

Dynamic Assessment and the "Interactive Examination"

Anders Jönsson, Nikos Mattheos, Gunilla Svingby and Rolf Attström

Malmö University, SE-205 06 MALMÖ, Sweden

anders.jonsson@mah.se // nikolaos.mattheos@mah.se // gunilla.svingby@mah.se // rolf.attstrom@mah.se

ABSTRACT

To assess own actions and define individual learning needs is fundamental for professional development. The development of self-assessment skills requires practice and feedback during the course of studies. The "Interactive Examination" is a methodology aiming to assist students developing their self-assessment skills. The present study describes the methodology and presents the results from a multicentre evaluation study at the Faculty of Odontology (OD) and School of Teacher Education (LUT) at Malmö University, Sweden. During the examination, students assessed their own competence and their self-assessments were matched to the judgement of their instructors (OD) or to their examination results (LUT). Students then received a personal task, which they had to respond to in written text. After submitting their response, the students received a document representing the way an "expert" in the field chose to deal with the same task. They then had to prepare a "comparison document", where they identified differences between their own and the "expert" answer. Results showed that students appreciated the examination in both institutions. There was a somewhat different pattern of self-assessment in the two centres, and the qualitative analysis of students' comparison documents also revealed some interesting institutional differences.

Keywords

Assessment, Self-assessment, Oral health education, Teacher education

Introduction

One of the major challenges for profession-directed higher education today, is not only to equip students with knowledge and skills, but also to help them develop into independent learners, able to cope with an ever increasing amount of information and learning needs. The basis of this process lies in the individual's ability to continuously assess his or her actions and define individual learning needs accordingly. Research has shown that the ability to assess ourselves, especially within professional settings, is not a quality we are born with, but rather a metacognitive skill which can be learned, improved and excelled (Brown et al., 1997). It is also shown that not all professionals have developed this ability to a satisfactory degree and might consequently be unable to identify shortcomings in their own professional competence (Hays et al., 2002; Ngan & Amini, 1998; Reisine, 1996). But if professional education is supposed to foster reflecting and self-assessing practitioners, the students must be given the opportunity to practice these skills (Yeh, 2004), as well as be assessed on them. To assess the students' self-assessment skills is of central importance since the assessment has a very strong influence on students' learning (Brown et al., 1997).

Examination schemes in profession-directed education traditionally provide educators with a thorough insight into students' profession-related skills and competences, but little is known about students' ability to self-assess their proficiency, to define their own learning objectives, and independently direct their competence development during their professional life. A structured assessment methodology focused on such metacognitive skills at the side of traditionally examined skills and knowledge, would therefore be a very important tool in higher education.

De la Harpe and Radloff (2000) give several examples of both qualitative, like learning logs and interviews, as well as quantitative methods, mainly Likert scale based, to assess metacognitive skills. Most of these methods are, however, not integrated in the learning activities in an authentic manner, and the authors also point to the fact that "Students may be reluctant to engage in activities that focus on learning rather than on course content and may not devote the time and effort needed to complete assessment tasks effectively" (p. 177).

To avoid this pitfall, the "Interactive Examination", a structured assessment methodology developed and evaluated in the Faculty of Odontology at Malmö University, Sweden, has included the assessment of self-assessment skills in a regular examination. The methodology aims to evaluate students' content specific skills and competences in parallel to their self-assessment skills, while expanding and supplementing the learning process. The self-assessment skills are assessed with both quantitative as well as qualitative means. Also, the methodology makes use of modern

information- and communication technology in order to facilitate training and feedback without necessarily increasing the workload of the personnel (Mattheos et al., 2004b).

The present study aims to describe the model of the Interactive Examination and present the results from a multicentre evaluation study with undergraduate students in the Faculty of Odontology (OD) and School of Teacher Education (LUT) at Malmö University. It should be emphasized from the start, however, that this study does not aim for direct comparison of the two student groups, as differences in educational context and experimental settings would make this task meaningless. Rather, what is attempted is a “parallel execution”, where differences and similarities in the two institutions can be identified, leading to improvements of the methodology, as well as giving rise to new questions for further investigation.

Material and method

General Principle of the “Interactive Examination”

In principle, the methodology is based on six explicit stages:

1. *Quantitative self-assessment.* At the beginning of the process, the students assess their own competence through a number of Likert-scale questions, graded from 1 (poor) to 6 (excellent). In addition there are three open text fields, where the students can elaborate further on their self-assessment. When possible, the self-assessments are compared with the instructors’ judgements of students’ competence, and feedback is given – a process that to some extent can be automatized by the software. The purpose of this comparison is to highlight differences between student’s and instructor’s judgement, and not to constitute a judgement per se. Possible deviations between self-assessment and instructor’s assessment are only communicated to the students as a subject for reflection or a possible discussion issue with the instructor.

