

Assessment formats in dental medicine: An overview

Abstract

Aim: At the annual meeting of German dentists in Frankfurt am Main in 2013, the Working Group for the Advancement of Dental Education (AKWLZ) initiated an interdisciplinary working group to address assessments in dental education. This paper presents an overview of the current work being done by this working group, some of whose members are also actively involved in the German Association for Medical Education's (GMA) working group for dental education. The aim is to present a summary of the current state of research on this topic for all those who participate in the design, administration and evaluation of university-specific assessments in dentistry.

Method: Based on systematic literature research, the testing scenarios listed in the National Competency-based Catalogue of Learning Objectives (NKLZ) have been compiled and presented in tables according to assessment value.

Results: Different assessment scenarios are described briefly in table form addressing validity (V), reliability (R), acceptance (A), cost (C), feasibility (F), and the influence on teaching and learning (EI) as presented in the current literature. Infoboxes were deliberately chosen to allow readers quick access to the information and to facilitate comparisons between the various assessment formats. Following each description is a list summarizing the uses in dental and medical education.

Conclusion: This overview provides a summary of competency-based testing formats. It is meant to have a formative effect on dental and medical schools and provide support for developing workplace-based strategies in dental education for learning, teaching and testing in the future.

Susanne Gerhard-Szep¹

Arndt Güntsch²

Peter Pospiech³

Andreas Söhnel⁴

Petra Scheutzel⁵

Torsten Wassmann⁶

Tugba Zahn⁷

1 Goethe-Universität,
Carolinum Zahnärztliches
Universitäts-Institut gGmbH,
Poliklinik
Zahnerhaltungskunde,
Frankfurt am Main,
Deutschland

2 Marquette University School
of Dentistry, Department of
Surgical Sciences,
Milwaukee, USA und
Universitätsklinikum Jena,
Zentrum für Zahn-, Mund-
und Kieferheilkunde, Jena,
Deutschland

3 Universität Würzburg,
Poliklinik für Zahnärztliche
Prothetik, Würzburg,
Deutschland

4 Universitätsmedizin
Greifswald, Poliklinik für
Zahnärztliche Prothetik,
Alterszahnheilkunde und
medizinischer
Werkstoffkunde, Greifswald,
Deutschland

5 Universitätsklinikum
Münster, Poliklinik für
Prothetische Zahnmedizin &
Biomaterialien, Münster,
Deutschland

6 Universitätsmedizin
Göttingen, Poliklinik für
Zahnärztliche Prothetik,
Göttingen, Deutschland

7 Goethe-Universität,
Carolinum Zahnärztliches
Universitäts-Institut gGmbH,
Poliklinik für Zahnärztliche
Prothetik, Frankfurt am Main,
Deutschland

1. Starting point

A concerted connection between teaching and testing (constructive alignment) is crucial to impart dental competencies during university study [1].

Also connected with the definition of competency-based learning objectives are the appropriate testing formats that measure the requisite combination of knowledge, practical skills and ability to engage in professional decision-making for each particular task (see Figure 1).

2. Method

A survey of the literature was undertaken between January 17 and December 17, 2014, in the databases of the German National Library (DNB), MEDLINE using the PubMed interface, Excerpta Medica Database (EMBASE), Education Resource Information Centre (ERIC), Cochrane Library, Science Citation Index, and Google Scholar. The search was conducted automatically and supplemented manually. In addition, available dissertations, open-access publications by German Medical Science (GMS), BEME (Best Evidence Medical and Health Professional Education), including German-language conference proceedings, such as the AKWLZ and GMA, were evaluated. The search terms included the German equivalents for “MCQ”; “MEQ”; “multiple choice”; “MC”; “multiple-choice questionnaire”; “SMP”; “structured oral examination”; “SOE”; “key feature”; “OSCE”; “OSPE”; “standardized patient”; “CEX”; “miniCEX”; “entrustable professional activities”; “DOPS”; “portfolio”, “multi-source feedback” in combination with “AND” and “dental”; “medicine”; “education”; “assessment.”

In an initial step, literature was selected based on title and abstract in accordance with pre-defined inclusion and exclusion criteria (inclusion criteria: published 1966-2013 in German or English with topical relevance; exclusion criteria: failure to meet inclusion criteria, full text not available in English or German, lack of topical relevance). The selected publications were then evaluated in terms of their relevance to the issue at hand and excluded, as required.

The articles were analyzed, and the results were described according to categories based on the value of assessments [2]. These categories cover the parameters of validity (V), reliability (R), acceptance (A), cost (C), feasibility (F) and influence on teaching and learning (EI). The results were then organized according to the evaluation parameters above. These criteria were further developed in 2011 by the working group headed by Norcini, and the parameters of “equivalence” (assessments conducted at different sites) and “catalytic effect” (consequences for the medical school) were added [3]. In this overview, both of these additional parameters were included and discussed in text form. The focus of this analysis has been carried out using the six criteria listed above (V, R, A, C, F, EI) in the form of tables to allow for clarity and comparisons.

3. Results

In total, n=223 publications were identified using the search strategy outlined above and drawn upon as the basis for the following analyses.

4. Discussion

Structured oral examinations and the **multiple-choice questionnaire (MCQ)** are suited for testing theoretical knowledge, meaning descriptive knowledge of competency level 1 [4], [5], [6], [7]. The MCQ is a written assessment with several response options (closed questions), of which a single choice or multiple ones (multiple choice, multiple select) can be the right answer. After a brief introduction of the content and question come the response options that include the correct answer(s) and distracters (wrong answers). Multiple-choice exams can be paper-based, combined with computer-assisted grading, or even administered entirely at computer workstations [4].

Use in medical and dental education

- MCQs are presently used in both medical and dental study programs [6], [8].
- The most important preliminary and final examinations include multiple-choice questions: the preliminary exam in natural science (NVP), preliminary dental exam (ZVP), dental exam (ZP), first and second state medical exams (ÄP). Moreover, MCQs are specifically found in all pre-clinical and clinical subjects in both study programs; this type of question represents one of the most traditional, predominating assessment formats [4], [8].

To assess factual knowledge, the MCQ offers a cost-efficient testing format with high reliability and validity if the questions correspond to the quality criteria. With MCQs it is possible to objectively test a large amount of content in a short period of time. However, this type of assessment can lead to superficial learning of facts.

Multiple-choice Questionnaire

Validity

- Classified as high [9]
- Quality criteria for questions must be met to have sufficient validity [10].
- A high construct validity can be achieved if questions are subjected to a review process (e.g. via Item Management System [IMS]) [11].

Reliability

- Classified as high [9], [11]
- A minimum of 40 high-quality questions are needed to yield α Cronbach's α of 0.8 [6].

Acceptance

- Scoring is objective [4].

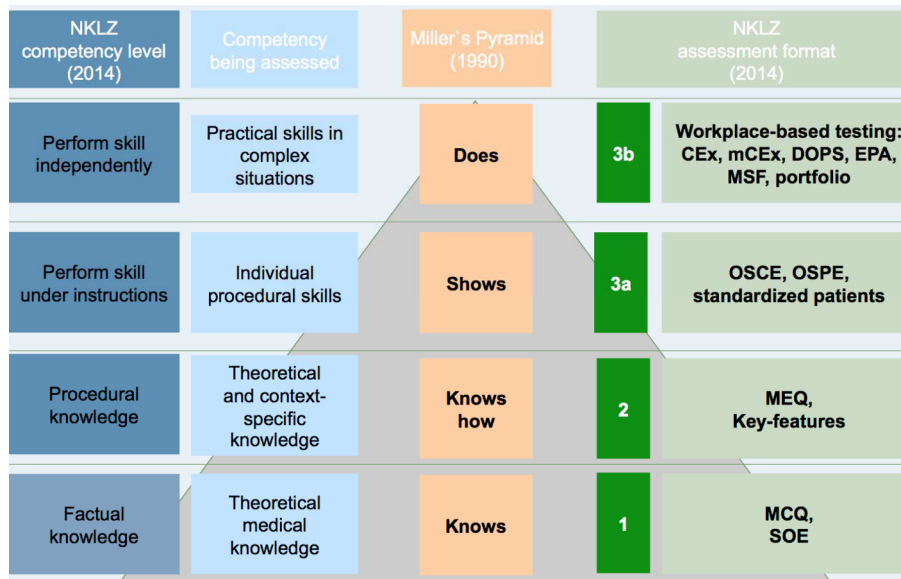


Figure 1: Examples of assessment scenarios depending on competency level according to the requirements in Miller and the National Competency-based Catalogue of Learning Objectives for Undergraduate Dental Education (NKLZ).

- MCQs are considered fair if what has been taught corresponds with what is tested [12].
- The possibility of passing by giving “strategic” responses, guessing, or picking up on cues is viewed critically by teachers [13], [14], [15], [16].

Cost

- In light of the numbers and frequency of tests, it is an effective assessment format [9], [17].
- A broad range of content can be assessed on one test [5].
- Proportionally low costs [18]
- Positive cost-benefit ratio
- An existing question pool can be kept current at relatively little cost [19].

Feasibility

- Effort is primarily involved in generating questions; administering and grading tests require much less time and resources.
- Creating the question pool is associated with not insignificant costs [4].
- Online assessment with digital scoring is possible [5].
- A question pool shared by multiple universities increases efficiency via synergies (e.g. IMS) [20].

Influence on teaching and learning

- Can lead to superficial learning [21]
- Theoretical knowledge is more important than practical skills [4].
- Correct responses are already given making passive recognition possible [14].

Structured oral examinations (SOE) are oral assessments which are conducted by an individual examiner or a panel of examiners.

Structured Oral Examination

Validity

- Direct dependence on the degree of structuredness [22]
- Validity increases with planning, design and conditions of testing [23], [24].
- Validity is more dependent on the examiners than the method.

Reliability

- Increases with the number of questions, length of assessment and decreases in the face of strongly differentiated scoring [10]
- Reliability and objectivity increase with several examiners [10], [17].
- Absolute verification of reliability is practically impossible [10].
- With a Cronbach's α of 0.65-0.88 [25], [26], [27], SOEs come out ahead of conventional exams (Cronbach's α of 0.24-0.50) [25], [28], [29].

Acceptance

- Performance-inhibiting stress, anxiety and other disruptive factors play a larger role compared to MCQs [12].
- Acceptance by teachers and students is reduced by:
 - Intensive supervision by examiner
 - Justification of scores
 - Limited information during limited time
- Questions or objections possible on the part of the student with no written test to refer to; the difference between content and type of response can lead to this [12].

Cost

- More cost-intensive than MCQ exams [10]
- Relativizes itself on high-stakes exams: emphasis is on reliability and validity, not on cost-effectiveness. [30], [31]

Feasibility

- More effort is required compared with MCQ, high financial burden resulting from need for staff and rooms / logistics [10].

Influence on teaching and learning

- Alongside facts, clinical reasoning, professional thinking, self-confidence and self-assurance can be assessed [12], [22].
- Since students adapt their behavior to fit a test [4], [5], [18], extensive preparation can be assumed.

If an examination is taken before a panel, the examiners consult and agree on their evaluation of the examinee's performance. Ideally, the final grades are assigned according to a blueprint governing exam content [7].

The SOE is a testing format that enables assessment of competency level 1 (NK LZ) and beyond within the scope of usual interactions in dental care. However, the higher expenses connected with the greater need for time and personnel should be noted, as well as the potential for performance-inhibiting stress in examinees.

Use in medical and dental education

- Oral examinations with different degrees of structuredness are used in dental and medical study programs [8].
- The most important preliminary and final assessments (high-stakes exams) in both study programs (NVP, ZVP, ZP, first and second ÄP) include SOEs in various settings. Furthermore, the SOE is represented in all pre-clinical and clinical subjects in both study programs; it represents one of the traditional, predominating assessment formats [4], [32].

Assessments that do not just measure factual knowledge (=descriptive knowledge: knows) [33], [34]), but also capture the ability to apply theoretical knowledge in a specific context to solve a problem or reach a clinical decision (= procedural knowledge: knows how), require a special testing format that is indeed capable of representing this skill. It must be noted that the ability to solve problems or reason is highly specific to context and always depends on the particular context-related factual knowledge [2], [35]. In addition to the SOE, other assessment formats for evaluating procedural knowledge are the written **modified essay question** (MEQ) and key features exam. These involve case-based, written assessments that evaluate active knowledge recall, problem-solving and higher order cognitive skills while simulating clinical situations in which decisions are made in the course of a physical examination, diagnosis and therapy. A patient's history is presented in stages, after each of which several questions are responded to in writing or by

selecting the best of several possible responses. Previous questions are partially explained in the following sections making it impermissible to flip back and forth between pages.

Use in medical and dental education

- Developed in Great Britain in the 1970s for the membership examination of the Royal College of General Practitioners [36], [37], [38], [39], [40], [41].
- Used internationally in the field of medicine, from undergraduate education to post-graduate training [42], [43], [44], [45], [46], [47], [48], [49], [50].
- Used in Germany as an undergraduate testing format and as a written exam that replaces the state examination [51], [52] in model study programs (Witten/Herdecke, Cologne, Bochum, etc.).
- Hardly any examples of use in dental education; potential areas of application include assessing problem-solving skills within POL and independent learning using case-based, problem-based learning [53], practical, case-based testing with virtual patient cases (e.g. in connection with procedures for handling acute toothache in endodontics) [54].

The MEQ represents a reliable instrument to assess context-specific, procedural knowledge in clinical situations if several basic rules are adhered to: 1. inclusion of the largest number of cases possible; 2. quality control of the pre-defined grading criteria for the write-in (WI) format by several evaluators; 3. computer-based short-menu (SM) or long-menu (LM) response format. Through the simulation of decision making in a clinical setting with questions that build off of each other, learning paired with feedback becomes part of the test experience. The MEQ format represents a significant addition to the written tests commonly used at present in dental education, but it is connected with distinctly higher costs than simply running down a list of MCQs to measure purely factual knowledge.

Modified Essay Question

Validity

- Higher validity than for the MCQ format through case-based, context-rich question format [48], [55], [56]
- Contradictory results for correlation (γ) between MEQs and the results of the final exam (NBME) and post-graduate performance in the first year of professional medical practice: γ 0.3/0.3–0.26 [57], $\gamma=0.51$ [56].

Reliability

- Reliability (Cronbach's α)=0.57–0.91 [38] depends on multiple factors [38], [39], [40], [47], [48], [58], [59]:
- Quality of the predetermined performance scale
- Response format (open-ended responses poorer than selecting from a given list)
- Number of cases and questions
- Number of graders

- → e.g. increase of Cronbach's α from 0.7 to 0.8 by increasing the number of questions from 7 to 12 or increasing the number of graders from 1 to 4 [40].

Acceptance

- Students generally rate the MEQ positively [41], [51] since the MEQ format reflects practice more closely than the MCQ [60].
- Teachers/examiners: greater effort involved in creating tests, coordination challenges [51]

Cost

- Drafting and grading an MEQ are very time consuming and requires personnel [36], [41], [51].
- Efforts can be minimized in terms of grading by using a computer-based testing format [61].

Feasibility

- Generating and grading MEQs is distinctly more involved than for MCQs; difficult to design questions that actually measure the ability to solve problems or make clinical decisions and do not simply test factual knowledge [37], [41], [42], [43], [44], [45], [46], [50], [52], [53], [54], [62].

Influence on teaching and learning

- MEQs simulate clinical reasoning processes enabling feedback and learning during the test [39], [51], [60].

In the **key features exam** (KFE) a case unfolds in a specific clinical situation about which multiple questions are asked focusing very closely on only those critical actions or decisions (key features) that are central to the key feature problem or those that are often done incorrectly [34], [63]. Key feature cases are developed in eight defined steps [34], [64], [65]:

1. identification of the domain or context;
2. selection of a clinical situation;
3. identification of the critical elements of the situation (key features [KF] of the problem);
4. selection and description of the clinical scenario (case vignette);
5. drafting of the questions about the key features of the problem (1-3 question per KF);
6. determination of the response format (open-ended text = write-in, selection = short menu or long menu);
7. generation of the evaluation scale; and
8. content validation.

Use in medical and dental education

- The KF assessment format proposed by Bordage und Page was developed to replace the commonly used written assessment of procedural knowledge using patient management problems (PMP) in medical specialty examinations [64], [65].
- Transfer to undergraduate education by Hatala & Norman [66], used worldwide since in medical education as a written assessment format to evaluate con-

text-specific procedural knowledge during the study phase and post-graduate education [67], [68].

- Recognized testing format in the German-speaking countries in the field of medicine (see the detailed information on the design and implementation of assessments published by the medical schools at the Universities of Bern and Graz [34], [60], [69].
- Studies and reports on the use of the KFE as a written assessment at German medical schools, including internal medicine (Universities of Freiburg, Heidelberg, and Munich [70], Universities of Heidelberg, Tübingen [71]), hematology and oncology (University of Düsseldorf [72]), communication skills (University of Witten-Herdecke [73]).
- Extensive pilot project in veterinary medicine at the school of veterinary medicine at the University of Hanover [74].
- Only a few reports of KF problems used as a written assessment format in dental education [75], [76].

The key feature exam is a valid and reliable instrument for assessing context-specific, procedural knowledge in connection with solving a clinical problem and represents a meaningful addition to the written testing formats currently used in dental education. KFEs can also be used in independent learning with virtual patient cases. For practical reasons, the computer-based format with the long menu response format is preferable to the paper-based version. It is also easier to hinder examinees from returning to previous pages or turning the pages out of order. To increase reliability, it is better to use many short KF cases (at least 15) with a maximum of three questions each than to use fewer, more in-depth cases with four or more questions.

Key Features Exam

Validity

- High content validity (92-94%) when graded by teachers/examiners [63], [65], [67].
- Piloting and regular review of the key features by students, teachers/examiners is a pre-requisite for high content validity [34], [63], [65].
- When a LM format is intended, a WI format is recommended for the pilot to improve the quality of the LMs (supplement missing answers and distracters) [34].
- Correlation between KFE scores and other assessment scores (e.g. MCQ) is only moderate ($\gamma=0.35-0.54$, [66], [70] which can be explained by the reference to different competency levels.

Reliability

- Reliability of the KF format is higher than for the PMP format [65].
- Due to greater case specificity [48], reliability is directly dependent on the number of KF problems (KFP=cases) → number of cases should be as high as possible; number of questions on each case should not exceed three items, since four or more reduces reliability [77].

- The selected response format appears to influence reliability, when the same number of KF cases are used:
- 15 KFPs with 1-4 questions, 2h length, WI format: Cronbach's $\alpha=0.49$ [66]
- 15 KFPs with 3-5 questions, 1.5h length, computer-based LM format: Cronbach's $\alpha=0.65$ [70] $\rightarrow \alpha=0.75$ is possible with 25 KFs!

Acceptance

- Students: relatively high acceptance [74], [78]: evaluated as realistic and supportive of practical learning.

Cost

- Generation and validation of a KFE involves great amounts of time and staff [67].

Feasibility

- Generating KFEs is more difficult and requires more time than an MCQ [60], [69].
- Necessary testing time depends on the selected response format: LM>WI>SM>MC [79].
- The advantages of LM response format (lower cueing effect than MCQ/SM, higher inter-rater reliability than WI) can be realized by using computer-based testing with a moderate testing time [70], [72], [79].
- Testing time for 15 KFPs with 3-5 questions is 90 minutes for a computer-based exam [70] and 120 minutes for a paper-based test with a WI response format [66].
- Practical examples exist [68], [71], [75], [76], [80].

Influence on teaching and learning

- KFE format is closer to a real patient situation, promotes the learning of clinically relevant material and practical case-based learning [81].

While study programs in dental medicine do impart advanced theoretical knowledge, they also require students to develop manual skills. Consequently, suitable assessment formats are needed to measure not only factual and procedural knowledge but also to give students an opportunity to demonstrate their practical abilities (shows how, [33]) and to evaluate this objectively. Simply "knows how" is raised a level to "shows how".

When creating such assessments, the learning objectives should be selected in advance and only those which represent a practical competency level should be employed. Standardization of test and examiner allows for an objective assessment of student performance. Suitable assessment formats for this are **objective structured clinical examinations** (OSCE), objective structured practical examinations (OSPE) and the use of simulated, or standardized, patients (SP).

An OSCE is appropriate for evaluating practical skills and the ability to communicate [14]. Students pass through different stations where particular practical skills are demonstrated (including partial treatments) or mock medical consultations are conducted. Evaluations are documented using a checklist created by a group of ex-

perts according to how the exam content is weighted. Test time per station is around five minutes; two minutes need to be planned for the examinee to change stations and for the examiner to make final notes or give feedback. Use in medical and dental education

- Widely used internationally in all clinical subjects since its introduction.
- Can be used in undergraduate and post-graduate programs [82], [83], [84], [85].
- There are many examples of use in dental disciplines: pre-clinical phase [86], [87], [88], orthodontics [89], [90], oro-maxillofacial surgery [91], [92], [93], restorative dentistry [87], [94], [95], [96], parodontology [97], clinical prosthetics [86], pediatric dentistry [98], radiology [99], microbiology [94], [97].
- Interdisciplinary OSCE [87], [94], [100], [101].
- Integration of an OSCE in the preliminary dental exam [102].
- Also used in dental education to evaluate communication skills [103], [104], problem-solving skills, and critical thinking [105].
- If possible, feedback should be included as part of the exam.

The OSCE is a reliable and valid testing format to assess individual competencies; it enjoys a high level of acceptance by students and teachers.

Objective Structured Clinical Examination

Validity

- Predictive validity
- Significant correlation between OSCE and performance on practical tests and scores on preliminary practical medical exams $p<0.01$ [87]
- No correlation between OSCE and MCQ [105]
- High content and construct validities [4], [106]
- High face validity [107]
- Acceptable predictive validity [108]
- Caution required if students have a language problem or suffer from high levels of stress [109]
- Attention must be paid to the blueprint [106], [110]
- Determine content areas early [110]
- Define questions within the content areas [110]

Reliability

- Cronbach's α between 0.11-0.97 [4]
- High reliability among OSCEs, with fewer than $n=10$ stations it being approximately 0.56, with more than $n=10$ stations, 0.74 [111]
- Varying recommendations on station number:
- at least 19 [4]
- 14-18 for 5-10 minutes each [106], [112]
- Stations with an SP should be assessed for at least 15 minutes [110]
- The more examiners, the higher the values.
- [111], [113] Method of evaluation critical: high values for global assessments, combinations of global assess-

ments and checklists are good, only checklists alone are least suitable

- Post-OSCE tests increase the reliability [110].

Acceptance

- Students: high acceptance, appropriate testing format for functional skills [96]
- Teachers/examiners: high acceptance [112], [114], [115], [116]

Cost

- 181 Euros/examinee
- 86-130 Euros/examinee [4], [117]
- 2.5h/examinee [118]
- \$15-200/examinee [117], [119], [120]
- Higher costs when including SPs [121]
- \$21-200/examinee [122]

Feasibility

- Testing format demands great amounts of time and resources [106], [119]
- Thorough preparation needed:
- Establishing shared structures helps on interdisciplinary OSCEs [100].
- Evaluation by external examiners is recommended.
- Ensure the quality of SPs
- Station content should be selected to match the OSCE scenario.
- Peer reviews pre- and post-OSCE (psychometric analysis with difficulty, discrimination, etc. is recommended)
- Take the extent of the examiner's experience, field of expertise, sex, and level of fatigue into consideration [3], [106], [112].
- Practical examples exist [95], [102].

Influence on teaching and learning

- Positive influence on learning [106], [108], [123]
- Stimulates learning [112]
- Learning at the stations has little to do with the reality of patients [112].
- Allot time for feedback [110]

Due to the extensive preparation involved before and after its administration, interdisciplinary cooperation is recommended to minimize this disadvantage. OSCEs can be substituted for previously used assessment formats or supplement them in meaningful ways. A sufficient number of stations ($n > 10$), a blueprint, peer review of station content and the scoring criteria, as well as a balance among the modes of evaluation (global, checklist, combination), training the examiners and, if needed, conducting a pilot OSCE should be taken into account when designing an OSCE. A special type of OSCE is embodied in the objective structured practical examination (OSPE) during which practical skills, knowledge and/or interpretation of data are demonstrated in a non-clinical situation [124]. These assessments can be conducted in labs or simulated stations in SimLab. In contrast to the

OSCE, an entire process can be evaluated through to the end result (for instance, a dental filling).

It is possible to confidently assess practical skills and/or the interpretation of clinical data with the OSPE. This format involves a reliable and valid assessment method to evaluate individual competencies; The OSPE enjoys a high level of acceptance by students and teachers.

Objective Structured Practical Examination

Validity

- High validity, $\gamma > 7$
- High construct validity

Reliability

- High reliability among the stations, Cronbach's $\alpha = 0.8$ [125]
- Inter-rater reliability ICC > 0.7
- High inter-rater reliability with equivalent levels of experience and knowledge among examiners, $\gamma = 0.79-0.93$; $p < 0.001$

Acceptance

- Students: high acceptance [126], [127]
- felt to be a "fair test" [128]
- preferred over traditional exam formats [126]
- Teachers: relevant, fair, objective and reliable testing format

Cost

- No information available

Feasibility

- Requires extensive planning and teamwork [128]

Influence on teaching and learning

- Individual competencies can be assessed, the need to demonstrate factual and procedural knowledge influences learning behavior [128].
- Makes strengths and weaknesses in practical skills discernible [129]
- Stimulates learning [129]
- Positive learning experience [130]

Defined grading criteria for each step within a process are necessary.

Use in medical and dental education

- OSPEs are administered around the world in medicine, including pharmacology [128], physiology, forensic medicine [130], and dentistry [131], [132].
- In Germany they are primarily used in the pre-clinical phase of dental education [133].

Simulated, or standardized, patients in dental education are specially trained (lay) actors who are capable of acting out common clinical pictures or typical occasions for dental consultations. They are used for both practicing and assessing doctor-patient consultations and examination techniques; the use of an SP also provides opportunities to learn how to conduct physical examinations

and acquire better communication skills. It is also possible to incorporate SPs into assessments, most frequently in OSCE scenarios.

Standardized patients can be used to assess doctor-patient interactions and examination techniques. They are especially suited for evaluation of clinical competencies and communication skills within the scope of an OSCE. When implementing this, the complexity of the case should be tailored to match the testing scenario.

Standardized Patients

Validity

- Assesses clinical competencies [134]

Reliability

- Consistent examination
- (No significant differences between exam cohorts and time points) [135]

Acceptance

- Use of standardized patients (SP) within the scope of an OSCE station [136]

Cost

- 10-18 Euro/examinee [136]

Feasibility

- Case complexity can be controlled and adjusted to reflect educational level [137]
- Faculty members can determine relevant learning objectives and coordinate role creation.
- Greater need for time and staff to select and train SPs and to monitor for quality [137]
- Checklists to record all SP observations of the doctor-patient consultation [138]
- Practical examples exist [139].

Influences on teaching and learning

- Improves students' clinical skills [140]

Use in medical and dental education

- This method has been used in clinical education since the 1960s [138].
- Patient contact can be simulated under standardized conditions [139].
- SPs can also provide feedback and critique the examinee's abilities [139].

The term "workplace-based assessment" (WBA) encompasses a wide variety of testing scenarios meant to assess practical skills associated with treating patients in complex situations.

