The use of the objective structured clinical examination (OSCE) in dental education
Schoonheim-Klein, M.E.

Citation for published version (APA):
Chapter 7

General Discussion
Chapter 7

Preface
In this thesis five studies are described that are all focused on the use of the Objective Structured Clinical Examination (OSCE) in dental education. The studies were carried out to investigate the utility of the OSCE in dental education in order to improve the clinical dental assessment in general. The underlying questions concerned the acceptability, the educational effect of the OSCE, the validity (including standard setting), and reliability of the OSCE.

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 relation of the language ability of Non-Native students with 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? (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 that incompetent students pass and competent students fail a dental OSCE? (Chapter 6, Validity)

The results of these studies will be summarized and their implication for clinical assessment in dental education will be discussed and combined with results from other studies to reach conclusions about the use of the OSCE in dental education. This chapter concludes with recommendations for future work.

General discussion
This thesis provided evidence for the use of the dental OSCE as an assessment method of clinical competences. Aspects of the utility (U), acceptability (A), educational effect (E), validity (V) and reliability (R) and standard setting as part of validity were investigated (U = A x E x V x R x C) (1). Although the financial implications/costs were not investigated in this thesis, this aspect will be explored together with the discussion of the results, concerning the desirable aspects of clinical assessment.

Acceptability (A)
Assessment methods that are highly reliable and valid may be time-consuming or complex to organize and resistance may develop when changes are introduced into an organisation. Indeed, acceptability is an indispensable component of the utility of OSCEs. In Chapter 2 the acceptance of the use of the OSCE for the assessment of clinical competence was supported by the increasing positive attitude of staff and students towards the OSCE. This was in accordance with other studies.
in dental education (2, 3). The implementation strategy applied by Argyris (4) using “stepwise” behaviour change and information, participation and commitment as tools, was effective in introducing the OSCE into a dental school. Staff considered the OSCE a good test for clinical competences and relevant for clinical practice. The students had a positive attitude towards the OSCE, although their attitude was less positive than the attitude of staff. The change from a formative to a summative OSCE caused a less positive attitude of the students towards the OSCE. Although not included in this thesis, all subsequent 3rd year OSCEs were evaluated by staff and students. The attitudes towards subsequent OSCEs were in accordance with the early evaluations. At present, the students are more familiar with the format of the OSCE and their evaluated attitude towards the OSCE is becoming increasingly positive. In the 1970’s the same acceptability changes have been reported on the multiple choice examinations (5). Without the acceptance of staff and students the OSCE is doomed to disappear. Evaluation of the attitudes of staff and students towards assessment methods is therefore a tool in quality assurance of dental education.

**Educational effect (E)**

Evidence for the educational effect of the OSCE was described in Chapter 3. Here it was shown that assessment drives learning. No effects of the implementation of an (undergraduate periodontal) OSCE were observed in study strategies, but the implementation of a dental OSCE appears to stimulate learning, resulting in greater achievement of specific clinical competence and a greater level of realistic self-assessment. Although the results indicate that no effects of the implementation of an OSCE in undergraduate periodontal education were observed in study strategies during the initial OSCEs, students’ study strategies did change while preparing for subsequent OSCEs, when they were more familiar with the format. At the students’ request, several departments organized revising courses about competences tested by the OSCE. i.e. Orthodontics and Communication. Students attended these courses to have more success in the OSCE. In this way, using the OSCE as assessment method stimulates learning in the long run also.

The implementation of an OSCE appears to stimulate learning, resulting in greater achievement of specific clinical competence and a greater level of realistic self-assessment. This is a positive effect because independent self-assessment may be an important stimulus to further learning and professional development (6) and a component of life-long learning (7).

It might be expected that students also learned from feedback after the examination. Immediate feedback can be highly influential on students’ learning and performance. As Hodder et al.(8) claimed, it can improve the students’ competence. For this reason a formative OSCE in the second year has recently been implemented in the curriculum in Amsterdam (ACTA). In this OSCE the students are provided with immediate feedback after each station. Further evaluation of this OSCE format is in progress. In the summative setting of an OSCE, however, this immediate feedback is not feasible. Therefore the students and staff were provided with feedback about students’ strengths and weaknesses after two days. However, after having passed the OSCE, the student is not motivated to improve his weaknesses, unless he needs to show his abilities again in another test. This was elucidated by the better performance of the students in the P-OSCE better in the
next end-of-year OSCE. Once again it appeared that assessment drives learning. This hidden curriculum, constructed by the assessment program (9), is the reason for the necessary alignment of the competences in the curriculum with the assessment program. This alignment needs serious attention in dental education.

**Validity (V)**

Does the OSCE measure what it is supposed to measure? Evidence for the validity of a dental OSCE is extensively described by Brown et al. (10). Chapter 4 of this thesis studied a bias, a threat to validity. It was concluded that the language background of the Non-Native students is an influential factor on this examination: 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 the OSCE performance of Non-Native students. This was in accordance with the findings in the UK (11, 12) and in Australia by Hays et al. (13). They concluded that one of the most powerful predictors of poor performance in OSCEs was living in an environment where English was not spoken at home. This language disability is therefore a threat to the **validity**, that need to be prevented (14). Since more and more free international movement of persons within the EU will happen with the lifting of international borders, the language barriers might be an increasingly interfering problem with learning and assessment in dental schools. It is therefore recommended that students with problems in language ability get additional tuition and practice.