2. *Personal task.* After the completion of the initial self-assessment, students receive a personal task in the form of a problem which they might experience during their professional life. This is an interactive part of the examination, where the interaction takes place between the student and the different affordances provided (such as links, pictures, background data etc.). The students have to come up with a solution strategy and elaborate their choices in written text.

3. *Comparison task.* After the personal task, the students receive a document representing the way an “expert” in the field chose to deal with the same task. This “expert” answer does not correspond to the best or the only solution, but rather to a justified rationale from an experienced colleague, which remains open to discussion. The “expert” documents have been written in advance and the students are given access to them as they submit their responses to the personal task. This is a way of dealing with the problem of providing timely feedback to a large number of students, but the “expert” answers also provide a kind of social interaction, although in a fixed (or “frozen”) form. The stance taken here is thus that, although interaction is needed in order for learning to take place, this interaction does not necessarily involve direct communication or collaboration between humans (cf. Wiberg in this issue), but the interaction could also be mediated by technology.

By the aid of the “expert” answer, the students can, according to the concept of “the zone of proximal development” (Vygotsky, 1978), potentially reach further than they can on their own, thus making the assessment *dynamic*. Dynamic assessment means that interaction can take place, and feedback can be given, during the assessment or examination, which separates it from more “traditional assessments” (Swanson & Lussier, 2001). In this way, dynamic assessment provides the possibility to learn from the assessment, but also to assess the student’s potential (“best performance”), rather than (or together with) his or her “typical performance” (Gipps, 2001). Empirical studies has shown that dynamic assessment indeed help to improve student performance, and also that low-performing students are those who benefit the most, thus making the difference between high- and low-performing students less pronounced (Swanson & Lussier, 2001).

After receiving the “expert” document, the students must, within a week, prepare a comparison document, where they identify differences between their own and the “expert” answer. The students are also expected to reflect on the reasons for these differences and try to identify own needs for further learning. This comparison document is a part

of the qualitative self-assessment in the Interactive Examination, which, in contrast to the quantitative self-assessment, is used for summative purposes as well.

4. *Evaluation.* After the examination the students evaluate the whole experience through a standardized form. At this point students have no feedback whether they have successfully completed the exam or not.

5. *Assessment of students.* The students are assessed on the basis of: (1) their competence and knowledge on course specific objectives, and their ability to relate theoretical knowledge to displayed scenarios and critical thinking, as expressed in their personal task, as well as (2) their ability to reflect on their choices, identify weaknesses and define future learning objectives, as expressed in the comparison document.

When poor performance is demonstrated in any of the above fields, students are assigned additional tasks. In this way students cannot “fail” the exam completely, but might be requested to practice and improve the respective skills, until a satisfactory level of competence is reached.

6. *Personalized feedback.* One month after the examination, individual feedback is sent electronically to all students. This feedback includes comments on students’ self-assessment and how it relates to the judgement of the clinical instructor, as well as comments on the personal task and the comparison document. Finally the feedback contains suggestions for future tasks if necessary.

Current Experimental Settings in OD and LUT

The current experimental settings, as presented below, show how the Interactive Examination was applied in the autumn 2004 to undergraduate students at OD and LUT, both which are faculties at Malmö University. OD was founded in 1946 and provides undergraduate education in Dentistry, Dental Technology, and Dental Hygiene. The Interactive Examination is used within the dentistry programme, where 40 students are accepted every autumn. Within this programme, Problem-Based Learning (PBL) has been used since 1990 (Malmö University, 2006). LUT, with approximately 8.000 students, is the largest faculty at the university, and the undergraduate education covers the whole range from pre-school teaching to the upper secondary level. The undergraduate education at LUT is organized in five major areas, or fields of knowledge, and the Interactive Examination is used within the field called Science, Environment and Society (Malmö University, 2007). The Interactive Examination was made available over the Internet through e-learning platforms, making it possible for the students to do the examination at any place they found suitable.