The **clinical evaluation exercise** (CEX) involves a workplace-based assessment in the clinical setting that stretches over a longer period of time (several hours to days) and covering treatment processes during which an examinee conducts a consultation with a single patient recording a patient health history and carrying out a physical examination. A maximum of two assessors

should participate, but do not generally have to be present the entire time. Often the data is collected from the patient without the assessor being present. This assessment format, also known as the tCEX (traditional CEX), represents a single event measure.

Use in medical and dental education

- Originally developed in the 1960s as an assessment in internal medicine by the American Board of Internal Medicine (ABIM), it replaced the oral examination as the standard method in 1972 [141], [142].
- Replaced by the mini-CEX around 1995 [143], [144].
- No documented examples of use in dental education are found in the literature.

This assessment format is an instrument of low validity and poor reliability for testing practical skills in complex situations. It is possible to improve the assessment by using the greatest number of patients possible (cases), the greatest number of assessors possible, and the most structured evaluation instruments possible. In addition, providing feedback as part of this testing format should be mandatory. Overall, it can be asserted that in dental education the CEX is a reasonable assessment format for measuring practical competencies in complex situations only if the previously mentioned attempts at improvement have been made.

Clinical Evaluation Exercise

Validity

- Insufficient content validity; does not completely cover curricular learning objectives [145]
- Simulated situation, does not correspond with the reality of medical practice since it is too long and detailed [144]

Reliability

- Questionable reliability since only few exercises can be done due to the great amount of time needed [146]
- Low inter-rater reliability [147]
- Cronbach's α is 0.24 for one case and even for two cases only 0.39 [141].

Acceptance

- Low level of acceptance since it is very dependent on the assessor [148]

Cost

- Less costly than the OSCE because real patients are used who do not need to be trained [145]

Feasibility

- Relatively simple since no special preparation is necessary [141]
- Practical examples exist [142], [143].

Influence on teaching and learning

- Patient-oriented, real-life situations [141]

The **mini-clinical evaluation exercise** (mCEX) is a patient-centered assessment format in the clinical setting that, in contrast to the CEX, requires a shorter amount of time and always includes feedback (approximately 15 minutes of assessment and 10 minutes of feedback). This testing format can be described as having three phases: observation, documentation and feedback. Over the course of the assessment, several assessors observe the examinee and evaluate what they see according to pre-defined criteria. Medical care is given to more than one patient under normal circumstances with a focus on communication and clinical examination [144]. Evaluations are generally formulated according to defined criteria valid for each examinee. These criteria can consist of a rating scale and/or short written comments. The difficulty remains in terms of the different patients undergoing physical examination. Viewed according to Miller's pyramid, a high level of practical skill is attained. Strictly speaking, it is a structured clinical observation.

Mini-Clinical Evaluation Exercise

Validity

- Higher validity than CEX [149]
- Acceptable validity and reliability have been demonstrated [146], [150].
- Able to validly differentiate between competency levels (first year, second year, etc.) [151]

Reliability

- Low inter-rater reliability [149]
- A minimum of 10 evaluations are necessary to yield reliable results; a larger number is better [151]
- At least 12-14 evaluations are recommended per year if there are different assessors to increase inter-rater reliability [152].
- Reliability of $G=0.4$ for 10 evaluations; $G=0.8$ for 50 evaluations [151]
- Dependent on number of assessors: if there is one examiner, a minimum of eight observations of different patients are necessary for a reliability of 0.8, in the case of two, four are necessary, and for three examiners, three observations [153].
- Nine items are better than five to cover differences in competencies [154].

Acceptance

- High level of satisfaction for students and teachers [151], [155], [156]
- Implementation is at present slow, since it involves something new [156].
- Partially problematic due to discrepancies between self-assessment and assessment by another [157].

Cost

- Substantial expense as a consequence of the amount of time needed [158], [159].

Feasibility

- Observations of authentic doctor-patient interactions by different educators in different situations; feedback on different clinical pictures at different locations each with a different focus [155]
- Thorough planning is necessary because giving feedback takes 8-17 minutes [155], [160].
- Relatively simple to implement with enough flexibility in the dental setting [161]
- Practical examples exist [162].

Influence on teaching and learning

- Improvement in competency through regular feedback from experts [163]
- Examiner/examinee receive feedback or a clear impression of clinical work making targeted mentoring possible [156].
- Giving constructive feedback must be learned and practiced; teaching skills are needed [164].
- No new discoveries or knowledge in comparison with traditional evaluation procedures [158]
- No influence in comparison with control groups [153]
- Learning objectives must reflect teaching content [165].
- Predictive validity between OSCE and mCEX cannot be demonstrated [165].

This assessment format is frequently referred to as the mCEX (mini-CEX) and represents a single event measure. Use in medical and dental education

- Developed in 1995 by Norcini [144]; replaced the tCEX in the 1990s.
- Reliability depends heavily on the number of assessors and cases [151], [153].
- Several documented instances in the literature of use in dental medicine (Dental Foundation Training in Great Britain), however, often without any precise information on the evaluation instruments [161], [162].

The mCEX is a valid and reliable instrument to assess practical skills in complex situations. Options for improvement include 1. increasing the number of response items (nine are better than five) or increasing the number of observations (a minimum of 10 observations are needed) and 2. offering train-the-teacher programs (for instance in the form of video demonstrations and role playing). Longitudinal use is recommended with implementation conceivable in a wide variety of different settings (including high-stakes exams). The mCEX format is a good testing format for use in dental education to measure practical competencies in dental medicine.

Entrustable professional activities close the gap between the theory of competency-based education and patient-centered practice in a clinical context [166]. This method first became known for its use in the area of post-graduate education; since 2013 it has also appeared in undergraduate medical education [167], [168]. The integration of theoretical and practical knowledge to solve complex problems is assessed (e.g. anamnesis, clinical examina-

tion of a patient in connection with different reasons for seeking medical advice) using existing competency-based roles, such as those defined by CanMeds or ACGME. During the assessment it is determined whether the examinee is able to perform the activity while receiving directions, under supervision, with occasional assistance, or independently [169], [170]. As a result, different performance levels can be identified [171]. It is not individual learning objectives that are assessed, but rather an overall activity centering on a patient [172]. In order to differentiate EPAs from general learning objectives, it is recommended that following sentence be completed: One day, the doctor/dentist will be expected to do (insert particular activity) without direct supervision [166]. According to its definition, an EPA should include activities that are important to daily practice, very often are subject to error when being performed, and integrate multiple competencies [172], [173]. Consequently, an EPA consists of diverse roles, each role, in turn, of multiple learning objectives, and each learning objective of different performance levels. The assessment can be a direct or indirect observation and include feedback. It is crucial that the observed performance of the examinee is combined with the performance evaluation over a defined period of time.

Entrustable Professional Activities

Validity

- High face validity [174]

Reliability

- Low inter-rater reliability [175]

Acceptance

- Potential for wide acceptance [166]
- Helps those learning to develop their own study schedule [176]
- Helps the entire faculty to maintain transparency in education [176]

Costs

- No information available

Feasibility

- Initially requires intensive, well thought-out preparation while EPAs are being designed [177]
- 20-30 EPAs are recommended for a degree program [177]
- Practical examples exist [178], [179]

Influence on teaching and learning

- EPAs require numerous competencies in an integrated, holistic manner [177].
- Methods of evaluation that focus on the required degree of supervision [180]
- Feedback is vital [174].
- Support from the faculty is necessary [175].

- Enables a broad (panoramic) view of the educational program [174].

A commonly reported combination is that of the mCEX with MSF (Multi-source feedback). Strictly speaking, this involves a multiple event measure.

Use in medical and dental education

- Introduced in the Netherlands by ten Cate in 2005; since then it has been used in the fields of surgery, family medicine, internal medicine, neurology, emergency medicine, pediatrics, urology, and is used widely by the Royal Australian and New Zealand College of Psychiatrists [178], [179].
- Initially in the pilot phase in German medical education [165].
- No documented instances of use in dental medicine

EPAs are a relatively new, little researched instrument for assessing practical skills in complex situations. The implementation of EPAs requires extensive and well thought-out preparation when determining the focus. To the extent possible, a maximum of 30 interdisciplinary EPAs per curricular unit should be defined drawing upon input from university instructors and practicing physicians or dentists. EPAs create a realistic link between competency-based learning objectives and higher level activities. Train-the-teacher programs (with practice giving feedback) should improve implementation. Longitudinal use is recommended. Implementation is conceivable in a wide variety of settings, including high-stakes exams. The EPA format represents an innovative approach with great future potential in terms of assessing practical skills in complex situations in dental education.

Similar to the mCEX, **Directly Observed Procedural Skills** (DOPS) entail a short workplace-based assessment in a clinical setting that includes feedback (approximately 15 minutes of assessment and 10 minutes of feedback). This also involves a three-phase assessment in which observation, documentation and feedback occur. Treatment given to (multiple) patients under conditions typical to a medical practice, as with the mCEX, but with a focus on manual skills and interventions observed by several assessors and evaluated according to defined criteria. This assessment format also represents a single event measure.

Use in medical and dental education

- Originally introduced in the United Kingdom by the General Medical Council in 2002 [144].
- Use reported in the fields of general medicine, surgery, and internal medicine [181].
- International reports of use in dentistry in Iran (universities of Shiraz and Mashad) and at Kings College in London [182], [183].

DOPS is a valid and reliable instrument to evaluate practical skills in complex situations. It is possible to improve this format by having three assessors intervene during two observations, conducting at least two observations, and by holding train-the-teacher sessions. Overall, longitudinal use is recommended. Implementation is

conceivable in diverse settings, including high-stakes exams. The DOPS format is a very reasonable testing format to capture practical skills in complex situations during dental education.

Directly Observed Procedural Skills

Validity

- High face validity [181]
- Formative assessment tool [182]
- Significantly different from MCQ; provides different assessments of student performance [182]
- Separate assessment tool that does not enable an overall evaluation; a system with different possibilities is needed [184].
- DOPS efficiently evaluates practical skills [182].

Reliability

- To achieve a high reliability, at least three assessors should observe a student during two different case scenarios [181].
- $G=0.81$ [185]
- Internal consistency is 0.94 and inter-rater reliability is 0.81
- Students do not view it as suitable for improving inter-rater reliability [186].
- Substantial differences between the assessors can influence the validity of the results if there has not been strict standardization [187].
- Good reliability and consensus among assessors is possible [188].
- Fewer assessors are needed in comparison with the mCEX [160].
- Fewer assessors and cases are needed in comparison with the mCEX [181].
- Higher item correlation values than for the mCEX: 0.7-0.8 versus 0.5-0.8 [150], [189]
- Reliability depends on the case [181].
- Reliability independent of process [160]

Acceptance

- High acceptance by students [186]
- Examinees find the scenarios to be stressful, but appreciate the feedback [190].

Cost

- Substantial expense is to be expected [159], [191].

Feasibility

- Great amount of time needed [163], [182]
- Great amount of time needed for preparing DOPS, including giving feedback [160]
- To increase the learning effect, it is necessary to give feedback directly after the assessment and to address strengths and weaknesses [192].
- Assessors must be trained in advance [12].
- It is feasible to use only one assessor [193].

Influence on teaching and learning

- Examinees perceive a positive influence on independence and the learning process [186].
- DOPS assessment improves practical clinical skills [192].
- Positive effect through directly observing the learner [192]
- Promotes an in-depth approach to learning in the clinical context [21]
- Positive influence on student reflections [181]
- Seventy percent of those observed believe that DOPS is helpful for improving practical skills [194].
- Compared to control groups there are significantly better results for DOPS regarding practical skills [195].
- Can also be used in peer arrangements in the pre-clinical and clinical context [183]

The **Portfolio** as an assessment tool is a pre-defined, objectives-centered collection of student learning activities with assigned self-reflection exercises, as well as feedback [20]. Portfolio contents are developed in alignment with the learning process; the following aspects can be taken into consideration: personal experiences (what was done, seen, written, created?), learning process (awareness that what has been experienced is relevant to future medical or dental practice), documentation (certificates, etc.), future goals regarding learning (looking ahead), and learning environments [196]. Portfolios are a multiple event measure.

Use in medical and dental education

- Portfolio-based learning was introduced in 1993 by the Royal College of General Practitioners, Portfolio assessing described by Shulman in 1998 [197], [198].
- Publications in the fields of general medicine, otorhinolaryngology, internal medicine, pediatrics, public health at universities in Maastricht (NL), Nottingham (GB), and Arkansas (USA) [196].
- Found in German medical education in Cologne [196].
- International reports of use in dentistry [199], [200], [201].

The portfolio entails a highly valid and reliable instrument for evaluating practical skills in complex situations, one that assesses collected, cumulative information about performance and development. Possibilities for optimization exist when more than one neutral grader is used, the student's mentor is not one of these graders, and train-the-teacher sessions on giving feedback are held. Longitudinal use is recommended. Implementation is conceivable in diverse setting, including high-stakes exams. The portfolio format represents a valuable assessment format to evaluate practical skills in complex situations in dental education.

Portfolio

Validity

- Good validity if there is an appropriate selection of all required competency areas [202], [203].

Reliability

- Cronbach's α is 0.8 with four graders [204]
- Cronbach's α is 0.8 with 15 portfolio entries and two graders [202].
- Use of a clear, competency-based master plan, clear grading criteria, inclusion of guidelines and experienced graders for development and evaluation [202], [203]
- Uniform and consistent grading is difficult [200].

Acceptance

- Portfolios are viewed as time consuming, a source of anxiety and not very effective [205].
- The acceptance of portfolios decreases the longer students spend time on them [205].

Cost

- No information available

Feasibility

- A portfolio typically includes seven case reports, two presentations, three self-reflections [202].
- Typical content includes diagnoses and treatment plans [202].
- Problematic since there is a conflict when portfolios are used for both assessment and learning [205].
- Difficulties being self-critical and honest [205]
- Conducting interviews with students about portfolio content improved feasibility [206]
- Practical examples exist [199], [201].

Influence on teaching and learning

- Allows the assessment of competencies that could not otherwise be measured [200]
- Portfolio content must be aligned with the learning objectives [202].
- Increases self-knowledge and encourages critical thinking [205]
- Improves the ability to learn independently and connects theory with practice [205]
- Time consuming for grader and student [200], [207]
- Students receive constructive feedback [207].
- Calibration and validation are critically important [200].
- Provides cumulative information on performance and progress [205]
- When it is known that the portfolio will be graded, students attempt to fulfill expectations which, in turn, affects the portfolio's content and educational value [205].
- Positive effects are heavily dependent on the support, direction, time commitment and feedback given by the teacher [205].

Multi-source feedback, also known as 360-degree feedback (MSF, multi-rater feedback), involves a workplace-based assessment in a clinical setting incorporating different groups of people associated with that particular work setting and the examinee (peers, dentists, nursing staff, patients, administrators, etc.). The focus of the ob-

servations is on professional conduct and teamwork, as well as the examinee taking responsibility as the person in charge [208], [209]. These aspects are observed by several assessors and evaluated according to defined criteria. The "supervisor" is given a special role in this testing scenario: this person collects all the results and gives them to the examinee. As a result, the individuals who have given feedback remain anonymous. The student receives a comprehensive picture based on all the input from different sources. High acceptance is achieved through selection of the assessors. Narrative comments and metric rating scales can be combined. This format entails a multiple event measure.

Use in medical and dental education

- Used in medicine since 1970, widespread in North America (Canada and USA), Europe (England, Holland), and Asia [210], [211].
- Reports of use in the fields of general medicine, internal medicine, surgery, gynecology, psychiatry, pathology, and radiology, etc. [210].
- Used in dental medicine by the Royal College of Surgeons of England, University of Bristol, UK Committee of Postgraduate Dental Deans.
- Validated instruments exist for evaluation (PAR: Physicians Achievement Review, SPRAT: Sheffield Peer Assessment Tool).

This method consists of a highly valid and reliable instrument for evaluating practical skills in complex situations.

Multisource Evaluations

Validity

- Can make it easier to evaluate inter-personal and communicative skills in particular [212]
- Good validity [213]

Reliability

- Review: to reach a value of 0.9 minimum for Cronbach's α , eight medical assessors, eight non-medical assessors and 25 patients must participate [210]
- High internal consistency ($=0.8$) with five assessors on two observed occasions [214]
- To reach a value of 0.8 for Cronbach's α , a minimum of 11 assessors must participate [215].
- Value for Cronbach's α is 0.98 [216].
- Problematic due to the number of assessors required [217]

Acceptance

- Rated 4.5 by examinees on a scale of 1-7 [214]
- Rated 5.3 by assessors on a scale of 1-7 [214]
- Evaluations are possibly too positive since anonymization is not fully trusted [217]

Cost

- Expense needs to be taken into account before implementation [159].

Feasibility

- Rated 4.4 by examinees on a scale of 1-7 [214]
- Rated 5.1 by assessors on a scale of 1-7 [214]
- Evaluations are generally verified via questionnaires making the process simple [159].
- To achieve a valid assessment, a certain number of evaluations are necessary; however, not all are possible to do [217].
- Ideally, feedback is gathered over a longer period of time [217].
- Can be easily implemented, even in a busy hospital [211], [218]

Influence on teaching and learning

- General improvement in clinical work, communication with co-workers and patients [219]
- Rated 4.2 by examinees on a scale of 1-7 [214]
- Rated 4.4 by assessors on a scale of 1-7 [214]
- Improvement of the evaluation process, advantage of receiving more detailed information and being exposed to different perspectives [217]
- Varying results: improvement in communication and conduct after receiving 360° feedback [220].
- Immensely time consuming and no improvement in assessment as a consequence of the feedback [221]
- It is possible to identify weak performers at an early stage [218].
- Feedback from SPs for students also possible [222].

Belonging to the success factors are a clear definition of the objectives and the sources of feedback. An important role is played by the selection of the assessors, credibility of the assessors and their familiarity with the situation under evaluation, along with the anonymity of the individuals supplying the feedback. This format can be optimized by using approximately five assessors for two observed situations and holding train-the-teacher sessions concerning constructive feedback. The combination of external feedback with self-evaluation by the examinee can be helpful, as can be jointly determining specific learning objectives for the future, including the discussion and documentation of concrete learning opportunities and supports. Longitudinal use is recommended. Implementation is also conceivable in diverse setting, including high-stakes exams. The MSF format represents a valuable assessment format for evaluating practical skills in complex situations in dental education.

5. Conclusion

The range of assessment methods presented in this overview significantly broadens the spectrum of already established university-specific exams—mostly MCQs and (structured) oral exams. Each of the methods outlined here meets different requirements and thus covers different competency levels. This must be taken into particular consideration by those who are involved in designing,

administering and evaluating assessments in dental medicine.

When developing and implementing a curriculum, not only the choice of assessment format is critical but also noting the general functions of an exam, which in turn has an effect on the curriculum [223]: assessments can be summative or formative. Summative assessments usually come at the end of a semester or after a skill has been taught in order to evaluate learning outcomes. Formative assessments are reflective of the learning process itself and do not determine whether a student passes or fails a course or is ultimately successful in displaying the mastery of a particular competency. Such an assessment shows students their current level of proficiency and is supposed to support the learning process through reflection by students on their weaknesses. Purely formative assessments are few in the face of limited staffing resources and time constraints, but are an ideal tool for fostering the learning process.

Within the scope of drafting the NKLZ it became clear that in the future other assessment formats will be needed in addition to the established methods such as oral examinations and MC exams; these new formats will need to measure required practical skills in dental medicine, not just in the Skills Lab, but also in patient treatment. Each assessment format should correspond with the targeted competency levels.

The presentation of the assessment formats in this overview enables quick orientation within each method and makes reference to relevant literature for those who wish to know more. Including even more detailed information on each of the assessment formats would have compromised the intended character of this article as an overview. Along with theoretical knowledge of an assessment format, it is important to engage in direct exchange with colleagues in higher education who are already following a particular method. For this reason, it is desirable, and perhaps the task of the relevant working groups, to establish a network of professionals who have already gathered experience with special assessment formats and who are willing to make themselves available to those with questions. Depending upon demand, continuing education programs could emerge from such a network providing substantial assistance in implementing new assessment formats.

6. Outlook

With the new licensing regulations for dentists (Approbationsordnung), German dental education will be brought up to date and more closely linked to medical education. The assessment methods mentioned as examples in the NKLZ and outlined in this paper demonstrate the various options for assessing at the competency level. After experience has been gathered with university examinations in dental education and following scientific analysis of these testing methods, additional appropriate assessment methods should be included in the licensing requirements

for dentistry. These should also be used to improve the quality of the state examinations.

Together with the introduction of the NKLZ, compiling experience in organizing, preparing, administering, conducting and evaluating the assessment formats profiled here will be an important task in the coming years, whereby dental medicine can make good use of the competencies under development for medical students since 2002. Dental medicine can also bring to bear its own experience and expertise in the assessment of practical skills. Our shared goal should be to continue developing assessment formats for the different competency levels in dental and medical education in cooperation with the German medical schools.

Acknowledgements

The authors wish to extend their gratitude to all those who have helped to write, edit and finalize this article. Special thanks to the executive board of AKWLZ, especially Prof. P. Hahn, MME (University of Freiburg) and Prof. H.-J. Wenz, MME (University of Kiel), for the detailed feedback and suggestions for improvement.

Competing interests

The authors declare that they have no competing interests.

Authors

Authors are listed in alphabetical order.

References

- Biggs J. Enhancing teaching through constructive alignment. *High Educ.* 1996;32(3):347–364. DOI: 10.1007/BF00138871
- van der Vleuten CP, Verwijnen GM, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 1996;18(2):103–109. DOI: 10.3109/01421599609034142
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206–214. DOI: 10.3109/0142159X.2011.551559
- Chenot JF, Ehrhardt M. Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med.* 2003;79:1-7.
- Examination and Assessments: Academic Integrity [Internet]. Imperial College London. [cited 2015 Jan 17]. Zugänglich unter/available from: <https://workspace.imperial.ac.uk/registry/Public/Procedures%20and%20Regulations/Policies%20and%20Procedures/Examination%20and%20Assessment%20Academic%20Integrity.pdf>
- Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. *GMS Z Med Ausbild.* 2014;31(3):Doc34. DOI: 10.3205/zma000926
- Nationaler Kompetenzbasierter Lernzielkatalog Zahnmedizin (NKLZ) [Internet]. [cited 2016 June 14]. Zugänglich unter/available from: http://www.nklz.de/files/nklz_katalog_20150706.pdf
- Möltner A, Schultz JH, Briem S, Böker T, Schellberg D, Jünger J. Grundlegende testtheoretische Auswertungen medizinischer Prüfungsaufgaben und ihre Verwendung bei der Aufgabenrevision. *GMS Z Med Ausbild.* 2005;22(4):Doc138. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000138.shtml>
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med Educ.* 1985;19(3):238–247. DOI: 10.1111/j.1365-2923.1985.tb01314.x
- Roloff S. Mündliche Prüfungen [Internet]. 1 p. [cited 2016 June 14]. Zugänglich unter/available from: http://www.hochschuldidaktik.net/documents_public/20121127-Roloff-MuendlPruef.pdf
- Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian.* 2005;12(1):19–24. DOI: 10.1016/S1322-7696(08)60478-3
- Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Adv Health Sci Educ.* 2010;15(2):277–289. DOI: 10.1007/s10459-008-9111-9
- Harden RM, Lever R, Wilson GM. Two systems of marking objective examination questions. *Lancet.* 1969;293(7584):40–42. DOI: 10.1016/S0140-6736(69)90999-4
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ.* 1975;1(5955):447–451. DOI: 10.1136/bmj.1.5955.447
- Lennox B. Marking multiple-choice examinations. *Br J Med Educ.* 1967;1(3):203–211. DOI: 10.1111/j.1365-2923.1967.tb01698.x
- McCarthy WH. An assessment of the influence of cueing items in objective examinations. *J Med Ed.* 1966;41(3):263–266. DOI: 10.1097/00001888-196603000-00010
- Hart IR, Competence OCOAC, Harden RM, Centre RCOPASOCSMEAR, médecins et chirurgiens du Canada des CR. *Further Developments in Assessing Clinical Competence.* Boston: Can-Heal Publications; 1987.
- Van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol.* 2010;24(6):703–719. DOI: 10.1016/j.bpobgyn.2010.04.001
- Schoonheim-Klein ME, Habets LL, Aartman IH, van der Vleuten CP, Hoogstraten J, van der Velden U. Implementing an Objective Structured Clinical Examination (OSCE) in dental education: effects on students' learning strategies. *Eur J Dent Educ.* 2006;10(4):226–235. DOI: 10.1111/j.1600-0579.2006.00421.x
- Fischer MR, Holzer M, Jünger J. Prüfungen an den medizinischen Fakultäten - Qualität, Verantwortung und Perspektiven. *GMS Z Med Ausbild.* 2010;27(5):Doc66. DOI: 10.3205/zma000703

21. Cobb KA, Brown G, Jaarsma DADC, Hammond RA. The educational impact of assessment: a comparison of DOPS and MCQs. *Med Teach*. 2013;35(11):e1598–1607.
22. Elmer A, Grifka J. Vergleich von Prüfungsmethoden in der klinischen Ausbildung. *Gesundheitswesen (Suppl Med Ausbild)*. 1998;15(Suppl1):14–17. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/1/Elmer_A,_Grifka_J.pdf
23. Sadaf S, Khan S, Ali SK. Tips for developing a valid and reliable bank of multiple choice questions (MCQs). *Educ Health*. 2012;25(3):195–197. DOI: 10.4103/1357-6283.109786
24. Wenzel A, Kirkevang L. Students' attitudes to digital radiography and measurement accuracy of two digital systems in connection with root canal treatment. *Eur J Dent Educ*. 2004;8(4):167–171. DOI: 10.1111/j.1600-0579.2004.00347.x
25. Yang JC, Laube DW. Improvement of reliability of an oral examination by a structured evaluation instrument. *J Med Educ*. 1983;58(11):864–872.
26. Hottinger U, Krebs R, Hofer R, Feller S, Bloch R. Strukturierte mündliche Prüfung für die ärztliche Schlussprüfung–Entwicklung und Erprobung im Rahmen eines Pilotprojekts. Bern: Universität Bern; 2004.
27. Wass V, Wakeford R, Neighbour R, van der Vleuten C, Royal College of General Practitioners. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ*. 2003;37(2):126–131. DOI: 10.1046/j.1365-2923.2003.01417.x
28. Schubert A, Tetzlaff JE, Tan M, Ryckman JV, Mascha E. Consistency, inter-rater reliability, and validity of 441 consecutive mock oral examinations in anesthesiology: implications for use as a tool for assessment of residents. *Anesthesiology*. 1999;91(1):288–298. DOI: 10.1097/0000542-199907000-00037
29. Kearney RA, Puchalski SA, Yang HYH, Skakun EN. The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Can J Anaesth*. 2002;49(3):232–236. DOI: 10.1007/BF03020520
30. Postgraduate Medical Education and Training Board. *Developing and Maintaining an Assessment System*. London: General Medical Council; 2007. S.1.
31. van der Vleuten CP. *Assessment of the Future* [Internet]. [cited 2016 June 14]. Zugänglich unter/available from: <http://www.youtube.com/watch?v=bvFbmTRVjIE>
32. Möltner A, Schellberg D, Briem S, Böker T, Schultz JH, Jünger J. Wo Cronbachs alpha nicht mehr reicht. *GMS Z Med Ausbild*. 2005;22(4):Doc137. Zugänglich unter/available from: <http://www.egms.de/de/journals/zma/2005-22/zma000137.shtml>
33. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63–67. DOI: 10.1097/00001888-199009000-00045
34. Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild*. 2006;23(3):Doc50. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000269.shtml>
35. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–949. DOI: 10.1016/S0140-6736(00)04221-5
36. Knox J. What is.... a Modified Essay Question? *Med Teach*. 1989;11(1):51–57. DOI: 10.3109/01421598909146276
37. Knox JD, Bouchier IA. Communication skills teaching, learning and assessment. *Med Educ*. 1985;19(4):285–289. DOI: 10.1111/j.1365-2923.1985.tb01322.x
38. Feletti GI. Reliability and validity studies on modified essay questions. *J Med Educ*. 1980;55(11):933–941. DOI: 10.1097/00001888-198011000-00006
39. Rabinowitz HK, Hojat M. A comparison of the modified essay question and multiple choice question formats: their relationship to clinical performance. *Fam Med*. 1989;21(5):364–367.
40. Lockie C, McAleer S, Mulholland H, Neighbour R, Tombleson P. Modified essay question (MEQ) paper: perestroika. *Occas Pap R Coll Gen Pract*. 1990;(46):18–22.
41. Feletti GI, Smith EK. Modified essay questions: are they worth the effort? *Med Educ*. 1986;20(2):126–132. DOI: 10.1111/j.1365-2923.1986.tb01059.x
42. van Bruggen L, Manrique-van Woudenberg M, Spierenburg E, Vos J. Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect Med Educ*. 2012;1(4):162–171. DOI: 10.1007/s40037-012-0024-1
43. Irwin WG, Bamber JH. The cognitive structure of the modified essay question. *Med Educ*. 1982;16(6):326–331. DOI: 10.1111/j.1365-2923.1982.tb00945.x
44. Weinman J. A modified essay question evaluation of pre-clinical teaching of communication skills. *Med Educ*. 1984;18(3):164–167. DOI: 10.1111/j.1365-2923.1984.tb00998.x
45. Khan MU, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci*. 2011;5(1):39–43.
46. Bodkha P. Effectiveness of MCQ, SAQ and MEQ in assessing cognitive domain among high and low achievers. *IJRRMS*. 2012;2(4):25–28.
47. Wallerstedt S, Erickson G, Wallerstedt SM. Short Answer Questions or Modified Essay questions–More Than a Technical Issue. *Int J Clin Med*. 2012;3:28. DOI: 10.4236/ijcm.2012.31005
48. Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. *Acad Med*. 1993;68(4):244–249. DOI: 10.1097/00001888-199304000-00002
49. Ferguson KJ. Beyond multiple-choice questions: Using case-based learning patient questions to assess clinical reasoning. *Med Educ*. 2006;40(11):1143. DOI: 10.1111/j.1365-2929.2006.02592.x
50. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ*. 2007;7(1):49. DOI: 10.1186/1472-6920-7-49
51. Wild D, Rützler M, Haarhaus M, Peters K. Der Modified Essay Question (MEQ)-Test an der medizinischen Fakultät der Universität Witten/Herdecke. *Gesundheitswesen (Suppl Med Ausbild)*. 1998;15(Suppl2):65–69. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/2/Wild_D,_R%C3%BCtzler_M,_Haarhaus_M,_Peters_K.pdf
52. Peters K, Scheible CM, Rützler M. MEQ – angemessen und praktikabel? Jahrestagung der Gesellschaft für Medizinische Ausbildung - GMA. Köln, 10.-12.11.2006. Düsseldorf, Köln: German Medical Science; 2006. Doc06gma085. Zugänglich unter/available from <http://www.egms.de/en/meetings/gma2006/06gma085.shtml>
53. O'Neill PN. Assessment of students in a problem-based learning curriculum. *J Dent Educ*. 1998;62(9):640–643.

