CREATING A LEARNING ENVIRONMENT BY USING SELF-, PEER- AND CO-ASSESSMENT

Received 21 July 1998; accepted (in revised form) 16 September 1998

ABSTRACT. To develop the skills and competencies required in professional organisations, students have to reflect on their own behaviour. Many current assessment practices in higher education do not answer this need. The recent interest in new assessment forms, such as self-, peer-, and co-assessment, can be seen as a means to tackle this problem. In the present article, a review of the literature provides answers to two questions: (1) How are self-, peer- and co-assessment applied in higher education? and (2) What are the effects of the use of these forms of assessment on the quality of the learning environment? Analyses of 62 studies showed that self-, peer- and co-assessment can be effective tools in developing competencies needed as a professional. These forms of assessment are often used in combination with each other. Implementation of these forms of assessments accelerates the developments of a curriculum based on competencies (knowledge as a tool) rather than knowledge (as a goal) and leads towards the integration of instruction and assessment in higher education. As such, this development of a learning environment contributes to the education of responsible and reflective professionals.

KEY WORDS: co-assessment, higher education, learning environment, peer-assessment, self-assessment

INTRODUCTION

It is widely recognised that the main goal of professional higher education is to help students to develop into 'reflective practitioners' who are able to reflect critically upon their own professional practice (Falchikov & Boud, 1989; Kwan & Leung, 1996; Schön, 1987). Students in modern organisations should be able to analyse information, to improve their problem-solving skills and communication, and to reflect on their own role in the learning process.

The need for lifelong learning in modern society will increase (Sambell & McDowell, 1998) as it becomes more recognised that the acquisition of knowledge and skills cannot be restricted to the phase of initial education; rather, it has to be a process continuing throughout one's entire working life. Traditional testing methods do not fit such goals as lifelong learning, reflective thinking, being critical, evaluating oneself and problem solving (Dochy & Moerkerke, 1997).

Alternatives in assessment have received much attention in the last decade and several forms of more authentic assessment have been introduced in higher education (Birenbaum & Dochy, 1996). The skills of self-, peer- and co-assessment are important in the development of autonomous, responsible and reflective individuals (Sambell & McDowell, 1998). Assessment procedures not only should serve as a tool for crediting students with recognised certificates, but also should be used to monitor progress and, if needed, to direct students to remedial learning activities. Research showed that the nature of assessment tasks influences the approaches to learning which students adopt (Beckwith, 1991). The existing assessment approaches can have effects contrary to those desired. Eisner (in Boud, 1995) identified the features of the new assessment in education:

- Assessment tasks need to reflect the tasks that students will encounter in the world outside schools, not merely those limited to the schools themselves.
- Assessment tasks should not be restricted to the solutions that students formulate, but also reveal how students go about solving a problem.
- Assessment tasks should reflect the values of the intellectual community from which tasks are derived.
- Assessment tasks need not be limited to a solo performance.
- Assessment tasks should have more than one acceptable solution to a problem and more than one acceptable answer to a question.
- Assessment tasks should have curricular relevance, but not be limited to the curriculum as taught.
- Assessment tasks should permit the student to select a form of representation that he or she chooses to display what has been learned.

The view that the assessment of students’ achievements is something which happens at the end of a process of learning is no longer widespread. Assessment is now represented as a tool for learning (Arter, 1996; Dochy & McDowell, 1997). The present article focuses on one new dimension of assessment innovation, namely, the changing place and function of the assessor. Alternative assessment also questions if the assessor must be the teacher, and whether self and peers can be introduced as assessors in different settings. This article provides a literature review which focuses on forms of self-, peer- and co-assessment from the points of view of their applicability in higher education and their effect on the learning environment.

This article specifically aims at finding out how self-, peer- and co-assessment are used in research and educational practice. Two research questions are addressed. How are self-, peer- and co-assessment applied
in higher education? What are the effects of the use of these forms of assessment on the quality of the learning environment?

