

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

Abstract

Objective: Since 2011, the new national final examination in human medicine has been implemented in Switzerland, with a structured clinical-practical part in the OSCE format. From the perspective of the national Working Group, the current article describes the essential steps in the development, implementation and evaluation of the Federal Licensing Examination Clinical Skills (FLE CS) as well as the applied quality assurance measures. Finally, central insights gained from the last years are presented.

Methods: Based on the principles of action research, the FLE CS is in a constant state of further development. On the foundation of systematically documented experiences from previous years, in the Working Group, unresolved questions are discussed and resulting solution approaches are substantiated (planning), implemented in the examination (implementation) and subsequently evaluated (reflection). The presented results are the product of this iterative procedure.

Results: The FLE CS is created by experts from all faculties and subject areas in a multistage process. The examination is administered in German and French on a decentralised basis and consists of twelve interdisciplinary stations per candidate. As important quality assurance measures, the national Review Board (content validation) and the meetings of the standardised patient trainers (standardisation) have proven worthwhile. The statistical analyses show good measurement reliability and support the construct validity of the examination. Among the central insights of the past years, it has been established that the consistent implementation of the principles of action research contributes to the successful further development of the examination.

Conclusion: The centrally coordinated, collaborative-iterative process, incorporating experts from all faculties, makes a fundamental contribution to the quality of the FLE CS. The processes and insights presented here can be useful for others planning a similar undertaking.

Keywords: national final examination, licensing examination, summative assessment, OSCE, action research

Christoph Berendonk¹
Christian Schirlo²
Gianmarco Balestra³
Raphael Bonvin⁴
Sabine Feller¹
Philippe Huber⁵
Ernst Jünger²
Matteo Monti⁴
Kai Schnabel¹
Christine Beyeler¹
Sissel Guttormsen¹
Sören Huwendiek¹

1 University of Bern, Medical Faculty, Institute of Medical Education, Bern, Switzerland

2 University of Zurich, Medical Faculty, Zurich, Switzerland

3 University of Basel, Medical Faculty, Basel, Switzerland

4 University of Lausanne, Medical Faculty, Lausanne, Switzerland

5 University of Geneva, Medical Faculty, Geneva, Switzerland

Introduction

In 2011, the changeover from the previous State Examination to the new Federal Licensing Examination in human medicine came into effect. The impetus and legal basis for this new national final examination was the Federal Law on Medical Professions (MedBG) adopted in June 2006 [https://www.admin.ch/opc/de/official-compilation/2007/4031.pdf cited at 14.12.13]. The educational objectives laid down in the MedBG refer

to the “Swiss Catalogue of Learning Objectives for Undergraduate Medical Training” (SCLO) [1], which is mandatory for all five Swiss medical faculties and which lists and specifies the required competencies in detail. In February 2007, a national steering group was set up by the Federal Office of Public Health, which was entrusted with the responsibility for the strategic orientation of the new Federal Licensing Examination (FLE) in human medicine. This supreme steering committee resolved that the future FLE in human medicine should be an interdisciplinary exam-

ination with two parts. The first part is a written examination with a Multiple Choice (MC) format. The second part (Clinical Skills Examination) consists of a structured clinical-practical examination with standardised patients (SP) and is based on the principle of the Objective Structured Clinical Examination (OSCE) [2].

The overarching development process of the entire FLE in human medicine, from the concept to the first implementation, including the legal framework, the discourse about the available scientific evidence on the theme of summative assessments, and the development of the cross-examination blueprints, is described in detail elsewhere and is not included in this article [3]. The current work is concerned with the development of the FLE Clinical Skills (CS) as it has been administered since 2011. Structured clinical-practical examinations as part of licensing examinations at the end of undergraduate training are known from Canada [4] and the USA [5], and have been successfully conducted there for several years. Licensing examinations are also common in Europe, but a national structured clinical-practical examination constitutes a novelty in the European context. Although there is a wealth of literature concerned with the most diverse of aspects of an OSCE [6], barely any studies have been published which investigate the specific characteristics of developing national clinical-practical examinations. Therefore, the aim of the current article is to describe – from the perspective of the Working Group CS – essential steps for the development, implementation and evaluation of the examination, as well as quality assurance measures and central insights gained from the last three years.

Methods

The national Working Group Clinical Skills (WG CS) is constituted of members of the five medical faculties (Basel, Bern, Geneva, Lausanne, Zurich) and the Institute of Medical Education of the University of Bern. All members have many years of practical experience with the OSCEs conducted at the individual faculties. The procedure chosen by the WG CS for the further concept development of the FLE CS is based on the methodological principles of action research [7]. Key elements of this method are the iterative development process (planning-implementation-reflection) and the active, democratic participation of all stakeholders [8]. The primary focus is not on an individual scientific method, but on mutually defined questions, with the aim of continuously improving a product or process (in this case the FLE CS). Characteristic for action research are scientific questions grounded in everyday life, which lead to the generation of knowledge that is of relevance and interest outside of the local context [9].

The active participation of the WG CS members is ensured by eight half-day sessions per year, with the role of chairperson rotating from session to session. The knowledge of the individual Working Group members regarding

processes and pitfalls within OSCEs is of fundamental importance. In the phase of *planning*, the individual members contribute their own experiences; together, questions and goals are defined, possible problem areas are identified, and possible solutions are sought. In this phase, reference is made to the existing scientific evidence on the respective theme. Furthermore, in terms of corresponding questions, contact is sought with the National Board of Medical Examiners of the USA and the Medical Council of Canada, in order to benefit from their long-standing experience with national clinical-practical exams. Prior to the *implementation*, the strategy developed by the WG CS is submitted to the Swiss Board of Examiners for approval. In the case of uncertainties regarding the selected strategy, the concept is first of all piloted in a faculty-based OSCE, and is only then implemented in the FLE CS. In order to evaluate the implemented strategy, a multitude of different qualitative and quantitative methods is employed. In the phase of *reflection*, the experiences gathered and results obtained are discussed in the WG CS with regard to questions or goals. If there is a further need for action, the initial concept is revised and the cycle starts again from the beginning. Each step of the FLE CS goes through this three-stage process of planning-implementation-reflection. As it would be beyond the scope of the current article to describe the whole process for each step, only the results of these processes are presented here.

Results

The processes of exam development, implementation, evaluation and quality assurance as well as the central insights presented in the following cannot, in reality, be considered as separate entities. Rather, they are interwoven and mutually influence one another. This division is made here to facilitate understanding.

Exam development

Case development: The initial step of case development consists in identifying a theme which is suitable for the Clinical Skills (CS) examination. The foundation for possible themes is the “problems as starting points” chapter of the Swiss Catalogue of Learning Objectives and the requirements of the blueprint. Once a theme has been determined, an experienced clinician, selected according to the case theme, writes a first case draft, which is evaluated and commented on in terms of formal criteria by methodological experts. This draft is then developed further in the framework of a workshop together with a second clinician (from a different faculty and subject area), as well as a CS coach (person with experience of the CS exam format). In these workshops, not only is the actual case content developed, but also the weighting of the individual assessment criteria is defined. This team approach ensures that in the case development, the necessary medical subject competence is present (clinician)

and the exam-format-specific aspects are taken into account (CS coach). The practicability of the developed cases is tested in the workshop itself with standardised patients (SP). These test runs provide important hints, which often have a substantial influence on the further case development.