54. Geerlings G, van de Poel AC. De gestructureerde open Vraag: Een Mogelijkheid tot Patientensimulatie binnen Hetonderwijs in de Endodontologie [The modified essay question: a possibility for patient simulation in endodontic education]. *Ned Tijdschr Tandheelkd.* 1984;91(7-8):305–308.
55. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309–317. DOI: 10.1111/j.1365-2929.2005.02094.x
56. Schwartz RW, Donnelly MB, Sloan DA, Young B. Knowledge gain in a problem-based surgery clerkship. *Acad Med.* 1994;69(2):148–151. DOI: 10.1097/00001888-199402000-00022
57. Rabinowitz HK. The modified essay question: an evaluation of its use in a family medicine clerkship. *Med Educ.* 1987;21(2):114–118. DOI: 10.1111/j.1365-2923.1987.tb00676.x
58. Stratford P, Pierce-Fenn H. Modified essay question. *Phys Ther.* 1985;65(7):1075–1079.
59. Norman GR, Smith EK, Powles AC, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Med Educ.* 1987;21(4):297–304. DOI: 10.1111/j.1365-2923.1987.tb00367.x
60. Bloch R, Hofer D, Krebs R, Schläppi P, Weis S, Westkämper R. Kompetent prüfen. Handbuch zur Planung, Durchführung und Auswertung von Facharztprüfungen Medizinische Fakultät Universität Bern, Wien: Institut für Aus-, Weiter- und Fortbildung; 1999.
61. Lim EC, Seet RC, Oh VM, Chia BL, Aw M, Quak SH, Onk BK. Computer-based testing of the modified essay question: the Singapore experience. *Med Teach.* 2007;29(9):e261–268. DOI: 10.1080/01421590701691403
62. Palmer EJ, Devitt PG. A method for creating interactive content for the iPod, and its potential use as a learning tool: Technical Advances. *BMC Med Educ.* 2007;7(1):32. DOI: 10.1186/1472-6920-7-32
63. Bordage G, Brailovsky C, Carretier H, Page G. Content validation of key features on a national examination of clinical decision-making skills. *Acad Med.* 1995;70(4):276–281. DOI: 10.1097/00001888-199504000-00010
64. Bordage G, Page G. An alternative approach to PMPs: The "key features" concept. In: Hart IR, Harden RM (Hrsg). Further developments in assessing clinical competence. Montreal: Can-Heal; 1987. S.59–75.
65. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med.* 1995;70(3):194–201. DOI: 10.1097/00001888-199503000-00009
66. Hatala R, Norman GR. Adapting the Key Features Examination for a clinical clerkship. *Med Educ.* 2002;36(2):160–165. DOI: 10.1046/j.1365-2923.2002.01067.x
67. Trudel JL, Bordage G, Downing SM. Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. *Ann Surg.* 2008;248(2):252–258. DOI: 10.1097/SLA.0b013e31818233d3
68. Ali SK, Bordage G. Validity of key features for a family medicine pilot exam at the College of Physicians and Surgeons Pakistan. *J Coll Phys Surg Pakistan.* 1995;5(6):256–260.
69. Bernhardt J, Griesbacher T, Ithaler D, Kresse A, Öttl K, Roller-Wirnsberger R, Vogl S. Kürzübersicht gängiger Prüfungsformate. Graz: Medizinische Universität Graz; 2012. Zugänglich unter/available from: http://www.medunigraz.at/fileadmin/lehren/planen-organisieren/pdf/Kurzuebersicht_Pruefungsformate_v7.pdf
70. Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J. A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach.* 2005;27(5):450–455. DOI: 10.1080/01421590500078471
71. Nikendei C, Mennin S, Weyrich P, Kraus B. Effects of a supplementary final year curriculum on students' clinical reasoning skills as assessed by key-feature examination. *Med Teach.* 2009;31(9):e438–443. DOI: 10.1080/01421590902845873
72. Rotthoff T, Baehring T, Dicken H-D, Fahrion U, Richter B, Fischer MR, Scherbaum WA. Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial. *BMC Med Educ.* 2006;6(1):50. DOI: 10.1186/1472-6920-6-50
73. Zupanic M, Iblher P, Töpfer J, Gartmeier M, Bauer J, Prenzel M, Möller G, Hoppe-Seyler T, Karsten G, Fischer MR. Key Feature-Assessment kommunikativer Leistungen: Weiterent-wicklung und quantitative Evaluation Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). München, 05.-08.10.2011. German Medical Science GMS Publishing House; 2011. Doc11gma024. DOI: 10.3205/11gma024
74. Schaper E, Tipold A, Ehlers JP. Use of key feature questions in summative assessment of veterinary medicine students. *Ir Vet J.* 2013;66(1):3. DOI: 10.1186/2046-0481-66-3
75. TU Dresden. Studienordnung für den Studiengang Zahnmedizin vom 08.09.2011 [Internet]. Dresden; 2011 [cited 2015 Feb 24]. Available from: https://tu-dresden.de/die_tu_dresden/fakultaeten/medizinische_fakultaet/studium/zm/recht-zm/index_html#StudOrgZM
76. Gerhardt-Szep S, Hahn P. Key feature - Fallerstellung (Master of Medical Education, Modul V). Heidelberg: Universität Heidelberg; 2008. Zugänglich unter/available from: http://archiv.ub.uni-heidelberg.de/volltextserver/11537/1/Manuscript_KF.pdf
77. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ.* 2006;40(7):618–623. DOI: 10.1111/j.1365-2929.2006.02511.x
78. Huwendiek S, Mennin SP, Nikendei C. Medical education after the Flexner report. *N Engl J Med.* 2007;356(1):90. DOI: 10.1056/NEJMc062922
79. Schuwirth LW, van der Vleuten CP, de Kock CA, Peperkamp AG, Donkers HH. Computerized case-based testing: A modern method to assess clinical decision making. *Med Teach.* 1996;18(4):294–299. DOI: 10.3109/01421599609034180
80. Huwendiek S, Reichert F, Brass K, Bosse H-M, Heid J, Möltner A, Haag M, Leven FJ, Hoffmann GF, Jünger J, Tönshoff B. Etablierung von fallbasiertem computerunterstütztem Prüfen mit langen Auswahl Listen: Ein geeignetes Instrument zur Prüfung von Anwendungswissen. *GMS Z Med Ausbild.* 2007;24(1):Doc51. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000345.shtml>
81. Huwendiek S, Heid J, Möltner A, Haag M, Tönshoff B. E-Learning und E-Prüfung mit virtuellen Patienten in der Medizin. Heidelberg: Universitätsklinikum Heidelberg; 2008. Zugänglich unter/available from: http://www.his-he.de/veranstaltung/dokumentation/Workshop_E-Pruefung/pdf/TOPO4.pdf
82. Ananthakrishnan N. Microteaching as a vehicle of teacher training—its advantages and disadvantages. *J Postgrad Med.* 1993;39(3):142–143.
83. Arnold RC, Walmsley AD. The use of the OSCE in postgraduate education. *Eur J Dent Educ.* 2008;12(3):126–130. DOI: 10.1111/j.1600-0579.2008.00469.x
84. Taguchi N, Ogawa T. OSCEs in Japanese postgraduate clinical training Hiroshima experience 2000-2009. *Eur J Dent Educ.* 2010;14(4):203–209. DOI: 10.1111/j.1600-0579.2009.00610.x

85. Pugh D, Touchie C, Wood TJ, Humphrey-Murto S. Progress testing: is there a role for the OSCE? *Med Educ*. 2014;48(6):623–631. DOI: 10.1111/medu.12423
86. Curtis DA, Lind SL, Brear S, Finzen FC. The correlation of student performance in preclinical and clinical prosthodontic assessments. *J Dent Educ*. 2007;71(3):365–372.
87. Eberhard L, Hassel A, Bäumer A, Becker F, Beck-Mußotter J, Bömicke W, Corcodel N, Cosgarea R, Eiffler C, Giannakopoulos NN; Kraus T, Mahabadi J, Rues S, Schmitter M, Wolff D, Wege KC. Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. *Eur J Dent Educ*. 2011;15(3):172–178. DOI: 10.1111/j.1600-0579.2010.00653.x
88. Graham R, Bitzer LA, Anderson OR. Reliability and Predictive Validity of a Comprehensive Preclinical OSCE in Dental Education. *J Dent Educ*. 2013;77(2):161–167.
89. Fields H, Rowland M, Vig K, Huja S. Objective structured clinical examination use in advanced orthodontic dental education. *Am J Orthod Dentofacial Orthop*. 2007;131(5):656–663. DOI: 10.1016/j.ajodo.2007.01.013
90. Derringer KA. Undergraduate orthodontic assessment and examination in UK dental schools. *Br Dent J*. 2006;201(4):225–229. DOI: 10.1038/sj.bdj.4813884
91. Macluskey M, Durham J, Balmer C, Bell A, Cowpe J, Dawson L, et al. Dental student suturing skills: a multicentre trial of a checklist-based assessment. *Eur J Dent Educ*. 2011;15(4):244–249. DOI: 10.1111/j.1600-0579.2010.00665.x
92. Hoefler SH, Schuebel F, Sader R, Landes C. Development and implementation of an objective structured clinical examination (OSCE) in CMF-surgery for dental students. *J Craniomaxillofac Surg*. 2013;41(5):412–416. DOI: 10.1016/j.jcms.2012.11.007
93. Landes CA, Hoefler S, Schuebel F, Ballon A, Teiler A, Tran A, Weber R, Walcher F, Sader R. Long-term prospective teaching effectivity of practical skills training and a first OSCE in Cranio Maxillofacial Surgery for dental students. *J Craniomaxillofac Surg*. 2014;42(5):e97–e104. DOI: 10.1016/j.jcms.2013.07.004
94. Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self- and peer-assessment. *Eur J Dent Educ*. 2008;12(1):2–7.
95. Kupke J, Wicht MJ, Stützer H, Derman SH, Lichtenstein NV, Noack MJ. Does the use of a visualised decision board by undergraduate students during shared decision-making enhance patients' knowledge and satisfaction? – A randomised controlled trial. *Eur J Dent Educ*. 2012;17(1):19–25. DOI: 10.1111/eje.12002
96. Hammad M, Oweis Y, Taha S, Hattar S, Madarati A, Kadim F. Students' Opinions and Attitudes After Performing a Dental OSCE for the First Time: A Jordanian Experience. *J Dent Educ*. 2013;77(1):99–104.
97. Mossey PA, Newton JP, Stirrups DR. Scope of the OSCE in the assessment of clinical skills in dentistry. *Br Dent J*. 2001;190(6):323–326. DOI: 10.1038/sj.bdj.4800961
98. Boone WJ, McWhorter AG, Seale NS. Purposeful assessment techniques (PAT) applied to an OSCE-based measurement of competencies in a pediatric dentistry curriculum. *J Dent Educ*. 2001;65(11):1232–1237.
99. Lele SM. A Mini-OSCE for Formative Assessment of Diagnostic and Radiographic Skills at a Dental College in India. *J Dent Educ*. 2011;75(12):1583–1589.
100. Schoonheim-Klein M, Walmsley AD, Habets L. An implementation strategy for introducing an OSCE into a dental school. *Eur J Dent Educ*. 2005;9(4):143–149. DOI: 10.1111/j.1600-0579.2005.00379.x
101. Licari FW, Knight GW. Developing a group practice comprehensive care education curriculum. *J Dent Educ*. 2003;67(12):1312–1315.
102. Ratzmann A, Wiesmann U, Kordaß B. Integration of an Objective Structured Clinical Examination (OSCE) into the dental preliminary exams. *GMS Z Med Ausbild*. 2012;29(1):Doc09. DOI: 10.3205/zma000779
103. Ogawa T, Taguchi N, Sasahara H. Assessing communication skills for medical interviews in a postgraduate clinical training course at Hiroshima University Dental Hospital. *Eur J Dent Educ*. 2003;7(2):60–65. DOI: 10.1034/j.1600-0579.2002.00273.x
104. Cannick GF, Horowitz AM, Garr DR, Reed SG, Neville BW, Day TA, Woolson RF, Lackland DT. Use of the OSCE to evaluate brief communication skills training for dental students. *J Dent Educ*. 2007;71(9):1203–1209.
105. Dennehy PC, Susarla SM, Karimbux NY. Relationship between dental students' performance on standardized multiple-choice examinations and OSCEs. *J Dent Educ*. 2008;72(5):585–592.
106. 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. DOI: 10.3109/0142159X.2013.818634
107. Deis N, Narciß E, Rahe J, Schüttelpelz-Braun K. Objektive standardisierte praktische Prüfungen zur Messung von praktischen Fertigkeiten und berufsrelevanten Kompetenzen. *Z Gesundheit Sport*. 2012;2(2):25–33.
108. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess*. 2011;15(1):i–xxi–1–162.
109. Brand HS, Schoonheim-Klein M. Is the OSCE more stressful? Examination anxiety and its consequences in different assessment methods in dental education. *Eur J Dent Educ*. 2009;13(3):147–153. DOI: 10.1111/j.1600-0579.2008.00554.x
110. Nikendei C, Jünger J. OSCE-praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild*. 2006;23(3):Doc47. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000266.shtml>
111. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ*. 2011;45(12):1181–1189. DOI: 10.1111/j.1365-2923.2011.04075.x
112. Schoonheim-Klein M, Muijtens A, Muijtens A, Habets L, Manogue M, van der Vleuten C, Hoogstraten J, Van der Velden U. On the reliability of a dental OSCE, using SEM: effect of different days. *Eur J Dent Educ*. 2008;12(3):131–137. DOI: 10.1111/j.1600-0579.2008.00507.x
113. Norcini JJ, Maihoff NA, Day SC, Benson JA. Trends in medical knowledge as assessed by the certifying examination in internal medicine. *JAMA*. 1989;262(17):2402–2404. DOI: 10.1001/jama.1989.03430170064029
114. Hofer M, Jansen M, Soboll S. [Potential improvements in medical education as retrospectively evaluated by candidates for specialist examinations]. *Dtsch med Wochenschr*. 2006;131(8):373–378. DOI: 10.1055/s-2006-932527
115. Gesellschaft für Medizinische Ausbildung, Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für Fakultäts-interne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. *GMS Z Med Ausbild*. 2008;25(1):Doc74. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml>

116. Davenport ES, Davis JE, Cushing AM, Holsgrove GJ. An innovation in the assessment of future dentists. *Br Dent J*. 1998;184(4):192–195.
117. Rau T, Fegert J, Liebhardt H. How high are the personnel costs for OSCE? A financial report on management aspects. *GMS Z Med Ausbild*. 2011;28(1):Doc13. DOI: 10.3205/zma000725
118. Kropmans TJ, O'Donovan BG, Cunningham D, Murphy AW, Flaherty G, Nestel D, Dunne FPI. An Online Management Information System for Objective Structured Clinical Examinations. *CIS*. 2011;28;5(1):38-48.
119. Barman A. Critiques on the Objective Structured Clinical Examination. *Ann Acad Med Singap*. 2005;34(8):478–482.
120. Stillman PL, Swanson DB, Smee S, Stillman AE, Ebert TH, Emmel VS, Gaslowitz J, Green HL, Hamolsky M, Hatem C, et al. Assessing clinical skills of residents with standardized patients. *Ann Intern Med*. 1986;105(5):762–771. DOI: 10.7326/0003-4819-105-5-762
121. Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. *Fam Med*. 2008;40(8):574–578.
122. Carpenter JL. Cost analysis of objective structured clinical examinations. *Acad Med*. 1995;70(9):828–833. DOI: 10.1097/00001888-199509000-00022
123. Duerson MC, Romrell LJ, Stevens CB. Impacting faculty teaching and student performance: nine years' experience with the Objective Structured Clinical Examination. *Teach Learn Med*. 2000;12(4):176–182. DOI: 10.1207/S15328015TLM1204_3
124. Harden RM, Cairncross RG. Self assessment. *Med Teach*. 1980;2(3):145–148. DOI: 10.3109/01421598009081185
125. Kundu D, Das HN, Sen G, Osta M, Mandal T, Gautam D. Objective structured practical examination in biochemistry: An experience in Medical College, Kolkata. *J Nat Sci Biol Med*. 2013;4(1):103–107. DOI: 10.4103/0976-9668.107268
126. Abraham RR, Raghavendra R, Surekha K, Asha K. A trial of the objective structured practical examination in physiology at Melaka Manipal Medical College, India. *Adv Physiol Educ*. 2009;33(1):21–23. DOI: 10.1152/advan.90108.2008
127. Adome RO, Kitutu F. Creating an OSCE/OSPE in a resource-limited setting. *Med Educ*. 2008;42(5):525–526. DOI: 10.1111/j.1365-2923.2008.03045.x
128. Nayak V, Bairy KL, Adiga S, Shenoy S, Magazine BC, Amberkar M, Kumari MK. OSPE in Pharmacology: Comparison with the conventional Method and Students' Perspective Towards OSPE. *Br Biomed Bull*. 2014;2(1):218-222.
129. Wani P, Dalvi V. Objective Structured Practical Examination vs Traditional Clinical Examination in Human Physiology: Student's perception. *Int J Med Sci Public Health*. 2013;2(3):522–547. DOI: 10.5455/ijmsph.2013.080320133
130. Menezes RG, Nayak VC, Binu VS, Kanchan T, Rao PP, Baral P, Lobo SW. Objective structured practical examination (OSPE) in Forensic Medicine: students' point of view. *J Forensic Leg Med*. 2011;18(8):347–349. DOI: 10.1016/j.jflm.2011.06.011
131. Huth KC, Baumann M, Kollmuss M, Hickel R, Fischer MR, Paschos E. Assessment of practical tasks in the Phantom course of Conservative Dentistry by pre-defined criteria: a comparison between self-assessment by students and assessment by instructors. *Eur J Dent Educ*. 2015. DOI: 10.1111/eje.12176 [Epub ahead of print] DOI: 10.1111/eje.12176
132. Banerjee R, Chandak A, Radke U. Bringing objectivity to assessment in Preclinical Prosthodontics: The student's perspective on OSPE. *JETHS*. 2014;1(2):30-33.
133. Schmitt I, Möltner A, Bärmeier J, Gärtner K, Dopfer S, Kuschel B, Kunkel F, Heidemann D, Gerhardt-Szép S. Wie viele Prüfer braucht ein OSCE? Düsseldorf: AKWLZ; 2013. Zugänglich unter/available from: <https://www.akwlz.de/documents/262773/490615/K8+-+Schmitt+-+Pr%E2%80%99er+OSPE.pdf/dc030606-f303-4e81-b85c-601177f11e0c?version=1.0>
134. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Med Teach*. 2003;25(3):262–270. DOI: 10.1080/0142159031000100300
135. Colliver JA, Barrows HS, Vu NV, Verhulst SJ, Mast TA, Travis TA. Test security in examinations that use standardized-patient cases at one medical school. *Acad Med*. 1991;66(5):279–282. DOI: 10.1097/00001888-199105000-00011
136. Ortwein H, Fröhmel A, Burger W. Einsatz von Simulationspatienten als Lehr-, Lern- und Prüfungsform. *Psychother Psychosom Med Psychol*. 2006;56(01):23–29. DOI: 10.1055/s-2005-867058
137. Collins J, Harden RM. AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Med Teach*. 1998;20(6):508–521. DOI: 10.1080/01421599880210
138. Barrows H. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med*. 1993;68(6):443–451. DOI: 10.1097/00001888-199306000-00002
139. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach*. 2009;31(6):477–486. DOI: 10.1080/01421590903002821
140. Hendrickx K, De Winter B, Tjalma W, Avonts D, Peeraer G, Wyndaele JJ. Learning intimate examinations with simulated patients: The evaluation of medical students' performance. *Med Teach*. 2009;31(4):E139–147. DOI: 10.1080/01421590802516715
141. Norcini JJ. The death of the long case? *BMJ*. 2002;324(7334):408–409. DOI: 10.1136/bmj.324.7334.408
142. Norcini J. The validity of long cases. *Med Educ*. 2001;35(8):720–721. DOI: 10.1046/j.1365-2923.2001.01006.x
143. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. 1995;123(10):795–799. DOI: 10.7326/0003-4819-123-10-199511150-00008
144. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. 2003;138(6):476–481. DOI: 10.7326/0003-4819-138-6-200303180-00012
145. Thornton S. A literature review of the long case and its variants as a method of assessment. *Educ Med J*. 2012;4(1):5–11.
146. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. 2002;77(9):900–904. DOI: 10.1097/00001888-200209000-00020
147. Herbers JE, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. 1989;4(3):202–208. DOI: 10.1007/BF02599524
148. Yousuf N. Mini clinical evaluation exercise: validity and feasibility evidences in literature. *Educ Med J*. 2012;4(1):e100-107.
149. Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Med Educ*. 2009;43(4):326–334. DOI: 10.1111/j.1365-2923.2008.03275.x

150. Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med.* 2003;78(10 Suppl):S33–35. DOI: 10.1097/00001888-200310001-00011
151. Alves De Lima A, Barrero C, Baratta S, Castillo Costa Y, Bortman G, Carabajales J, Conde D, Galli A, Degrange G, Van der Vleuten C. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Med Teach.* 2007;29(8):785–790. DOI: 10.1080/01421590701352261
152. Norcini JJ. The mini clinical evaluation exercise (mini-CEX). *Clin Teach.* 2005;2(1):25–30. DOI: 10.1111/j.1743-498X.2005.00060.x
153. Alves de Lima A. Assessment of clinical competence: Reliability, Validity, Feasibility and Educational Impact of the mini-CEX. Maastricht: Universit at Maastricht; 2013. Zug anglich unter/available from: http://www.icba.com.ar/profesionales/pdf/aal/Thesis_Alberto_Alves_de_Lima_170x240_v10.pdf
154. Cook DA, Beckman TJ. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. *Adv Health Sci Educ.* 2009;14(5):655–664. DOI: 10.1007/s10459-008-9147-x
155. Alves de Lima AE, Conde D, Aldunate L, van der Vleuten CP. Teachers' experiences of the role and function of the mini clinical evaluation exercise in post-graduate training. *Int J Med Educ.* 2010;1:68–73. DOI: 10.5116/ijme.4c87.8e13
156. Berendonk C, Beyeler C, Westk amper R, Giger M. Strukturiertes Feedback in der  rztlichen Weiterbildung: Mini-CEX und DOPS. *Schweiz  rztz.* 2008;89(32):1337–1340.
157. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005;80(10 Suppl):S46–54. DOI: 10.1097/00001888-200510001-00015
158. Brazil V, Ratcliffe L, Zhang J, Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department—does cost outweigh value? *Med Teach.* 2012;34(12):1017–1023. DOI: 10.3109/0142159X.2012.719653
159. Magnier KM, Dale VH, Pead MJ. Workplace-based assessment instruments in the health sciences. *J Vet Med Educ.* 2012;39(4):389–395. DOI: 10.3138/jvme.1211-118R
160. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ.* 2008;42(4):364–373. DOI: 10.1111/j.1365-2923.2008.03010.x
161. Prescott-Clements L, van der Vleuten CP, Schuwirth LW, Hurst Y, Rennie JS. Evidence for validity within workplace assessment: the Longitudinal Evaluation of Performance (LEP). *Med Educ.* 2008;42(5):488–495. DOI: 10.1111/j.1365-2923.2007.02965.x
162. Deshpande S, Chahande J. Impact of computer-based treatment planning software on clinical judgment of dental students for planning prosthodontic rehabilitation. *Adv Med Educ Pract.* 2014;5:269–274. DOI: 10.2147/AMEP.S66264
163. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach.* 2006;28(2):117–128. DOI: 10.1080/01421590600622665
164. Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med.* 2004;19(5 Pt 2):558–561. DOI: 10.1111/j.1525-1497.2004.30134.x
165. Montagne S, Rogausch A, Gemperli A, Berendonk C, Jucker-Kupper P, Beyeler C. The mini-clinical evaluation exercise during medical clerkships: are learning needs and learning goals aligned? *Med Educ.* 2014;48(10):1008–1019. DOI: 10.1111/medu.12513
166. Berberat PO, Harendza S, Kadmon M, Gesellschaft f ur Medizinische Ausbildung, GMA-Ausschuss f ur Weiterbildung. Entrustable professional activities - visualization of competencies in postgraduate training. Position paper of the Committee on Postgraduate Medical Training of the German Society for Medical Education (GMA). *GMS Z Med Ausbild.* 2013;30(4):Doc47. DOI: 10.3205/zma000890
167. Colleges AOAM. Core Entrustable Professional Activities for Entering Residency. [Internet] [cited 2016 June 14] 2014. Zug anglich unter/available from: <https://members.aamc.org/eweb/upload/Core%20EPA%20Curriculum%20Dev%20Guide.pdf>
168. Ten Cate O. Trusting graduates to enter residency: what does it take? *J Grad Med Educ.* 2014;6(1):7–10. DOI: 10.4300/JGME-D-13-00436.1
169. Ten Cate O, Snell L, Carraccio C. Medical competence: the interplay between individual ability and the health care environment. *Med Teach.* 2010;32(8):669–675. DOI: 10.3109/0142159X.2010.500897
170. Jones MD, Rosenberg AA, Gilhooly JT, Carraccio CL. Perspective: Competencies, Outcomes, and Controversy—Linking Professional Activities to Competencies to Improve Resident Education and Practice. *Acad Med.* 2011;86(2):161–165. DOI: 10.1097/ACM.0b013e31820442e9
171. Mulder H, Ten Cate O, Daalder R, Berkvens J. Building a competency-based workplace curriculum around entrustable professional activities: The case of physician assistant training. *Med Teach.* 2010;32(10):e453–459. DOI: 10.3109/0142159X.2010.513719
172. Chang A, Bowen JL, Buranosky RA, Frankel RM, Ghosh N, Rosenblum MJ, Thompson S, Green ML. Transforming primary care training—patient-centered medical home entrustable professional activities for internal medicine residents. *J Gen Intern Med.* 2013;28(6):801–809. DOI: 10.1007/s11606-012-2193-3
173. Scheele F, Teunissen P, van Luijk S, Heineman E, Fluit L, Mulder H, Meiningen A, Wijnen-Meijer M, Glas G, Sluiter H, Hummel T. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach.* 2008;30(3):248–253. DOI: 10.1080/01421590801993022
174. Englander R, Carraccio C. From theory to practice: making entrustable professional activities come to life in the context of milestones. *Acad Med.* 2014;89(10):1321–1323. DOI: 10.1097/ACM.0000000000000324
175. Aylward M, Nixon J, Gladding S. An entrustable professional activity (EPA) for handoffs as a model for EPA assessment development. *Acad Med.* 2014;89(10):1335–1340. DOI: 10.1097/ACM.0000000000000317
176. Boyce P, Spratt C, Davies M, McEvoy P. Using entrustable professional activities to guide curriculum development in Psychiatry training. *BMC Med Educ.* 2011;11(1):96. DOI: 10.1186/1472-6920-11-96
177. Ten Cate O. Competency-based education, entrustable professional activities, and the power of language. *J Grad Med Educ.* 2013;5(1):6–7. DOI: 10.4300/JGME-D-12-00381.1
178. Hauer KE, Kohlwes J, Cornett P, Hollander H, Cate ten O, Ranji SR, Soni K, Iobst W, O'Sullivan P. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ.* 2013;5(1):54–59. DOI: 10.4300/JGME-D-12-00060.1

179. Shaughnessy AF, Sparks J, Cohen-Osher M, Goodell KH, Sawin GL, Joseph Gravel J. Entrustable Professional Activities in Family Medicine. *J Grad Med Educ.* 2013;5(1):112–118. DOI: 10.4300/JGME-D-12-00034.1
180. Ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542–547. DOI: 10.1097/ACM.0b013e31805559c7
181. Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). *J Coll Physicians Surg Pak.* 2013;23(1):77–82.
182. Bazrafkan L. Comparison of the Assessment of Dental Students'. *J Med Ed.* 2009;13(1, 2):3-8.
183. Tricio J, Woolford M, Thomas M, Lewis-Greene H, Georghiou L, Andiappan M, Escudier M. Dental students' peer assessment: a prospective pilot study. *Eur J Dent Educ.* 2014;19(3):140-148. DOI: 10.1111/eje.12114
184. Abraham RR, Upadhy S, Torke S, Ramnarayan K. Student perspectives of assessment by TEMM model in physiology. *Adv Physiol Educ.* 2005;29(2):94–97. DOI: 10.1152/advan.00051.2004
185. Barton JR, Corbett S, van der Vleuten CP, English Bowel Cancer Screening Programme, UK Joint Advisory Group for Gastrointestinal Endoscopy. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointest Endosc.* 2012;75(3):591–597. DOI: 10.1016/j.gie.2011.09.053
186. Akbari M, Shamsabadi RM. Direct Observation of Procedural Skills (DOPS) in Restorative Dentistry: Advantages and Disadvantages in Student's Point of View. *Iran J Med Educ.* 2013;13(3):212-220.
187. Andersen RM, Davidson PL, Atchison KA, Hewlett E, Freed JR, Friedman J-A, et al. Pipeline, profession, and practice program: evaluating change in dental education. *J Dent Educ.* 2005;69(2):239–248.
188. Hamdy H, Prasad K, Williams R, Salih FA. Reliability and validity of the direct observation clinical encounter examination (DOCEE). *Med Educ.* 2003;37(3):205–212. DOI: 10.1046/j.1365-2923.2003.01438.x
189. Torre DM, Simpson DE, Elnicki DM, Sebastian JL, Holmboe ES. Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. *Teach Learn Med.* 2007;19(3):271–277. DOI: 10.1080/10401330701366622
190. Cohen SN, Farrant PBJ, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol.* 2009;161(1):34–39. DOI: 10.1111/j.1365-2133.2009.09097.x
191. Center of Innovation in Professional Health Education and Research (CIPHER). Review of work-based assessment methods. Sydney: University of Sydney; 2007.
192. Roghieh N, Fateme H, Hamid S, Hamid H. The effect of formative evaluation using "direct observation of procedural skills" (DOPS) method on the extent of learning practical skills among nursing students in the ICU. *Iran J Nurs Midwifery Res.* 2013;18(4):290–293.
193. Hamilton KES, Coates V, Kelly B, Boore JRP, Cundell JH, Gracey J, et al. Performance assessment in health care providers: a critical review of evidence and current practice. *J Nurs Manag.* 2007;15(8):773–791. DOI: 10.1111/j.1365-2934.2007.00780.x
194. Morris A, Hewitt J, Roberts CM. Practical experience of using directly observed procedures, mini clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgrad Med J.* 2006;82(966):285–288. DOI: 10.1136/pgmj.2005.040477
195. Sh S, Pooladi A, BahramRezaie M, Farhadifar F, Khatibi R. Evaluation of the Effects of Direct Observation of Procedural Skills (DOPS) on clinical externship students' learning level in obstetrics ward of kurdistan university of medical sciences. *J Med Ed.* 2009;13(1):29-33.
196. Stosch C, Wichelhaus AS, Matthes J. Die Portfolio-Methode: Modernes Assessment auf dem Prüfstand. *GMS Z Med Ausbild.* 2006;23(3):Doc43. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000262.shtml>
197. Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, Moreley D, Pollard D, Ashcroft T, Popovic C, Sayers J. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. *BEME Guide No. 11. Med Teach.* 2009;31(4):282–298. DOI: 10.1080/01421590902889897
198. Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: *BEME Guide No 12. Med Teach.* 2009;31(4):299–318. DOI: 10.1080/01421590902883056
199. Pocock I. A new route for dental graduates. *Dent Update.* 2007;34(1):59.
200. Kramer GA, Albino JEN, Andrieu SC, Hendricson WD, Henson L, Horn BD, Neumann LM, Young SK. Dental student assessment toolbox. *J Dent Educ.* 2009;73(1):12–35.
201. Gadbury-Amyot CC, McCracken MS, Woldt JL, Brennan RL. Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *J Dent Educ.* 2014;78(5):657–667.
202. Michels NR, Driessen EW, Muijtjens AM, Van Gaal LF, Bossaert LL, de Winter BY. Portfolio assessment during medical internships: How to obtain a reliable and feasible assessment procedure? *Educ Health.* 2009;22(3):313.
203. O'sullivan PS, Reckase MD, McClain T, Savidge MA, Clardy JA. Demonstration of portfolios to assess competency of residents. *Adv Health Sci Educ.* 2004;9(4):309–323. DOI: 10.1007/s10459-004-0885-0
204. Melville C, Rees M, Brookfield D, Anderson J. Portfolios for assessment of paediatric specialist registrars. *Med Educ.* 2004;38(10):1117–1125. DOI: 10.1111/j.1365-2929.2004.01961.x
205. McMullan M. Students' perceptions on the use of portfolios in pre-registration nursing education: a questionnaire survey. *Int J Nurs Stud.* 2006;43(3):333–343. DOI: 10.1016/j.ijnurstu.2005.05.005
206. Burch VC, Seggie JL. Use of a structured interview to assess portfolio-based learning. *Med Educ.* 2008;42(9):894–900. DOI: 10.1111/j.1365-2923.2008.03128.x
207. Kadagad P, Kotrashetti SM. Portfolio: a comprehensive method of assessment for postgraduates in oral and maxillofacial surgery. *J Maxillofac Oral Surg.* 2013;12(1):80–84. DOI: 10.1007/s12663-012-0381-7
208. Brett JF, Atwater LE. 360 degree feedback: accuracy, reactions, and perceptions of usefulness. *J Appl Psychol.* 2001;86(5):930–942. DOI: 10.1037/0021-9010.86.5.930
209. Lepsinger R, Lucia AD. *The Art and Science of 360 Degree Feedback.* New York: John Wiley & Sons; 2009.

210. Donnon T, Ansari AI A, Alawi AI S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med.* 2014;89(3):511–516. DOI: 10.1097/ACM.000000000000147
211. Zhao Y, Zhang X, Chang Q, Sun B. Psychometric characteristics of the 360° feedback scales in professionalism and interpersonal and communication skills assessment of surgery residents in China. *J Surg Educ.* 2013;70(5):628–635. DOI: 10.1016/j.jsurg.2013.04.004
212. Joshi R, Ling FW, Jaeger J. Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Acad Med.* 2004;79(5):458–463. DOI: 10.1097/00001888-200405000-00017
213. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ.* 2005;330(7502):1251–1253. DOI: 10.1136/bmj.38447.610451.8F
214. Murphy DJ, Bruce DA, Mercer SW, Eva KW. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Adv Health Sci Educ.* 2009;14(2):219–232. DOI: 10.1007/s10459-008-9104-8
215. Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med.* 1993;68(9):680–687. DOI: 10.1097/00001888-199309000-00014
216. Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. *Can J Psychiatry.* 2008;53(8):525–533.
217. Chandler N, Henderson G, Park B, Byerley J, Brown WD, Steiner MJ. Use of a 360-degree evaluation in the outpatient setting: the usefulness of nurse, faculty, patient/family, and resident self-evaluation. *J Grad Med Educ.* 2010;2(3):430–434. DOI: 10.4300/JGME-D-10-00013.1
218. Hesketh EA, Anderson F, Bagnall GM, Driver CP, Johnston DA, Marshall D, Needham G, Orr G, Walker K. Using a 360 degrees diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Med Teach.* 2005;27(3):219–233. DOI: 10.1080/01421590500098776
219. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Med Educ.* 2014;14(1):76. DOI: 10.1186/1472-6920-14-76
220. Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, Dewitt TG, Britto MT. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med.* 2007;161(1):44–49. DOI: 10.1001/archpedi.161.1.44
221. Weigelt JA, Brasel KJ, Bragg D, Simpson D. The 360-degree evaluation: increased work with little return? *Current Surgery.* 2004;61(6):616–626. DOI: 10.1016/j.cursur.2004.06.024
222. Garry A, Stirling K. Achieving 360° student feedback using SPaCE. *Clin Teach.* 2012;9(4):222–227. DOI: 10.1111/j.1743-498X.2012.00550.x
223. Fabry G. *Medizindidaktik: ein Handbuch für die Praxis.* Karlsruhe: Huber; 2008.

Corresponding author:

PD Dr. med. dent Susanne Gerhard-Szep, MME
Goethe-Universität, Carolinum Zahnärztliches
Universitäts-Institut gGmbH, Poliklinik
Zahnerhaltungskunde, Frankfurt am Main, Deutschland
s.szep@em.uni-frankfurt.de

Please cite as

Gerhard-Szep S, Güntsch A, Pospiech P, Söhnel A, Scheutzel P, Wassmann T, Zahn T. Assessment formats in dental medicine: An overview. *GMS J Med Educ.* 2016;33(4):Doc65. DOI: 10.3205/zma001064, URN: urn:nbn:de:0183-zma0010641

This article is freely available from

<http://www.egms.de/en/journals/zma/2016-33/zma001064.shtml>

Received: 2015-10-23

Revised: 2016-03-24

Accepted: 2016-05-09

Published: 2016-08-15

Copyright

©2016 Gerhard-Szep 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/>.

Prüfungsformate für die Zahnmedizin: eine Übersicht

Zusammenfassung

Ziele: Auf Initiative des Arbeitskreises für die Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ) wurde 2013 auf dem Deutschen Zahnärztetag in Frankfurt am Main eine interdisziplinäre Arbeitsgruppe zum Thema „Prüfungen in der Zahnmedizin“ gegründet. Diese Übersicht stellt die Zusammenfassung der aktuellen Arbeit dieser AG dar, deren Mitglieder zum Teil auch in der Arbeitsgruppe Zahnmedizin der Gesellschaft für Medizinische Ausbildung (GMA) aktiv mitwirken. Ziel der vorliegenden Übersichtsarbeit ist es, allen Interessierten, die sich mit der Planung, Durchführung und Auswertung von fakultätsinternen Prüfungen im Fach Zahnmedizin beschäftigen, den aktuellen Forschungsstand zum Thema darzustellen und zusammenzufassen.

Methoden: Basierend auf einer systematischen Literaturrecherche wurde anlehnend am sogenannten Nutzwertindex von Prüfungen eine tabellarische Darstellung der im NKLZ aufgeführten Szenarien realisiert.

Ergebnisse: Unterschiedliche Prüfungsszenarien wurden nach einer kurzen Beschreibung bezüglich ihrer Validität / Gültigkeit (V), der Reliabilität / Zuverlässigkeit (R), der Akzeptanz (A), der Kosten (C), der Durchführbarkeit (F) und des Einflusses auf Lernen und Lehren (EI) nach aktuellem Stand der Literatur tabellarisch dargestellt. Die Darstellungsform der Infoboxen wurde bewusst gewählt, um den Interessierten einen schnellen Einstieg in und Vergleich der einzelnen Prüfungsszenarien zu ermöglichen. Am Ende jeder Prüfungsbeschreibung befinden sich eine stichwortartige Sammlung der jeweiligen Anwendung in der Zahn(Medizin) und ein Fazit für die Anwendung.

Folgerungen: Die vorliegende Übersichtsarbeit beinhaltet eine Zusammenfassung zum Thema von kompetenzbasierten Prüfungsszenarien. Sie soll einen formativen Effekt auf die (zahn)medizinischen Fakultäten haben und sie darin unterstützen, die für die Zahnmedizin typischen Schwerpunkte des arbeitsplatzbasierten Lernens, Lehrens und Prüfens in der Zukunft weiterzuentwickeln.

Susanne Gerhard-Szep¹

Arndt Güntsch²

Peter Pospiech³

Andreas Söhnel⁴

Petra Scheutzel⁵

Torsten Wassmann⁶

Tugba Zahn⁷

1 Goethe-Universität,
Carolinum Zahnärztliches
Universitäts-Institut gGmbH,
Poliklinik
Zahnerhaltungskunde,
Frankfurt am Main,
Deutschland

2 Marquette University School
of Dentistry, Department of
Surgical Sciences,
Milwaukee, USA und
Universitätsklinikum Jena,
Zentrum für Zahn-, Mund-
und Kieferheilkunde, Jena,
Deutschland

3 Universität Würzburg,
Poliklinik für Zahnärztliche
Prothetik, Würzburg,
Deutschland

4 Universitätsmedizin
Greifswald, Poliklinik für
Zahnärztliche Prothetik,
Alterszahnheilkunde und
medizinischer
Werkstoffkunde, Greifswald,
Deutschland

5 Universitätsklinikum
Münster, Poliklinik für
Prothetische Zahnmedizin &
Biomaterialien, Münster,
Deutschland

6 Universitätsmedizin
Göttingen, Poliklinik für
Zahnärztliche Prothetik,
Göttingen, Deutschland

7 Goethe-Universität,
Carolinum Zahnärztliches
Universitäts-Institut gGmbH,
Poliklinik für Zahnärztliche
Prothetik, Frankfurt am Main,
Deutschland

1. Ausgangslage

Eine konzertierte Verknüpfung von Prüfungen und Lehre („constructive alignment“) ist für die Implementierung zahnmedizinischer Kompetenzen in der Ausbildung erforderlich [1].

Mit der Definition kompetenzorientierter Lernziele ist zudem verbunden, dass die entsprechenden Prüfungsformate die für eine konkrete Aufgabe erforderliche zielgerichtete Verknüpfung von Wissen, praktischen Fertigkeiten und professionellen Haltungen erfassen (siehe Abbildung 1).

2. Methodik

Eine Literaturrecherche wurde im Zeitraum vom 17.01.2014 bis 17.12.2014 in den Literaturdatenbanken der Deutschen Nationalbibliothek (DNB), MEDLINE mit der PubMed-Oberfläche, Excerpta Medica Database (EMBASE), Education Resource Information Centre (ERIC), Cochrane Library, Science Citation Index, Google Scholar durchgeführt und durch eine Handrecherche ergänzt. Es wurden außerdem verfügbare Dissertationen, Open-Access-Publikationen der German Medical Science (GMS), der BEME (Best Evidence Medical and Health Professional Education) inklusive deutschsprachiger Tagungsberichte (conference proceedings) beispielsweise der AKWLZ und der GMA evaluiert. Die Suchwörter lauteten jeweils „MCQ“; „MEQ“; „Multiple Choice“; „MC“; „Multiple Choice Questionnaire“; „SMP“; „Strukturierte Mündliche Prüfung“; „SOE“; „Structured oral examination“; „Key feature“; „OSCE“; „OSPE“; „standardized patient“; „CEX“; „miniCEX“; „entrustable professional activities“; „DOPS“; „Portfolio“, „multisource feedback“ in der Kombination mit „AND“ und „dental“; „medicine“; „education“; „assessment“.

In einem ersten Schritt wurde die Literatur anhand des Titels und des Abstracts gemäß vorher definierter Ein- und Ausschlusskriterien selektiert (Einschlusskriterien Publikationszeitraum 1966 bis 2013, Publikationssprache deutsch oder englisch, thematische Relevanz. Ausschlusskriterien: Nicht-Erfüllung der Einschlusskriterien, kein Volltext in Englisch oder Deutsch erhältlich, fehlende thematische Relevanz für die Fragestellung). Anschließend wurden die verbliebenen Publikationen hinsichtlich ihrer Relevanz für die vorliegende Thematik geprüft und gegebenenfalls ausgeschlossen.

Die Artikel wurden analysiert und die Ergebnisse anlehend am sogenannten Nutzwertindex von Prüfungen [2] in Kategorien zusammengefasst beschrieben. Hierzu gehörten die Parameter der Validität/Gültigkeit (V), der Reliabilität/Zuverlässigkeit (R), der Akzeptanz (A), der Kosten (C), der Durchführbarkeit (F) und der Einfluss auf Lernen und Lehren (EI). Anschließend wurden die Ergebnisse in Form der oben genannten Bewertungsparameter zusammengefasst. Diese Kriterien wurden 2011 von der Arbeitsgruppe von Norcini weiterentwickelt und die Parameter „Äquivalenz“ (beispielsweise hinsichtlich Prüfungen

an unterschiedlichen Standorten) und „Katalytischer Effekt“ (beispielsweise hinsichtlich Folgeeffekte bezogen auf die Fakultät) hinzugefügt [3]. In dem vorliegenden Review wurden die beiden Letztgenannten fakultativ den bereits recherchierten Ergebnissen in Textform hinzugefügt. Der Schwerpunkt der Auswertung erfolgte anhand der weiter oben genannten sechs Kriterien (V, R, A, C, F, EI) in einem tabellarischen Format, um eine effektive Übersichtlichkeit und Vergleichbarkeit zu gewährleisten.

3. Ergebnisse

Insgesamt wurden n=223 Publikationen mithilfe der oben beschriebenen Suchstrategie identifiziert und als Grundlage für die folgenden Analysen herangezogen.

4. Diskussion

Um theoretisches Faktenwissen (d.h. deskriptives Wissen der Kompetenzebene 1) zu prüfen, bieten sich die Prüfungsformate der strukturierten mündlichen Prüfung und des **Multiple Choice Questionnaire** an [4], [5], [6], [7]. Beim Mehrfachantwortauswahlverfahren MCQ handelt es sich um ein schriftliches Verfahren mit mehreren vorgegebenen Antwortmöglichkeiten (geschlossene Fragen), von denen eine (single choice) oder mehrere (multiple choice, multiple select) richtig sein können. Nach kurzem Fragen-Itemstamm (Einleitung/Fragestellung) folgen Antwortmöglichkeiten bestehend aus der/den richtige/n Antwort/en und Distraktoren (falsche Antwortmöglichkeiten). MC-Klausuren können sowohl rein papierbasiert, kombiniert analog-digital mit computergestützter Auswertung der ausgefüllten Bögen oder aber auch vollständig digital an Computerarbeitsplätzen durchgeführt werden [4].

Anwendung in der Human-/Zahnmedizin

- MCQs finden aktuell in den Studiengängen der Zahn- und Humanmedizin Anwendung [6], [8].
- die wichtigsten Abschluss- und Zwischenprüfungen (Naturwissenschaftliche Vorprüfung/NVP, Zahnärztliche Vorprüfung/ZVP, Zahnärztliche Prüfung/ZP, 1. Ärztliche Prüfung/ÄP und 2. Ärztliche Prüfung/ÄP) bestehen u.a. aus MCQ. Weiterhin sind insbesondere MCQ in allen vorklinischen und klinischen Fächern beider Studiengänge vertreten, sie stellen eines der „klassischen“, vorherrschenden Prüfungsformate dar [4], [8].

Zur Prüfung von Faktenwissen bieten MCQ ein kosteneffizientes Prüfungsformat mit hoher Reliabilität und Validität, insofern die eingesetzten Fragen den Gütekriterien entsprechen. Es ist mit MCQ möglich, objektiv viele Inhalte in kurzer Zeit zu prüfen. Das Lernen oberflächlicher Fakten jedoch kann gefördert werden.

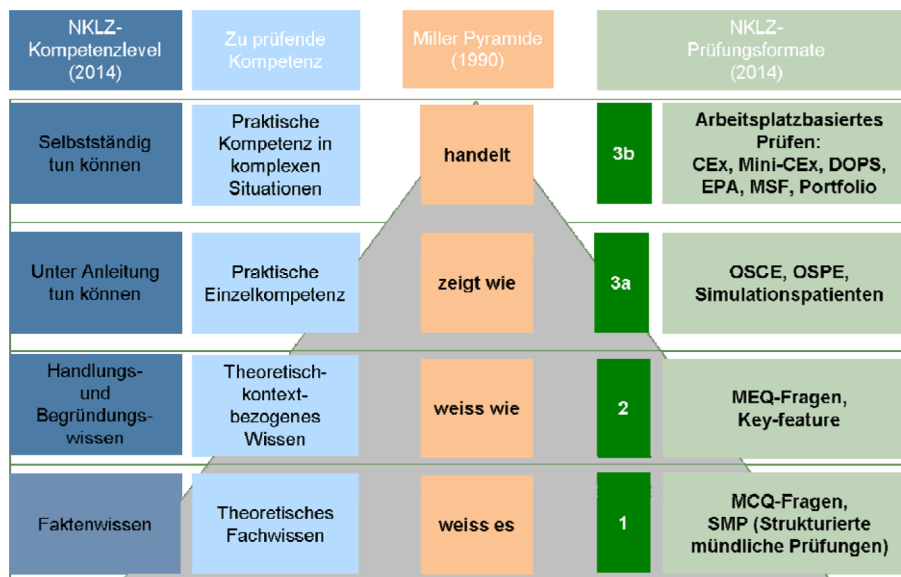


Abbildung 1 Beispiele für Prüfungsszenarien in Abhängigkeit von den Kompetenzebenen in Anlehnung an die Vorgaben von Miller und des NKLZ (Die Zahlen 1, 2, 3a und 3b kennzeichnen die verschiedenen Kompetenzlevel).

Multiple Choice Questionnaire

Validität/Gültigkeit

- als hoch einzustufen [9].
- Die Gütekriterien für Fragen müssen für eine ausreichende Validität erfüllt sein [10].
- Unterliegen die Fragen einem Reviewprozessen (z.B. via Item Management System, IMS), sichert dies eine hohe Konstruktvalidität [11].

Reliabilität

- als hoch einzustufen [9], [11].
- Ein Cronbachs α von 0,8 ist mit mindestens 40 hochwertigen Fragen zu erreichen [6].

Akzeptanz

- Die Auswertung ist objektiv [4].
- MCQ werden dann als fair empfunden, wenn Gelehrtes und Geprüftes sich entsprechen [12].
- Die Möglichkeit „taktisch“ zu kreuzen oder durch Raten bzw. Cues („Hinweise in der Aufgabe“) zu bestehen, wird von Lehrenden kritisch gesehen [13], [14], [15], [16].

Kosten

- Gemessen an der Anzahl der Prüfungen pro Zeit, eine effektive Prüfungsform [9], [17].
- Prüfen von breit gefächerten Inhalten in einer Prüfung [5].
- Verhältnismäßig geringe Kosten [18].
- Guter Kosten-Nutzen-Index.
- Ein bereits erstellter Fragenpool kann mit verhältnismäßig geringem Aufwand aktuell gehalten werden [19].

Durchführbarkeit

- Der Arbeitsaufwand wird in die Erstellungsphase verlagert, Durchführung und Auswertung verlangen weniger Zeit und Ressourcen.
- Das Generieren des Fragenpools ist mit nicht unerheblichem Aufwand verbunden [4].
- Online Prüfung mit digitaler Auswertung möglich [5].
- Ein fakultätsübergreifender Fragenpool steigert die Effizienz durch Synergien (z.B. IMS) [20].

Einfluss auf Lernen und Lehren

- Kann zu oberflächlichem Lernen führen [21].
- Theoretischer Kenntnisstand ist wichtiger, als praktische Fertigkeiten [4].
- Die korrekten Antworten werden bereits vorgeschlagen, passives Wiedererkennen ist dadurch möglich [14].

Die **strukturierten mündliche Prüfungen** (SMP, Structured Oral Examinations, SOE) sind mündliche Fachprüfungen, die durch einzelne Fachprüferinnen bzw. Fachprüfer oder als Kollegialprüfung vor einer Kommission (mehrere Prüferinnen und Prüfer / Kollegium) abgelegt werden.