2. Methods

In order to answer the research questions, a literature search was conducted using the following sources:

- The database of the Educational Resources Information Center (ERIC). This database was searched, according to the key words 'self-assessment', 'peer- assessment' and 'co-assessment'. The search was conducted for publications from 1987 to 1998.
- The database of Current Contents on Disk. The years 1996 to 1998 were searched in order to find recent studies which had not yet been included in ERIC. The same key words were used as in the ERIC search.
- Through the so-called snowball method, the references in all the above materials were checked for other studies.
- The Internet was searched with the Alta Vista search engine, and this identified a number of WWW addresses on assessment containing abstracts of conference papers, information of workshops, etc.

From these sources, all abstracts of the hits were printed and their content was analysed. Three criteria were used to determine whether literature would be included in our study. First, the assessment form had to be predominantly self-, peer- or co-assessment. Portfolio assessment and performance assessment, for example, were not central themes, although there was often a strong relationship with self-, peer- and co-assessment. Second, the literature about the relationship between assessment and learning/reflection was included in this search to find answers to the second research about the learning environment. Third, the subjects of study had to be students in higher education. For instance, studies dealing with peer assessment of university personnel were excluded.

The search of ERIC resulted in 191 articles. The articles reporting the application of assessment in higher education were selected for this study by analysing titles and abstracts. This resulted in 34 articles. The search of Current Contents on Disks resulted in 53 articles. After selection based on the above criteria, 10 articles were retained. The snowball method led to the selection of another 18 articles. In total, 62 articles were selected for further analysis. The search on Internet led to the selection of one relevant WWW address.
The 62 studies were analysed in detail and recorded in tables according to the authors and the year reported, the title, the aim of the study, the type and number of subjects, the strengths (of the assessment form discussed), the weaknesses (of the assessment form discussed), the treatment/instrument, the assessment (criteria), the value of statistics reported and the conclusions of the study.

The studies were classified first into empirical and non-empirical reports. In most of the articles, no statistics were reported. A study was considered empirical when quantitative data were gathered or statistical analyses were reported. Non-empirical studies were included because the information in these reports was too valuable to exclude them from the search.

Then, the studies were classified into self-, peer- and co-assessment. However, because these forms of assessment are sometimes strongly interrelated, it was decided to divide the studies into five categories: studies of self-assessment, studies of peer-assessment, studies using self- and peer-assessment, studies using self-, peer- and co-assessment, and more general studies about the relationship between assessment and learning or reflection on learning processes. The studies were coded by the three authors independently. Comparing the coding schemes revealed that there were no differences in classifying the studies and showed that there were: 14 studies about self-assessment, 10 studies about peer assessment, 7 studies about self/peer assessment, 19 studies about self/peer/co-assessment, and 12 studies about the relationship between assessment and learning/reflection.

In the present article, a narrative review of the literature is used. This form of conventional literature review implies careful reading of separate studies and integrating them. Of course, because patterns in the results are sought, this integration is an intuitive undertaking (Knoors, Dochy & Moerkerke, 1995; Slavin, 1986). Only one of the selected studies included a control group and an experimental group. As a consequence, a statistical meta-analysis could not be done.

3. How are Self-, Peer- and Co-assessment Applied in Higher Education?

In this section, the results are described related to the first research question using four separate subsections for the different combinations of self-, peer-, and co-assessment. These subsections begin in a definition. Then, the main findings are presented and the way in which the assessment forms are used in practice is outlined. Finally, a conclusion ends each subsection.
3.1. Self-Assessment

Self-assessment refers to the involvement of learners in making judgements about their own learning, particularly about their achievements and the outcomes of their learning (Boud & Falchikov, 1989). Self-assessment is not a new technique, but a way of increasing the role of students as active participants in their own learning (Boud, 1995), and is mostly used for formative assessment in order to foster reflection on one’s own learning processes and results (Sluijsmans, Dochy & Moerkerke, in press).

