The use of the objective structured clinical examination (OSCE) in dental education

Klein, M.E.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 8

Summary
This thesis is concerned with the quality of the assessment in dentistry and, in particular, the use of the Objective Structured Clinical Examination (OSCE) in the assessment of student’s clinical competences in dental education. The OSCE will be described later in this chapter.

Issues in the design of clinical assessment and OSCE

In medical and dental education new competency based curricula were implemented with new educational approaches, i.e. Student centered and competency based. In accordance with these new competency based curricula also competency based assessment methods have to be developed. These assessment methods can be used for formative purposes: the student should receive feedback to aid their learning. In contrast to this formative function the assessment method has a summative purpose, when it is used for qualifications to assure that the student has the desired level of competence. Besides for the benefit of the student, the competence based assessment serves also the benefit of the dental school organization and the benefit of the patient.

In the last decades the used methods of clinical assessment in dentistry were found not satisfactory. Some of the shortcomings were subjectivity, no consequent approach, no explicit criteria and only numerical requirements and mere testing of senso-motor skills without testing of the communication and other chairside skills.

To enhance assessment and provide a greater degree of objectivity and consistency and to make the clinical assessment more authentic, Harden in 1975 developed an Objective Structured Clinical Examination (OSCE) and implemented this form of clinical assessment of competence into the medical curriculum. Since then OSCEs have been extensively used in medical education.

In an OSCE, the examinee circulates through a series of test-stations. At each station students are asked to perform a clinical task, which does test elements of clinical competence. In dentistry, such examples include diagnostic, clinical and communication skills, all based on relevant knowledge. To minimize subjective bias and inconsistent approaches, the students all face the same pre-defined clinical tasks, the same time limit (often 5 min) and the same standardized checklists. Their behavior is directly observed and scored by observers. Because a wide range of skills can be tested, it is assumed, that a reliable overall view of the clinical competencies of the students can be obtained. In this way the OSCE can give a measure of clinical competence through observable behaviors. Manikins and simulated patients can be used to allow large numbers of students to be tested on the same clinical problem. Marking can be completed during the OSCE procedure and in formative OSCEs it is possible to provide immediate feedback. Drawbacks of an OSCE can be the risk of trivializing the skills. Although the logistics of the administration and the number of staff members needed as examiners can be expensive and cumbersome, the OSCE could allow testing of competencies that used to be ignored. In summary it can be concluded, that the OSCE seems a good method to test objectively elements of competence with explicit criteria, with a more standardized consequent approach, and with the possibility to test different aspects of communication and chairside skills.
Although no single assessment method can test all the competencies, the OSCE as part of an assessment program seems best used to test the competence of students in the stage where they "show how" they can perform in a simulated environment and how they can apply their basic knowledge and their developing clinical reasoning in this simulated context of the examination. In CHAPTER 1 a model for evaluating the utility of assessment methods is introduced. The utility (U) or value of assessment methods of competence can be captured in 5 criteria: acceptability (A), educational effect (E), validity (V), reliability(R), and costs (C) (U=A x E x V x R x C). The utility of a test depends on the purpose of the test and its properties.

CHAPTER 1 provides a review of de literature on the utility of the OSCEs for different purposes in dental education. The review demonstrated that the evidence on the utility of dental OSCEs is very "skinny" and only a few studies are available that investigated aspects of reliability of dental OSCEs. There is one study with evidence for the validity of a well designed dental OSCE. No evidence was available to test the educational effect of the dental OSCE, the feasibility and costs. Also no summative OSCEs was investigated, with the use of standard setting methods to determine a pass/fail standard.

Therefore the general aim of this thesis was to investigate the utility of the OSCE in the assessment of clinical competence of dental students.

In order to meet the aim the following specific research questions were formulated:
1. Will an implementation strategy based on participation information and commitment be effective and will attitudes change positively towards the OSCE? (Chapter 2, Acceptability)
2. Will the implementation of the OSCE in dental education improve students' learning of clinical competence? (Chapter 3, Educational effect)
3. What is the effect of the language ability of Non-Native students on their performance in the OSCE? (Chapter 4, Validity)
4. Will lengthening of the time of an OSCE station enhance the performance of Non-Native students? (Chapter 4, Validity)
5. Does a dental OSCE administered over multiple days result in reliable passing scores? Is the mean score of the students on day 1, lower than those on subsequent days? (Chapter 5, Reliability)
6. How many stations in a dental OSCE are required for a sufficiently reliable decision about the student's performance? (Chapter 5, Reliability)
7. Which standard setting method is the optimal instrument to prevent incompetent students from passing and competent students from failing a dental OSCE? (Chapter 6, Validity)