Strukturierte mündliche Prüfungen

Validität/Gültigkeit

- direkt abhängig vom Grad der Strukturierung [22].
- steigt mit Planung, Konzeption und Bedingungen der Prüfung [23], [24].
- Validität mehr von den Anwendern (Prüfer), als von der Methode abhängig.

Reliabilität

- steigt mit Anzahl der Fragen, der Dauer der Prüfung und sinkt mit stark differenzierender Bewertung [10].
- Reliabilität und Objektivität steigen bei einer Kollegialprüfung = mehrere Prüfer [10], [17].

- Absolute Überprüfung der Reliabilität praktisch nicht möglich [10].
- Strukturierte mündliche Prüfungen liegen mit einem Cronbachs α von 0,65 – 0,88 [25], [26], [27] über konventionellen Prüfungen mit einem Cronbachs α von 0,24 – 0,50 [25], [28], [29].

Akzeptanz

- Leistungshemmender Stress, Angst und andere Störfaktoren spielen eine, verglichen mit MCQ, größere Rolle [12].
- die Akzeptanz von Lehrenden und Lernenden wird reduziert durch:
 - Intensive Betreuung durch Prüfer,
 - die Begründung von Noten
 - begrenzte Hinweise in begrenzter Zeit
- Fragen, eventueller Einspruch des Prüflings ohne vorliegende schriftliche Prüfung, die Differenzierung zwischen Inhalt und Art der Antwort sind u.a. dazu geeignet [12].

Kosten

- Kostenintensiver als MCQ [10]
- Relativiert sich bei high-stakes Prüfungen: hier stehen Reliabilität und Validität im Vordergrund, nicht die Kosteneffizienz [30], [31].

Durchführbarkeit

- Aufwändiger als MCQ, hohe finanzielle Belastung bedingt durch Personal- und Raumbedarf/Logistik [10].

Einfluss auf Lernen und Lehren

- Neben reinen Fakten können auch klinische Entscheidungsfindung, professionelles Denken, Selbstvertrauen und Selbstbewusstsein bewertet werden [12], [22].
- Da Studierende ihr Lernverhalten an die Prüfung anpassen [4], [5], [18] ist von einer umfassenderen Vorbereitung auszugehen.

Findet eine Kollegialprüfung statt, wird gemeinsam über die Bewertung der Leistung beraten und abgestimmt. Die abschließende Bewertung erfolgt idealerweise unter Anwendung eines inhaltlichen Blueprints (Erwartungshorizont) [7].

SMP sind ein Prüfungsformat, das die Abfrage auf Kompetenzniveau 1 (entsprechend NKLZ) und darüber hinaus im natürlichen Umfeld zahnärztlicher Interaktion ermöglicht. Hier sind jedoch die höhere finanzielle Belastung, bedingt durch hohen zeitlichen und personellen Aufwand, sowie die Beobachtungsmöglichkeit leistungshemmenden Stresses bei KandidatInnen zu beachten.

Anwendung in der Human-/Zahnmedizin

- mündliche Prüfungen unterschiedlichen Strukturierungsgrads finden in den Studiengängen der Zahn- und Humanmedizin Anwendung [8].
- die wichtigsten Abschluss- und Zwischenprüfungen (high-stakes Prüfungen) in beiden Studiengängen (NVP, ZVP, ZP, 1. ÄP und 2. ÄP) bestehen u.a. SMP unterschiedlicher Settings. Weiterhin sind SMP in allen

vorklinischen und klinischen Fächern beider Studiengänge vertreten, sie stellen eines der „klassischen“, vorherrschenden Prüfungsformate dar [4], [32].

Prüfungen, die nicht nur reines Faktenwissen (= deskriptives Wissen: „knows“, [33], [34]), sondern auch die Fähigkeit, theoretisches Wissen in einem spezifischen Kontext zur Lösung eines Problems bzw. zur klinischen Entscheidungsfindung einzusetzen (= prozedurales Wissen/Handlungs- und Begründungswissen: „knows how“), erfassen sollen, erfordern ein besonderes Prüfungsformat, das diese Fähigkeit auch tatsächlich abbilden kann. Hierbei ist grundsätzlich zu beachten, dass die Fähigkeit zur Problemlösung bzw. Entscheidungsfindung sehr kontextspezifisch ist und immer auch vom jeweiligen kontextbezogenen Faktenwissen abhängt [2], [35]. Als Prüfungsformate zur Erfassung von Handlungs- und Begründungswissen kommen neben strukturierten mündlichen Prüfungen schriftliche Prüfungen im **Modified-Essay-Question** (MEQ)- oder Key-Feature (KF)-Format in Frage. Dabei handelt es sich um fallbasierte schriftliche Prüfungen, welche sowohl aktiv reproduzierbares Wissen als auch Problemlösungs- bzw. Entscheidungskompetenz überprüfen soll und die klinische Situation eines Entscheidungsprozesses im Zusammenhang mit Untersuchung, Diagnose und Therapie simuliert. Hierbei wird eine Patientengeschichte erarbeitet, indem auf kurze, aufeinander aufbauende Situationsbeschreibungen jeweils einige wenige (2-3) kurze Fragen folgen, die durch Freitexteingabe bzw. Auswahlmöglichkeit unter mehreren vorgegebenen Antworten beantwortet werden müssen. Vorstehende Fragen werden zum Teil durch den nachfolgenden Text erklärt, weshalb ein Vor- und Zurückblättern nicht erlaubt ist.

Anwendung in der Human-/Zahnmedizin

- in den 70er Jahren für die allgemeinärztliche Facharztprüfung des „Royal College of General Practitioners“ in Großbritannien entwickelt [36], [37], [38], [39], [40], [41].
- im Bereich der Humanmedizin sowohl im Studium als auch in der postgradualen Weiterbildung weltweit verbreitet [42], [43], [44], [45], [46], [47], [48], [49], [50].
- in Deutschland im Medizinstudium neben der Anwendung als studienbegleitendes Prüfungsformat auch im Einsatz als schriftliche, das Staatsexamen ersetzende Prüfung [51], [52] innerhalb von Reformstudiengängen (u.a. Witten/Herdecke, Köln, Bochum).
- in der Zahnmedizin bisher kaum Anwendungsbeispiele; mögliche Einsatzgebiete im Zahnmedizinstudium: Überprüfung der Problemlösungs-Kompetenz im Rahmen von POL und Selbststudium im Rahmen fallbasierter problemorientierten Lernens [53], handlungsorientiertes fallbasiertes Prüfen mit virtuellen Patientenfällen (z.B. im Zusammenhang mit dem Vorgehen bei akutem Zahnschmerz in der Endodontie) [54]

Bei Beachtung einiger Grundsätze (1. möglichst große Zahl von Fällen, 2. beim sogenannten WI-(Write-In)-For-

mat: Qualitätskontrolle des vorgegebenen Lösungshorizontes durch mehrere BewerterInnen, 3. computerbasiertes SM (Short Menu)- oder Long-Menu-Antwortformat) stellt das MEQ-Prüfungsformat ein reliables Instrument zur Bewertung von kontextbezogenem, prozeduralem Wissen in klinischen Entscheidungssituationen dar, wobei durch die Simulation des klinischen Entscheidungsprozesses mit aufeinander aufbauenden Fragen schon während der Prüfung ein Lern- und Feedbackeffekt eintritt. Das MEQ-Format stellt somit eine sinnvolle Erweiterung der bisher in der Zahnmedizin üblichen schriftlichen Prüfungsformate dar, ist allerdings mit deutlich höherem Aufwand verbunden als die Abfrage reinen Faktenwissens mittels MCQ.

Modified Essay Questions

Validität/Gültigkeit

- Durch fallbasiertes, kontextreiches Fragenformat höhere Validität als bei MCQ-Format [48], [55], [56]
- Widersprüchliche Ergebnisse zur Korrelation (γ) von MEQ mit Ergebnis der Abschlussprüfung (NBME) bzw. postgradualer Performance im 1. Berufsjahr als Arzt: $\gamma=0,3/0,03-0,26$ [57], $\gamma=0,51$ [56]

Reliabilität

- Reliabilität (Cronbachs α) = 0,57-0,91 [38], abhängig von mehreren Einflussfaktoren [38], [39], [40], [47], [48], [58], [59]:
- Qualität des vorgegebenen Lösungshorizontes
- Antwortformat (Freitextantwort schlechter als Auswahl aus vorgegebener Liste)
- Anzahl von Fällen und Fragen
- Anzahl der Bewerter
- → z.B. Steigerung von Cronbachs α von 0,7 auf 0,8 durch Erhöhung der Fragenanzahl von 7 auf 12 oder Erhöhung der Bewerterzahl von 1 auf 4 [40]

Akzeptanz

- Studierende bewerten MEQ i.d.R. positiv [49], [51], da Fragenformat des MEQ näher an der Praxis als ein MCQ-Test [60]
- Lehrende/Prüfende: Hoher Aufwand bei Prüfungserstellung, Schwierigkeit der Koordination [51]

Kosten

- Erstellung und Korrektur der MEQ-Prüfung sehr zeit- und personalintensiv [36], [41], [51]
- Aufwand bei der Bewertung kann durch computerbasiertes Testformat verringert werden [61]

Durchführbarkeit

- Erstellung von MEQs deutlich anspruchsvoller als bei MCQs; Schwierigkeit, Fragen zu entwickeln, mit denen tatsächlich Problemlösungsfähigkeit bzw. klinische Entscheidungskompetenz erfasst und nicht nur Faktenwissen abgefragt wird [37], [41], [42], [43], [44], [45], [46], [50], [52], [53], [54], [62]

Einfluss auf Lernen und Lehren

- MEQ-Prüfung spielt klinischen Entscheidungsprozess durch, deswegen bereits in der Prüfung Feedback- und Lerneffekt [39], [51], [60]

Bei der **Key-feature**-Prüfung wird vor dem Hintergrund einer spezifischen klinischen Situation ein Fall entwickelt, in dessen Kontext mehrere Fragen gestellt werden, die sich ganz gezielt nur auf solche kritischen Aktionen oder Entscheidungen (= „Key-Features“) beziehen, die zur Lösung des klinischen Problems in der dargestellten Schlüsselsituation (= „Key Feature-Problem“) entscheidend sind, bzw. häufig falsch gemacht werden [34], [63]. Die Entwicklung eines KF-Falles erfolgt in 8 definierten Schritten [34], [64], [65]:

1. Festlegung der Domäne bzw. des Kontextes,
2. Auswahl einer klinischen Problemsituation (Rahmenbedingung),
3. Identifikation der kritischen Entscheidungen in dieser Situation (KFs des Problems),
4. Auswahl und Beschreibung des klinischen Szenarios (Fallvignette),
5. Schreiben der Fragen zu den Key-Features des Problems (1-3 Fragen pro KF),
6. Festlegung des Antwortformates (Freitext = Write-In, Auswahl = Short Menu oder Long Menu),
7. Erstellung des Bewertungsschlüssels und
8. Inhaltsvalidierung.

Anwendung in der Human-/Zahnmedizin

- KF-Prüfungsformat von Bordage und Page als Ersatz für bis dahin übliche schriftliche Überprüfung prozeduralen Wissens mittels Patienten-Management-Problems (PMP) für den Einsatz in Facharztprüfung entwickelt [64], [65].
- durch Hatala & Norman [66] Transfer in die studentische Ausbildung, seitdem im Bereich der Humanmedizin als schriftliches Prüfungsformat zur Bewertung von kontextbezogenem prozeduralem Wissen sowohl im Studium als auch in der postgradualen Weiterbildung weltweit verbreitet [67], [68].
- auch im deutschsprachigen Raum anerkanntes Prüfungsformat im Bereich der Humanmedizin (siehe u.a. ausführliche Anleitungen der Medizinischen Universitäten Bern & Graz zur Planung und Durchführung von Prüfungen [34], [60], [69].
- Studien und Erfahrungsberichte zum Einsatz des KF-Formates bei schriftlicher Leistungsüberprüfung im Medizinstudium in Deutschland u.a. für die Bereiche: Innere Medizin (Uni Freiburg/ Heidelberg/München [70], Uni Heidelberg/Tübingen [71], Hämatologie & Onkologie (Uni Düsseldorf [72], Kommunikationstraining (Uni Witten-Herdecke [73]).
- im Bereich Tiermedizin umfangreiches Pilotprojekt an der TiHo Hannover [74].
- nur vereinzelte Hinweise auf Einsatz von KF-Problemen als schriftliches Prüfungsformat im Rahmen des Zahnmedizinstudiums [75], [76].

Key-Feature-Prüfungen sind ein valides und reliables Instrument zur Bewertung von kontextbezogenem prozeduralem Wissen in Zusammenhang mit der Lösung eines klinischen Problems und stellen eine sinnvolle Erweiterung der bisher in der Zahnmedizin üblichen schriftlichen Prüfungsformate dar. KFs können darüber hinaus auch zum Selbststudium mit virtuellen Patientenfällen eingesetzt werden. Aus Praktikabilitätsgründen ist die computerbasierte Form im Long-Menu-Antwortformat der papierbasierten Prüfungsform vorzuziehen. Hiermit ist außerdem das Unterbinden des Vor- und Zurückblätterns leichter zu realisieren. Zur Steigerung der Reliabilität sind viele kurze KF-Fälle (mindestens 15) mit maximal 3 Fragen besser als wenige umfangreiche Fälle mit 4 oder mehr Fragen.

Key-Feature-Prüfung

Validität/Gültigkeit

- Hohe Inhaltsvalidität (92-94%) bei Beurteilung durch Lehrende / Prüfende [63], [65], [67].
- Voraussetzung für hohe Inhaltsvalidität ist Pilottestung & regelmäßiges Review der KF's durch Studierende, Lehrende/Prüfer [34], [63], [65].
- Bei geplantem LM-Antwortformat empfiehlt sich für Pilotphase das WI-Format, um Qualität des LMs zu verbessern (Ergänzen fehlender Antworten & Distraktoren) [34]
- Korrelation von KF-Testergebnissen und anderen Prüfungsergebnissen (z.B. MCQ) nur mittelmäßig ($r=0,35-0,54$, [66], [70]), was durch den Bezug auf unterschiedliche Kompetenzebenen erklärt werden kann.

Reliabilität

- Reliabilität des KF-Formates höher als beim PMP-Format [65]
- Wegen hoher Fallspezifität [48] ist Reliabilität unmittelbar von Anzahl der KF-Probleme (=Fälle) abhängig → Anzahl der KF-Probleme (=Fälle) sollte möglichst hoch, Anzahl der Fragen zu Einzelfall sollte 3 Items nicht überschreiten, da ab 4 Items Reliabilität abnimmt [77].
- Bei gleicher Anzahl von KF-Fällen scheint auch gewähltes Antwortformat Einfluss auf Reliabilität zu haben:
- 15 KF-Probleme mit 1 – 4 Fragen, 2h Dauer, WI-Format: Cronbachs $\alpha=0,49$ [66]
- 15 KF-Probleme mit 3 – 5 Fragen, 1,5h Dauer, computerbasiertes LM-Format: Cronbachs $\alpha=0,65$ [70] → mit 25 KF's ist $\alpha=0,75$ erreichbar!

Akzeptanz

- Studierende: relativ hohe Akzeptanz [74], [78]: Bewertung als realitätsnah und die praktische Ausbildung unterstützend.

Kosten

- Erstellung und Validierung der KF-Prüfung zeit- und personalintensiv [67]

Durchführbarkeit

- Erstellung von KFs schwieriger und zeitintensiver als MCQs [60], [69].
- Benötigte Prüfungszeit abhängig vom gewählten Antwortformat: LM >WI >SM >MC [79]
- Durch computerbasierte Umsetzung können Vorteile des LM-Antwortformates (geringerer Cueing-Effekt als MCQ/ SM, höhere Interrater-Reliabilität als WI) bei überschaubarer Prüfungszeit genutzt werden [70], [72], [79].
- Prüfungszeit für 15 KF-Probleme mit 3 - 5 Fragen beträgt bei computerbasiertem Test 90 Min. [70] und bei papierbasiertem Test mit WI-Antwortformat 120 Min. [66].
- Anwendungsbeispiele [68], [71], [75], [76], [80]

Einfluss auf Lernen und Lehren

- KF-Prüfungsformat ist näher an realer Patientensituation, fördert das Lernen klinisch relevanter Sachverhalte und das handlungsorientierte fallbasierte Lernen [81]

Während der zahnmedizinischen Ausbildungen wird nicht nur theoretisches Wissen auf hohem Niveau vermittelt, sondern Studierende müssen auch manuelle Fertigkeiten entwickeln. Es bedarf daher geeigneter Prüfungsformate, die nicht nur Fakten- und Handlungswissen erfassen, sondern bei denen Studierende ihre praktische Kompetenzen zeigen können („shows how“, [33]) und diese objektiv bewertet werden. Das reine „Wissen wie etwas geht“, wird hier auf die Stufe „zeigen wie etwas geht“ gehoben.

Bei der Erstellung solcher Prüfungen sollte im Vorfeld eine Selektion der Lernziele erfolgen und nur diese ausgewählt werden, die eine praktische Kompetenzebene aufweisen. Eine Standardisierung seitens Prüfer und Prüfung erlaubt eine objektive Beurteilung der gezeigten Leistungen der Studierenden. Als Prüfungsformate bieten sich hierbei **Objective Structured Clinical Examinations (OSCE)**, **Objective Structured Practical Examinations (OSPE)** und der Einsatz von Simulations- bzw. standardisierten Patienten (SP) an.

Zur Überprüfung praktischer Fertigkeiten und kommunikativer Kompetenzen eignet sich eine OSCE [14]. Die Studierenden durchlaufen verschiedene Stationen, bei denen jeweils praktische Tätigkeiten (auch Teilabschnitte einer Behandlung) oder Patientengespräche durchgeführt werden. Die Beurteilung erfolgt mithilfe einer Checkliste, die von einem Expertengremium nach Wichtung der Prüfungsinhalte erstellt wurde. Die reine Prüfungszeit pro Station beträgt ca. 5 Minuten, zum Stationswechsel (Prüfling) und zur Nachbereitung und/oder Feedback (Prüfer) sollten ca. 2 Minuten eingeplant werden.

Anwendung in der Human-/Zahnmedizin

- in der Medizin seit Einführung der OSCE weltweit in allen klinischen Fächern verbreitet.
- sowohl in der Graduierten- als auch Postgraduierten-ausbildung einsetzbar [82], [83], [84], [85].
- in der Zahnmedizin finden sich in vielen Disziplinen Anwendungsbeispiele: Vorklinik [86], [87], [88], Kieferorthopädie [89], [90], Mund-Kiefer-Gesichtschirurgie [91], [92], [93], Zahnerhaltungskunde [87], [94], [95], [96], Parodontologie [97], klinische Prothetik [86], Kinderzahnheilkunde [98], Radiologie [99], Mikrobiologie [94], [97].
- interdisziplinäre OSCE [87], [94], [100], [101].
- Integration einer OSCE in das zahnmedizinische Physikum [102].
- zur Evaluierung von kommunikativen Fertigkeiten [103], [104], Problemlösungsfähigkeit, kritischem Denken [105] auch in der Zahnmedizin angewandt.
- Sollte nach Möglichkeit in Zusammenhang mit Feedback durchgeführt werden.

Zur Überprüfung praktischer Einzelkompetenzen steht mit der OSCE eine reliable und valide Prüfungsmethode zur Verfügung, die sich bei Studierenden und Lehrenden einer hohen Akzeptanz erfreut.

Objective Structured Clinical Examination

Validität/Gültigkeit

- Prädiktive Validität
- signifikante Korrelation zwischen OSCE und Abschneiden bei praktischen Tests sowie praktischen Physikumsresultaten $p < 0,01$ [87]
- keine Korrelation zwischen OSCE und MCQ [105]
- Hohe Inhalts- und Konstruktvalidität [4], [106]
- hohe Augenscheinvalidität [107]
- akzeptable Vorhersagevalidität [108]
- Cave bei Sprachproblemen der Studierenden und hohem Stresslevel [109]
- Blueprinting beachten [106], [110]
- Festlegung der Problemfelder am Anfang beachten [110]
- Definition von Aufgaben innerhalb der Problemfelder [110]

Reliabilität

- Cronbachs α zwischen 0,11–0,97 [4]
- Hohe Reliabilität innerhalb der OSCEs, bei weniger als $n=10$ Stationen ca. 0,56, bei mehr als $n=10$ Stationen 0,74 [111]
- Unterschiedliche Empfehlungen zu Stationsanzahl:
- mindestens 19 [4]
- 14-18 für jeweils 5-10 Minuten [106], [112]
- Stationen mit SP mindestens 15 Minuten lang prüfen [110]
- Je mehr Prüfer desto höhere Werte
- [111], [113] Bewertungsmodus ausschlaggebend: Hohe Werte bei globalen Bewertungen, auch Kombi

aus globalen Bewertungen und Checkliste gut, nur Checkliste am wenigsten geeignet),

- Post-OSCE-Tests erhöhen die Reliabilität [110]

Akzeptanz

- Studierende: hohe Akzeptanz, geeignete Prüfungsform für operative Fertigkeiten [96]
- Lehrende / Prüfende: Hohe Akzeptanz [112], [114], [115], [116]

Kosten

- 181 €/Prüfling
- 86-130 € /Prüfling [4], [117]
- 2,5 h/Prüfling [118]
- 15-200 Dollar/Prüfling [117], [119], [120]
- Höhere Kosten bei Einbindung von SP [121]
- 21-200 \$ pro Prüfling [122]

Durchführbarkeit

- Zeitintensive und ressourcenintensive Prüfungsform [106], [119]
- Gute Vorplanung nötig:
- gemeinsame Strukturen aufbauen hilft bei interdisziplinären OSCEs [100]
- Evaluation durch externe Prüfer empfohlen
- Qualität von SPs sicherstellen
- Stationsinhalt sollte passend zum OSCE-Szenario gewählt werden
- Peer-review prä- und post OSCE (Psychometrische Analyse mit Schwierigkeit, Trennschärfe etc. empfohlen)
- Erfahrung des Prüfers, Fachrichtung, Geschlecht und Ermüdung der Prüfer zu beachten [4], [106], [112]
- Anwendungsbeispiele [95], [102]

Einfluss auf Lernen und Lehren

- Positiver Einfluss aufs Lernen [106], [108], [123]
- stimuliert das Lernen [112]
- Stationslernen hat wenig mit Realität am Patienten zu tun [112]
- Feedback einplanen [110]

Aufgrund der zeit- und ressourcenintensiven Vorbereitung und Durchführung, empfiehlt sich eine fächerübergreifende Zusammenarbeit, um diese Nachteile zu minimieren. OSCEs können bisherige Prüfungsformen substituieren oder sinnvoll ergänzen. Eine ausreichende Anzahl der Stationen ($n > 10$), ein Blueprinting, ein Peer-review der Stationsinhalte und des Bewertungskataloges sowie eine Balancierung der Bewertungsmodi (global, Checkliste, Kombination), die Schulung der Prüfer und ggf. die Durchführung eines Pilot-OSCE sollten bei der Erstellung einer OSCE berücksichtigt werden. Eine Spezialform der OSCE stellen sogenannte „objektive strukturierte praktische Prüfungen“ (auch OSPE genannt) dar, bei denen praktische Fertigkeiten, Wissen und/oder Interpretation von Daten in nicht-klinischer Situation erfolgt [124]. Diese können in Laboren oder an den Simulationsplätzen im SimLab durchgeführt werden. Im Unterschied zum OSCE können im Rahmen einer OSPE ganze Arbeitsprozesse

bis hin zu einem „Endprodukt“ (bspw. Füllung) geprüft werden.

Mit dem **OSPE** kann man praktische Fertigkeiten und/oder die Interpretation klinischer Daten sicher überprüfen. Es handelt sich dabei um eine reliable und valide Prüfungsmethode zur Überprüfung praktischer Einzelkompetenzen, die sich bei Studierenden und Lehrenden einer hohen Akzeptanz erfreut.

Objective Structured Practical Examination

Validität/Gültigkeit

- Hohe Validität $\gamma > 7$
- hohe Konstruktvalidität

Reliabilität

- Hohe Reliabilität innerhalb der Stationen Cronbachs $\alpha > 0,8$ [125]
- Interrater Reliabilität ICC $> 0,7$
- hohe Interrater Reliabilität bei gleichem Erfahrungs- und Wissenstand der Prüfer $\gamma = 0,79$ bis $0,93$; $p < 0,001$

Akzeptanz

- Studierende: Hohe Akzeptanz [126], [127]
- wird als „faire Prüfung“ empfunden [128]
- wird traditionellen Prüfungsformen vorgezogen [126]
- Lehrende: relevante, faire, objektive und reliable Prüfungsform

Kosten

- keine Angaben

Durchführbarkeit

- erfordert intensive Planung und Teamarbeit [128]

Einfluss auf Lernen und Lehren

- individuelle Kompetenzen können überprüft werden, Fakten- und Handlungswissen praktisch demonstrieren zu müssen beeinflusst das Lernverhalten [128]
- zeigt Stärken und Schwächen praktischer Fertigkeiten auf [129]
- Lernstimulus [129]
- positive Lernerfahrung [130]

Ein definierter Bewertungskatalog zu den Einzelschritten innerhalb des Arbeitsprozesses ist notwendig.

Anwendung in der Human-/Zahnmedizin

- OSPE's werden in Medizin, bspw. Pharmakologie [128], Physiologie, Forensische Medizin [130] und Zahnmedizin [131], [132] weltweit durchgeführt.
- In Deutschland primär in der präklinischen zahnmedizinischen Ausbildung eingesetzt [133].

Simulationspatienten sind für die zahnmedizinische Ausbildung gezielt geschulte (Laien-) Schauspieler, die Krankheitsbilder oder typische Gesprächsanlässe des zahn-/medizinischen Alltags (auch standardisiert) darstellen können. Einsatzgebiete reichen von der Übung oder Prüfung beispielsweise von Arzt-Patienten-Gesprächen,

Untersuchungstechniken, bis hin zum Training der körperlichen Untersuchung, zum Erlernen kommunikativer Kompetenzen beispielsweise auch im Rahmen von Prüfungen (häufig im OSCE-Szenario).

Simulationspatienten können zur Prüfung von Zahnarzt-Patienten-Gesprächen und von Untersuchungstechniken eingesetzt werden. Zur Evaluation klinischer und kommunikativer Kompetenzen vor allem im Rahmen des OSCEs sind sie besonders geeignet. Dabei sollte die Komplexität des Falles auf das Prüfungsszenario abgestimmt sein.

Simulationspatienten

Validität/Gültigkeit

- Zur Evaluation klinischer Kompetenzen [134]

Reliabilität

- Konsistente Prüfung
- (keine signifikanten Unterschiede zwischen Prüfungsgruppen und -zeitpunkten) [135]

Akzeptanz

- Einsatz von Simulationspatienten im Rahmen einer OSCE -Station [136]

Kosten

- 10-18 €/Prüfling [136]

Durchführbarkeit

- Komplexität der Fälle kann kontrolliert und dem Ausbildungsstand angepasst werden [137]
- Fakultätsmitglieder können relevante Lernziele bestimmen und das Rollendesign darauf abstimmen
- hoher Zeit- und Personalaufwand für SP-Selektion, Ausbildung und Monitoring für Qualitätsmanagement [137]
- Checkliste, in die alle Beobachtungen des Arzt-Patienten-Gesprächs vom SP eingetragen werden [138]
- Anwendungsbeispiele [139]

Einfluss auf Lernen und Lehren

- Steigerung der klinischen Fertigkeiten der Studierenden [140]

Anwendung in der Human-/Zahnmedizin

- Bereits seit den 1960er Jahren eingesetzte Methode in der klinischen Ausbildung [138].
- Patientenkontakt unter standardisierten Bedingungen simulierbar [139].
- SP können neben ihrer Rolle als Patient auch Feedback geben oder die Kompetenzen des Prüflings einschätzen [139].