Boud and Falchikov (1989) classified the literature of self-assessment under three headings: conceptual, practical qualitative, and quantitative. One of the most important parts in the conceptual framework is the literature about the reflective practitioner (Schön, 1987). The practical qualitative group includes the processes involved in introducing and using self-assessment in different situations. The quantitative group focuses on studies of student self-ratings compared to the ratings of students by teachers. Boud and Falchikov (1989) analysed studies from 1932 to 1988 and reported the over-rating and the under-rating of students. They related these findings to the different abilities of students. The finding was that good students tended to under-rate themselves and that weaker students over-rated themselves. Students in higher levels of classes could better predict their performance than students in lower levels of classes. Griffee (1995) also investigated the question of whether there is a difference in student self-assessment between first-year, second-year and third-year classes in a university department. The general answer to this question was that there was no difference. All classes tended to rate themselves lower at the beginning of the school year and higher as the semester progressed. As the semester progressed, students gained more confidence in their ability to perform.

Several studies obviously show that the ability of students to rate themselves improves in the light of feedback or development over time (Birenbaum & Dochy, 1996; Boud & Falchikov, 1989; Griffee, 1995). Moreover, students’ interpretations are not just dependent on the form of the assessment process, but on how these tasks are embedded within the total context of the subject and within their total experience of educational life.

In educational practice, different instruments are used for self-assessment. Harrington (1995) used three different self-assessment instruments. One was simply a listing of abilities with definitions and directions to indicate those areas that you feel are your best or strongest. A second approach is to apply a Likert scale to a group of designated abilities. (For example, “in comparison to others of the same age, my art
ability is excellent, above average, average, below average, or poor"). Another approach is, for each ability, to provide different examples of the ability's applications so that individuals rate their performance level from high to low, and subsequently these are summed to obtain a total score. The self-assessment forms that Harrington described are cheaper and less time intrusive than traditional ways of assessing students (Nevo, 1995).

An electronic interactive advice system for self-assessment is provided by Gentle (1994). The aim of this system is to see how accurately students are able to assess their own work without the involvement of their supervisor. The system is based on question-and-answer screens for 38 skills. These skills are arranged into the four sections of (1) approach to the project – effort, time management, etc., (2) quality of day-to-day work, (3) quality of the description of the work and (4) quality of presentation. The procedure is as follows. “The user moves a cursor on a continuous scale of performance on that aspect of the work. The middle and end points on the scale are picked out by written statements to help the user and there is also a full advice screen available for each question. This feature makes this system much more than just an assessment program, since it includes large tranches of practical assistance, useful at any point in the project work. The output also provides much more than a mark; the five best and the five weakest points, selected by their weighted contribution to the mark, are extracted and displayed” (Gentle, 1994, p. 1159). Results of the use of the system show that students can assess themselves to within five percentage points. Students become more aware of the quality of their own work. They can predict their own mark and, while they are doing this, they reflect on their behaviour (reflective practitioner). Because the students reflect more often than once on their work, this will lead to a higher quality of the products. According to Gentle, the system is less time consuming than the conventional self-assessment because the supervisor has a minor part in the assessment.

In research conducted by Hassmén et al. (1996), 128 women learned the correct answers on a specific task by either performing or observing. Participants took either a performance or a written test, with or without making self-assessments about how sure they were that their selected answer was correct. Findings of the research support the hypothesis that those participants who engage in overt self-assessment while learning obtain a higher percentage of correct responses during learning trials than those who learn without self-assessments.

This is also illustrated in a study reporting successful language learning. McNamara and Deane (1995) designed a variety of activities that foster self-assessment. Three of them are writing letters to the teacher, keeping
a daily language learning log, and preparing an English portfolio. These activities can help students to identify their strengths and weaknesses in English, to document their progress, and to identify effective language learning strategies and materials. They also become aware of the language learning contexts that work best for them, and they establish goals for future independent learning. The idea of self-assessment for use with portfolios is described by Keith (1996), who suggests self-assessment assignments which ask students to report on their own learning. Assignments include sharing preconceptions about teaching and learning, comparing goals, creating a community of learners, generating student explanations and improving communication, group quizzes, challenging thinking dispositions, posttest evaluations and collaborative assessing. The roots of all the described assignments lie in collaborative learning. Keith finds that the most influential variable for effective learning is the amount of meaningful energy that the students put in. The assignments have to encourage students to feel responsible for their own learning.