The acceptability of the dental OSCE
Since resistance could develop when changes are introduced into an organisation, the use of a strategy for the implementation of such change will help to diminish opposition and may therefore result in the co-operation of staff and their departments. The investigation started by studying the effectiveness of an implementation strategy by measuring attitudes of both staff and students, towards the OSCE as a new form of clinical assessment in a dental school (ACTA). CHAPTER 2
Chapter 8

describes an implementation strategy of the OSCE. “Stepwise” behaviour change (with information, participation and commitment as tools) was used as a strategy to minimize protective behaviour to the introduction of an OSCE. After lectures on assessment, 59 staff members participated in a mini OSCE with 8 test-stations, playing both the role of a student and observer. A questionnaire, designed to test attitudes and commitment towards the new OSCE, was completed after the examination. Six months later, 22 staff of all departments had developed and run a pilot OSCE for 44 students. The same questionnaire was answered by staff and students. A year later, another OSCE for all 103 3rd year students was designed, organised and evaluated with full co-operation of the clinical teaching staff.

It was found that the staff total attitude is growing positively (p=0.001). Student's total attitude is lower than staff (p<0.001) The results of the survey after the mini-staff-OSCE and pilot and final OSCE were favourable in terms of the acceptance of use of an OSCE for the assessment of clinical competences. Therefore it can be concluded that the implementation strategy appears to have been successful. The objective of gaining the co-operation of staff and departments and avoiding resistance to change was achieved.

The educational impact of the dental OSCE

In CHAPTER 3 the effect of the implementation of an OSCE on students’ learning strategies was investigated and also the effect on the competence to manage periodontal diseases in patients was investigated. The implemented OSCE was expected to be superior to the existing Written Exam in fostering the acquisition of clinical competencies in terms of study strategies that are more oriented towards clinical practice, longer study time, greater clinical proficiency, and more realistic self-assessment.

For this study, after a clinical course in periodontology, 72 third year dental students were assessed summatively, either using a Written Examination or an OSCE (P-OSCE). The students were informed beforehand about the assessment formats. The self-assessed clinical competence, study time and strategies (i.e. practice with a manikin, peers, and patient case) were evaluated by means of a questionnaire. After a comprehensive dental care course, all 72 students were assessed by an overall end-of-year OSCE, in which three periodontal stations were included: ‘measuring pockets’, ‘educating patients’ and ‘tracing an X-ray with bone-loss’. The competence of the previous Written Exam group and the P-OSCE group was investigated by determining the mean scores and pass-fail scores of three periodontal test-stations as well as the total score of the end-of-year OSCE. The degree of realistic self-assessment was studied by correlating the self-assessed competencies as evaluated by means of the questionnaire with the total score of the end-of-year OSCE.

Results showed that Self-assessed clinical competence, study time and study strategies showed no differences between the P-OSCE and the Written Exam-group. The clinical competence determined in the test-station ‘measuring pockets’ in the end of year Overall OSCE was higher for the P-OSCE group when compared with the Written Exam group; the two groups performed equally well in the test station ‘educating patients’, whereas the performance in ‘tracing an X-ray with bone-loss’ was better in the Written Exam group. This group also had a higher total score in
the end-of-year OSCE. The degree of realistic self-assessment was higher in the P-OSCE group than in the Written Exam group: in the P-OSCE group the self-assessed clinical competencies correlated significantly with the total score of the Overall End-of-year OSCE.

These results indicate that no effects of the implementation of an OSCE in undergraduate periodontal education were observed in study strategies, but the implementation of an OSCE in undergraduate periodontal education appears to stimulate learning, resulting in greater achievement of specific clinical competence and a greater level of realistic self-assessment.

The native language of students

Chapter 4 describes a bias study as a possible threat to the validity of the OSCE. The purpose of this study was to investigate whether Non-Native dental students perceive drawbacks in their education and examination experience because of their lack of language proficiency and to test whether prolonging OSCE test-station duration improves their performance. Dental students (n=345) completed a questionnaire about their native country, their language background and the possible drawback they perceived in dental education and examination because of their language proficiency. Students were marked as ‘Native’, when they were born in the Netherlands with Dutch as native language or ‘Non-Native’ when they were born outside the Netherlands, raised with a non-Dutch native language, or raised bilingually.