Structure of a Clinical Skills station: The history-taking and physical examination to be performed by the candidate, as well as the differential diagnoses and the further diagnostic and therapeutic procedures (so-called ASM criteria: anamnesis, status, management) are assessed by means of case-specific checklists. The communication between candidate and SP (so-called CC criteria), by contrast, is rated on a generic scale which comprises four aspects and is identical for all stations [10], [11].

Review process: In the next step, the cases are presented to the national Review Board. This consists of representatives of the five medical faculties and representatives from general medicine. The task of the Review Board is to evaluate the cases in terms of their correctness, relevance and complexity. Following a critical discussion, the Review Board specifies the corresponding need for amendment or approves the case for the further development process.

Standardisation: Thereafter, the cases are made available to the SP trainers. The SP trainers of all five faculties develop a uniform understanding of the cases in joint meetings. Each case is discussed, acted out and videotaped. The video documentation enables complex situations (e.g. presentation of the patient's mental health) or the degree of a certain symptom (e.g. neurological deficit) to be precisely defined. Within these meetings, it is also determined how, for instance, make-up can best be applied to represent a cutaneous efflorescence or how the room has to be set up. Where necessary, the experiences from the SP trainer meetings can be further specified or adapted in the role descriptions.

Translation: Once the role descriptions have been finalised in the SP trainer meetings, the case is translated from the original German or French into the respective other language. To achieve as high a degree of uniformity as possible, the cases are each translated by the same two physicians into their native language.

Quality assurance: With this centrally coordinated, multistage process of exam development, involving content- and format-specific experts from all five faculties, it is ensured that the developed cases fulfil the format-specific requirements of an OSCE, pass through a national validation process, and can be employed at all faculties under standardised conditions. In terms of content validation, the task and function of the Review Board is of central importance. The national SP trainer meetings, in turn, have a decisive influence on the standardisation process.

Exam implementation

The FLE CS is conducted once a year at all five medical faculties on a decentralised basis. Depending on the

number of candidates at the respective faculty, the examination is offered on two or three days. A different set of cases is used on each examination day; however, the examination sets are identical at all faculties on each respective day. The individual candidate completes 12 stations, each lasting for 13 minutes (with an additional two minutes rotation time between the stations and three 15-minute breaks). There is one physician-examiner in each room, and thus candidates are assessed by 12 different examiners in total. In addition to the case-specific ASM and generic CC criteria, the examiners assign two judgements on a global rating scale. These global judgements serve the purpose of calculating the pass mark.

Quality assurance: Detailed information (including video example) on the exam procedure is provided to the candidates in good time before the examination [<http://www.bag.admin.ch/themen/berufe/07918/07919/> cited 14.12.13]. Clinicians who are assessing in the FLE CS for the first time are provided with orientation about the OSCE format at an information event and are instructed regarding how to handle the assessment grid. During and after the examination, the examiners have the possibility to give content-based feedback on the cases in general and on the assessment criteria specifically. Additionally, individual members of the WG CS visit the five faculties to get an idea of the exam administration at the different sites.

Exam evaluation

In a first step, all comments of the examiners are evaluated, and the difficulty and discrimination of the assessment criteria are calculated. On this basis, criteria which show content-related or formal deficiencies are eliminated. Following the elimination, the total score for a candidate at one station is calculated from the sum of the weighted checklist points of the ASM criteria added to the weighted score of the communication scale (total score = $0.75 \cdot \text{ASM} + 0.25 \cdot \text{CC}$). The 12 stations all contribute equally to the total exam score. The candidates' total exam score is calculated as the mean value across the 12 station scores. The candidates' results are subsequently standardised in order to make the exam results comparable over the three days. In a next step, by means of the borderline regression method [12], [13], the pass mark per station is calculated. On this basis, the Board of Examiners determines the definitive pass mark. Table 1 shows the number of candidates separated according to groups with Swiss or non-European diploma and repeaters as well as the corresponding pass rates.

Quality indices: In table 2, some of the statistical quality indices of the FLE CS since 2011 are compiled. These results are in accordance with expectations insofar as the two exam formats MC and CS capture different aspects of medical competence but they (naturally) share the same basis. Moreover, the lower correlation of CC with MC (compared to ASM with MC) supports the construct validity of the examination.

Table 1: Number of candidates and pass rate for the Federal Licensing Examination Clinical Skills 2011-2013

Year	Diploma from a Swiss university	Diploma from a non-European university	Repeaters
2011	758 (98.3%)	7 (42.8%)	0 (-)
2012	785 (98.6%)	16 (50.0%)	16 (62.5%)
2013	786 (97.6%)	16 (56.3%)	17 (58.8%)

Number of candidates who took part in the Federal Licensing Examination Clinical Skills in 2011-2013. The percentage of candidates who passed the examination is indicated in brackets.

Table 2: Statistical quality indices of the Federal Licensing Examination Clinical Skills 2011-2013

Year	Reliability*	Correlation**		
		MC vs. CS	MC vs. CS ASM	MC vs. CS CC
2011	0.84-0.88	0.48	0.46	0.39
2012	0.86-0.90	0.52	0.51	0.36
2013	0.86-0.89	0.55	0.54	0.39

* Cronbach's alpha per day, minimum and maximum value respectively. Calculated on the level of the criteria.
 ** Correlations (Pearson) between the Multiple Choice (MC) and Clinical Skills (CS) examination, and between the MC examination and the areas ASM (Anamnesis-Status-Management) and CC (Communication), respectively, of the CS examination (all correlations are significant at $p < 0.01$)

Quality assurance: The data preparation and statistical analyses of the examination are undertaken according to established guidelines [14] centrally by one institution (Institute of Medical Education of the University of Bern). Additionally, the meticulous analysis of the examiners' comments on the individual cases, which enables the elimination of flawed assessment criteria, contributes further to the quality of the examination. The low number of eliminated items is, in turn, a measure for the quality of the exam development process (in 2013, 22 of the 807 criteria (2.7%) were eliminated).

Central insights from the last three years

The iterative process with planning-implementation-reflection is considered by the WG CS to be an essential factor for the successful further development of the FLE CS. To illustrate this process using an example: The examiners' feedback during the first two implementations indicated that dichotomous criteria are only partially suitable for assessing complex skills. Based on this experience, the members of the WG CS discussed possible alternative assessment criteria (planning). The goal was to adapt the assessment grid so that it can evaluate not only *whether* but also *how* a certain activity is performed. Based on these reflections, as well as consultation of the relevant literature, in the 2013 examination, rating scales were piloted (implementation). Results demonstrated that the correlations between scores on the global rating scale and scores achieved on the adapted checklist were higher at stations which contained these adapted rating scales compared to stations which contained only dichotomous criteria (reflection). The correlation of global judgement and total score is a criterion for the quality of an OSCE station. Accordingly, from 2014, where possible, adapted rating scales will be employed in the FLE CS for the assessment of complex skills.

Essential insights with regard to the individual steps of exam development, implementation and evaluation are described in the following. The centrally coordinated, collaborative process, with case development workshops, Review Board and national SP trainer meetings, and the incorporation of a large number of clinicians from all

faculties and subject areas, is an essential factor of *exam development*. This consensus-oriented process makes a decisive contribution to the individual case as well as to the quality of the examination as a whole.

Since the first *exam implementation* in 2011, the examiners have evaluated the cases in terms of their medical content-based quality. Additionally, since 2013, the quality of SP performance has been captured by means of a standardised assessment grid. The use of this grid, with its explicit measurement criteria, which is given to the SPs in advance, has contributed to a further increase in the quality of SP performance as well as a further harmonisation between the individual sites.