Due to these findings, a program was developed to test the communication skills of all students in the first and second year of the dental curriculum. As a result of these tests, the students diagnosed with a language problem were urgently advised to attend special language courses at the University of Amsterdam, to improve their language ability. All these improvements are aimed at the improvement of the learning of the students and the improvement of their communication with patients. The first evaluations of these actions were disappointing. Of the 27 students with diagnosed language disabilities only 3 attended these courses. Perhaps because of the overcrowded dental curriculum it is difficult for these students to attend these language classes. More pressure and planning might be necessary, or a high level of language proficiency might be added as a selection criterion for entering dental education.

There were more threats to validity apart from the language of the students. The perceived stress of the students and gender - a popular topic in education all over the world (12, 15) - were reported in this thesis as other biases. In chapter 3 it was shown that female students outperformed males in communication and diagnostic stations. In chapter 4 the OSCE experience directed male students to more realistic self assessment. In contrast, women tended to underestimate themselves. Unfortunately it is one thing to observe human behavior, but quite another to explain it. Perhaps the differences in upbringing of male and female has effect on their communication abilities. May be the change in focus in dental education from sensori motor skills to competences including communication and diagnostics, was not as much to the benefit of male students as it was to female students. More research from many perspectives will be necessary to truly understand these gender effects.
The other possible drawback of the OSCE that was reported by students is the stress perceived by students. Students reported stress caused by being observed and by the timed schedule. These interactive aspects of the OSCE can create high levels of student anxiety, that may have an impact on the performance among students unfamiliar with the format (16). For students’ performance to be reflective of their actual competence, strategies must be implemented to desensitize students to the OSCE testing format. Therefore in the 1st year and 2nd year the students are tested with a formative OSCE. This probably contributed to a better evaluation of the OSCE by students in the 3rd year. Examinations are perceived as threatening situations by a small number of students, who sometimes developed a phenomenology of psychological and physiological symptoms (17). While “examination and grades” produced the most academic stress (18), different types of examinations were associated with different levels of state and test anxiety (19). This stress is a concern and subject to further research.

Standard setting as evidence of validity
The validity of the OSCE is also influenced by the decisions about passing or failing after the OSCE based on the standard setting. Chapter 6 investigates which standard setting method is the optimal instrument to prevent that incompetent students pass and competent students fail a dental OSCE in different compensation models. The results showed that the Borderline Regression method applied for a partial compensatory model provides defensible pass/fail standards and seems to be the optimal choice for dental OSCEs.

This is the first reported study about the use of standard setting methods in dental education. The present study confirmed results reported by Kramer et al. (20) and Wood et al (21), in medical education. The Borderline Regression standard setting method proved to be credible and reliable. The method is less time consuming, therefore more cost-effective, than the Angoff method and has the added credibility of being based on direct observation. Therefore it can be concluded that the Borderline Regression method is the standard setting method of preference for OSCEs in health education.

With the evidence of these findings the students, the institution and the patients can be more confident that the right decisions about passing and failing students are taken, when using the OSCE format of assessment. The methods presented in Chapter 6 can also be used to help implement standard setting procedures for other assessment methods in dental education.

Reliability (R)
Unreliable assessment scores do not provide a good picture of the strengths and weaknesses of the students and do not support accurate pass/fail decisions. Consequently patients cannot be sure the students have received a proper education (22).

Because of case specificity, many stations are needed for a reliable decision in OSCEs in medical education, as concluded by Swanson et al.(23). The performance in one station is not predictive of the performance in another station, i.e. the performance in communication with a patient is not a predictor for placing a good temporary restoration. Chapter 5 revealed that it is
reliable to administer the OSCE on different days, when testing large numbers of students. In order to make reliable decisions of passing and failing a summative dental OSCE, at least 17 stations are needed. Clearly, wide sampling of stations is at the heart of obtaining reliable scores in OSCEs, also in dental education.

It is suggested that a written component of post-test OSCE stations with questions about preceding stations will enhance the reliability (24). When using fewer stations, the loss of reliability due to the use of fewer stations in an OSCE can be fully compensated by lengthening the written test. Wass (25) commented that the components testing different aspects of knowledge and clinical skills must be carefully balanced to ensure both content validity and parity between items and test length. Also Brown et al. (10) used post-test stations and they found reliable OSCE results. Norcini (22) warned that the purpose of the test should be taken into account: “If the end of year evaluation is intended to assess broad competence, then a test aimed at both clinical and knowledge skills is required. However, if the purpose of the test is to ensure an appropriate level of clinical competence, then the addition of Multiple Choice Questions may not be warranted” (p.501). This alignment of learning goal and assessment format is a recurrent topic in studies on quality of clinical assessment.