A sample of 108 students was assessed by an OSCE testing a periodontal course with 9 test-stations. Test-station topics were: 1. history taking, 2. measuring attachment level, 3. educating patients, 4. tracing a radiograph, 5. root-planing, 6. writing a prescription 7. diagnostics and prognostics, 8. differential diagnostics, and 9. writing a referral letter. The first five test-stations mentioned were of 5 minutes duration. The other four test-stations were provided in two modes: either with a short (5 min) or longer (10 min) version. Every student took at random two long and two short test-stations.

Results showed that in the group of 345 questionnaire responders, Non-native students (n=116) perceived significantly more drawback in education and examination than Native students (n=229) (p<0.001). When Non-Native students speak Dutch at home, around 38% of them reported perceived drawbacks in education, whereas when they speak their native language at home, around 60% reported perceived drawbacks in education (p=0.005). In the periodontal OSCE (n=108) the Native group (n=70) had significantly higher total scores than the Non-Native (n=38), (p=0.009, d=0.53). The Non-Native group had significantly lower mean scores in the communication station “educating patients” (p=0.034 d=0.42). Prolonged test-station duration from 5 minutes to 10 minutes had no positive effect in all experimental test-stations in the Native and Non-Native group. Female students in the Native group out-performed male in a communication test-stations. Female students in Native and Non-Native groups were found to be more successful in “tracing bone loss on radiographs”.

In summary it was concluded, that the Non-native students perceived a drawback in dental education and examination because of their language proficiency in Dutch, which is confirmed by their actual OSCE performance. Prolonging the time for a test-station did not improve OSCE
performance of Non-Native students. It is recommended that students with problems in language ability get additional tuition and practice.

The Reliability of the dental OSCE

The OSCE for large groups of students was administered per OSCE on four different days of one week. Chapter 5 reports a reliability study. In four OSCE administrations, 463 students of the year 2005 and 2006 took the summative OSCE after a dental course in comprehensive dentistry. The OSCEs had 16-18 five-minutes stations, and was administered per OSCE on four different days of one week.

The first aim of this study was to investigate the reliability of a dental OSCE administered over multiple days in the week and the second aim was to assess the number of test-stations, required for a sufficiently reliable decision in three score interpretation-perspectives of a dental OSCE administered over multiple days. ANOVA was used to test for examinee performance variation across days. Generalizability theory was used for reliability analyses. Reliability was studied from three interpretation perspectives: for relative (norm) decisions, for absolute (domain) and pass-fail (mastery) decisions. As indicator of reproducibility of test scores in this dental OSCE, the Standard Error of Measurement (SEM) was used. The benchmark of SEM was set at < 0.51. This is corresponding to a 95% Confidence Interval (CI) of < 1 on the original scoring scale that ranged from 1-10. Mean weighted total OSCE score appeared 7.14 on a 10-point scale. With the pass-fail score set at 6.2 for the four OSCEs, 90% of the 463 students passed. There was no significant increase in scores over the different days the OSCE was administered. “Wished” variance due to students was 6.3% .Variance due to interaction between student and stations and residual error was 66.3%, more than two times larger than variance due to station’s difficulty (27.4%). The SEM-norm was 0.42 with a 95% CI of ± 0.83 and the SEM-domain was 0.50, with a CI of ±0.98. In order to make reliable relative decisions (SEM <0.51) the use of minimal 12 stations is necessary, and for reliable absolute and pass-fail decisions, the use of minimal 17 stations is necessary in this dental OSCE.

It appeared reliable, when testing large numbers of students, to administer the OSCE on different days. In order to make reliable decisions, for this dental OSCE minimum 17 stations are needed. Clearly, wide sampling of stations is at the heart of obtaining reliable scores in OSCEs, also in dental education.

Standard setting

As evidence for validity of dental OSCEs, CHAPTER 6 describes a study to elucidate which standard setting method is the optimal instrument to prevent that incompetent students pass and competent students fail a dental OSCE. An OSCE with 14 test-stations was used to assess the performance of 119 dental 3rd year students in a training group practice. Three standard setting methods were applied: the Angoff I method, the modified Angoff II with reality check and the Borderline Regression (BR) method to establish the pass/fail standard per station. For the final decision about passing or failing the complete OSCE, the 3 methods were compared in a Total
Compensatory (TC), a Partial Compensatory (PC) within clusters of competence, and a Non-Compensatory (NC) model. The reliability of the pass/fail standard of the 3 methods was indicated by the Root Mean Square Error (RMSE). As a criterion measure, a sample of the students (n=89) was rated in the clinic by their instructors and accordingly these students were divided into two groups: competent and incompetent students. The students’ clinical rating (considered for this study as “true qualification”) was compared with the pass-fail classification resulting from the OSCE. Undeserved passing of an incompetent student (false positive) was considered as more damaging than failing of a competent student (false negative).