In terms of the *exam evaluation*, the need emerged for additional quantitative parameters, which help to better judge the quality of the individual cases. With a set of statistical parameters described by Pell and others, different quality aspects of a station can be evaluated [15]. These parameters provide important information for the future process of case development, as they provide hints regarding an improvement of the assessment criteria, or where a standardisation is particularly demanding.

Discussion

The current article presents the essential steps of exam development, implementation and evaluation, and provides central insights gained from the first three administrations of the Federal Licensing Examination Clinical Skills (FLE CS), from the perspective of the Working Group Clinical Skills.

A multistage procedure is recommended in international guidelines for the *development* of OSCE stations [16]. The centrally coordinated, collaborative-iterative process described here, with the participation of subject- and format-specific experts from all faculties, drives this principle forward and is an essential quality factor of the FLE CS. The *implementation* and *evaluation* of the FLE CS draws on the validated national Clinical Skills examinations from North America. For instance, the number of stations per candidate is identical to the United States Medical Licensing Examination Step 2 Clinical Skills

[<http://www.usmle.org/> cited 14.12.13], [17]. Additionally, the FLE CS has the following features in common with the Medical Council of Canada Qualifying Examination Part 2 [<http://mcc.ca/examinations/mccqe-part-ii/> cited 14.12.13], [18]:

1. The candidates' performance is assessed by clinicians of the respective teaching hospitals or universities;
2. The pass mark per station is calculated on the basis of expert assessments (global judgements);
3. The medical content and communicational aspects are added to form an overall score; and
4. The examination is offered in two languages.

Moreover, with regard to the pass rates, parallels between the US-American and the Swiss examination are apparent [19].

Among the central insights of the past years, one particular factor emerged: The consistent implementation of the principles of action research, in conjunction with the active participation of the many stakeholders, is essential for the successful further development of the FLE CS. On the whole, the FLE CS benefits from the experience of the OSCE examinations conducted at the individual faculties. However, this influence is not unidirectional. Rather, the faculty-based OSCEs have also benefited from the experiences with the FLE CS. Whereas, until recently, two to three persons at each faculty dealt locally with the OSCEs in isolation, since the introduction of the FLE CS, a Swiss-wide "community of practice" [20] has developed, which deals with the theme in a professional manner, cultivates a lively exchange, and strives to develop the structured clinical-practical examination further. The further development of this exam format, in particular for the use in licensing examinations, is not complete. Other institutions are also endeavouring to develop their national structured clinical-practical exams further [21].

At present, there are only a small number of countries worldwide with a national, structured clinical-practical examination [22]. Accordingly, only a small number of scientific projects have dealt with the development, implementation and evaluation of such examinations. The current article makes a first contribution to closing this gap. The essential steps and central insights documented here can be useful for interest groups, in particular with regard to the planning of national Clinical Skills examinations. However, it is also conceivable that the procedure we have described can be applied in 'university consortia', in order to offer a high-quality examination which exceeds the resources and expertise of one individual educational establishment.

The current work is limited to the perspective of the Working Group CS. It would be desirable in the future to also incorporate the experiences of candidates, lecturers and further stakeholders into the process. A further limitation of this work is that with such an extensive project, it is barely possible to provide a comprehensive and detailed description of all employed methods and the results obtained.

Conclusion

Based on the methodological foundation of action research, essential steps were presented which enable the successful implementation and further development of a national Clinical Skills examination, with which a candidate's practical abilities can be assessed in a fair, reliable and valid manner. The processes and experiences documented in the current article can also be useful for other organisations planning a similar undertaking.

Acknowledgements

A national Clinical Skills examination can only be accomplished through the commitment of countless persons. Our thanks go in particular to the case authors, members of the national Review Board, the Standardised Patient trainers, the Standardised Patients and the persons in charge at each site. We would also like to thank the clinicians for their work as examiners. Such an examination would also not be conceivable without the good collaboration and constant support of the Board of Examiners and the Federal Office of Public Health. Special thanks go also to Patrick Jucker-Kupper for the critical review of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Bürgi H, Rindlisbacher B, Bader C, Bloch R, Bosman F, Gasser C, Gerke W, Humair JP, Im Hof V, Kaiser H, Lefebvre D, Schläppi P, Sottas B, Spinass GA, Stuck AE. Swiss Catalogue of Learning Objectives for Undergraduate Medical Training. Bern: Universität Bern; 2008. Zugänglich unter/available from: <http://sco.smifk.ch>
2. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J.* 1975;1(5955):447-451. DOI: 10.1136/bmj.1.5955.447
3. Guttormsen S, Beyeler C, Bonvin R, Feller S, Schirlo C, Schnabel K, Schurter T, Berendonk C. The new licencing examination for human medicine: from concept to implementation. *Swiss Med Wkly.* 2013;143:w13897. Zugänglich unter/available from: <http://www.smw.ch/content/smw-2013-13897/>
4. Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L, Lacombe G, Baumber J, Poldre P, Levasseur L. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. *Acad Med.* 1992;67(8):487-494. DOI: 10.1097/00001888-199208000-00001
5. De Champlain A, Swygert K, Swanson DB, Boulet JR. Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Acad Med.* 2006;81(10 Suppl):S17-20.

6. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach*. 2013;35(9):e1437-1446.
7. Meyer J. Qualitative research in health care. Using qualitative methods in health related action research. *BMJ*. 2000;320(7228):178-181.
8. Edler A. Action research in medical education: a shifting paradigm or old wine in new skins? *Clin Teach*. 2009;6:139-410.
9. Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. *BMJ*. 2008;337:a567.
10. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ*. 2003;37(11):1012-1016. DOI: 10.1046/j.1365-2923.2003.01674.x
11. Scheffer S. Validierung des Berliner Global Rating" (BGR) -ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinisch-praktischer Prüfungen (OSCE). Berlin: Charité - Universitätsmedizin Berlin, Medizinische Fakultät; 2009.
12. Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Objective structured clinical examinations. *Med Educ*. 2003;37(2):132-139. DOI: 10.1046/j.1365-2923.2003.01429.x
13. Wood TJ, Humphrey-Murto SM, Norman GR. Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci Educ Theory Prac*. 2006;11(2):115-122. DOI: 10.1007/s10459-005-7853-1
14. Downing S. Statistics of testing. In: Downing S, Yudkowsky R (Hrsg). *Assessment in Health Professions Education*. London: Routledge; 2009. S.93-117.
15. Pell G, Fuller R, Homer M, Roberts T, International Association for Medical E. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Med Teach*. 2010;32(10):802-811. DOI: 10.3109/0142159X.2010.507716
16. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach*. 2013;35(9):e1447-1463.
17. Gilliland WR, La Rochelle J, Hawkins R, Dillon GF, Mechaber AJ, Dyrbye L, Papp K, Durning SJ. Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach*. 2008;30(3):325-327. DOI: 10.1080/01421590801953026
18. Woloschuk W, McLaughlin K, Wright B. Predicting performance on the Medical Council of Canada Qualifying Exam Part II. *Teach Learn Med*. 2013;25(3):237-241. DOI: 10.1080/10401334.2013.797351
19. Boulet JR, Swanson DB, Cooper RA, Norcini JJ, McKinley DW. A comparison of the characteristics and examination performances of U.S. and non-U.S. citizen international medical graduates who sought Educational Commission for Foreign Medical Graduates certification: 1995-2004. *Acad Med*. 2006;81(10 Suppl):S116-119.
20. Bereiter C. Design Research for Sustained Innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Sci Soc*. 2002;9(9):321-327.
21. Haist SA, Katsufakis PJ, Dillon GF. The evolution of the United States Medical Licensing Examination (USMLE): enhancing assessment of practice-related competencies. *JAMA*. 2013;310(21):2245-2246. DOI: 10.1001/jama.2013.282328
22. McCrorie P, Boursicot KA. Variations in medical school graduating examinations in the United Kingdom: are clinical competence standards comparable? *Med Teach*. 2009;31(3):223-229. DOI: 10.1080/01421590802574581

Corresponding author:

Dr. med. Christoph Berendonk, MME
University of Bern, Medical Faculty, Institute of Medical Education, Konsumstraße 13, CH-3010 Bern, Switzerland
christoph.berendonk@iml.unibe.ch

Please cite as

Berendonk C, Schirlo C, Balestra G, Bonvin R, Feller S, Huber P, Jünger E, Monti M, Schnabel 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, URN: urn:nbn:de:0183-zma0009828

This article is freely available from

<http://www.egms.de/en/journals/zma/2015-32/zma000982.shtml>

Received: 2014-01-14

Revised: 2014-10-30

Accepted: 2015-01-12

Published: 2015-10-15

Copyright

©2015 Berendonk 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/>.