The platforms used are non-commercial and both have been developed locally to meet the requirements of the specific learning activities to take place, as well as to facilitate the research conducted. At LUT a platform called ALHE (Accessibility and Learning in Higher Education) has been used, and this will be the one presented more thoroughly in this article. ALHE is in many respects a conventional educational platform with both asynchronous (discussion forums and e-mail) and synchronous (online chat) communication tools, but it also includes some quite specific features. For example, students’ discussions are logged and displayed in a way to help the students reflect upon their own dialogic pattern and mutual knowledge building (e.g. who communicates with whom and to what extent, what kind of contributions have the students made), but also to facilitate research on the same issues. Another feature is the use of questionnaires, where the results can be exported to data sheets (such as Microsoft Excel or SPSS) for further analysis. Furthermore, ALHE is built to allow for the addition of new modules, and the Interactive Examination is such a module that has been implemented into the main platform. In the teacher interface, the files necessary for the examination (e.g. movies and “expert” documents) can be uploaded in a sequence similar to the methodology, and the files are then accessible in the student interface as hyperlinks. This is then complemented with the use of questionnaires in the quantitative self-assessment and the student evaluation as described below.

1. *Quantitative self-assessment.* Both students of OD and LUT started the Interactive Examination through a questionnaire-based self-assessment in specific course directed competencies.

Each OD student was assigned to one of six clinical instructors. The clinical instructors held regular meetings, where they revised the learning objectives for the 3rd semester and the quality assessment criteria for clinical work. As a result, 11 specific self-assessment fields were prepared for the students, reflecting 11 basic competencies. In the

Interactive Examination, the students self-assessed their competence in the 11 knowledge and skill areas through an on-line form connected to a database. The self-assessment was carried out through ordinal scales marked from one to six, with six marked as “excellent” and one being “poor”. The same form had been used by the clinical instructors when assessing students’ clinical competence at the end of the semester. The results from students’ self-assessment were later compared to those originating from their clinical instructors.

The self-assessment form LUT students completed was based on a scoring guide, or rubric, which was developed for this particular examination. The criteria in the rubric were equivalents to the self-assessment questions, making possible a comparison of students’ self-assessment to their actual results. The questionnaire had 13 questions relating to basic teacher competencies and three questions regarding reflection and self-assessment skills.

2. *Personal task.* The OD students received a clinical patient case, accompanied with the possibility to access relevant images and diagnostic data. Their task was to identify the problem (diagnosis) and propose a treatment plan. Figure 1 illustrates the personal task, on the left side links to movies and other affordances are visible just below the Swedish title “Interaktiv examination”. The three forms described in the text are seen in the central portion of the screen. The right hand picture shows how one of the movies is displayed in the browser by using Flash software. Movies were also available as mpeg-files, with higher resolution and higher-quality sound, for those students not using a dial-up internet connection.

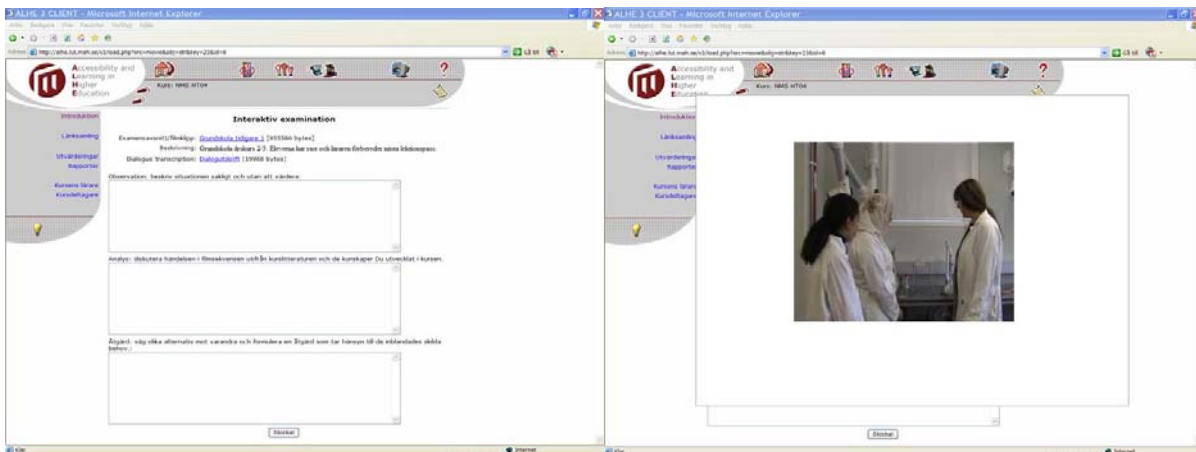


Figure 1. Screenshots from the student interface in the ALHE version of the Interactive Examination

The LUT students watched short movie sequences showing different problematic situations in a classroom context. Along with the movie sequences, the students could access some background data for the situation displayed, as well as the dialogue in text format. Of the total examination time, about one hour per movie was allocated for this part. With the movies as a starting point, the students filled in three different forms on the screen (see Figure 1):

- i. Describe the situation objectively and without prejudice,
- ii. Analyze the displayed situation on the basis of relevant literature and knowledge developed in the course, and
- iii. Consider different alternatives and give a proposal of how the teacher in the movie sequence should act.