Anderson and Freiberg (1995) used an audiotape self-assessment instrument for student teachers to reflect on their teaching. This instrument—called the Low Inference Self-Assessment Measure (LISAM)—has been developed to let student teachers analyse their instruction. Ten secondary student teachers completed four stages in the study. In the first stage, students learned to record themselves during a lesson. In the second stage, students were trained to analyse their own audiotapes. In the third stage, findings and suggestions for effective use of the LISAM were discussed. The students set goals for future use of the self-assessment instrument. In the last stage, there was an interview with every student teacher. Anderson and Freiberg describe three reasons why the LISAM is practical and effective: (1) the use of LISAM makes student teachers more independent, provides feedback and stimulates them to reflect on their own teaching; (2) student teachers can practice LISAM immediately; and (3) the LISAM teaching behaviours are observable and alterable.

Generally, next to addressing the instruments used for self-assessment, we could address the content. At the content level, it is striking that self-assessments are mostly used to foster skills and abilities (in addition to knowledge) and that assessments are used in a formative or diagnostic way (Birenbaum & Dochy, 1996). For example, students at Alverno College have to develop problem solving as one of the eight abilities in order to graduate (Loacker & Jensen, 1988). At the heart of the educational process at Alverno stands assessment, which is seen as a natural part of encouraging, directing and providing for development of abilities. Because self-assessment is required to be integrated with students’ problem-solving
process, students show increasing understanding of inter-relationships of ability, content and context. Students take responsibility for their learning as a dynamic, continuing process. They gradually internalise their practice of both problem solving and self-assessment abilities.

Overall, it can be concluded that research reports positive findings concerning the use of self-assessment in educational practice. Students in higher education are well able to self-assess accurately (see Gentle, 1994), and this ability improves with feedback and development over time. Moreover, students who engage in self-assessment tend to score higher on tests. Self-assessment, used in most cases to promote the learning of skills and abilities, leads to more reflection on one’s own work, higher quality of products, responsibility for one’s own learning, and increasing understanding of problem solving. Instruments for self-assessment vary from Likert scales, ability listings and written tests to portfolios, audiotape assessments or electronic interactive systems.

3.2. Peer Assessment

Falchikov (1995) defines peer assessment as the process whereby groups of individuals rate their peers. This exercise might entail previous discussion or agreement over criteria. It could involve the use of rating instruments or checklists, which have been designed by others, before the peer assessment exercise, or be designed by the user group to meet their particular needs.

More theoretical studies found in the search come with goals and forms of peer assessment. Somervell (1993) found that peer assessment engages students in making judgements about the work or the performance of other students. At one end of the spectrum, it could involve them giving feedback of a qualitative nature or, at the other, it might involve them in marking. The assessment could be formative or summative and could form part of a larger scheme in which peer feedback is given prior to self-assessment by the recipient (http://141.163.121.36/TeachingLearning/Ethos/Assessment/PeerAss1.html). Peer assessment is not only a grading procedure, but also part of a learning process in which skills are developed. Peer assessment can be seen as a part of the self-assessment process and serves to inform self-assessment. The contribution of other students can be a very useful input into the self-assessment process. Students have an opportunity to observe their peers throughout the learning process and often have more detailed knowledge of the work of others than do their teachers (Somervell, 1993). Keaten and Richardson (1992) reported that peer assessment can foster high levels of responsibility among students, but that the students must be fair and
accurate with the judgements which they make regarding their peers.