Die neue klinisch-praktische Schlussprüfung in Humanmedizin der Schweiz: Essentielle Schritte der Prüfungserstellung, Durchführung und Auswertung sowie zentrale Erkenntnisse aus Sicht der nationalen Arbeitsgruppe Clinical Skills

Zusammenfassung

Zielsetzung: Seit 2011 wird in der Schweiz eine neue nationale Schlussprüfung in Humanmedizin mit einem strukturierten klinisch-praktischen Teil im OSCE Format durchgeführt. Der vorliegende Beitrag beschreibt, aus Sicht der nationalen Arbeitsgruppe, die essentiellen Schritte der Erstellung, Durchführung und Auswertung der Eidgenössischen Prüfung Clinical Skills (EP CS) sowie der jeweils eingesetzten Qualitätssicherungsmaßnahmen. Schliesslich werden zentrale Erkenntnisse der vergangenen Jahre vorgestellt.

Methodik: Basierend auf den Prinzipien der Aktionsforschung wird die EP CS kontinuierlich weiter entwickelt. Auf der Grundlage systematisch dokumentierter Erfahrungen aus den Vorjahren werden in der Arbeitsgruppe anstehende Fragen diskutiert, daraus resultierende Lösungsansätze konkretisiert (Planung), in der Prüfung umgesetzt (Implementation) und anschliessend evaluiert (Reflexion). Die vorgestellten Ergebnisse sind das Produkt dieses iterativen Vorgehens.

Ergebnisse: Die EP CS wird durch Fachexperten aller Fakultäten und Fachrichtungen in einem mehrstufigen Prozess erstellt. Die Prüfung wird dezentral in Deutsch und Französisch durchgeführt und besteht aus zwölf interdisziplinären Stationen pro Kandidat. Als wichtige Qualitätssicherungsmaßnahmen haben sich das nationale Review Board (inhaltliche Validierung) und die Treffen der Schauspielpatienten-Trainer (Standardisierung) erwiesen. Die statistischen Auswertungen zeigen eine gute Messzuverlässigkeit und weisen auf die Konstruktvalidität der Prüfung hin. Zu den zentralen Erkenntnissen der vergangenen Jahre gehört, dass die konsequente Umsetzung der Prinzipien der Aktionsforschung zur erfolgreichen Weiterentwicklung der Prüfung beiträgt.

Schlussfolgerung: Der zentral koordinierte, kollaborativ-iterative Prozess mit Einbindung von Fachexperten aus allen Fakultäten trägt wesentlich zur Qualität der EP CS bei. Die hier vorgestellten Prozesse und Erkenntnisse können für andere nützlich sein, welche Ähnliches vorhaben.

Schlüsselwörter: nationale Schlussprüfung, lizenzierende Prüfung, summatives Assessment, OSCE, Aktionsforschung

Christoph Berendonk¹
Christian Schirlo²
Gianmarco Balestra³
Raphael Bonvin⁴
Sabine Feller¹
Philippe Huber⁵
Ernst Jünger²
Matteo Monti⁴
Kai Schnabel¹
Christine Beyeler¹
Sissel Guttormsen¹
Sören Huwendiek¹

1 Universität Bern,
Medizinische Fakultät,
Institut für Medizinische
Lehre, Bern, Schweiz

2 Universität Zürich,
Medizinische Fakultät,
Zürich, Schweiz

3 Universität Basel,
Medizinische Fakultät, Basel,
Schweiz

4 Universität Lausanne,
Medizinische Fakultät,
Lausanne, Schweiz

5 Universität Genf,
Medizinische Fakultät, Genf,
Schweiz

Einleitung

Im Jahr 2011 wurde in der Schweiz die Umstellung vom bisherigen Staatsexamen auf die neue Eidgenössische Prüfung Humanmedizin vollzogen. Anstoss und gesetzli-

che Grundlage für diese neue nationale Schlussprüfung ist das im Juni 2006 verabschiedete Medizinalberufegesetz (MedBG) [<https://www.admin.ch/opc/de/official-compilation/2007/4031.pdf> zitiert am 14.12.13]. Die im MedBG festgehaltenen Ausbildungsziele verweisen auf den für alle fünf schweizerischen Medizinischen Fakultäten verbindlichen Lernzielkatalog „Swiss Catalogue

of Learning Objectives for Undergraduate Medical Training“ (SCLO) [1], der im Detail die verlangten Kompetenzen auflistet und präzisiert. Im Februar 2007 wurde vom Bundesamt für Gesundheit eine nationale Steuerungsgruppe ins Leben gerufen, welcher die Verantwortung für die strategische Ausrichtung der neuen Eidgenössischen Prüfung (EP) Humanmedizin übertragen wurde. Dieses oberste Steuerungsorgan entschied, dass die zukünftige EP Humanmedizin eine fächerübergreifende Prüfung mit zwei Teilen sein soll. Im ersten Teil wird im Multiple Choice (MC) Format schriftlich geprüft. Der zweite Teil (Clinical Skills Prüfung) besteht aus einer strukturierten klinisch-praktischen Prüfung mit Schauspielpatienten (SP) und basiert auf dem Prinzip der Objective Structured Clinical Examination (OSCE) [2].

Der übergreifende Entwicklungsprozess der gesamten EP Humanmedizin vom Konzept bis zur ersten Durchführung, mit den rechtlichen Rahmenbedingungen, dem Diskurs über die vorhandene wissenschaftliche Evidenz zum Thema summative Prüfungen und der Entwicklung des prüfungsübergreifenden Blueprints wird an anderer Stelle ausführlich beschrieben und ist nicht Bestandteil dieses Artikels [3]. Die vorliegende Arbeit befasst sich mit der Entwicklung der EP Clinical Skills (CS), wie sie seit 2011 durchgeführt wird.

Strukturierte klinisch-praktische Prüfungen als Teil lizenzierender Examina am Ende der Ausbildung sind in Kanada [4] und den USA [5] bekannt und werden dort seit mehreren Jahren erfolgreich durchgeführt. Lizenzierende Prüfungen sind in Europa ebenfalls verbreitet, doch stellt eine nationale strukturierte klinisch-praktische Prüfung im europäischen Kontext ein Novum dar. Zwar existiert eine Fülle von Literatur, die sich mit den verschiedensten Aspekten einer OSCE-Prüfung beschäftigt [6], aber es sind kaum Studien publiziert, welche sich mit den Besonderheiten der Entwicklung von nationalen klinisch-praktischen Prüfungen auseinandersetzen. Ziel des vorliegenden Beitrags ist deshalb, die - aus Sicht der Arbeitsgruppe CS - essentiellen Schritte der Prüfungserstellung, Durchführung, Auswertung sowie Qualitätssicherungsmaßnahmen und zentralen Erkenntnisse der letzten drei Jahre darzustellen.