3. *Expert response and comparison document.* Both OD and LUT students received a text representing the way a qualified colleague chose to deal with the same problem. All students had to come up with a written reflection as directed by the previously described principles.

4. *Evaluation.* After the examination the students evaluated the whole experience through a standardized form. The form included ten fields to which students could respond to on an ordinal scale from 1-9, as well as some multiple choice questions and free text fields. Free text comments were possible at the side of all fields. Eight fields were identical in the two centres, and two fields were similar. Along with practical issues, the form contained questions about the examination as a learning experience from the students’ point of view and the perceived relevance for their future profession.

5. *Assessment of the students.* The students were assessed on an “acceptable”/“not acceptable” basis, depending on their performance in the written task and the comparison document. In the Faculty of Odontology one assessor evaluated all personal tasks and comparison documents, whereas six clinical instructors (4 female - 2 male) provided judgements for comparison with students’ initial self-assessment. The personal task was assessed through specific discipline related evaluation criteria, while the comparison document was assessed through a specific scoring guide (Table 1).

At LUT, a scoring rubric covering the personal task as well as the comparison document was developed to provide information to both assessor and students what was to be assessed (Table 2). The students had access to the rubric well before the examination to enable a discussion of the assessment criteria with their instructors. The guide was also thought to make possible a more reliable assessment, despite the complexity of the task. All examinations were assessed by an external assessor.

6. *Personalized feedback.* One month after the examination, individual feedback was sent to all students. Feedback to the OD students included their performance in the written task, commentary on their comparison document, as well as suggestions for future learning.

For the LUT students, examination results were provided for each criterion in the scoring rubric. This very specific feedback showed both which criteria to give more attention in the future, as well as the direction of that attention to steer future learning. The students, as well as the researcher, could also easily compare the initial self-assessment to the actual results.

Sample

Both studies were carried out in the autumn 2004, with a cohort of first year student teachers in science and mathematics and second year dental students respectively. While all dental students were included in the study (n=34, 18 female- 16 male), some student teachers did not show up for the exam (n=171 out of 174, 103 female- 68 male). Also some student answers are missing in the first part of the examination, due to technical problems, making the LUT sample somewhat smaller for the self-assessment (n=166). All students in both centres were exposed to the Interactive Examination for the first time.

Statistical Analysis

Students’ responses on the self-assessment fields were compared for agreement to those from their clinical instructors (OD), or to the actual examination results (LUT), using a two tailed Wilcoxon’s signed rank test. This test is a non-parametric analogue to the t-test and is used to determine whether two paired sets of data differ significantly from each other. The comparison was carried out for each individual student and also independently for each of the self-assessment fields. A frequency analysis was performed for the total of students’ and instructors’ scores in the assessment forms. The potential influence of gender or instructor on the examination scores, as well as the pattern of the self-assessment (higher, lower or in agreement), was investigated with regression analysis. Non-parametric linear regression was used to correlate gender, group and examination scores with students’ pattern of self-assessment.

Qualitative Analysis

The qualitative analysis of the students’ answers to the comparison task aims to answer the following questions:

1. What kinds of differences or similarities between student and “expert” answers were identified?
2. Which reasons, for the identified differences, are stated?
3. Which weaknesses in their own competence as a teacher or dentist, and which learning needs, are identified by the students?