Unter dem Oberbegriff Arbeitsplatzbasiertes Prüfen (AbA = Arbeitsplatzbasiertes Assessment) werden unterschiedlichste Prüfungsszenarien zu praktischen Kompetenzen am Patienten in komplexen Situationen zusammengefasst.

Beim **Clinical Evaluation Exercise** (CEX) handelt es sich um eine arbeitsplatzbasierte Prüfung im klinischen Be-

handlungsumfeld, die sich über einen längeren Zeitraum erstreckt (mehrere Stunden bis Tage) und Behandlungsabläufe, bei welchen in der Regel ein einziger Patient unter Praxisbedingungen zum Beispiel hinsichtlich Anamneseerhebung und Untersuchung bewertet wird. Zum Einsatz kommen maximal zwei Prüfer, die in der Regel nicht die gesamte Zeit am Patienten vor Ort anwesend sind. Häufig werden die vom Prüfling ermittelten Daten mit dem Patienten somit ohne die Anwesenheit des Prüfers erhoben. Diese Prüfungsform wird auch tCEX (traditional CEX) genannt und stellt eine sogenannte single event measure dar.

Anwendung in der Human-/Zahnmedizin

- in 60er Jahren ursprünglich von der ABIM (American Board of Internal Medicine) für Internisten entwickelt und löste im Jahr 1972 die mündliche Prüfung als Standardmethode ab [141], [142].
- seit ca. 1995 durch Mini-CEX ersetzt [143], [144].
- in der Zahnmedizin bisher keine dokumentierten Anwendungsbeispiele in der Fachliteratur.

Es handelt sich somit um ein gering valides und gering reliables Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen. Optimierungsmöglichkeiten bestehen darin, dass 1. eine möglichst große Anzahl von Patienten (Fällen), 2. eine möglichst große Anzahl von Prüfern und 3. möglichst strukturierte Evaluationsinstrumente zum Einsatz kommen. Zudem sollte bei dieser Prüfungsform ein Feedback obligat erfolgen. Insgesamt kann man festhalten, dass das CEX-Format nur unter Nutzung der möglichen Optimierungsvarianten ein sinnvolles Prüfungsformat für die Abbildung von praktischen Kompetenzen in komplexen Situationen in der zahnmedizinischen Ausbildung darstellt.

Clinical Evaluation Exercise

Validität/Gültigkeit

- unzureichende Inhaltsvalidität, deckt Lernziele im Curriculum unvollständig ab [145]
- künstliche Situation, entspricht nicht der Realität unter Praxisbedingungen, da zu lange und ausführlich [144]

Reliabilität

- fragliche Zuverlässigkeit, da infolge des hohen Zeitbedarfs in der Regel nur wenige CEX durchgeführt werden können [146]
- geringe Interrater-Reliabilität [147]
- Cronbachs α bei einem Fall 0,24 und auch bei zwei Fällen lediglich 0,39 [141]

Akzeptanz

- Geringe Akzeptanz, da stark abhängig vom Prüfer [148]

Kosten

- Günstiger als OSCE, weil echte Patienten, die nicht extra geschult werden müssen [145]

Durchführbarkeit

- Relativ einfach, da keine besondere Vorbereitung nötig [141]
- Anwendungsbeispiele [142], [143]

Einfluss auf Lernen und Lehren

- Patientenbasierte, reale Situation [141]

Beim **Mini-Clinical Evaluation Exercise** (mCEX) handelt es sich um eine patientenbasierte Prüfungsform im klinischen Behandlungsumfeld, die sich, im Unterschied zum CEX-Prüfungsformat über einen kurzen Zeitraum erstreckt und stets Feedback beinhaltet (ca. 15 Minuten Prüfung + ca. 10 Minuten Feedback). Man kann diese Prüfungsform als dreistufig beschreiben, indem 1. beobachtet, 2. das Ergebnis festgehalten und 3. Feedback erteilt wird. Dabei wird die Prüfung von mehreren Prüfern beobachtet und nach strukturierten Vorgaben bewertet, wobei die inkludierten Behandlungsabläufe an mehreren Patienten unter Praxisbedingungen mit Schwerpunkt Kommunikation und klinische Untersuchung durchgeführt werden [144]. Die Bewertung erfolgt prinzipiell anhand zuvor festgelegter Kriterien und gilt für jeden Prüfling. Diese Kriterien können eine Ratingskala beinhalten und / oder durch Freitextkommentare definiert werden. Die Schwierigkeit bleibt bezüglich unterschiedlicher zu untersuchenden Patienten. Aus Sicht der Prüfungspyramide nach Miller steht man jedoch auf einer hohen Anwendungsebene. Es handelt sich streng genommen um eine strukturierte klinische Beobachtung.

Mini-Clinical Evaluation Exercise

Validität/Gültigkeit

- Höhere Validität als CEX [149]
- akzeptable Validität und Reliabilität nachgewiesen [146], [150]
- ist in der Lage, unterschiedliche Kompetenzebenen (1. Jahr, 2. Jahr etc.) valide zu unterscheiden [151]

Reliabilität

- geringe Interrater-Reliabilität [149]
- mindestens 10 Evaluationen sind nötig, um verlässliche Ergebnisse zu erhalten; eine größere Anzahl ist besser [151]
- es werden mindestens 12-14 Bewertungen pro Jahr empfohlen bei unterschiedlichen Bewertern um die Interrater-Reliabilität zu steigern [152]
- Reliabilität von $G=0,4$ bei 10 Bewertungen, $G=0,8$ bei 50 Bewertungen [151]
- Abhängig von Prüferanzahl: Bei einem Prüfer mindestens 8 Beobachtungen an unterschiedlichen Patienten, bei 2 Prüfern 4, bei 3 Prüfern 3 Beobachtungen nötig, um eine Reliabilität von 0,8 zu erlangen [153]
- Neun Items besser als 5 Items um Unterschiede in Kompetenzen abzubilden [154]

Akzeptanz

- Hohe Zufriedenheit sowohl seitens der Lehrenden als auch der Lernenden [151], [155], [156]
- Durchführung z.T. erst zögerlich, da Neuerung [156]
- Zum Teil problematisch wegen Diskrepanz zwischen Selbst- und Fremdeinschätzung [157]

Kosten

- Entstehung erheblicher Kosten infolge des hohen Zeitbedarfs [158], [159]

Durchführbarkeit

- Beobachtung von realen Arzt-Patienten-Interaktionen durch verschiedene Weiterbildner in unterschiedlichen Situationen; Feedback zu verschiedenen Krankheitsbildern an unterschiedlichen Orten mit unterschiedlichem Fokus [155]
- Gute Planung nötig, denn Feedback dauert 8-17 Minuten [155], [160]
- Relativ einfache Implementierung mit genügender Flexibilität in dentaler Umgebung [161]
- Anwendungsbeispiele [162]

Einfluss auf Lernen und Lehren

- Kompetenzsteigerung durch regelmäßiges Feedback durch Experten [163]
- Prüfer/Prüfling erhält Feedback bzw. ein konkretes Bild der klinischen Arbeit. Somit gezielte Förderung möglich [156]
- Erteilen eines konstruktiven Feedbacks muss erlernt und geübt werden, pädagogische Fähigkeiten sind nötig [164]
- Keine neuen Erkenntnisse gegenüber traditionellen Bewertungsverfahren [158]
- Kein Einfluss im Vergleich zur Kontrollgruppe [153]
- Verknüpfung zwischen der Definition von Lernzielen und den identifizierten Lerninhalten erforderlich [165]
- Prädiktive Validität zwischen OSCE und mCEX nicht nachweisbar [165]

Diese Prüfungsform wird auch häufig mCEX (mini CEX) genannt und stellt eine sogenannte single event measure dar.

Anwendung in der Human-/Zahnmedizin

- 1995 entwickelt von Norcini [144], löste in 90er Jahren tCEX ab.
- Reliabilität stark abhängig von Prüfer- und Fallanzahl [151], [153].
- in der Zahnmedizin mehrere dokumentierte Anwendungsbeispiele in der Fachliteratur (Dental Foundation Training in Großbritannien), jedoch häufig ohne genauen Angaben hinsichtlich der Evaluationsinstrumente [161], [162].

Beim mCEX handelt es sich um ein valides und reliables Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen. Optimierungsmöglichkeiten bestehen 1. in der Erhöhung der Anzahl der Antwortitems (9 besser als 5) bzw. der Anzahl der Beobachtungen

(mindestens 10 Beobachtungen sinnvoll) und 2. im Angebot von Train-the-Teacher-Veranstaltungen (beispielsweise in Form von Videodemonstrationen und Rollenspielen). Es wird eine longitudinale Anwendung empfohlen, wobei die Implementierung in diversen Settings (auch in High-Stakes-Examina) vorstellbar ist. Das mCEX-Format stellt ein sinnvolles Prüfungsformat für die Abbildung von praktischen Kompetenzen in komplexen Situationen im zahnärztlichen Ausbildungsszenario dar.

Anvertraubare professionelle Tätigkeiten (Entrustable Professional Activities) schließen die Lücke zwischen der Theorie eines kompetenzbasierten Trainings und einer patientenzentrierten Umsetzung im klinischen Kontext [166]. Bekannt wurde dieses Szenario anfangs aus dem Bereich der ärztlichen Weiterbildung (postgraduate), seit 2013 auch für die medizinische undergraduate-Ausbildung [167], [168]. Bewertet wird die Integration verschiedener Wissens- und Fertigkeitinhalte für die Lösung komplexer Aufgaben (z.B. Anamneseerhebung, klinische Untersuchung am Patienten im Zusammenhang mit unterschiedlichen Konsultationsanlässen) unter Einbeziehung existierender kompetenzbasierter Rollen, wie von CanMeds bzw. ACGME vorgegeben. Bei der Überprüfung wird eingestuft, ob der Geprüfte diese Tätigkeit unter Anleitung, Supervision, gelegentlicher Hilfestellung oder eigenständig durchführen kann [169], [170]. Damit können unterschiedliche Leistungsniveaustufen abgebildet werden [171]. Es werden keine einzelnen Lernziele, sondern eine vollständige Aktivität am Patienten geprüft [172]. Um EPAs von allgemeinen Lernzielen unterscheiden zu können, wird vorgeschlagen, den Satz in modifizierter Form: „Morgen wird es dem (Zahn)Arzt erlaubt sein, zu...“ zu vervollständigen [166]. Definitionsgemäß soll eine EPA Aktivitäten beinhalten, die in der täglichen Praxis wichtig, in der Ausführung stark fehlerbelastet sind und mehrere Kompetenzen integrieren [172], [173]. Somit besteht eine EPA aus diversen Rollen, jede Rolle aus mehreren Lernzielen und jedes Lernziel aus unterschiedlichen Leistungsniveaustufen. Die Überprüfung kann wahlweise eine direkte oder indirekte Beobachtung und Feedback beinhalten. Dabei ist es entscheidend, dass man die beobachteten Leistungen des Prüflings mit den tatsächlichen Leistungsbewertungen über einen bestimmten Zeitraum kombiniert.

Entrustable Professional Activities

Validität/Gültigkeit

- Hohe „face-validity“ [174]

Reliabilität

- Niedrige Interrater-Reliabilität [175]

Akzeptanz

- Potential für breite Akzeptanz [166]
- Hilft den Lernenden, einen eigenen Lernplan zu entwickeln [176]
- Hilft der gesamten Fakultät, eine Transparenz in der Ausbildung zu gewähren [176]

Kosten

- keine Angaben

Durchführbarkeit

- erfordert initial eine intensive, durchdachte Vorbereitung, bei der die EPA erstellt werden [177]
- es werden 20-30 EPAs für ein ganzes Ausbildungsprogramm empfohlen [177]
- Anwendungsbeispiele [178], [179]

Einfluss auf Lernen und Lehren

- EPAs verlangen zahlreiche Kompetenzen in einer integrativen, holistischen Art und Weise [177]
- Methode der Bewertung, die ihren Schwerpunkt auf den benötigten Grad der Supervision legt [180]
- Feedback ist ausschlaggebend [174]
- Benötigt Unterstützung seitens der Fakultät [175]
- Ermöglicht eine breite (panoramaweite) Sicht auf die Ausbildung [174]

Eine häufig beschriebene Kombination ist hierbei die von mCEX und MSF (Multi-Source-Feedback). Es handelt sich streng genommen um eine sogenannte multiple event measure.

Anwendung in der Human-/Zahnmedizin

- 2005 von ten Cate in den Niederlanden eingeführt, seither beschrieben in der Chirurgie, Familienmedizin, Inneren Medizin, Neurologie, Notfallmedizin, Pädiatrie, Urologie und in breiter Anwendung des „Royal Australian and New Zealand College of Psychiatrists“ [178], [179].
- in Deutschland im Medizinstudium erst in der Pilotphase [166].
- in der Zahnmedizin keine dokumentierten Anwendungsbeispiele.

Beim EPA handelt es sich um ein relativ neues, bisher wenig erforschtes Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen. Die Implementierung von EPAs erfordert initial eine intensiv durchdachte Vorbereitung bei der Festlegung der Schwerpunkte. Nach Möglichkeit sollten für einen Ausbildungsabschnitt maximal 30 EPAs interdisziplinär unter Einbeziehung von universitären und in Praxen niedergelassenen Meinungsbildnern definiert werden. EPAs verschaffen eine realistische Verknüpfung von kompetenzorientierten Lernzielen zu übergeordneten Aktivitäten. Train-the-Teacher-Veranstaltungen (mit Übungen zu Feedback) sollen die Implementierung optimieren. Eine longitudinale Anwendung wird empfohlen. Die Implementierung ist in diversen Settings (auch in High-Stakes-Examina) vorstellbar. Das EPA-Format stellt zusammenfassend eine innovative Herangehensweise für die Abbildung von praktischen Kompetenzen in komplexen Situationen im zahnärztlichen Ausbildungsszenario mit viel Zukunftspotential dar.

Beim **Directly Observed Procedural Skills** (DOPS) handelt es sich um eine arbeitsplatzbasierte Prüfung im klinischen Behandlungsumfeld, die sich über einen kurzen Zeitraum

erstreckt und Feedback beinhaltet (ca. 15 Minuten Prüfung und ca. 10 Minuten Feedback) ähnlich dem mCEX. Auch hier handelt es sich um eine dreistufige Prüfungsform, indem 1. beobachtet, 2. das Ergebnis festgehalten und 3. Feedback erteilt wird. Hierbei werden die Behandlungsabläufe an (mehreren) Patienten unter Praxisbedingungen wie bei der mCEX, aber mit den Schwerpunkten der manuell-technische Fertigkeiten bzw. Interventionen von mehreren Prüfern beobachtet und nach strukturierten Vorgaben bewertet. Diese Prüfungsform stellt eine ebenfalls sogenannte single event measure dar.

Anwendung in der Human-/Zahnmedizin

- ursprünglich eingeführt im Vereinigten Königreich vom General Medical Council 2002 [144].
- beschrieben in der Allgemeinen Medizin, Chirurgie und Innere Medizin [181].
- in der Zahnmedizin international beschrieben in Iran (Shiraz und Mashad Universität) und im Kings College in London/GB [182], [183].

Beim DOPS handelt es sich um ein valides und reliables Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen. Optimierungsmöglichkeiten bestehen darin, dass 1. drei Prüfer während zwei Beobachtungen intervenieren, dass 2. mindestens zwei Beobachtungen durchgeführt werden und dass 3. Train-the-Teacher-Veranstaltungen durchgeführt werden. Es wird insgesamt eine longitudinale Anwendung empfohlen. Die Implementierung ist in diversen Settings (auch in High-Stakes-Examina) vorstellbar. Das DOPS-Format stellt zusammenfassend ein sinnvolles Prüfungsformat für die Abbildung von praktischen Kompetenzen in komplexen Situationen im zahnärztlichen Ausbildungsszenario dar.

Directly Observed Procedural Skills

Validität/Gültigkeit

- Hohe „face-validity“ [181]
- Formatives assessment tool [182]
- Signifikante Unterschiede zu MCQ; liefert unterschiedliche Einschätzungen studentischer Leistungen [182]
- Einzelnes Assessment tool, kann nicht eine umfassende Bewertung ermöglichen, es wird ein System mit verschiedenen Möglichkeiten benötigt [184]
- DOPS evaluiert die praktischen Fähigkeiten effizient [182]

Reliabilität

- Um eine hohe Reliabilität zu erlangen, sollten mindestens drei Prüfer einen Prüfling beobachten während zwei unterschiedlicher Fallszenarien [181]
- $G=0,81$ [185]
- Interne Konsistenz=0,94 und Interrater Reliabilität=0,81
- in der Einschätzung der Studenten nicht gut geeignet, die inter-rater reliability zu verbessern [186]
- große Unterschiede zwischen den Prüfern können die Aussagekraft der Ergebnisse beeinflussen, solange

nicht eine strenge Standardisierung vorgenommen wurde [187]

- gute Reliabilität und Übereinstimmung zwischen Prüfern möglich [188]
- es werden weniger Prüfungen benötigt verglichen mit Mini-CEX [160]
- Es werden weniger Prüfer und Fälle benötigt als im Mini-CEX [181]
- Höhere Item-correlation-Werte als bei Mini-CEX=0.7-0.8 versus 0.5-0.8 [150], [189]
- Reliabilität abhängig vom Fall [181]
- Reliabilität unabhängig vom Prozess [160]

Akzeptanz

- hohe Akzeptanz bei Studierenden [186]
- Prüflinge empfinden das Szenario als sehr stressig, freuen sich aber über das erteilte Feedback [190]

Kosten

- Hohe Kosten zu erwarten [159], [191]

Durchführbarkeit

- hoher Zeitbedarf [160], [182]
- hoher Zeitbedarf für die Vorbereitung der DOPS inkl. Feedbackgabe [160]
- zur Steigerung des Lerneffekts ist es nötig, unmittelbar nach Durchführung Feedback zu geben und klar Stärken und Schwächen anzusprechen [192]
- Prüfer müssen im Vorfeld geschult werden [12]
- Durchführbarkeit gegeben, wenn nur ein Prüfer eingesetzt wird [193]

Einfluss auf Lernen und Lehren

- in der Wahrnehmung der Prüflinge ein positiver Einfluss auf Unabhängigkeit und Lernprozess [186]
- DOPS Assessment verbessert die klinischen praktischen Fähigkeiten [192]
- Positiver Effekt durch die direkte Beobachtung der lernenden Person [192]
- fördert einen tief gehenden Ansatz für Lernen im klinischen Kontext [21]
- positiver Einfluss auf die Reflexion beim Lernen [181]
- 70% der Prüflinge sind der Ansicht, dass DOPS hilfreich sind, um praktische Fertigkeiten zu verbessern [194]
- Gegenüber Kontrollgruppe signifikant besseres Resultat bei DOPS hinsichtlich praktischer Fertigkeiten [195]
- Auch in Peer-Arrangements im präklinischen und klinischen Kontext anwendbar [183]

Bei dem Portfolio-Prüfungsszenario handelt es sich um eine vorgegebene, an Lernzielen orientierte Sammlung von studentischen Lernaktivitäten mit geplanter Selbstreflexion inklusive eines Feedbacks [20]. Die Sammlung wird dem Lernprozess entsprechend konzipiert, wobei folgende Aspekte darin Berücksichtigung finden können:

1. Erfahrungen (was wurde getan, gesehen, geschrieben, erstellt?),

2. Lernprozess (das Bewusstsein, dass das Erfahrene für die zukünftige Tätigkeit oder Entwicklung bedeutsam ist),
3. Nachweise (Belege, Scheine),
4. Lernbedürfnis (Zukunftsorientierung) und
5. Lernbedingungen [196].

Es handelt sich um eine sogenannte multiple event measure.

Anwendung in der Human-/Zahnmedizin

- Portfolio based learning (portfoliobasiertes Lernen) eingeführt 1993 vom Royal College of General Practitioners, Portfolio assessing (portfoliobasiertes Prüfen) 1998 von Shulman beschrieben [197], [198].
- publiziert in der Allgemeinmedizin, HNO, Innere Medizin, Pädiatrie, Public Health an den Universitäten Maastricht (NL), Nottingham (GB), Arkansas (USA) [196].
- in Deutschland im Medizinstudium in Köln [196].
- in der Zahnmedizin international beschrieben [199], [200], [201].

Beim Portfolio handelt es sich um ein hoch valides und reliables Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen, das die gesammelten kumulativen Informationen über Leistung und Entwicklung bewertet. Optimierungsmöglichkeiten bestehen darin, dass man 1. mehrere unabhängige Prüfer implementiert, dass 2. der eigentliche Mentor nicht gleichzeitig Prüfer sein sollte und dass 3. Train-the-Teacher-Veranstaltungen (Feedback) durchgeführt werden. Eine longitudinale Anwendung wird empfohlen. Die Implementierung in diversen Settings (auch in High-Stakes-Examina) ist vorstellbar. Das Portfolio-Format stellt somit ein sinnvolles Prüfungsformat für die Abbildung von praktischen Kompetenzen in komplexen Situationen im zahnärztlichen Ausbildungsszenario dar.

Portfolio

Validität/Gültigkeit

- gute Validität, wenn eine geeignete Auswahl aus allen benötigten Kompetenzfeldern getroffen wird [202], [203]

Reliabilität

- Cronbachs α : 0.8 bei 4 Prüfern [204]
- Cronbachs α : 0.8 bei 15 Portfolioinhalten und 2 Prüfern [202]
- Nutzung eines kompetenzbasierten Masterplans, klare Bewertungskriterien, Richtlinien und erfahrene Bewerter bei der Entwicklung und Bewertung [202], [203]
- einheitliche Bewertung schwierig [200]

Akzeptanz

- Portfolios werden als zeitaufwendig, angsterzeugend und wenig effektiv eingestuft [205]
- Die Akzeptanz von Portfolios sinkt, je länger sich Studierende damit beschäftigen [205]

Kosten

- keine Angaben

Durchführbarkeit

- Typische Inhalte sind beispielsweise: 7 Fallberichte, 2 Präsentationen, 3 Selbstreflexionen [202]
- Typische Inhalte sind beispielsweise: Diagnosen, Behandlungspläne [202]
- problematisch, da ein Konflikt entsteht, wenn Portfolios zum Bewerten und zum Lernen eingesetzt werden [205]
- Schwierigkeiten, selbstkritisch und ehrlich zu sein [205]
- Interviews mit den Studierenden zum Inhalt des Portfolios verbessert die Durchführbarkeit [206]
- Anwendungsbeispiele [199], [201]

Einfluss auf Lernen und Lehren

- Ermöglicht, Fähigkeiten zu evaluieren, die mit anderen Werkzeugen nicht evaluiert werden können [200]
- Inhalt des Portfolios muss mit den Lernzielen übereinstimmen [202]
- steigert Selbsterkenntnis und regt zum kritischen Denken an [205]
- steigert die Fähigkeit, unabhängig zu lernen und verknüpft Theorie mit Praxis [205]
- hoher Zeitbedarf für Prüfer und Prüfling [200], [207]
- Ermöglicht für die Lernenden ein gutes Feedback [207]
- Kalibrierung und Validierung extrem wichtig [200]
- Liefert kumulative Informationen über Leistung und Entwicklung [205]
- Wenn bekannt ist, dass das Portfolio zum Bewerten benutzt wird, wirkt sich das auf den Inhalt und Lernwert aus, da versucht wird, Lehr-Erwartungen zu erfüllen [205]
- Effekt hängt stark von der Unterstützung, Anleitung, der Zeit und dem Feedback der Lehrenden ab [205]

Bei der **360°-Evaluation** oder Multisource Feedback (MSF, Multirater Feedback) handelt es sich um eine arbeitsplatzbasierte Prüfung im klinischen Behandlungsumfeld, wobei für die Beurteilung verschiedene Personengruppen aus dem Arbeitsumfeld des Prüflings involviert werden (Peers, Zahnärzte, Pflegepersonal, Patienten, Administratoren etc.). Den Schwerpunkt der Beobachtung bilden Professionalität und das Arbeiten im Team sowie Handeln als Verantwortungsträger [208], [209]. Diese werden von mehreren Prüfern beobachtet und nach strukturierten Vorgaben bewertet. Der sogenannte „Supervisor“ erhält eine besondere Rolle in diesem Prüfungsszenario: diese Person sammelt alle Ergebnisse und überbringt sie dem jeweiligen Prüfling. Damit bleiben die einzelnen Feedbackgeber anonym. Man erhält ein umfassendes Bild durch die zahlreichen Rückmeldungen und eine hohe Akzeptanz durch die Wahl der Beurteiler. Narrative Kommentare und metrische Bewertungsskalen können miteinander kombiniert werden. Es handelt sich um eine sogenannte multiple event measure.

Anwendung in der Human-/Zahnmedizin

- seit 1970 in der Medizin, weit verbreitet in Nordamerika (Kanada und USA) Europa (England, Holland) und Asien [210], [211].
- beschrieben in der Allgemeinen Medizin, Inneren Medizin, Chirurgie, Gynäkologie, Psychiatrie, Pathologie, Röntgenologie etc. [210].
- in der Zahnmedizin angewendet im Royal College of Surgeons of England, University Bristol, UK Committee of Postgraduate Dental Deans.
- validierte Instrumente für die Evaluation vorhanden (PAR: Physicians Achievement Review, SPRAT: Sheffield Peer Assessment Tool).

Es handelt sich um ein hoch valides und reliables Instrument zur Bewertung der praktischen Kompetenz in komplexen Situationen.