Methodik

Die nationale Arbeitsgruppe Clinical Skills (AG CS) konstituiert sich aus Mitgliedern¹ aus den fünf medizinischen Fakultäten (Basel, Bern, Genf, Lausanne, Zürich) und dem Institut für Medizinische Lehre der Universität Bern. Alle Mitglieder verfügen über langjährige praktische Erfahrungen mit den an den einzelnen Fakultäten durchgeführten OSCE Prüfungen. Das von der AG CS gewählte Vorgehen zur Konzept- und Weiterentwicklung der EP CS basiert auf den methodologischen Prinzipien der Aktionsforschung [7]. Schlüsselemente dieser Methode sind der iterative Entwicklungsprozess (Planung-Implementation-Reflexion) und die aktive demokratische Partizipation aller Beteiligten [8]. Dabei steht nicht eine einzelne wis-

enschaftliche Methode im Vordergrund, sondern gemeinsam definierte Fragestellungen mit dem Ziel ein Produkt resp. Prozess (in unserem Fall die EP CS) stetig zu verbessern. Charakteristisch für die Aktionsforschung ist die Kombination aus wissenschaftlichen Fragestellungen und konkreter Umsetzung im realen Alltag, die zur Generierung von Wissen führt, welches auch ausserhalb des lokalen Kontextes von Relevanz und Interesse ist [9].

Die aktive Partizipation der AG CS Mitglieder wird durch jährlich acht halbtägige Sitzungen mit jeweils rotierender Sitzungsleitung gewährleistet. Das breite Erfahrungswissen der einzelnen Arbeitsgruppenmitglieder um Abläufe und Fallstricke, die es bei einer strukturierten praktisch-klinischen Prüfung zu beachten gilt, ist von grundlegender Bedeutung für alle folgenden Schritte. In der Phase der *Planung* bringen die einzelnen Mitglieder ihre Erfahrungen ein, gemeinsam werden Fragestellungen bzw. Ziele definiert, mögliche Problemfelder identifiziert und nach Lösungsmöglichkeiten gesucht. In dieser Phase der Konzepterstellung wird immer Bezug auf die vorhandene wissenschaftliche Evidenz zum entsprechenden Thema genommen. Auch wird bei entsprechenden Fragestellungen der Kontakt zum National Board of Medical Examiners der USA und zum Medical Council of Canada gesucht, um von deren langjähriger Erfahrung mit nationalen strukturierten klinisch-praktischen Prüfungen zu profitieren. Vor der *Implementation* wird die von der AG CS erarbeitete Umsetzungsstrategie der Prüfungskommission zur Genehmigung vorgelegt. Bestehen Unsicherheiten bezüglich der gewählten Strategie, wird das Konzept zuerst in einer fakultären OSCE Prüfung pilotiert und erst anschliessend in der EP CS eingesetzt. Im Rahmen der Evaluation wird eine Vielzahl unterschiedlicher qualitativer und quantitativer Methoden eingesetzt. In der Phase der *Reflexion* wird in der AG CS über die gesammelten Erfahrungen und Evaluationsergebnisse in Bezug auf Fragestellungen bzw. Ziele diskutiert. Besteht weiterer Handlungsbedarf, wird das initiale Konzept überarbeitet und der Zyklus beginnt von vorne. Diesen dreistufigen Prozess mit Planung-Implementation-Reflexion durchlaufen alle die Prüfung betreffenden Schritte. Da die Darstellung des gesamten Prozesses für jeden einzelnen Schritt den Rahmen dieser Arbeit sprengen würde, werden hier nur die Ergebnisse dieser Prozesse vorgestellt.

Ergebnisse

Die im Folgenden vorgestellten Prozesse der Prüfungserstellung, Durchführung, Auswertung und Qualitätssicherung sowie die zentralen Erkenntnisse sind in der Realität nicht als separate Entitäten zu betrachten, sondern sind miteinander verwoben und beeinflussen sich gegenseitig. Diese Aufteilung wird aus Gründen der besseren Verständlichkeit vorgenommen.

Prüfungserstellung

Fallerstellung: Der initiale Schritt der Fallentwicklung besteht in der Identifizierung eines für die Clinical Skills (CS) Prüfung geeigneten Themas. Ausgangspunkt und Grundlage der möglichen Themen bilden die „problems as starting points“ des nationalen Lernzielkatalogs und die Anforderungen des Blueprints. Ist ein Thema bestimmt, schreibt ein entsprechend dem Fallthema ausgewählter erfahrener Kliniker einen ersten Fallentwurf, welcher von methodischen Experten auf formale Kriterien beurteilt und kommentiert wird. Dieser Entwurf wird anschliessend im Rahmen eines Workshops zusammen mit einem zweiten Kliniker (aus einer anderen Fakultät und anderen Fachrichtung) sowie einem CS-Coach (Person mit Erfahrung im CS Prüfungsformat) weiter entwickelt. In diesen Workshops wird nicht nur der eigentliche Fallinhalt entwickelt, sondern auch die Gewichtung der einzelnen Assessment-Kriterien festgelegt. Mit diesem Teamansatz wird sichergestellt, dass bei der Fallerstellung die nötige medizinische Fachkompetenz vorhanden ist (Kliniker) und die prüfungsformat-spezifischen Aspekte beachtet werden (CS-Coach). Die praktische Durchführbarkeit der entwickelten Fälle wird am Workshop selbst mit Schauspielpatienten (SP) getestet. Diese Testläufe bringen wichtige Hinweise, welche die weitere Fallentwicklung oft wesentlich beeinflussen.

Aufbau der Stationen: Die vom Kandidaten im Rahmen der Interaktion mit dem SP zu erhebende Anamnese und klinische Untersuchung sowie die darauf basierenden Differenzialdiagnosen und das weitere diagnostische und therapeutische Vorgehen (sogenannte ASM Kriterien: Anamnese, Status, Management) werden mittels fallspezifischer Checklisten bewertet. Die Kommunikation zwischen Kandidat und SP (sogenannte KK Kriterien) wird hingegen mit einer an allen Stationen identischen, vier Aspekte umfassenden, generischen Skala beurteilt [10], [11].

Review Prozess: In einem nächsten Schritt wird der Fall dem nationalen Review Board vorgelegt. Dieses besteht aus klinisch tätigen Vertretern der fünf medizinischen Fakultäten und Repräsentanten der Allgemeinmedizin. Aufgabe des Review Boards ist die Beurteilung der Fälle auf ihre fachliche Korrektheit, Relevanz und Komplexität. Nach kritischer Prüfung und Diskussion legt das Review Board den entsprechenden Änderungsbedarf fest oder gibt den Fall für den weiteren Entwicklungsprozess frei.

Standardisierung: Nach erfolgter nationaler Validierung durch das Review Board wird der Fall den SP-Trainern zur Verfügung gestellt. Die SP-Trainer aller fünf Fakultäten erarbeiten in gemeinsamen Treffen ein einheitliches Verständnis der Fälle. Jeder Fall wird besprochen, gespielt und videodokumentiert. Die Videodokumentation ermöglicht, komplexe Situationen (z.B. Darstellung der psychischen Verfassung des Patienten) oder das Ausmass eines bestimmten Symptoms (z.B. neurologisches Defizit) genau festzuhalten. Im Rahmen dieser Treffen wird auch festgelegt, wie beispielsweise eine kutane Effloreszenz am besten geschminkt wird oder wie die Raumausstattung

zu erfolgen hat. Bei Bedarf werden die Erfahrungen aus den SP-Trainer Treffen in den Rollenbeschreibungen weiter präzisiert oder angepasst.