Table 1. Criteria for grading OD students' comparison documents

Evaluation	Excellent (3 pts)	Acceptable (2 pts)	Not acceptable (1 pt)
<i>Comparison of content</i>	The student has identified most/all the important differences	The student has identified half of the major differences	The student has only identified very few or irrelevant differences
<i>Analysis explanation of the differences</i>	The student is able to analyze/attribute differences	The student can only partly analyze/attribute differences	The student does not attempt to analyze the differences.
<i>Defining learning objectives</i>	The student reaches the learning objectives deriving from the analysis of differences	The student provides learning objectives only partly relevant to his analysis of differences	The student does not reach learning objectives, or they are irrelevant to his analysis of differences

Table 2. Part of the LUT scoring guide

Evaluation	Acceptable	Excellent
Reflection: Can you use your own, as well as others', experiences as a basis for reflection and development?	The reflection identifies differences between the own and the other teacher's interpretation of the situation.	The reflection identifies most of, or all, relevant differences between the own and the other teacher's interpretation of the situation.
	The reflection presents some reason, or reasons, for the identified differences.	The reflection argues in favour of own standpoints on the basis of relevant literature.
	The reflection identifies shortcomings in own professional competence.	The reflection identifies shortcomings in own professional competence and states learning needs resulting from these shortcomings.

In a previous study (Mattheos et al., 2004b) the differences which students identified in the comparison of their own answers to the "expert", were categorized into differences in (1) form, (2) content, and (3) attitude towards the content. This classification is used in the current study as well, with the addition of a fourth category which was mainly present among LUT students: Differences in interpretation. Difference in interpretation occurs when the student and the "expert" has interpreted the same situation in totally different ways, so that attempts to compare the two are extremely difficult. One example is when a student interprets the situation as a gender issue, while the "expert" writes about the same situation from an assessment point of view. Due to the nature of the problem dealt with in the cases of the OD students, differences in interpretation were much less prevalent.

Results

Evaluation – Students' Attitudes

Students' acceptance of the methodology was positive, with the median values in most evaluation fields lying between 6 and 8 (OD), or around 6 (LUT), on a scale to 9 (Table 3). Based on their free text comments, OD students favoured the opportunity to reflect on one's own self-assessment and the contact with their educators in this type of assessment. Some students would prefer more timely and personal feedback.

Student teachers appear to have found this mode of examination interesting and instructive, especially the comparison part. Some mentioned enhanced motivation and engagement as a consequence of the authentic tasks. A few students wrote about less stress and anxiety as compared to more traditional modes of examination. Negative comments were mainly about lack of background information for the movie situations.

Table 3. Some results from the student evaluation

Question	Student response OD (median)	Student response LUT (median)
<i>How do you value the Interactive Examination as a learning experience? 1 (not effective) - 9 (very effective)</i>	8 (n = 33)	6 (n = 140)
<i>Was it clear what was expected from you in the Interactive Examination? 1 (very unclear) - 9 (very clear)</i>	6 (n = 33)	5 (n = 138)
<i>To what extend do you feel you got the chance to show what you know? 1 (very little) - 9 (very much)</i>	7 (n = 33)	6 (n = 140)
<i>How much do you think this type of examination can help you to prepare for your working tasks as a dentist/teacher? 1 (very little) - 9 (very much)</i>	6 (n = 33)	6 (n = 139)
<i>How difficult were the examination cases? 1 (very easy) - 9 (very difficult)</i>	6 (n = 33)	6 (n = 138)

Self-Assessment

The students' self-assessment forms provided a total of 369 scores from 34 students (OD) and 2647 scores from 166 students (LUT). The responding forms from the clinical instructors at the OD amounted to 374 scores, as some students had probably omitted some of the evaluation fields. The unmatched scores were excluded and the comparison was based on 369 scores from 34 students.

A total of 142 (38 %) dental student scores were higher than the judgement of the clinical instructors, while 88 (24 %) were lower and 139 (38 %) were in agreement. On an individual basis, 12 students' (35 %) judgement was significantly higher than that from their instructors ($p < 0.05$), while in 6 cases (17 %) it was lower. Multiple linear regression analysis revealed no relation between the students' gender, group, success on the exam or clinical instructor and their self-assessment pattern.

In the case of the LUT students, a total of 1898 (72%) student scores were higher than their respective examination results, 318 (12%) lower and 431 (16%) were in agreement. On an individual basis 130 out of 166 students (78%) significantly overestimated their score as compared to the examination results ($p < 0.05$). There was no relation between the students' gender or group and their self-assessment pattern. However, there was a relation between getting a high score on the exam and not significantly overestimate in the self-assessment ($\eta = 0.501$; $p < 0.001$).

Comparison Task

Even though the task was to identify differences, the majority of the student teachers have chosen to highlight similarities. Most students, however, identified differences as well. Slightly more than half of the answers pointed to differences in content, while about 7 out of 10 identified differences concerning attitudes to content. Differences in content were mainly about examples and specific details, and attitudes to content about what to focus, different values or sources of reference. A minority of students identified differences in form or interpretation. Few answers presented a reason for the identified difference. Among these, differences in experience, habit or routine are stated as main reasons. In line with this, more experience and practical training are suggested as learning needs.