360°-Evaluation**Validität/Gültigkeit**

- kann insbesondere die Evaluierung zwischenmenschlicher und kommunikativer Fähigkeiten erleichtern [212]
- gute Validität [213]

Reliabilität

- Review: Um Cronbachs Alpha-Werte von mindestens 0,9 zu erreichen müssten 8 medizinische Prüfer, 8 nicht-medizinische Prüfer und 25 Patienten beteiligt werden [210]
- Hohe Interne Konsistenz (=0,8) bei fünf Prüfern zu zwei beobachteten Gelegenheiten [214]
- Um Cronbachs Alpha-Werte von 0,8 zu erreichen müssten mindestens 11 Prüfer beteiligt werden [215]
- Cronbachs Alpha-Wert 0,98 [216]
- Problematisch wegen der erforderlichen Anzahl von Prüfern [217]

Akzeptanz

- Auf einer Skala 1-7 seitens der Prüflinge mit 4,5 bewertet [214]
- Auf einer Skala 1-7 seitens der Prüfer mit 5,3 bewertet [214]
- ggf. eine zu positive Einschätzung, da der Anonymisierung nicht getraut wird [217]

Kosten

- Kosten vor der Durchführung berücksichtigen [159]

Durchführbarkeit

- Auf einer Skala 1-7 seitens der Prüflinge mit 4,4 bewertet [214]
- Auf einer Skala 1-7 seitens der Prüfer mit 5,1 bewertet [214]
- In der Regel Überprüfung der Einschätzung mittels Fragebogen, daher einfach durchzuführen [159]

- für eine valide Auswertung wird eine gewisse Anzahl an Bewertungen benötigt, die nicht immer erzielt werden kann [217]
- idealerweise wird das Feedback über einen längeren Zeitraum erhoben [217]
- gut implementierbar, selbst in einem „busy hospital“ [211], [218]

Einfluss auf Lernen und Lehren

- Generelle Verbesserung in der klinischen Tätigkeit, in der Kommunikation mit Kollegen und Patienten [219]
- Auf einer Skala 1-7 seitens der Prüflinge mit 4,2 bewertet [214]
- Auf einer Skala 1-7 seitens der Prüfer mit 4,4 bewertet [214]
- Verbesserung des Evaluierungsprozesses, Gewinn weitergehender Informationen und verschiedener Perspektiven [217]
- Unterschiedliche Resultate: Verbesserung der Kommunikation und des Verhaltens nach einem 360-Grad-Feedback [220]
- großer Zeitaufwand und keine Verbesserung der Bewertungen als Folge des Feedbacks [221]
- man kann schwache Performer früh identifizieren [218]
- auch von SP an Studierende [222]

Zu den Erfolgsfaktoren gehören 1. eine klare Zieldefinition und 2. die Quelle des Feedbacks. Hierbei spielen die Wahl der Prüfer durch den Prüfling, die Glaubwürdigkeit und die Vertrautheit der Prüfer mit der zu beurteilenden Situation und die Anonymität der Einzelrückmeldungen eine wichtige Rolle. Optimierungsmöglichkeiten bestehen darin, dass 1. ca. fünf Prüfer zu zwei beobachteten Gelegenheiten eingesetzt und dass 2. Train-the-Teacher-Veranstaltungen (hinsichtlich der Fähigkeit konstruktives Feedback geben zu können) durchgeführt werden. Die Kombination der Fremdeinschätzung mit der Eigenbeurteilung des zu Beurteilenden kann hilfreich sein, ebenso wie die finale gemeinsame Festlegung von Lernzielen für die Zukunft inklusive der Besprechung (und Dokumentation) konkreter Lern- und Unterstützungsmöglichkeiten. Eine longitudinale Anwendung wird empfohlen. Auch die Implementierung in diversen Settings (auch in High-Stakes-Examina) ist vorstellbar. Das MSF-Format stellt somit ein sinnvolles Prüfungsformat für die Abbildung von praktischen Kompetenzen in komplexen Situationen im zahnärztlichen Ausbildungsszenario dar.

5. Schlussfolgerung

Das Repertoire der hier in der Übersicht aufgeführten Prüfungsmethoden erweitert das Spektrum der bereits etablierten fakultätsinternen Prüfungen – zumeist die MCQ-Fragen und (strukturierte) mündliche Prüfungen – erheblich. Jede der vorgestellten Methoden deckt unterschiedliche Anforderungen und somit verschiedene Kompetenzlevel ab. Dies ist besonders von denen zu beachten, die sich mit der Planung, Durchführung und

Auswertung der Prüfungen im Fach Zahnmedizin beschäftigen.

In der Planung und Umsetzung eines Curriculums ist nicht nur die Wahl des Prüfungsformates entscheidend, sondern auch die allgemeinen Funktionen einer Prüfung zu beachten, die wiederum einen Einfluss auf das Curriculum haben [223]: Prüfungen können summativ oder formativ eingesetzt werden. Summative Prüfungen werden meist am Ende einer Lehrveranstaltung bzw. nach der Vermittlung einer Kompetenz eingesetzt, um den Lernerfolg zu bewerten. Bei formativen Prüfungen steht eine reflexive Begleitung des Lernprozesses im Vordergrund und entscheidet nicht über Bestehen oder Durchfallen in einer Lehrveranstaltung bzw. Kompetenzvermittlung. Eine solche Prüfung spiegelt den Studierenden ihren aktuellen Wissensstand wider und soll den Lernprozess in der reflexiven Ermittlung von eigenen Schwächen unterstützen. Rein formative Prüfungen sind in Zeiten dünner Personaldecken und geringer zeitlicher Ressourcen leider nur bedingt anzutreffen, sind aber ein ideales Mittel, den Lernprozess zu fördern.

Im Rahmen der Erstellung des NKLZ wurde deutlich, dass in Zukunft neben den etablierten Prüfungsmethoden, wie mündliche Prüfungen oder MC-Klausuren, auch andere Prüfungsformate benötigt werden, die insbesondere die in der Zahnmedizin erforderlichen praktischen Fertigkeiten sowohl im Skills-Lab als auch in der Patientenbehandlung erfassen. Dabei sollte sich die jeweilige Prüfungsform an der angestrebten Kompetenzebene orientieren.

Die in dieser Übersichtsarbeit dargestellten Prüfungsformen ermöglichen einen schnellen Einstieg in die jeweilige Methode und verweisen auf weiterführende Literatur für den Interessierten, da eine ausführliche Darstellung jeder einzelnen Prüfungsform den Charakter dieser Übersichtsarbeit sprengen würde. Neben dem theoretischen Wissen um eine Prüfungsform, ist auch der direkte Erfahrungsaustausch mit Kollegen der eigenen oder anderen Fakultäten wichtig, die die entsprechende Methode bereits anwenden, um weitere Informationen zu erhalten. Von daher ist es wünschenswert und vielleicht auch eine Aufgabe der AG Prüfungen, ein Netzwerk von Kollegen aufzubauen, die Erfahrungen mit speziellen Prüfungsformen bereits gesammelt haben und als Ansprechpartner bei Fragen gerne zur Verfügung stehen. Aus diesem Netzwerk heraus könnten je nach Bedarf auch Fortbildungsangebote entwickelt werden, die bei der Umsetzung und Implementierung einer neuen Prüfungsform eine große Hilfe sein können.

6. Ausblick

Mit der zu erwartenden neuen Approbationsordnung für Zahnärzte wird die zahnmedizinische enger mit der medizinischen Lehre verknüpft und auf ein zeitgemäßes Niveau gebracht. Die im NKLZ beispielhaft genannten und in dieser Publikation ausgeführten Prüfungsmethoden zeigen die verschiedenen Möglichkeiten auf Kompetenzlevel zu prüfen. Nach einer gewissen Zeit des Erfahrungs-

sammeln mit universitären Prüfungen in der zahnmedizinischen Lehre und ihrer wissenschaftlichen Überprüfung sollen zusätzliche geeignete Prüfungsmethoden auch in die Vorgaben der zahnärztlichen Approbationsordnung aufgenommen werden und somit auch die Qualität der staatlichen Prüfungen optimieren.

Zusammen mit der Einführung des NKLZ ist das Sammeln von Erfahrungen bei der Organisation, Vorbereitung, Durchführung und Nachbereitung der vorgestellten Prüfungsformen für die nächsten Jahre eine wichtige Aufgabe, wobei sich hier die Zahnmedizin die bereits seit 2002 entwickelten Kompetenzen der Humanmediziner zu Nutze machen kann. Auf der anderen Seite kann sich die Zahnmedizin aber mit ihrem jahrelangen Wissen in der Überprüfung der praktischen Kompetenzen in den Austausch und entsprechende Erfahrungen in die Entwicklung einbringen. Unser gemeinsames Ziel sollte es sein, Prüfungsformate für die verschiedenen Kompetenzlevel für Zahn- und Humanmedizin in Zusammenarbeit der medizinischen Fakultäten weiterzuentwickeln.

Danksagung

Die Autoren danken allen Mitwirkenden, die mitgeholfen haben, dieses Review zu initiieren, zu erstellen, zu revidieren und zu finalisieren. Besonderer Dank gebührt dem Vorstand der AKWLZ, namentlich Frau Prof. P. Hahn, MME (Universität Freiburg) und Herrn Prof. H.-J. Wenz, MME (Universität Kiel), für die detaillierten Rückmeldungen und Verbesserungsvorschläge zum vorliegenden Manuskript.

Interessenkonflikt

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

Autorenschaft

Die Nennung der Autoren erfolgt in alphabetischer Reihenfolge.

Literatur

- Biggs J. Enhancing teaching through constructive alignment. *High Educ.* 1996;32(3):347–364. DOI: 10.1007/BF00138871
- van der Vleuten CP, Verwijnen GM, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 1996;18(2):103–109. DOI: 10.3109/01421599609034142
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206–214. DOI: 10.3109/0142159X.2011.551559
- Chenot JF, Ehrhardt M. Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med.* 2003;79:1–7.
- Examination and Assessments: Academic Integrity [Internet]. Imperial College London. [cited 2015 Jan 17]. Zugänglich unter/available from: <https://workspace.imperial.ac.uk/registry/Public/Procedures%20and%20Regulations/Policies%20and%20Procedures/Examination%20and%20Assessment%20Academic%20Integrity.pdf>
- Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. *GMS Z Med Ausbild.* 2014;31(3):Doc34. DOI: 10.3205/zma000926
- Nationaler Kompetenzbasierter Lernzielkatalog Zahnmedizin (NKLZ) [Internet]. [cited 2016 June 14]. Zugänglich unter/available from: http://www.nklz.de/files/nklz_katalog_20150706.pdf
- Möltner A, Schultz JH, Briem S, Böker T, Schellberg D, Jünger J. Grundlegende testtheoretische Auswertungen medizinischer Prüfungsaufgaben und ihre Verwendung bei der Aufgabenrevision. *GMS Z Med Ausbild.* 2005;22(4):Doc138. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000138.shtml>
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med Educ.* 1985;19(3):238–247. DOI: 10.1111/j.1365-2923.1985.tb01314.x
- Roloff S. Mündliche Prüfungen [Internet]. 1 p. [cited 2016 June 14]. Zugänglich unter/available from: http://www.hochschuldidaktik.net/documents_public/20121127-Roloff-MuendPruef.pdf
- Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian.* 2005;12(1):19–24. DOI: 10.1016/S1322-7696(08)60478-3
- Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Adv Health Sci Educ.* 2010;15(2):277–289. DOI: 10.1007/s10459-008-9111-9
- Harden RM, Lever R, Wilson GM. Two systems of marking objective examination questions. *Lancet.* 1969;293(7584):40–42. DOI: 10.1016/S0140-6736(69)90999-4
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ.* 1975;1(5955):447–451. DOI: 10.1136/bmj.1.5955.447
- Lennox B. Marking multiple-choice examinations. *Br J Med Educ.* 1967;1(3):203–211. DOI: 10.1111/j.1365-2923.1967.tb01698.x
- McCarthy WH. An assessment of the influence of cueing items in objective examinations. *J Med Ed.* 1966;41(3):263–266. DOI: 10.1097/00001888-196603000-00010
- Hart IR, Competence OCOAC, Harden RM, Centre RCOPASOCSMEAR, médecins et chirurgiens du Canada des CR. Further Developments in Assessing Clinical Competence. Boston: Can-Heal Publications; 1987.
- Van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol.* 2010;24(6):703–719. DOI: 10.1016/j.bpobgyn.2010.04.001
- Schoonheim-Klein ME, Habets LL, Aartman IH, van der Vleuten CP, Hoogstraten J, van der Velden U. Implementing an Objective Structured Clinical Examination (OSCE) in dental education: effects on students' learning strategies. *Eur J Dent Educ.* 2006;10(4):226–235. DOI: 10.1111/j.1600-0579.2006.00421.x

20. Fischer MR, Holzer M, Jünger J. Prüfungen an den medizinischen Fakultäten - Qualität, Verantwortung und Perspektiven. *GMS Z Med Ausbild.* 2010;27(5):Doc66. DOI: 10.3205/zma000703
21. Cobb KA, Brown G, Jaarsma DADC, Hammond RA. The educational impact of assessment: a comparison of DOPS and MCQs. *Med Teach.* 2013;35(11):e1598-1607.
22. Elmer A, Grifka J. Vergleich von Prüfungsmethoden in der klinischen Ausbildung. *Gesundheitswesen (Suppl Med Ausbild).* 1998;15(Suppl1):14-17. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/1/Elmer_A,_Grifka_J.pdf
23. Sadaf S, Khan S, Ali SK. Tips for developing a valid and reliable bank of multiple choice questions (MCQs). *Educ Health.* 2012;25(3):195-197. DOI: 10.4103/1357-6283.109786
24. Wenzel A, Kirkevang L. Students' attitudes to digital radiography and measurement accuracy of two digital systems in connection with root canal treatment. *Eur J Dent Educ.* 2004;8(4):167-171. DOI: 10.1111/j.1600-0579.2004.00347.x
25. Yang JC, Laube DW. Improvement of reliability of an oral examination by a structured evaluation instrument. *J Med Educ.* 1983;58(11):864-872.
26. Hottinger U, Krebs R, Hofer R, Feller S, Bloch R. Strukturierte mündliche Prüfung für die ärztliche Schlussprüfung - Entwicklung und Erprobung im Rahmen eines Pilotprojekts. Bern: Universität Bern; 2004.
27. Wass V, Wakeford R, Neighbour R, van der Vleuten C, Royal College of General Practitioners. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ.* 2003;37(2):126-131. DOI: 10.1046/j.1365-2923.2003.01417.x
28. Schubert A, Tetzlaff JE, Tan M, Ryckman JV, Mascha E. Consistency, inter-rater reliability, and validity of 441 consecutive mock oral examinations in anesthesiology: implications for use as a tool for assessment of residents. *Anesthesiology.* 1999;91(1):288-298. DOI: 10.1097/0000542-199907000-00037
29. Kearney RA, Puchalski SA, Yang HYH, Skakun EN. The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Can J Anaesth.* 2002;49(3):232-236. DOI: 10.1007/BF03020520
30. Postgraduate Medical Education and Training Board. *Developing and Maintaining an Assessment System.* London: General Medical Council; 2007. S.1.
31. van der Vleuten CP. Assessment of the Future [Internet]. [cited 2016 June 14]. Zugänglich unter/available from: <http://www.youtube.com/watch?v=bvFmTRVjIE>
32. Möltner A, Schellberg D, Briem S, Böker T, Schultz JH, Jünger J. Wo Cronbachs alpha nicht mehr reicht. *GMS Z Med Ausbild.* 2005;22(4):Doc137. Zugänglich unter/available from: <http://www.egms.de/de/journals/zma/2005-22/zma000137.shtml>
33. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63-67. DOI: 10.1097/00001888-199009000-00045
34. Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild.* 2006;23(3):Doc50. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000269.shtml>
35. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949. DOI: 10.1016/S0140-6736(00)04221-5
36. Knox J. What is.... a Modified Essay Question? *Med Teach.* 1989;11(1):51-57. DOI: 10.3109/01421598909146276
37. Knox JD, Bouchier IA. Communication skills teaching, learning and assessment. *Med Educ.* 1985;19(4):285-289. DOI: 10.1111/j.1365-2923.1985.tb01322.x
38. Feletti GI. Reliability and validity studies on modified essay questions. *J Med Educ.* 1980;55(11):933-941. DOI: 10.1097/00001888-198011000-00006
39. Rabinowitz HK, Hojat M. A comparison of the modified essay question and multiple choice question formats: their relationship to clinical performance. *Fam Med.* 1989;21(5):364-367.
40. Lockie C, McAleer S, Mulholland H, Neighbour R, Tombleson P. Modified essay question (MEQ) paper: perestroika. *Occas Pap R Coll Gen Pract.* 1990;(46):18-22.
41. Feletti GI, Smith EK. Modified essay questions: are they worth the effort? *Med Educ.* 1986;20(2):126-132. DOI: 10.1111/j.1365-2923.1986.tb01059.x
42. van Bruggen L, Manrique-van Woudenberg M, Spierenburg E, Vos J. Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect Med Educ.* 2012;1(4):162-171. DOI: 10.1007/s40037-012-0024-1
43. Irwin WG, Bamber JH. The cognitive structure of the modified essay question. *Med Educ.* 1982;16(6):326-331. DOI: 10.1111/j.1365-2923.1982.tb00945.x
44. Weinman J. A modified essay question evaluation of pre-clinical teaching of communication skills. *Med Educ.* 1984;18(3):164-167. DOI: 10.1111/j.1365-2923.1984.tb00998.x
45. Khan MU, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci.* 2011;5(1):39-43.
46. Bodkha P. Effectiveness of MCQ, SAQ and MEQ in assessing cognitive domain among high and low achievers. *IJRRMS.* 2012;2(4):25-28.
47. Wallerstedt S, Erickson G, Wallerstedt SM. Short Answer Questions or Modified Essay questions - More Than a Technical Issue. *Int J Clin Med.* 2012;3:28. DOI: 10.4236/ijcm.2012.31005
48. Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. *Acad Med.* 1993;68(4):244-249. DOI: 10.1097/00001888-199304000-00002
49. Ferguson KJ. Beyond multiple-choice questions: Using case-based learning patient questions to assess clinical reasoning. *Med Educ.* 2006;40(11):1143. DOI: 10.1111/j.1365-2929.2006.02592.x
50. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ.* 2007;7(1):49. DOI: 10.1186/1472-6920-7-49
51. Wild D, Rützler M, Haarhaus M, Peters K. Der Modified Essay Question (MEQ)-Test an der medizinischen Fakultät der Universität Witten/Herdecke. *Gesundheitswesen (Suppl Med Ausbild).* 1998;15(Suppl2):65-69. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/2/Wild_D,_R%C3%BCtzler_M,_Haarhaus_M,_Peters_K.pdf
52. Peters K, Scheible CM, Rützler M. MEQ - angemessen und praktikabel? Jahrestagung der Gesellschaft für Medizinische Ausbildung - GMA. Köln, 10.-12.11.2006. Düsseldorf, Köln: German Medical Science; 2006. Doc06gma085. Zugänglich unter/available from <http://www.egms.de/en/meetings/gma2006/06gma085.shtml>
53. O'Neill PN. Assessment of students in a problem-based learning curriculum. *J Dent Educ.* 1998;62(9):640-643.

54. Geerlings G, van de Poel AC. De gestructureerde open Vraag: Een Mogelijkheid tot Patientensimulatie binnen Hetonderwijs in de Endodontologie [The modified essay question: a possibility for patient simulation in endodontic education]. *Ned Tijdschr Tandheelkd.* 1984;91(7-8):305–308.
55. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309–317. DOI: 10.1111/j.1365-2929.2005.02094.x
56. Schwartz RW, Donnelly MB, Sloan DA, Young B. Knowledge gain in a problem-based surgery clerkship. *Acad Med.* 1994;69(2):148–151. DOI: 10.1097/00001888-199402000-00022
57. Rabinowitz HK. The modified essay question: an evaluation of its use in a family medicine clerkship. *Med Educ.* 1987;21(2):114–118. DOI: 10.1111/j.1365-2923.1987.tb00676.x
58. Stratford P, Pierce-Fenn H. Modified essay question. *Phys Ther.* 1985;65(7):1075–1079.
59. Norman GR, Smith EK, Powles AC, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Med Educ.* 1987;21(4):297–304. DOI: 10.1111/j.1365-2923.1987.tb00367.x
60. Bloch R, Hofer D, Krebs R, Schläppi P, Weis S, Westkämper R. Kompetent prüfen. Handbuch zur Planung, Durchführung und Auswertung von Facharztprüfungen Medizinische Fakultät Universität Bern, Wien: Institut für Aus-, Weiter- und Fortbildung; 1999.
61. Lim EC, Seet RC, Oh VM, Chia BL, Aw M, Quak SH, Onk BK. Computer-based testing of the modified essay question: the Singapore experience. *Med Teach.* 2007;29(9):e261–268. DOI: 10.1080/01421590701691403
62. Palmer EJ, Devitt PG. A method for creating interactive content for the iPod, and its potential use as a learning tool: Technical Advances. *BMC Med Educ.* 2007;7(1):32. DOI: 10.1186/1472-6920-7-32
63. Bordage G, Brailovsky C, Carretier H, Page G. Content validation of key features on a national examination of clinical decision-making skills. *Acad Med.* 1995;70(4):276–281. DOI: 10.1097/00001888-199504000-00010
64. Bordage G, Page G. An alternative approach to PMPs: The "key features" concept. In: Hart IR, Harden RM (Hrsg). *Further developments in assessing clinical competence.* Montreal: Can-Heal; 1987. S.59–75.
65. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med.* 1995;70(3):194–201. DOI: 10.1097/00001888-199503000-00009
66. Hatala R, Norman GR. Adapting the Key Features Examination for a clinical clerkship. *Med Educ.* 2002;36(2):160–165. DOI: 10.1046/j.1365-2923.2002.01067.x
67. Trudel JL, Bordage G, Downing SM. Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. *Ann Surg.* 2008;248(2):252–258. DOI: 10.1097/SLA.0b013e31818233d3
68. Ali SK, Bordage G. Validity of key features for a family medicine pilot exam at the College of Physicians and Surgeons Pakistan. *J Coll Phys Surg Pakistan.* 1995;5(6):256–260.
69. Bernhardt J, Griesbacher T, Ithaler D, Kresse A, Öttl K, Roller-Wirnsberger R, Vogl S. *Kürzübersicht gängiger Prüfungsformate.* Graz: Medizinische Universität Graz; 2012. Zugänglich unter/available from: http://www.medunigraz.at/fileadmin/lehren/planen-organisieren/pdf/Kurzuebersicht_Pruefungsformate_v7.pdf
70. Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J. A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach.* 2005;27(5):450–455. DOI: 10.1080/01421590500078471
71. Nikendei C, Mennin S, Weyrich P, Kraus B. Effects of a supplementary final year curriculum on students' clinical reasoning skills as assessed by key-feature examination. *Med Teach.* 2009;31(9):e438-443. DOI: 10.1080/01421590902845873
72. Rotthoff T, Baehring T, Dicken H-D, Fahrion U, Richter B, Fischer MR, Scherbaum WA. Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial. *BMC Med Educ.* 2006;6(1):50. DOI: 10.1186/1472-6920-6-50
73. Zupanic M, Iblher P, Töpfer J, Gartmeier M, Bauer J, Prenzel M, Möller G, Hoppe-Seyler T, Karsten G, Fischer MR. *Key Feature-Assessment kommunikativer Leistungen: Weiterent-wicklung und quantitative Evaluation Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA).* München, 05.-08.10.2011. German Medical Science GMS Publishing House; 2011. Doc11gma024. DOI: 10.3205/11gma024
74. Schaper E, Tipold A, Ehlers JP. Use of key feature questions in summative assessment of veterinary medicine students. *Ir Vet J.* 2013;66(1):3. DOI: 10.1186/2046-0481-66-3
75. TU Dresden. *Studienordnung für den Studiengang Zahnmedizin vom 08.09.2011* [Internet]. Dresden; 2011 [cited 2015 Feb 24]. Available from: https://tu-dresden.de/die_tu_dresden/fakultaeten/medizinische_fakultaet/studium/zm/recht-zm/index_html#StudOrgZM
76. Gerhard-Szep S, Hahn P. *Key feature - Fallerstellung (Master of Medical Education, Modul V).* Heidelberg: Universität Heidelberg; 2008. Zugänglich unter/available from: http://archiv.ub.uni-heidelberg.de/volltextserver/11537/1/Manuscript_KF.pdf
77. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ.* 2006;40(7):618–623. DOI: 10.1111/j.1365-2929.2006.02511.x
78. Huwendiek S, Mennin SP, Nikendei C. Medical education after the Flexner report. *N Engl J Med.* 2007;356(1):90. DOI: 10.1056/NEJMc062922
79. Schuwirth LW, van der Vleuten CP, de Kock CA, Peperkamp AG, Donkers HH. Computerized case-based testing: A modern method to assess clinical decision making. *Med Teach.* 1996;18(4):294–299. DOI: 10.3109/01421599609034180
80. Huwendiek S, Reichert F, Brass K, Bosse H-M, Heid J, Möltner A, Haag M, Leven FJ, Hoffmann GF, Jünger J, Tönshoff B. *Etablierung von fallbasiertem computerunterstütztem Prüfen mit langen Auswahl Listen: Ein geeignetes Instrument zur Prüfung von Anwendungswissen.* GMS Z Med Ausbild. 2007;24(1):Doc51. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000345.shtml>
81. Huwendiek S, Heid J, Möltner A, Haag M, Tönshoff B. *E-Learning und E-Prüfung mit virtuellen Patienten in der Medizin.* Heidelberg: Universitätsklinikum Heidelberg; 2008. Zugänglich unter/available from: http://www.his-he.de/veranstaltung/dokumentation/Workshop_E-Pruefung/pdf/TOPO4.pdf
82. Ananthakrishnan N. Microteaching as a vehicle of teacher training—its advantages and disadvantages. *J Postgrad Med.* 1993;39(3):142–143.
83. Arnold RC, Walmsley AD. The use of the OSCE in postgraduate education. *Eur J Dent Educ.* 2008;12(3):126–130. DOI: 10.1111/j.1600-0579.2008.00469.x
84. Taguchi N, Ogawa T. OSCEs in Japanese postgraduate clinical training Hiroshima experience 2000-2009. *Eur J Dent Educ.* 2010;14(4):203–209. DOI: 10.1111/j.1600-0579.2009.00610.x