Übersetzung: Nachdem in den SP-Trainer Treffen die Rollenbeschreibungen finalisiert worden sind, wird der Fall von der Originalsprache Deutsch resp. Französisch in die jeweils andere Sprache übersetzt. Um ein möglichst hohes Mass an Einheitlichkeit zu erreichen, werden alle Fälle jeweils von den zwei gleichen Ärzten in ihre Muttersprache übersetzt.

Qualitätssicherung: Mit diesem zentral koordinierten, mehrstufigen Prozess der Prüfungserstellung mit Beteiligung von inhalts- und formatspezifischen Experten aller fünf Fakultäten wird sichergestellt, dass die entwickelten Fälle fachlich korrekt sind, die formatspezifischen Anforderungen eines OSCE erfüllen, einen nationalen Validierungsprozess durchlaufen und an allen Fakultäten unter standardisierten Bedingungen eingesetzt werden können. Bezüglich der inhaltlichen Validierung ist die Aufgabe und Funktion des Review Boards von zentraler Bedeutung. Die nationalen SP-Trainer Treffen wiederum haben einen entscheidenden Einfluss auf den Standardisierungsprozess.

Prüfungsdurchführung

Die EP CS wird dezentral an den fünf medizinischen Fakultäten einmal jährlich durchgeführt. Abhängig von der Anzahl der Kandidaten an der jeweiligen Fakultät wird die Prüfung an zwei resp. drei Tagen angeboten. An jedem Prüfungstag wird mit einem unterschiedlichen Set von Fällen geprüft; die Prüfungssets des entsprechenden Tages sind aber an allen Fakultäten identisch. Der einzelne Kandidat absolviert 12 Stationen à 13 Minuten (zusätzlich zwei Minuten Rotationszeit zwischen den Stationen und drei mal 15 Minuten Pause). Die Leistung der Kandidaten in den 12 Stationen wird von erfahrenen Klinikern der jeweiligen Fakultät beurteilt. Neben der Bewertung mittels fallspezifischer und generischer Kriterien geben die Examinatoren zu jedem Kandidat zwei Gesamtbeurteilungen in Bezug auf ASM und KK ab. Diese sogenannten Globalurteile dienen dem Berechnen der Bestehensgrenze.

Qualitätssicherung: Detaillierte Informationen (inklusive Videobeispiel) zum Prüfungsablauf werden den Kandidaten rechtzeitig vor der Prüfung zur Verfügung gestellt [<http://www.bag.admin.ch/themen/berufe/07918/07919/> zitiert am 14.12.13]. Erstmalig an der EP CS prüfende Kliniker werden an einer Informationsveranstaltung über das OSCE-Format orientiert und in der Handhabung des Bewertungsrasters instruiert. Während und nach der Prüfung haben die Examinatoren die Möglichkeit, inhaltliche Rückmeldungen zu den Fällen im Allgemeinen und zu den Beurteilungskriterien im Speziellen zu geben. Einzelne Mitglieder der AG CS visitieren während der laufenden Prüfung die verschiedenen Standorte und können sich so ein Bild über die konkrete Umsetzung der EP CS vor Ort machen.

Prüfungsauswertung

In einem ersten Schritt werden sämtliche Kommentare der Examinatoren gesichtet und gewertet sowie die Schwierigkeits- und Trennschärfewerte der einzelnen Beurteilungskriterien berechnet. Auf dieser Grundlage werden Kriterien, welche inhaltliche oder formale Mängel aufweisen, eliminiert. Anschliessend wird das Punktetotal eines Kandidaten an einer Station errechnet aus der Summe der gewichteten Checklistenpunkte der ASM Kriterien addiert mit der gewichteten Punktschätzung der Kommunikationsskala ($\text{Punktetotal} = 0.75 \cdot \text{ASM} + 0.25 \cdot \text{KK}$). Die 12 Stationen tragen zum Prüfungstotal alle gleich bei. Das Prüfungstotal der Kandidaten berechnet sich als Mittelwert über die 12 Stationspunktetotale. Die Ergebnisse der Kandidaten werden anschliessend standardisiert, um die Prüfungsergebnisse über die drei Tage vergleichbar zu machen. In einem nächsten Schritt wird mittels der Borderline Regressionsmethode [12], [13] die Bestehensgrenze pro Station errechnet. Auf dieser Grundlage legt die Prüfungskommission die definitive Bestehensgrenze fest. Tabelle 1 zeigt die Anzahl Kandidaten gefächert nach Gruppen mit schweizerischem resp. aussereuropäischem Diplom und Repetenten sowie die entsprechenden Bestehensquoten.

Qualitätsindices: In Tabelle 2 sind die statistischen Qualitätsindices der EP CS seit 2011 zusammen gestellt. Diese Resultate entsprechen den Erwartungen dahingehend, dass mit den beiden Prüfungsformaten unterschiedliche Aspekte der medizinischen Kompetenz erfasst werden, die aber (selbstverständlich) eine gemeinsame Grundlage aufweisen. Zudem spricht die geringere Korrelation von KK mit MC (im Vergleich zu ASM mit MC) für die Konstruktvalidität der Gesamtprüfung MC und CS.

Qualitätssicherung: Die Datenaufbereitung und statistische Auswertung der Prüfung wird nach etablierten Richtlinien [14] zentral von einer Institution (Institut für Medizinische Lehre der Universität Bern) vorgenommen. Daneben trägt die minutiöse Analyse der Kommentare der Examinatoren zu den einzelnen Fällen, welche die Elimination inhaltlich oder formal inadäquater Beurteilungskriterien ermöglicht, zur Qualität der Prüfung bei. Die geringe Anzahl eliminerter Kriterien wiederum ist ein Mass für die Qualität des Prüfungserstellungsprozesses (2013 wurden 22 von 807 Kriterien (2.7%) eliminiert).

Zentrale Erkenntnisse der vergangenen drei Jahre

Der iterative Prozess mit Planung-Implementation-Reflexion wird von der AG CS als einer der entscheidenden Faktoren für die erfolgreiche Weiterentwicklung der EP CS angesehen. Um diesen Prozess zu veranschaulichen, soll dieser an einem Beispiel exemplarisch erläutert werden. Aufgrund von Examinatoren Rückmeldungen während den ersten beiden Durchführungen zeigte sich, dass sich dichotome Kriterien zur Beurteilung komplexer Fertigkeiten nur bedingt eignen. Basierend auf dieser Erfahrung diskutierten die Mitglieder der AG CS über

mögliche alternative Beurteilungskriterien (Planung). Mit einem adaptierten Bewertungsraster sollte nicht nur beurteilt werden können *ob* sondern auch *wie* eine bestimmte Tätigkeit ausgeführt wird. Basierend auf diesen Überlegungen, sowie Konsultation der relevanten Literatur, wurden in der Prüfung 2013 mehrstufige Beurteilungskriterien pilotiert (Implementation). Im Rahmen der Evaluation zeigte sich, dass bei Stationen, die solche mehrstufigen Beurteilungskriterien enthielten, die Korrelationen der Globalurteile und der erzielten Punktzahlen höher lagen als bei Stationen ohne diese mehrstufigen Kriterien (Reflexion). Die Korrelation von Globalurteil und Punktetotal ist ein Gütekriterium zur Qualitätsbeurteilung einer OSCE Station. Entsprechend werden ab 2014 in der EP CS, wo möglich, mehrstufige Kriterien zur Beurteilung komplexer klinischer Fertigkeiten eingesetzt.