As occurring at previous occasions of the Interactive Examination, the OD students seem to focus primarily on differences. Similarities were only briefly noticed and students scarcely attempted to elaborate on them. The majority of students chose to prioritise differences in the content and the attitude towards the content and only a few elaborated on form differences as well. Identification of learning needs was mainly directed by differences in the

content. Differences in the attitude were rarely reflected in the defined future learning objectives. On the other hand, the attitude differences were the field where students would most likely choose to defend their choice and argue against the response of the “expert”. Differences in the interpretation were something that was very rarely encountered by the OD students.

In the majority of cases the OD students proved to be very skilful in locating the weak points and gaps in their knowledge, as opposed to the LUT students. Six students (2 females - 4 males) failed to identify the actual problems with their essays and were assigned some additional tasks.

Discussion

The main focus of this study was the reflective process which is initiated through comparing your own work with that of someone else. This process is well rooted in students’ self-assessment ability, a necessary professional skill. The Interactive Examination is a methodology developed on this principle and has been carried out with several cohorts of students with promising results (Mattheos et al., 2004a; Mattheos et al., 2004b).

The present application of the Interactive Examination is unique in the sense that it brings together two different educational environments. As was emphasized initially, this study does not aim for direct comparison of the two student groups, but rather for a “parallel execution”, investigating the self-assessment pattern and the acceptance by the students in two different institutions. Furthermore, the study aimed to identify institutional differences and similarities, hopefully leading to improvements of the effectiveness and applicability of the methodology, as well as providing new insights into the respective institutional learning cultures.

The students appeared to receive this form of examination favourably in both institutions. Students’ appreciation and acceptance of the examination methodology, as well as the value of the reflective process, are seen as prerequisites in order to affect their learning. The positive experience of the students therefore justifies a further analysis and discussion of the results from the Interactive Examination.

In the studies reported here, there was a somewhat different pattern of self-assessment in the two centres. While a large portion of the scores at OD were in agreement with the judgements of the clinical instructors, the corresponding value at LUT was much lower. Also, whereas the dental students had a number of self-assessment scores both higher and lower than the instructors; the by far greatest number of scores from the student teachers were higher than the examination results. It should be kept in mind, however, that the self-assessment was somewhat different in the two centres. While the OD student scores were compared to their instructors’ judgement, the LUT students’ scores were compared to their examination results. This difference, along with other contextual factors (such as the formulation of the self-assessment questions), have a potential influence on the students’ self-assessment pattern. Also, as OD students and instructors have been spending a whole semester together, a “calibration” effect might bring their judgements closer to each other’s. This means that the differences between student teachers and dental students must be interpreted with caution. There are, however, striking differences in the frequency of scores in agreement, as well as in the distribution of higher and lower scores, warranting further discussion and research.

According to previous research on self-assessment, low-ability students many times overestimate their grade or score as compared to the teachers’ judgement, while high-ability students more often are in agreement with their teacher (Kruger & Dunning, 1999). Also, progress in the course of studies seems to affect the self-assessment skills, where students in the beginning of a course produce less reliable assessments of themselves (Topping, 2003). In OD there was no relation between the students’ success on the exam and their self-assessment pattern, while on LUT, there was a relation between self-assessment pattern and success on the exam. A possible explanation to this, is that the dental students both have progressed somewhat further in their education (they were on the 3rd semester of their studies), but might also be more homogenous and calibrated as a group. By the 3rd semester the dental students have already spent a considerable amount of time working together in an environment where peer learning and group dynamics have a significant role, as well as continuous contact with the instructors.

No other factor, such as students’ gender, group or clinical instructor, could be found relating to the self-assessment pattern in either OD or LUT. Studies reported in the literature on self-assessment shows no uniform results on gender differences in self-assessment skills either (Arnold et al., 1985; Ericson et al., 1997; Topping, 2003).