The BR method showed more acceptable results than the two Angoff methods. In terms of pass rate the BR method showed the highest pass rates: for the TC model the Angoff method I and II and the BR showed the pass rates: 86.6%, 86.6%, and 97.5% respectively. For the PC model the pass rates were: 30.3%, 34.5%, and 61.3%, and for the NC model the pass rates were 0.8%, 1.7% and 7.6%.

The BR method showed lower RMSEs (higher reliability): for the TC model the RMSEs were 1.3%, 1.0% and 0.3% for the Angoff I, Angoff II and BR method respectively, and for the PC model, the RMSE of the clusters of competence ranged for Angoffs I: 2.0%-3.7%; for Angoff II: 1.8%-2.2%, and for the BR method 0.6%-0.7%. In terms of incorrect decisions, the BR method had a higher loss due to incorrect decisions for the TC model than for the PC model. It was concluded that the Borderline Regression method applied for a partial compensatory model provides defensible pass/fail standards and seems to be the optimal choice for this type of dental OSCEs.

In Chapter 7, the general discussion, the utility of the OSCE with the five desirable components of clinical assessment, the acceptability, the educational effect, the validity and reliability, and costs were discussed and placed in a broader context.

The acceptability of the OSCE by staff remained positive and acceptability by students increased positively in time. Students are always less enthusiastic about exams than staff, as also appeared during the former introduction of MCQ in education. Continues evaluation of attitudes of staff and students is a tool for quality assurance of dental education.

In respect of the educational effect was discussed that also in the long run the OSCE showed an educational effect. At the students’ request, several departments organized revising courses about competences tested by the OSCE to have more success in the OSCE: apparently, the students changed their study strategies. As also shown in literature, the hidden curriculum, based on the assessment program was obvious and asked for alignment of the competences and the assessment program.

In this thesis the bias is caused by gender, language and stress was evident. In respect to validity, the bias caused by the mother language had consequences. Language tests were introduced for all dental students, to make them aware of possible disabilities and to motivate students with diagnosed language problems to attend language courses. How to align this language program with the dental curriculum is not yet solved. Gender effects continue to be the intriguing bias to validity also in dental education. Do we test fair and equal to male and female? Also the student's stress in examinations is a concern and subject to further research.
The validity of the OSCE is also influenced by the decisions taken after the OSCE based on the standard setting. The results of the study on standard setting were in accordance with results in medical research, which gave possibilities to further generalization of the results. The Borderline Regression Standard setting method studied in this thesis can now be considered as the preferable method for OSCEs.

For reliable decisions about passing or failing a summative dental OSCE 17 stations are needed. This implies an effort in observing 17 stations. It is suggested that a written component of post-test-OSCE stations could reduce this number with remaining equal reliability.

Although not investigated in this thesis, the costs of the OSCE were also explored in the general discussion. It is reasonable to compare the costs for this type of dental OSCE with the costs of “normal teaching”. For now the costs for the OSCE seem equal to the costs of regular teaching. The marking of OSCEs is less labour intensive for teachers than other forms of marking assessments. The balance in the cost-benefit equation must lie in an evaluation of the gains brought in terms of efficiency and value. Due to the other positive aspects of the OSCE in quality assessment and its educational effect together with its relevance for practice it is worth the effort.

Future research should be aimed at several challenges. Performance assessment in dental education, while students are treating real patients, is the challenge of the century including new methods like portfolio and its content. Methodologies used in this thesis might be of help. An additional challenge is quality assurance of integrated clinical assessment programs as a whole. How to investigate the utility of these methods and programs to make them evidence based? And last but not least, to bring evidence in a European or global convergence in clinical assessment in dental education is perhaps the most difficult challenge.

As described in the introduction of this thesis, the shortcomings of methods of clinical assessment in dental education were subjectivity, no consequent approach, no explicit criteria and only numerical requirements, only testing of sensori-motor skills without testing of the communication and other “chairside” skills.

The studies in this thesis provided evidence for the use of the dental OSCE as an objective, reliable and valid method for testing competence including communication skills. Although bias as stress and language needs more attention, the OSCE has educational impact and is well accepted. Last but not least the Borderline Regression method can be used with confidence when setting the pass/fail score of OSCEs. Other dental faculties and schools are advised to implement OSCEs together with the Borderline Regression standard setting method in their assessment programs, for the benefit of the student’s learning, the institution’s quality assessment and the patient’s protection.