Wesentliche Erkenntnisse in Bezug auf die einzelnen Schritte der Prüfungserstellung, Durchführung und Auswertung werden im Folgenden dargestellt.

Der zentral koordinierte, kollaborative Prozess der *Prüfungserstellung* mit Fallerstellungsworkshops, Review Board und nationalen SP-Trainer Treffen und die Einbindung einer grossen Zahl von Klinikern aus allen Fakultäten und Fachrichtungen sind wesentliche Faktoren, die zur inhaltlichen Qualität der Prüfung beitragen. Dieser auf Konsens hin ausgerichtete Prozess trägt entscheidend zur Robustheit des einzelnen Falles und zur Qualität der Prüfung als Ganzes bei.

Im Rahmen der *Prüfungsdurchführung* beurteilen die Examinatoren seit 2011 die Fälle auf ihre medizinisch-inhaltliche Qualität. Zusätzlich wird seit 2013 die Qualität der SP-Performanz mittels eines standardisierten Beurteilungsrasters erfasst. Der Einsatz dieses Rasters mit seinen expliziten Messkriterien, das den SP im Vorfeld abgegeben wird, hat dazu beigetragen, dass die Qualität der SP-Performanz noch gesteigert und zwischen den einzelnen Standorten weiter harmonisiert werden konnte. Auch in der *Prüfungsauswertung* entstand der Bedarf nach zusätzlichen quantitativen Parametern, welche helfen, die Qualität der einzelnen Fälle und ihren Einsatz an den einzelnen Prüfungsstandorten besser zu beurteilen. Mit einem von Pell und anderen beschriebenen Set von statistischen Parametern lassen sich unterschiedliche Qualitätsaspekte einer Station abschätzen [15]. Diese Parameter liefern wichtige Informationen für den zukünftigen Fallerstellungsprozess, da sie Hinweise geben, bei welchen Stationen eine Optimierung der Beurteilungskriterien angezeigt ist oder wo eine Standardisierung besonders anspruchsvoll ist.

Diskussion

Im vorliegenden Beitrag werden die essentiellen Schritte bei der Erstellung, Durchführung und Auswertung sowie die zentralen Erkenntnisse aus den ersten drei Durchführungen der Eidgenössischen Prüfung Clinical Skills (EP CS) aus Sicht der Arbeitsgruppe Clinical Skills vorgestellt.

Tabelle 1: Anzahl Kandidaten und Bestehensquote der Eidgenössischen Prüfung Clinical Skills 2011-2013

Jahr	Diplom einer Schweizer Universität	Diplom einer aussereuropäischen Universität	Repetenten
2011	758 (98.3%)	7 (42.8%)	0 (-)
2012	785 (98.6%)	16 (50.0%)	16 (62.5%)
2013	786 (97.6%)	16 (56.3%)	17 (58.8%)

Anzahl Kandidaten die an der Eidgenössischen Prüfungen Clinical Skills 2011-2013 teilgenommen haben. In Klammer ist der Prozentsatz der Kandidaten angegeben, welche die Prüfung bestanden haben.

Tabelle 2: Statistische Qualitätsindizes der Eidgenössischen Prüfung Clinical Skills 2011-2013

Jahr	Reliabilität*	Korrelation**		
		MC vs. CS	MC vs. CS ASM	MC vs. CS KK
2011	0.84-0.88	0.48	0.46	0.39
2012	0.86-0.90	0.52	0.51	0.36
2013	0.86-0.89	0.55	0.54	0.39

* Cronbach-Alpha pro Tag jeweils minimaler resp. maximaler Wert. Berechnet auf Ebene der Kriterien.
 ** Korrelationen (Pearson) zwischen der Multiple Choice (MC) und Clinical Skills (CS) Prüfung, so wie zwischen der MC Prüfung und den Bereichen ASM (Anamnese-Status-Management) resp. KK (Kommunikation) der CS Prüfung (alle Korrelationen sind signifikant mit $p < 0.01$)

Ein mehrstufiges Vorgehen zur Erstellung von OSCE Prüfungen wird in internationalen Guidelines grundsätzlich empfohlen [16]. Der zentral koordinierte, mehrstufige Prozess der *Prüfungserstellung* mit Beteiligung von fach- und formatspezifischen Experten aller Fakultäten führt diesen Grundsatz konsequent weiter und ist ein wesentlicher Qualitätsfaktor der EP CS. Die *Durchführung* und *Auswertung* der EP CS lehnt sich an die validierten und bewährten nationalen Clinical Skills Prüfungen aus Nordamerika an. So ist die Anzahl Stationen pro Kandidat identisch mit der United States Medical Licensing Examination Step 2 Clinical Skills [http://www.usmle.org/ zitiert am 14.12.13], [17]. Mit der strukturierten klinisch-praktischen Prüfung des Medical Council of Canada Qualifying Examination Part 2 [http://mcc.ca/examinations/mccqe-part-ii/ zitiert am 14.12.13], [18] hat die EP CS folgende Gemeinsamkeiten:

1. Die Leistung der Kandidaten wird durch Kliniker der entsprechenden Lehrkrankenhäuser resp. Universitäten bewertet;
2. auf der Basis von Expertenurteilen (Globalurteile) wird die Bestehensgrenze pro Station errechnet;
3. die inhaltlich-praktischen und die kommunikativen Aspekte werden zu einem Gesamtscore addiert und
4. die Prüfung wird in zwei Sprachen angeboten.

In Bezug auf die Bestehensquoten zeigen sich zudem Parallelen zwischen der US-amerikanischen und schweizerischen Prüfung [19].

Gesamthaft betrachtet steht eine Erkenntnis der vergangenen drei Jahre im Vordergrund: Die konsequente Umsetzung der Prinzipien der Aktionsforschung verbunden mit der aktiven Partizipation der unterschiedlichen Akteure trägt entscheidend zur erfolgreichen Weiterentwicklung der EP CS bei. Insgesamt profitiert die EP CS ganz erheblich von der Erfahrung der an den einzelnen Fakultäten durchgeführten OSCE-Prüfungen. Dieser Einfluss ist aber nicht unidirektional. Vielmehr haben die fakultären OSCE auch von den Erfahrungen mit der EP CS profitiert. Wo sich bis vor kurzem an jeder Fakultät lokal isoliert zwei bis drei Personen mit OSCE Prüfungen beschäftigten, hat sich seit Einführung der EP CS eine schweizweite „community of practice“ [20] entwickelt, die sich mit dem

Thema professionell auseinandersetzt, einen regen Austausch pflegt und bestrebt ist, die strukturierte klinisch-praktische Prüfung weiterzuentwickeln. Die Entwicklung dieses Prüfungsformates, insbesondere für den Einsatz in lizensierenden Schlussprüfungen, ist nicht abgeschlossen. Auch andere Institutionen sind daran, ihre nationalen strukturierten klinisch-praktischen Prüfungen weiterzuentwickeln [21].