The qualitative analysis of students' comparison documents provided some very interesting findings. Evidently, the dental students were primarily focused on differences with the "solution" provided, while student teachers seem to have extensively focused on similarities. Similarities tend to be only briefly mentioned, if at all, by dental students and are usually not accompanied by further arguments. It appears that the dental students treat the similarities as something well expected, almost self-evident, not worth of special attention and choose to focus on the explanation of differences instead. This attitude has been repeated almost in every cohort of dental students so far, to the extent that the assessors in the dental faculty have considered this a standard attitude, the reasons of which were never questioned. However, the execution of the Interactive Examination in the Teacher Education has brought a valuable insight in this field. In addition, the differences in interpretation are encountered in student teachers' documents, while they are rarely observed with dental students. Besides contextual factors, such as the nature of the expert document, one might consider many reasons as likely to have contributed to these differences:

Different nature of the assessed task. Diagnosis and treatment planning as encountered in the second year dental students' cases require a well defined array of knowledge fields and competences. Although controversies are very often encountered, the existence of specific guidelines and accepted practices, limits down the spectrum of viable choices as well as the importance of subjective factors. Dental students in their great majority identified the same main problem in each clinical case. With most students having the same starting point, it might be that differences are more likely to attract attention than similarities.

On the other hand, the task which student teachers were called to complete covered a wider area of subjects, including social and moral issues, where application of standards and guidelines is sometimes unclear. Furthermore, the cases could be approached from different points of view, defining different problems as starting points. This resulted in different intervention strategies and differences in interpretation.

Difference in the institutional learning cultures. It appeared that dental students tend to see more authority in the "qualified dentist" than student teachers see in a "qualified teacher". Dental students, at least at this early stage of their studies, seem to be less eager to question the opinions of the qualified dentist than student teachers of an experienced teacher. If this is true, it might reflect differences in how the students see their future role as "end products" of their education.

A dentist might represent for the dental students a very strictly defined set of competences, accompanied by a certain degree of "authority", which they are most likely uncomfortable to challenge. Student teachers, however, seem to adopt more of a peer attitude towards their qualified colleagues. This is reflected in the fact that some students has chosen to criticize the "expert". The criticism is mainly about views and values, but to some degree also about the interpretation and the examples chosen. In other cases the students regard their own solutions as being qualitatively better than the qualified teacher's. Here it is foremost choices of specific examples or actions taken that are considered better.

In future studies it would be very interesting to further investigate this assumption and see if there are certain differences between students of different profession-directed educations, in terms of how students perceive their development towards the "final product" of their studies.

Conclusions

The added value of this multicentre study is threefold. The first lies in the validation of the methodology. There is often a problem in estimating the quality of new modes of assessment, since they cannot always be evaluated on the basis of traditional psychometric criteria. Gielen et al. (2003) argue that "To do right to the basic assumptions of these assessment forms ["authentic assessment" and "performance assessment"], the traditionally used psychometric criteria need to be expanded, and additional relevant criteria for evaluating the quality of assessment need to be developed" (p. 38). In this widened set of criteria, referred to as "edumetric" criteria, the validity concept has been expanded to include the tasks used, considering authenticity and complexity in relation to the knowledge domain being assessed, but also consequences of the assessment such as the influence on students' learning or learning strategies (Sambell et al., 1997; Gielen et al., 2003).

Even though further research on the quality of the Interactive Examination is needed to better determine the consequences of the methodology, in terms of students' learning and learning strategies, efforts have been made to include features aiming specifically for self-assessment skills and thus trying to make the examination more valid for the proposed purpose (Frederiksen & Collins, 1989). As described earlier, two parts of the Interactive Examination involves self-assessment. First, the students estimate their own competence according to Likert-like questions, where the results are compared either to judgements from the instructor (OD) or the actual examination results (LUT). This comparison, however, does not constitute a judgement per se and possible deviations between self-assessment and results are used only to draw the students' attention to the difference and thus make reflection and learning possible. Secondly, the students compare their answers with the answer of an "expert". This comparison is assessed, and feedback is given, according to scoring criteria.

But addressing self-assessment skills is not the same thing as really capturing them. However, by investigating how the methodology is used and perceived in the two different institutions, an estimation of the validity can be made. Even though there are some institutional differences, displayed for instance in the different ways to handle the comparison task by LUT and OD students, the overall applicability of the methodology is similar in both centres and the students respond to it in an analogous manner. This indicates that the Interactive Examination might be a valid methodology for assessing students' self-assessment skills in authentic settings, and thus a potential tool for assisting the development of certain metacognitive skills in higher education.

The second added value of cross-sectional, multicentre studies such as this, is the provision of a better insight to students' self-assessment abilities. Future studies could investigate the longitudinal changes of students' self-assessment abilities throughout the curriculum. Such follow-up studies are necessary in order understand how these skills evolve and also to allow educators to design proper interventions in order to early identify and support students with weak self-assessment abilities.