The Costs (C)
The improvement by the use (post) test stations with a written content could also make the OSCE less costly, because the written component does not need an observer. Financial costs can play an important role in choices between assessment methods, but it seems not always a valid argument. Organizing an OSCE with 17 stations can be cumbersome, because of the need and cost of many observers in the stations. It requires commitment of assessors, patients and organisers. i.e. In Amsterdam (ACTA), when all the dental teachers, who are normally engaged in teaching in the dental training practice, can be involved in the OSCE, eight teachers can be examiners/observers. Two stations were observed by nurses from radiology and cross infection control. The other four recruited observers are consultants from departments in regular education. Three stations were on diagnostics and needed no direct observers, and one was rated with the computer. The standardized patients were recruited from administration staff, former teachers, and volunteers. Therefore the costs did not exceed costs of “normal” teaching during patient treatments. In this way real costs can be regulated, as Cusinamo et al. (26) also stated, by using academic staff as assessors as part of their normal educational task, and by building a database of simulated patients, who may perhaps become trained assessors. The majority of staff was enthusiastic (chapter 2) and in another study staff felt the outcomes to be worth the time they had spent (27).

It is reasonable to compare the financial costs of the assessment methods, like the OSCE with the “normal” teaching (28). In dental education in the Netherlands, the patient pays (75% of the normal price) for his treatment by students in the dental training practice. During assessment by an OSCE, however, there are no directly profitable patient’s treatments. The main purpose of the dental school is, however, to provide education, and assessment is an inextricable element of learning in this dental education. Inevitably, this will costs money. The Academic Centre for
Dentistry in Amsterdam (ACTA) is a dental school sponsored by the national government and the financial goal is based on a non-profit outcome. To state it differently: money for education is not comparable with income from patient treatments.

The balance in the cost-benefit equation must be placed in line with an evaluation of the gains brought in terms of efficiency and value (2). In a short time the performance of many students (in ACTA more than a 100) can efficiently be assessed and the marking of OSCEs is less labour intensive for teachers than other forms of marking assessments. 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.

**Generalization**

The studies in this thesis are on OSCEs in dental education in the Netherlands. Generalization of the results to other parts of health education in other countries in the world might be possible, but must be done with care. Different social environment and the differences between dental and medical and other health education areas can become confounders. The study of the educational effect of the OSCE in dental education was possible because of a curriculum change, which is not always easily duplicated. Every OSCE is different. An OSCE in general without a specified content is just a concept or an administration format (5). Therefore it is important that the validity, reliability, and acceptability of any newly developed OSCE, is properly evaluated.

**Implications for further research**

The role of the Objective Structured Clinical Examination (OSCE) in clinical assessment is to simulate clinical situations, to assess components of clinical competence and to provide feedback to students (10). The OSCE as a part of the assessment program of the curriculum is a credible method for assessment in the ‘shows how’ facet: behavioural performance in the simulated environment. The OSCE is not assessing real life situations, it uses a simulated context. It is like a play: with people acting as patients and dentists, and an examiner as an audience. The place of the OSCE in an assessment program is in the level of “shows how”. The OSCEs described in this study are assessing students in their 3rd year (of a 5-year curriculum). They provide students with a license to proceed to treat more complex patients. There is a need for evidence on other assessment methods used to test the performance of students in the dental training practice, when they are actually treating real patients. As Val Wass (29) also concluded: assessment at the apex of Miller’s pyramid, the “does”, is “the international challenge of the century” for all involved in clinical competence testing, which includes dental educators.

The methods used in this thesis might be beneficial for this research. In the reported study on the reliability of the OSCE, the Generalizibility theory (30) was applied. These analyses might also be applied in research on clinical performance assessment during patient treatments. Williams et al (31) and Carline et al.(32) reported this in medical education and they concluded that reliable ratings of students’ overall clinical skills, including overall clinical grades, can be achieved by collecting a minimum of seven to eleven independent observations by different examiners.
More research is needed to make definitive conclusions on how many observations of clinical performance are needed for a reliable assessment of dental student's clinical competence. Other challenges are to investigate the utility of the case presentations and to make judgments on the utility of the portfolio of the students.

As in other health areas, to ensure the quality of the assessment program of the dental curriculum, an assessment blueprint needs to be developed (5). Consequently an additional challenge is quality assurance of integrated clinical assessment programs as a whole.

Perhaps evidence-based assessment methods, e.g. the OSCE, can help to assure quality of assessment programs in dental education (33). In order to use the outcome of assessment in the quality assurance the methods need to be acceptable, reliable and valid, they need to have educational effect and be feasible and cost-effective. Therefore there is a need for more evidence on utility in assessment of methods and programs in dental education.

And last but not least the need for European convergence is another focus for further research (34). International cooperation in research on assessment methods is of the utmost importance. Since the competences of the European dentist have been described (34), international cooperation in assessment of these competences is needed to ascertain that the European dentist can work across borders and that the quality of patient's treatment is safeguarded. To bring evidence in a European or global convergence in clinical assessment in dental education is perhaps the most difficult challenge.

**Conclusion**

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 (35).

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.
References