Weltweit gibt es zum jetzigen Zeitpunkt nur wenige Staaten mit einer nationalen, strukturierten klinisch-praktischen Abschlussprüfung [22]. Entsprechend setzen sich nur wenige wissenschaftliche Arbeiten konkret mit der Erstellung, Durchführung und Auswertung solcher Prüfungen auseinander. Die vorliegende Arbeit liefert einen ersten Beitrag, um diese Lücke zu schliessen. Die hier dokumentierten essentiellen Schritte und zentralen Erkenntnisse können Interessengruppen dienen, insbesondere im Hinblick auf die Planung von nationalen klinisch-praktischen Prüfungen. Es ist aber auch durchaus vorstellbar, das von uns beschriebene Vorgehen in „Universitätsverbänden“ anzuwenden, um eine qualitativ hochwertige Prüfung anzubieten, welche die Ressourcen und Expertise einer einzelnen Ausbildungsstätte übersteigt.

Die vorliegende Arbeit beschränkt sich auf die Sichtweise der Arbeitsgruppe CS. Wünschenswert wäre, dass zukünftig auch die Erfahrungen von Kandidaten, Dozenten und weiteren Involvierten in den Gesamtprozess vermehrt einfließen. Eine weitere Einschränkung dieser Arbeit ist, dass eine umfassende und detaillierte Darstellung aller eingesetzten Methoden und der daraus resultierenden Ergebnisse bei einem so umfangreichen Projekt kaum möglich ist. Die dargestellte Vorgehensweise sollte es dem Leser jedoch ermöglichen, sich ein Bild der Gesamtsituation zu machen.

Schlussfolgerung

Basierend auf der methodologischen Grundlage der Aktionsforschung wurden essentielle Schritte vorgestellt, welche die erfolgreiche Durchführung und Weiterentwicklung einer nationalen strukturierten klinisch-praktischen

Prüfung ermöglichen, mit welcher sich die praktischen Fähigkeiten der Kandidaten auf faire, reliable und valide Weise beurteilen lassen. Die im vorliegenden Beitrag dokumentierten Prozesse und Erfahrungen können für andere Organisationen mit ähnlichen Vorhaben nützlich sein.

Danksagung

Eine nationale Clinical Skills Prüfung lässt sich nur durch Einsatz von unzähligen engagierten Personen bewerkstelligen. Unser ausgesprochen Dank geht insbesondere an die Fallautoren, Mitglieder des nationalen Review Boards, die Schauspielpatienten Trainer, die Schauspielpatienten und Standortverantwortlichen. Bedanken möchten wir uns auch bei den Klinikern für ihren Einsatz als Examinatoren. Ohne die gute Zusammenarbeit und stete Unterstützung durch die Prüfungskommission und das Bundesamt für Gesundheit wäre eine solche Prüfung ebenfalls nicht denkbar. Ein spezieller Dank geht auch an Patrick Jucker-Kupper für die kritische Sichtung des Manuskriptes.

Anmerkung

¹ Die männliche Form gilt in gleicher Weise für das weibliche Geschlecht.

Interessenkonflikt

Die Autoren erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

- Bürgi H, Rindlisbacher B, Bader C, Bloch R, Bosman F, Gasser C, Gerke W, Humair JP, Im Hof V, Kaiser H, Lefebvre D, Schläppi P, Sottas B, Spinas GA, Stuck AE. Swiss Catalogue of Learning Objectives for Undergraduate Medical Training. Bern: Universität Bern; 2008. Zugänglich unter/available from: <http://scolo.smfkk.ch>
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J.* 1975;1(5955):447-451. DOI: 10.1136/bmj.1.5955.447
- Guttormsen S, Beyeler C, Bonvin R, Feller S, Schirlo C, Schnabel K, Schurter T, Berendonk C. The new licencing examination for human medicine: from concept to implementation. *Swiss Med Wkly.* 2013;143:w13897. Zugänglich unter/available from: <http://www.smw.ch/content/smw-2013-13897/>
- Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L, Lacombe G, Baumber J, Poldre P, Levasseur L. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. *Acad Med.* 1992;67(8):487-494. DOI: 10.1097/00001888-199208000-00001
- De Champlain A, Swygert K, Swanson DB, Boulet JR. Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Acad Med.* 2006;81(10 Suppl):S17-20.
- Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach.* 2013;35(9):e1437-1446.
- Meyer J. Qualitative research in health care. Using qualitative methods in health related action research. *BMJ.* 2000;320(7228):178-181.
- Edler A. Action research in medical education: a shifting paradigm or old wine in new skins? *Clin Teach.* 2009;6:139-410.
- Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. *BMJ.* 2008;337:a567.
- Hodges B, McLroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ.* 2003;37(11):1012-1016. DOI: 10.1046/j.1365-2923.2003.01674.x
- Scheffer S. Validierung des Berliner Global Rating® (BGR) -ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinisch-praktischer Prüfungen (OSCE). Berlin: Charité - Universitätsmedizin Berlin, Medizinische Fakultät; 2009.
- Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Objective structured clinical examinations. *Med Educ.* 2003;37(2):132-139. DOI: 10.1046/j.1365-2923.2003.01429.x
- Wood TJ, Humphrey-Murto SM, Norman GR. Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci Educ Theory Prac.* 2006;11(2):115-122. DOI: 10.1007/s10459-005-7853-1
- Downing S. Statistics of testing. In: Downing S, Yudkowsky R (Hrsg). *Assessment in Health Professions Education.* London: Routledge; 2009. S.93-117.
- Pell G, Fuller R, Homer M, Roberts T, International Association for Medical E. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Med Teach.* 2010;32(10):802-811. DOI: 10.3109/0142159X.2010.507716
- Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach.* 2013;35(9):e1447-1463.
- Gilliland WR, La Rochelle J, Hawkins R, Dillon GF, Mechaber AJ, Dyrbye L, Papp K, Durning SJ. Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach.* 2008;30(3):325-327. DOI: 10.1080/01421590801953026
- Woloschuk W, McLaughlin K, Wright B. Predicting performance on the Medical Council of Canada Qualifying Exam Part II. *Teach Learn Med.* 2013;25(3):237-241. DOI: 10.1080/10401334.2013.797351
- Boulet JR, Swanson DB, Cooper RA, Norcini JJ, McKinley DW. A comparison of the characteristics and examination performances of U.S. and non-U.S. citizen international medical graduates who sought Educational Commission for Foreign Medical Graduates certification: 1995-2004. *Acad Med.* 2006;81(10 Suppl):S116-119.
- Bereiter C. Design Research for Sustained Innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Sci Soc.* 2002;9(9):321-327.
- Haist SA, Katsfrakis PJ, Dillon GF. The evolution of the United States Medical Licensing Examination (USMLE): enhancing assessment of practice-related competencies. *JAMA.* 2013;310(21):2245-2246. DOI: 10.1001/jama.2013.282328

22. McCrorie P, Boursicot KA. Variations in medical school graduating examinations in the United Kingdom: are clinical competence standards comparable? *Med Teach*. 2009;31(3):223-229. DOI: 10.1080/01421590802574581

Korrespondenzadresse:

Dr. med. Christoph Berendonk, MME
Universität Bern, Medizinische Fakultät, Institut für
Medizinische Lehre, Konsumstraße 13, CH-3010 Bern,
Schweiz
christoph.berendonk@iml.unibe.ch

Bitte zitieren als

Berendonk C, Schirlo C, Balestra G, Bonvin R, Feller S, Huber P, Jünger E, Monti M, Schnabel 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, URN: urn:nbn:de:0183-zma0009828

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2015-32/zma000982.shtml>

Eingereicht: 14.01.2014

Überarbeitet: 30.10.2014

Angenommen: 12.01.2015

Veröffentlicht: 15.10.2015

Copyright

©2015 Berendonk 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/>.