The last added value to be commented upon, relates to the use ICT in the Interactive Examination. Information and communication technology is used in several ways in the examination methodology, and for several reasons. For example, it makes possible a automatized comparison of the quantitative self-assessment and instructors' judgement, and it also provides the necessary interactivity in the personal task, where the dental students have access to relevant images and diagnostic data, and the student teachers watch movie sequences that has to be accompanied with links to background data and other affordances. In both cases the students need to access the "expert" document after submitting their personal task, while at the same time saving their answers to the personal task in a database available to both assessors and researchers. Most importantly, however, the use of technology makes it possible to make valid assessments of student competences in a way not possible without this technological support. For instance, the authenticity of the examination could not be brought about by a paper-and-pencil test (cf. Lam, Williams, & Chua, 2007), nor could the same effectiveness be achieved if the students were assessed while actually performing in practice – this is especially true for the teacher education with such a large number of students. A conclusion is thus that training and valid assessment of self-assessment skills can be facilitated through the Interactive Examination, and that this can be done without necessarily increasing staff numbers or workload. In addition, as the examination is available online, the methodology could easily be used for distance education purposes. The Internet accessibility was used by a majority of the LUT students who preferred to carry out the examination at home or, in a few cases, from other parts of the world (e.g. Afghanistan and Iceland).

References

- Arnold, L., Willoughby, T., & Calkins, E. (1985). Self-evaluation in undergraduate medical education: a longitudinal perspective. *Journal of Medical Education*, 60 (1), 21-28.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*, London: Routledge.
- De la Harpe, B., & Radloff, A. (2000). Informed teachers and learners: The importance of assessing the characteristics needed for lifelong learning. *Studies in Continuing Education*, 22 (2), 169-182.
- Ericson, D., Christersson, C., Manogue, M., & Rohlin, M. (1997). Clinical guidelines and self-assessment in dental education. *European Journal of Dental Education*, 1 (3), 123-128.

Fredriksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18 (9), 27-32.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In Segers M., Dochy F., & Cascallar E. (Eds.), *Optimizing new modes of assessment: In search of qualities and standards*, Dordrecht: Kluwer Academic Publishers, 37-54.

Gipps, C. (2001). Sociocultural aspects of assessment. In Svingby, G. & Svingby, S. (Eds.), *Bedömning av kunskap och kompetens*, Stockholm: Lärarhögskolan i Stockholm, PRIM-gruppen, 15-67.

Hays, R. B., Jolly, B. C., Caldon, L. J., McCrorie, P., McAvoy, P. A., McManus, I. C., & Rethans, J-J. (2002). Is insight important? Measuring capacity to change performance. *Medical Education*, 36 (10), 965-971.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77 (6), 1121-1134.

Lam, W., Williams, J. B., & Chua, A. Y. K. (2007). E-xams: harnessing the power of ICTs to enhance authenticity. *Educational Technology & Society*, 10 (3), 209-221.

Malmö University (2006). *About the Faculty*, Retrieved October 1, 2007, from, http://www.mah.se/templates/Page___13096.aspx.

Malmö University (2007). *About the School*, Retrieved October 1, 2007, from http://www.mah.se/templates/Page___13082.aspx.

Mattheos, N., Nattestad, A., Christersson, C., Jansson, H., & Attström, R. (2004a). The effects of an interactive software application on the self-assessment ability of dental students. *European Journal of Dental Education*, 8 (3), 97-104.

Mattheos, N., Nattestad, A., Falk Nilsson, E., & Attström, R. (2004b). The Interactive Examination: Assessing students' self-assessment ability. *Medical Education*, 38 (4), 378-389.

Ngan, P., & Amini, H. (1998). Self-confidence of general dentists in diagnosing malocclusion and referring patients to orthodontists. *Journal of Clinical Orthodontology*, 32 (4), 241-245.

Reisine, S. (1996). An overview of self-reported outcome assessment in dental research. *Journal of Dental Education*, 60 (6), 488-93.

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?" An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23 (4), 349-371.

Swanson, H. L. & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, 71 (2), 321-363.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In Segers M., Dochy F., & Cascallar E. (Eds.), *Optimizing new modes of assessment: In search of qualities and standards*, Dordrecht: Kluwer Academic Publishers, 55-87.

Vygotsky, L. S. (1978). *Mind in society*, Cambridge: Harvard University Press.

Yeh, Y-C. (2004). Nurturing reflective teaching during critical-thinking instruction in a computer simulation program. *Computers & Education*, 42 (2), 181-194.