore M. Assessment of multiple physician competencies in postgraduate training: utility of the structured oral examination. *Adv Health Sci Educ Theory Pract*. 2011;16(5):569-577. DOI: 10.1007/s10459-011-9275-6
13. Pabst R. Medical education and reform initiatives in Germany. *Acad Med*. 1995;70(11):1006-1011. DOI: 10.1097/00001888-199511000-00019
14. Chenot JF. Undergraduate medical education in Germany. *Medizinstudium in Deutschland*. *GMS Ger Med Sci*. 2009;7:Doc02. DOI: 10.3205/000061
15. Bloom B. Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay; 1956.
16. Kluge F. Etymologisches Woerterbuch der Deutschen Sprache. 25th ed. Berlin: De Gruyter; 2012. DOI: 10.1515/9783110223651
17. Joughin G. Dimensions of oral assessment. *Assess Eval High Educ*. 1998;23:367-378. DOI: 10.1080/0260293980230404
18. 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
19. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach*. 2003;25(6):569-584. DOI: 10.1080/0142159032000151907
20. Stern DT, Wojtczak A, Schwarz MR. The assessment of global minimum essential requirements in medical education. *Med Teach*. 2003;25(6):589-595. DOI: 10.1080/0142159032000151295

21. Jilg S, Moeltner A, Berberat P, Fischer MR, Breckwoldt J. How do supervising clinicians of a University Hospital and associated teaching hospitals rate the relevance of the key competencies within the CanMEDS roles framework in respect to teaching in clinical clerkships? *GMS Z Med Ausbild.* 2015;32(3):Doc33. DOI: 10.3205/zma000975
22. Oechsner W, Geiler S, Huber-Lang M. The impact of a dedicated training program for oral examiners at a medical school in Germany: a survey among participants from operative and non-operative disciplines. *Patient Saf Surg.* 2013;7(1):22. DOI: 10.1186/1754-9493-7-22
23. Tamblyn R, Abrahamowicz M, Dauphinee D, Wenghofer E, Jacques A, Klass D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA.* 2007;298(9):993-1001. DOI: 10.1001/jama.298.9.993
24. Pruskil S, Burgwinkel P, Georg W, Keil T, Kiessling C. Medical students' attitudes towards science and involvement in research activities: a comparative study with students from a reformed and a traditional curriculum. *Med Teach.* 2009;31(6):e254-e259. DOI: 10.1080/01421590802637925
25. Huber-Lang M, Neugebauer E. The dual role of academic surgeons as clinicians and researchers - an attempt to square the circle? *Patient Saf Surg.* 2011;5:16. DOI: 10.1186/1754-9493-5-16
26. Schickler A, Bruestle P, Biller S. The Final Oral/Practical State Examination at Freiburg Medical Faculty in 2012 – Analysis of grading to test quality assurance. *GMS Z Med Ausbild.* 2015;32(4):Doc39. DOI: 10.3205/zma000981
27. Bland CJ, Starnaman S, Wersal L, Moorhead-Rosenberg L, Zonia S, Henry R. Curricular change in medical schools: how to succeed. *Acad Med.* 2000;75(6):575-594. DOI: 10.1097/00001888-200006000-00006
28. Pruskil S, Deis N, Druener S, Kiessling C, Philipp S, Rockenbach K. Implementation of "social and communicative competencies" in medical education. The importance of curriculum, organizational and human resource development. *GMS Z Med Ausbild.* 2015;32(5):Doc50. DOI: 10.3205/zma000992
29. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach.* 2006;28(6):497-526. DOI: 10.1080/01421590600902976
30. Berendonk C, Schirlo C, Balestra G, Bonvin R, Feller S, Huber P, Jünger E, Monti M, Schanbel K, Beyeler C, Guttormsen S, Huwendiek S. The new final clinical skills examination in human medicine in Switzerland: essential steps of exam development, implementation and evaluation, and central insights from the perspective of the national Working Group. *GMS Z Med Ausbild.* 2015;32(4):Doc40. DOI: 10.3205/zma000982
31. Guttormsen S, Beyeler C, Bonvin R, Feller S, Schirlo C, Schnabel K, Schurter T, Berendonk C. The new licensing examination for human medicine: from concept to implementation. *Swiss licensing examination for human medicine. Swiss Med Wkly.* 2013;43:w13897.
32. Dwyer T, Glover Takahashi S, Kennedy Hynes M, Herold J, Wasserstein D, Nousiainen M, Ferguson P, Wadey V, Murnaghan ML, Leroux T, Semple J, Hodges B, Ogilvie-Harris D. How to assess communication, professionalism, collaboration and the other intrinsic CanMEDS roles in orthopedic residents: use of an objective structured clinical examination (OSCE). *Can J Surg.* 2014;57(4):230-236. DOI: 10.1503/cjs.018813
33. Nguyen LN, Tardioli K, Roberts M, Watterson J. Development and incorporation of hybrid simulation OSCE into in-training examinations to assess multiple CanMEDS competencies in urologic trainees. *Can Urol Assoc J.* 2015;9(1-2):32-36. DOI: 10.5489/cuaj.2366
34. Kadmon M, Bender MJ, Adili F, Arbab D, Heinemann MK, Hofmann HS, König S, Küper MA, Obertacke U, Rennekampff HO, Rolle U, Rücker M, Sader R, Tingart M, Tolksdorf MM, Tronnier V, Will B, Walcher F; Chirurgische Arbeitsgemeinschaft Lehre der Deutschen Gesellschaft für Chirurgie (CAL). Competency-based medical education: National Catalogue of Learning Objectives in surgery. *Chirurg.* 2013;84(4):277-285. DOI: 10.1007/s00104-012-2449-9
35. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? *Anat Sci Educ.* 2009;2(5):199-204. DOI: 10.1002/ase.102
36. Nikendei C, Kraus B, Schrauth M, Briem S, Juenger J. Ward rounds: how prepared are future doctors? *Med Teach.* 2008;30(1):88-91. DOI: 10.1080/01421590701753468

Korrespondenzadresse:

Prof. Dr. med. Markus Huber-Lang
 Universitätsklinikum Ulm, Medizinische Fakultät, Institut
 für klinische und experimentelle Trauma-Immunologie,
 Helmholtzstr. 8/1, 89081 Ulm, Deutschland, Tel: +49
 (0)731/500-54800, Fax: +49 (0)731/500-5418
 markus.huber-lang@uniklinik-ulm.de

Bitte zitieren als

Huber-Lang M, Palmer A, Grab C, Boeckers A, Boeckers TM, Oechsner W. Visions and reality: the idea of competence-oriented assessment for German medical students is not yet realised in licensing examinations. *GMS J Med Educ.* 2017;34(2):Doc25. DOI: 10.3205/zma001102, URN: urn:nbn:de:0183-zma0011024

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2017-34/zma001102.shtml>

Eingereicht: 19.08.2016

Überarbeitet: 21.02.2017

Angenommen: 07.03.2017

Veröffentlicht: 15.05.2017

Copyright

©2017 Huber-Lang et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.