85. Pugh D, Touchie C, Wood TJ, Humphrey-Murto S. Progress testing: is there a role for the OSCE? *Med Educ.* 2014;48(6):623–631. DOI: 10.1111/medu.12423
86. Curtis DA, Lind SL, Brear S, Finzen FC. The correlation of student performance in preclinical and clinical prosthodontic assessments. *J Dent Educ.* 2007;71(3):365–372.
87. Eberhard L, Hassel A, Bäumer A, Becker F, Beck-Mußotter J, Bömicke W, Corcodel N, Cosgarea R, Eiffler C, Giannakopoulos NN; Kraus T, Mahabadi J, Rues S, Schmitter M, Wolff D, Wege KC. Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. *Eur J Dent Educ.* 2011;15(3):172–178. DOI: 10.1111/j.1600-0579.2010.00653.x
88. Graham R, Bitzer LA, Anderson OR. Reliability and Predictive Validity of a Comprehensive Preclinical OSCE in Dental Education. *J Dent Educ.* 2013;77(2):161–167.
89. Fields H, Rowland M, Vig K, Huja S. Objective structured clinical examination use in advanced orthodontic dental education. *Am J Orthod Dentofacial Orthop.* 2007;131(5):656–663. DOI: 10.1016/j.ajodo.2007.01.013
90. Derringer KA. Undergraduate orthodontic assessment and examination in UK dental schools. *Br Dent J.* 2006;201(4):225–229. DOI: 10.1038/sj.bdj.4813884
91. Macluskey M, Durham J, Balmer C, Bell A, Cowpe J, Dawson L, et al. Dental student suturing skills: a multicentre trial of a checklist-based assessment. *Eur J Dent Educ.* 2011;15(4):244–249. DOI: 10.1111/j.1600-0579.2010.00665.x
92. Hoefler SH, Schuebel F, Sader R, Landes C. Development and implementation of an objective structured clinical examination (OSCE) in CMF-surgery for dental students. *J Craniomaxillofac Surg.* 2013;41(5):412–416. DOI: 10.1016/j.jcms.2012.11.007
93. Landes CA, Hoefler S, Schuebel F, Ballon A, Teiler A, Tran A, Weber R, Walcher F, Sader R. Long-term prospective teaching effectivity of practical skills training and a first OSCE in Cranio Maxillofacial Surgery for dental students. *J Craniomaxillofac Surg.* 2014;42(5):e97–e104. DOI: 10.1016/j.jcms.2013.07.004
94. Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self- and peer-assessment. *Eur J Dent Educ.* 2008;12(1):2–7.
95. Kupke J, Wicht MJ, Stützer H, Derman SH, Lichtenstein NV, Noack MJ. Does the use of a visualised decision board by undergraduate students during shared decision-making enhance patients' knowledge and satisfaction? – A randomised controlled trial. *Eur J Dent Educ.* 2012;17(1):19–25. DOI: 10.1111/eje.12002
96. Hammad M, Oweis Y, Taha S, Hattar S, Madarati A, Kadim F. Students' Opinions and Attitudes After Performing a Dental OSCE for the First Time: A Jordanian Experience. *J Dent Educ.* 2013;77(1):99–104.
97. Mossey PA, Newton JP, Stirrups DR. Scope of the OSCE in the assessment of clinical skills in dentistry. *Br Dent J.* 2001;190(6):323–326. DOI: 10.1038/sj.bdj.4800961
98. Boone WJ, McWhorter AG, Seale NS. Purposeful assessment techniques (PAT) applied to an OSCE-based measurement of competencies in a pediatric dentistry curriculum. *J Dent Educ.* 2001;65(11):1232–1237.
99. Lele SM. A Mini-OSCE for Formative Assessment of Diagnostic and Radiographic Skills at a Dental College in India. *J Dent Educ.* 2011;75(12):1583–1589.
100. Schoonheim-Klein M, Walmsley AD, Habets L. An implementation strategy for introducing an OSCE into a dental school. *Eur J Dent Educ.* 2005;9(4):143–149. DOI: 10.1111/j.1600-0579.2005.00379.x
101. Licari FW, Knight GW. Developing a group practice comprehensive care education curriculum. *J Dent Educ.* 2003;67(12):1312–1315.
102. Ratzmann A, Wiesmann U, Kordaß B. Integration of an Objective Structured Clinical Examination (OSCE) into the dental preliminary exams. *GMS Z Med Ausbild.* 2012;29(1):Doc09. DOI: 10.3205/zma000779
103. Ogawa T, Taguchi N, Sasahara H. Assessing communication skills for medical interviews in a postgraduate clinical training course at Hiroshima University Dental Hospital. *Eur J Dent Educ.* 2003;7(2):60–65. DOI: 10.1034/j.1600-0579.2002.00273.x
104. Cannick GF, Horowitz AM, Garr DR, Reed SG, Neville BW, Day TA, Woolson RF, Lackland DT. Use of the OSCE to evaluate brief communication skills training for dental students. *J Dent Educ.* 2007;71(9):1203–1209.
105. Dennehy PC, Susarla SM, Karimbux NY. Relationship between dental students' performance on standardized multiple-choice examinations and OSCEs. *J Dent Educ.* 2008;72(5):585–592.
106. 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. DOI: 10.3109/0142159X.2013.818634
107. Deis N, Narciß E, Rahe J, Schüttelpelz-Braun K. Objektive standardisierte praktische Prüfungen zur Messung von praktischen Fertigkeiten und berufsrelevanten Kompetenzen. *Z Gesundheit Sport.* 2012;2(2):25–33.
108. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess.* 2011;15(1):i–xxi–1–162.
109. Brand HS, Schoonheim-Klein M. Is the OSCE more stressful? Examination anxiety and its consequences in different assessment methods in dental education. *Eur J Dent Educ.* 2009;13(3):147–153. DOI: 10.1111/j.1600-0579.2008.00554.x
110. Nikendei C, Jünger J. OSCE-praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild.* 2006;23(3):Doc47. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000266.shtml>
111. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ.* 2011;45(12):1181–1189. DOI: 10.1111/j.1365-2923.2011.04075.x
112. Schoonheim-Klein M, Muijtens A, Muijtens A, Habets L, Manogue M, van der Vleuten C, Hoogstraten J, Van der Velden U. On the reliability of a dental OSCE, using SEM: effect of different days. *Eur J Dent Educ.* 2008;12(3):131–137. DOI: 10.1111/j.1600-0579.2008.00507.x
113. Norcini JJ, Maihoff NA, Day SC, Benson JA. Trends in medical knowledge as assessed by the certifying examination in internal medicine. *JAMA.* 1989;262(17):2402–2404. DOI: 10.1001/jama.1989.03430170064029
114. Hofer M, Jansen M, Soboll S. [Potential improvements in medical education as retrospectively evaluated by candidates for specialist examinations]. *Dtsch med Wochenschr.* 2006;131(8):373–378. DOI: 10.1055/s-2006-932527
115. Gesellschaft für Medizinische Ausbildung, Kompetenzzentrum Prüfungen Baden-Württemberg, Fischer MR. Leitlinie für Fakultäts-interne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. *GMS Z Med Ausbild.* 2008;25(1):Doc74. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml>

116. Davenport ES, Davis JE, Cushing AM, Holsgrove GJ. An innovation in the assessment of future dentists. *Br Dent J*. 1998;184(4):192–195.
117. Rau T, Fegert J, Liebhardt H. How high are the personnel costs for OSCE? A financial report on management aspects. *GMS Z Med Ausbild*. 2011;28(1):Doc13. DOI: 10.3205/zma000725
118. Kropmans TJ, O'Donovan BG, Cunningham D, Murphy AW, Flaherty G, Nestel D, Dunne FPI. An Online Management Information System for Objective Structured Clinical Examinations. *CIS*. 2011;28;5(1):38-48.
119. Barman A. Critiques on the Objective Structured Clinical Examination. *Ann Acad Med Singap*. 2005;34(8):478–482.
120. Stillman PL, Swanson DB, Smee S, Stillman AE, Ebert TH, Emmel VS, Gaslowitz J, Green HL, Hamolsky M, Hatem C, et al. Assessing clinical skills of residents with standardized patients. *Ann Intern Med*. 1986;105(5):762–771. DOI: 10.7326/0003-4819-105-5-762
121. Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. *Fam Med*. 2008;40(8):574–578.
122. Carpenter JL. Cost analysis of objective structured clinical examinations. *Acad Med*. 1995;70(9):828–833. DOI: 10.1097/00001888-199509000-00022
123. Duerson MC, Romrell LJ, Stevens CB. Impacting faculty teaching and student performance: nine years' experience with the Objective Structured Clinical Examination. *Teach Learn Med*. 2000;12(4):176–182. DOI: 10.1207/S15328015TLM1204_3
124. Harden RM, Cairncross RG. Self assessment. *Med Teach*. 1980;2(3):145–148. DOI: 10.3109/01421598009081185
125. Kundu D, Das HN, Sen G, Osta M, Mandal T, Gautam D. Objective structured practical examination in biochemistry: An experience in Medical College, Kolkata. *J Nat Sci Biol Med*. 2013;4(1):103–107. DOI: 10.4103/0976-9668.107268
126. Abraham RR, Raghavendra R, Surekha K, Asha K. A trial of the objective structured practical examination in physiology at Melaka Manipal Medical College, India. *Adv Physiol Educ*. 2009;33(1):21–23. DOI: 10.1152/advan.90108.2008
127. Adome RO, Kitutu F. Creating an OSCE/OSPE in a resource-limited setting. *Med Educ*. 2008;42(5):525–526. DOI: 10.1111/j.1365-2923.2008.03045.x
128. Nayak V, Bairy KL, Adiga S, Shenoy S, Magazine BC, Amberkar M, Kumari MK. OSPE in Pharmacology: Comparison with the conventional Method and Students' Perspective Towards OSPE. *Br Biomed Bull*. 2014;2(1):218-222.
129. Wani P, Dalvi V. Objective Structured Practical Examination vs Traditional Clinical Examination in Human Physiology: Student's perception. *Int J Med Sci Public Health*. 2013;2(3):522–547. DOI: 10.5455/ijmsph.2013.080320133
130. Menezes RG, Nayak VC, Binu VS, Kanchan T, Rao PP, Baral P, Lobo SW. Objective structured practical examination (OSPE) in Forensic Medicine: students' point of view. *J Forensic Leg Med*. 2011;18(8):347–349. DOI: 10.1016/j.jflm.2011.06.011
131. Huth KC, Baumann M, Kollmuss M, Hickel R, Fischer MR, Paschos E. Assessment of practical tasks in the Phantom course of Conservative Dentistry by pre-defined criteria: a comparison between self-assessment by students and assessment by instructors. *Eur J Dent Educ*. 2015. DOI: 10.1111/eje.12176 [Epub ahead of print] DOI: 10.1111/eje.12176
132. Banerjee R, Chandak A, Radke U. Bringing objectivity to assessment in Preclinical Prosthodontics: The student's perspective on OSPE. *JETHS*. 2014;1(2):30-33.
133. Schmitt I, Möltner A, Bärmeier J, Gärtner K, Dopfer S, Kuschel B, Kunkel F, Heidemann D, Gerhardt-Szep S. Wie viele Prüfer braucht ein OSCE? Düsseldorf: AKWLZ; 2013. Zugänglich unter/available from: <https://www.akwlz.de/documents/262773/490615/K8+-+Schmitt+-+Pr%E2%80%99er+OSPE.pdf/dc030606-f303-4e81-b85c-601177f11e0c?version=1.0>
134. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Med Teach*. 2003;25(3):262–270. DOI: 10.1080/0142159031000100300
135. Colliver JA, Barrows HS, Vu NV, Verhulst SJ, Mast TA, Travis TA. Test security in examinations that use standardized-patient cases at one medical school. *Acad Med*. 1991;66(5):279–282. DOI: 10.1097/00001888-199105000-00011
136. Ortwein H, Fröhmel A, Burger W. Einsatz von Simulationspatienten als Lehr-, Lern- und Prüfungsform. *Psychother Psychosom Med Psychol*. 2006;56(01):23–29. DOI: 10.1055/s-2005-867058
137. Collins J, Harden RM. AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Med Teach*. 1998;20(6):508–521. DOI: 10.1080/01421599880210
138. Barrows H. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med*. 1993;68(6):443–451. DOI: 10.1097/00001888-199306000-00002
139. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach*. 2009;31(6):477–486. DOI: 10.1080/01421590903002821
140. Hendrickx K, De Winter B, Tjalma W, Avonts D, Peeraer G, Wyndaele JJ. Learning intimate examinations with simulated patients: The evaluation of medical students' performance. *Med Teach*. 2009;31(4):E139–147. DOI: 10.1080/01421590802516715
141. Norcini JJ. The death of the long case? *BMJ*. 2002;324(7334):408–409. DOI: 10.1136/bmj.324.7334.408
142. Norcini J. The validity of long cases. *Med Educ*. 2001;35(8):720–721. DOI: 10.1046/j.1365-2923.2001.01006.x
143. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. 1995;123(10):795–799. DOI: 10.7326/0003-4819-123-10-199511150-00008
144. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. 2003;138(6):476–481. DOI: 10.7326/0003-4819-138-6-200303180-00012
145. Thornton S. A literature review of the long case and its variants as a method of assessment. *Educ Med J*. 2012;4(1):5–11.
146. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. 2002;77(9):900–904. DOI: 10.1097/00001888-200209000-00020
147. Herbers JE, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. 1989;4(3):202–208. DOI: 10.1007/BF02599524
148. Yousuf N. Mini clinical evaluation exercise: validity and feasibility evidences in literature. *Educ Med J*. 2012;4(1):e100-107.
149. Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Med Educ*. 2009;43(4):326–334. DOI: 10.1111/j.1365-2923.2008.03275.x

150. Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med.* 2003;78(10 Suppl):S33–35. DOI: 10.1097/00001888-200310001-00011
151. Alves De Lima A, Barrero C, Baratta S, Castillo Costa Y, Bortman G, Carabajales J, Conde D, Galli A, Degrange G, Van der Vleuten C. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Med Teach.* 2007;29(8):785–790. DOI: 10.1080/01421590701352261
152. Norcini JJ. The mini clinical evaluation exercise (mini-CEX). *Clin Teach.* 2005;2(1):25–30. DOI: 10.1111/j.1743-498X.2005.00060.x
153. Alves de Lima A. Assessment of clinical competence: Reliability, Validity, Feasibility and Educational Impact of the mini-CEX. Maastricht: Universitat Maastricht; 2013. Zuganglich unter/available from: http://www.icba.com.ar/profesionales/pdf/aal/Thesis_Alberto_Alves_de_Lima_170x240_v10.pdf
154. Cook DA, Beckman TJ. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. *Adv Health Sci Educ.* 2009;14(5):655–664. DOI: 10.1007/s10459-008-9147-x
155. Alves de Lima AE, Conde D, Aldunate L, van der Vleuten CP. Teachers' experiences of the role and function of the mini clinical evaluation exercise in post-graduate training. *Int J Med Educ.* 2010;1:68–73. DOI: 10.5116/ijme.4c87.8e13
156. Berendonk C, Beyeler C, Westkamper R, Giger M. Strukturiertes Feedback in der arztlischen Weiterbildung: Mini-CEX und DOPS. *Schweiz Arzte.* 2008;89(32):1337–1340.
157. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005;80(10 Suppl):S46–54. DOI: 10.1097/00001888-200510001-00015
158. Brazil V, Ratcliffe L, Zhang J, Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department—does cost outweigh value? *Med Teach.* 2012;34(12):1017–1023. DOI: 10.3109/0142159X.2012.719653
159. Magnier KM, Dale VH, Pead MJ. Workplace-based assessment instruments in the health sciences. *J Vet Med Educ.* 2012;39(4):389–395. DOI: 10.3138/jvme.1211-118R
160. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ.* 2008;42(4):364–373. DOI: 10.1111/j.1365-2923.2008.03010.x
161. Prescott-Clements L, van der Vleuten CP, Schuwirth LW, Hurst Y, Rennie JS. Evidence for validity within workplace assessment: the Longitudinal Evaluation of Performance (LEP). *Med Educ.* 2008;42(5):488–495. DOI: 10.1111/j.1365-2923.2007.02965.x
162. Deshpande S, Chahande J. Impact of computer-based treatment planning software on clinical judgment of dental students for planning prosthodontic rehabilitation. *Adv Med Educ Pract.* 2014;5:269–274. DOI: 10.2147/AMEP.S66264
163. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach.* 2006;28(2):117–128. DOI: 10.1080/01421590600622665
164. Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med.* 2004;19(5 Pt 2):558–561. DOI: 10.1111/j.1525-1497.2004.30134.x
165. Montagne S, Rogausch A, Gemperli A, Berendonk C, Jucker-Kupper P, Beyeler C. The mini-clinical evaluation exercise during medical clerkships: are learning needs and learning goals aligned? *Med Educ.* 2014;48(10):1008–1019. DOI: 10.1111/medu.12513
166. Berberat PO, Harendza S, Kadmon M, Gesellschaft fur Medizinische Ausbildung, GMA-Ausschuss fur Weiterbildung. Entrustable professional activities - visualization of competencies in postgraduate training. Position paper of the Committee on Postgraduate Medical Training of the German Society for Medical Education (GMA). *GMS Z Med Ausbild.* 2013;30(4):Doc47. DOI: 10.3205/zma000890
167. Colleges AOAM. Core Entrustable Professional Activities for Entering Residency. [Internet] [cited 2016 June 14] 2014. Zuganglich unter/available from: <https://members.aamc.org/eweb/upload/Core%20EPA%20Curriculum%20Dev%20Guide.pdf>
168. Ten Cate O. Trusting graduates to enter residency: what does it take? *J Grad Med Educ.* 2014;6(1):7–10. DOI: 10.4300/JGME-D-13-00436.1
169. Ten Cate O, Snell L, Carraccio C. Medical competence: the interplay between individual ability and the health care environment. *Med Teach.* 2010;32(8):669–675. DOI: 10.3109/0142159X.2010.500897
170. Jones MD, Rosenberg AA, Gilhooly JT, Carraccio CL. Perspective: Competencies, Outcomes, and Controversy—Linking Professional Activities to Competencies to Improve Resident Education and Practice. *Acad Med.* 2011;86(2):161–165. DOI: 10.1097/ACM.0b013e31820442e9
171. Mulder H, Ten Cate O, Daalder R, Berkvens J. Building a competency-based workplace curriculum around entrustable professional activities: The case of physician assistant training. *Med Teach.* 2010;32(10):e453–459. DOI: 10.3109/0142159X.2010.513719
172. Chang A, Bowen JL, Buranosky RA, Frankel RM, Ghosh N, Rosenblum MJ, Thompson S, Green ML. Transforming primary care training—patient-centered medical home entrustable professional activities for internal medicine residents. *J Gen Intern Med.* 2013;28(6):801–809. DOI: 10.1007/s11606-012-2193-3
173. Scheele F, Teunissen P, van Luijk S, Heineman E, Fluit L, Mulder H, Meiningen A, Wijnen-Meijer M, Glas G, Sluiter H, Hummel T. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach.* 2008;30(3):248–253. DOI: 10.1080/01421590801993022
174. Englander R, Carraccio C. From theory to practice: making entrustable professional activities come to life in the context of milestones. *Acad Med.* 2014;89(10):1321–1323. DOI: 10.1097/ACM.0000000000000324
175. Aylward M, Nixon J, Gladding S. An entrustable professional activity (EPA) for handoffs as a model for EPA assessment development. *Acad Med.* 2014;89(10):1335–1340. DOI: 10.1097/ACM.0000000000000317
176. Boyce P, Spratt C, Davies M, McEvoy P. Using entrustable professional activities to guide curriculum development in Psychiatry training. *BMC Med Educ.* 2011;11(1):96. DOI: 10.1186/1472-6920-11-96
177. Ten Cate O. Competency-based education, entrustable professional activities, and the power of language. *J Grad Med Educ.* 2013;5(1):6–7. DOI: 10.4300/JGME-D-12-00381.1
178. Hauer KE, Kohlwes J, Cornett P, Hollander H, Cate ten O, Ranji SR, Soni K, Iobst W, O'Sullivan P. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ.* 2013;5(1):54–59. DOI: 10.4300/JGME-D-12-00060.1

179. Shaughnessy AF, Sparks J, Cohen-Osher M, Goodell KH, Sawin GL, Joseph Gravel J. Entrustable Professional Activities in Family Medicine. *J Grad Med Educ.* 2013;5(1):112–118. DOI: 10.4300/JGME-D-12-00034.1
180. Ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542–547. DOI: 10.1097/ACM.0b013e31805559c7
181. Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). *J Coll Physicians Surg Pak.* 2013;23(1):77–82.
182. Bazrafkan L. Comparison of the Assessment of Dental Students'. *J Med Ed.* 2009;13(1, 2):3-8.
183. Tricio J, Woolford M, Thomas M, Lewis-Greene H, Georgiou L, Andiappan M, Escudier M. Dental students' peer assessment: a prospective pilot study. *Eur J Dent Educ.* 2014;19(3):140-148. DOI: 10.1111/eje.12114
184. Abraham RR, Upadhy S, Torke S, Ramnarayan K. Student perspectives of assessment by TEMM model in physiology. *Adv Physiol Educ.* 2005;29(2):94–97. DOI: 10.1152/advan.00051.2004
185. Barton JR, Corbett S, van der Vleuten CP, English Bowel Cancer Screening Programme, UK Joint Advisory Group for Gastrointestinal Endoscopy. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointest Endosc.* 2012;75(3):591–597. DOI: 10.1016/j.gie.2011.09.053
186. Akbari M, Shamsabadi RM. Direct Observation of Procedural Skills (DOPS) in Restorative Dentistry: Advantages and Disadvantages in Student's Point of View. *Iran J Med Educ.* 2013;13(3):212-220.
187. Andersen RM, Davidson PL, Atchison KA, Hewlett E, Freed JR, Friedman J-A, et al. Pipeline, profession, and practice program: evaluating change in dental education. *J Dent Educ.* 2005;69(2):239–248.
188. Hamdy H, Prasad K, Williams R, Salih FA. Reliability and validity of the direct observation clinical encounter examination (DOCEE). *Med Educ.* 2003;37(3):205–212. DOI: 10.1046/j.1365-2923.2003.01438.x
189. Torre DM, Simpson DE, Elnicki DM, Sebastian JL, Holmboe ES. Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. *Teach Learn Med.* 2007;19(3):271–277. DOI: 10.1080/10401330701366622
190. Cohen SN, Farrant PBJ, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol.* 2009;161(1):34–39. DOI: 10.1111/j.1365-2133.2009.09097.x
191. Center of Innovation in Professional Health Education and Research (CIPHER). Review of work-based assessment methods. Sydney: University of Sydney; 2007.
192. Roghieh N, Fateme H, Hamid S, Hamid H. The effect of formative evaluation using "direct observation of procedural skills" (DOPS) method on the extent of learning practical skills among nursing students in the ICU. *Iran J Nurs Midwifery Res.* 2013;18(4):290–293.
193. Hamilton KES, Coates V, Kelly B, Boore JRP, Cundell JH, Gracey J, et al. Performance assessment in health care providers: a critical review of evidence and current practice. *J Nurs Manag.* 2007;15(8):773–791. DOI: 10.1111/j.1365-2934.2007.00780.x
194. Morris A, Hewitt J, Roberts CM. Practical experience of using directly observed procedures, mini clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgrad Med J.* 2006;82(966):285–288. DOI: 10.1136/pgmj.2005.040477
195. Sh S, Pooladi A, BahramRezaie M, Farhadifar F, Khatibi R. Evaluation of the Effects of Direct Observation of Procedural Skills (DOPS) on clinical externship students' learning level in obstetrics ward of kurdistan university of medical sciences. *J Med Ed.* 2009;13(1):29-33.
196. Stosch C, Wichelhaus AS, Matthes J. Die Portfolio-Methode: Modernes Assessment auf dem Prüfstand. *GMS Z Med Ausbild.* 2006;23(3):Doc43. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000262.shtml>
197. Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, Moreley D, Pollard D, Ashcroft T, Popovic C, Sayers J. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. *BEME Guide No. 11. Med Teach.* 2009;31(4):282–298. DOI: 10.1080/01421590902889897
198. Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Med Teach.* 2009;31(4):299–318. DOI: 10.1080/01421590902883056
199. Pocock I. A new route for dental graduates. *Dent Update.* 2007;34(1):59.
200. Kramer GA, Albino JEN, Andrieu SC, Hendricson WD, Henson L, Horn BD, Neumann LM, Young SK. Dental student assessment toolbox. *J Dent Educ.* 2009;73(1):12–35.
201. Gadbury-Amyot CC, McCracken MS, Woldt JL, Brennan RL. Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *J Dent Educ.* 2014;78(5):657–667.
202. Michels NR, Driessen EW, Muijtjens AM, Van Gaal LF, Bossaert LL, de Winter BY. Portfolio assessment during medical internships: How to obtain a reliable and feasible assessment procedure? *Educ Health.* 2009;22(3):313.
203. O'sullivan PS, Reckase MD, McClain T, Savidge MA, Clardy JA. Demonstration of portfolios to assess competency of residents. *Adv Health Sci Educ.* 2004;9(4):309–323. DOI: 10.1007/s10459-004-0885-0
204. Melville C, Rees M, Brookfield D, Anderson J. Portfolios for assessment of paediatric specialist registrars. *Med Educ.* 2004;38(10):1117–1125. DOI: 10.1111/j.1365-2929.2004.01961.x
205. McMullan M. Students' perceptions on the use of portfolios in pre-registration nursing education: a questionnaire survey. *Int J Nurs Stud.* 2006;43(3):333–343. DOI: 10.1016/j.ijnurstu.2005.05.005
206. Burch VC, Seggie JL. Use of a structured interview to assess portfolio-based learning. *Med Educ.* 2008;42(9):894–900. DOI: 10.1111/j.1365-2923.2008.03128.x
207. Kadagad P, Kotrashetti SM. Portfolio: a comprehensive method of assessment for postgraduates in oral and maxillofacial surgery. *J Maxillofac Oral Surg.* 2013;12(1):80–84. DOI: 10.1007/s12663-012-0381-7
208. Brett JF, Atwater LE. 360 degree feedback: accuracy, reactions, and perceptions of usefulness. *J Appl Psychol.* 2001;86(5):930–942. DOI: 10.1037/0021-9010.86.5.930
209. Lepsinger R, Lucia AD. *The Art and Science of 360 Degree Feedback.* New York: John Wiley & Sons; 2009.

210. Donnon T, Ansari AI A, Alawi AI S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med.* 2014;89(3):511–516. DOI: 10.1097/ACM.000000000000147
211. Zhao Y, Zhang X, Chang Q, Sun B. Psychometric characteristics of the 360° feedback scales in professionalism and interpersonal and communication skills assessment of surgery residents in China. *J Surg Educ.* 2013;70(5):628–635. DOI: 10.1016/j.jsurg.2013.04.004
212. Joshi R, Ling FW, Jaeger J. Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Acad Med.* 2004;79(5):458–463. DOI: 10.1097/00001888-200405000-00017
213. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ.* 2005;330(7502):1251–1253. DOI: 10.1136/bmj.38447.610451.8F
214. Murphy DJ, Bruce DA, Mercer SW, Eva KW. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Adv Health Sci Educ.* 2009;14(2):219–232. DOI: 10.1007/s10459-008-9104-8
215. Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med.* 1993;68(9):680–687. DOI: 10.1097/00001888-199309000-00014
216. Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. *Can J Psychiatry.* 2008;53(8):525–533.
217. Chandler N, Henderson G, Park B, Byerley J, Brown WD, Steiner MJ. Use of a 360-degree evaluation in the outpatient setting: the usefulness of nurse, faculty, patient/family, and resident self-evaluation. *J Grad Med Educ.* 2010;2(3):430–434. DOI: 10.4300/JGME-D-10-00013.1
218. Hesketh EA, Anderson F, Bagnall GM, Driver CP, Johnston DA, Marshall D, Needham G, Orr G, Walker K. Using a 360 degrees diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Med Teach.* 2005;27(3):219–233. DOI: 10.1080/01421590500098776
219. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Med Educ.* 2014;14(1):76. DOI: 10.1186/1472-6920-14-76
220. Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, Dewitt TG, Britto MT. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med.* 2007;161(1):44–49. DOI: 10.1001/archpedi.161.1.44
221. Weigelt JA, Brasel KJ, Bragg D, Simpson D. The 360-degree evaluation: increased work with little return? *Current Surgery.* 2004;61(6):616–626. DOI: 10.1016/j.cursur.2004.06.024
222. Garry A, Stirling K. Achieving 360° student feedback using SPaCE. *Clin Teach.* 2012;9(4):222–227. DOI: 10.1111/j.1743-498X.2012.00550.x
223. Fabry G. *Medizindidaktik: ein Handbuch für die Praxis.* Karlsruhe: Huber; 2008.

Korrespondenzadresse:

PD Dr. med. dent Susanne Gerhard-Szep, MME
Goethe-Universität, Carolinum Zahnärztliches
Universitäts-Institut gGmbH, Poliklinik
Zahnerhaltungskunde, Frankfurt am Main, Deutschland
s.szep@em.uni-frankfurt.de

Bitte zitieren als

Gerhard-Szep S, Güntsch A, Pospiech P, Söhnel A, Scheutzel P, Wassmann T, Zahn T. Assessment formats in dental medicine: An overview. *GMS J Med Educ.* 2016;33(4):Doc65. DOI: 10.3205/zma001064, URN: urn:nbn:de:0183-zma0010641

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2016-33/zma001064.shtml>

Eingereicht: 23.10.2015

Überarbeitet: 24.03.2016

Angenommen: 09.05.2016

Veröffentlicht: 15.08.2016

Copyright

©2016 Gerhard-Szep 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/>.