Peer evaluation is also an alternative term to peer assessment (Weaver & Cotrell, 1986). Peer evaluation “emphasizes skills, encourages involvement, focuses on learning, establishes a reference, promotes excellence, provides increased feedback, fosters attendance, and teaches responsibility” (Weaver & Cotrell, 1986, p. 25). Dancer and Dancer (1992) indicate that research studies have not shown the validity of peer rating. Peers are prone to produce ratings based on uniformity, race and friendship if there is no extensive training in peer rating. Based on this assumption, it is sometimes important to determine an individual’s contribution to a group project.

Different forms of assessment are distinguished by Kane and Lawler (1978). Peer ranking involves each group member ranking all of the others from best to worst on one or more factors. Peer nomination involves each group member being the highest in the group on a particular characteristic or dimension of performance. Peer rating involves each group member rating each other group member on a given set of performance or personal characteristics, using any one of several kinds of rating scales.

The more practically oriented studies focus on one of the main advantages of peer assessment, namely, fairness (construction of assessment criteria, more objective assessment by more subjects, assessment of output and process such as assessing student contribution). Conway et al. (1993) indicate that students found group projects more interesting than traditional methods of teaching. Because the fairness of the assessment was found to be the only negative aspect of this type of working, peer assessment was introduced. First, each group’s presentation was assessed by the other members of the group. Secondly, the students assessed the contribution of their fellow group members to the work of the project. The aim of the study was to examine ways in which students can be awarded individual marks, which reflect personal effort, for group projects. Conway et al. found good elements in the scheme of Goldfinch and Raeside (1990) and simplified this scheme by combining the elements which could be very effective. The method which Conway et al. used is not outlined here, but the results showed that students felt that peer assessment is a good method and fair enough. Students felt that they should play a part in the assessment in order to make it more objective.

Cutler and Price (1995) describe an investigation in which presentations and seminars, built into each of the three years of the geography program, are peer assessed against a set of criteria. Self-appraisal forms are also a part of the assessment procedure. The majority of the students were happy and confident about being assessed by their peers. Half of the students felt that their assessment of their peers was accurate. A third of the students
thought that they had improved in confidence, organisation of materials and use of voice.

Boud (1995) investigated students' ability to make assessments of themselves and their peers in a first year class entitled 'The Legal System' taught by the second author at the University of New South Wales. A series of instructions was provided for each student to give guidelines for formulating assessment criteria. Once the students had constructed the criteria that they would use, they made an assessment of both their own performance in class and that of the other students in the class. There were two methods of scaling used. Method A involves a scale of 1, 2, 3 ... 10 with 5 as the pass mark. Method B involves a scale of \(-2, -1, 0, 1, 2\) with the rule that there should be equal numbers of students above and below the mean of 0. At the end of the exercise, the self-mark, the peer-mark and the teacher-mark were available for each student on each criterion using both methods. Students found constructing the assessment criteria very useful, they rated themselves more highly than they were rated by their peers, and students rated themselves less highly than the teacher using method A but higher using method B. In general, there was a very high level of agreement between the marks given by peers and those given by the teacher.

Experience from peer-assessment revealed that, as a formative assessment method and as a part of the learning process, it can be seen as valuable. Students are more involved, both in the learning and in the assessment process. They find peer assessment fair enough and accurate. However, peer assessment also can involve: friendship marking (resulting in over-marking); collusive marking (resulting in a lack of differentiation within groups); decibel marking (where individuals dominate groups and get the highest marks); and parasite marking (where students fail to contribute but benefit from group marks) (Pond et al., 1995). These problems can be prevented by combining peer assessment with self-assessment or co-assessment. We observed that this is exactly the reason why the majority of studies involved these combinations of assessment forms. The following sections illustrate this.

Experiences revealed that peer-assessment, as a formative assessment method and as a part of the learning process, can be valuable because students are more involved both in learning and in the assessment process and because they find it fair and accurate. Disadvantages of peer assessment, such as friendship marking and decibel marking, are mostly solved by using combinations of peer-assessment with self- and co-assessment, as further explained later in this article.
3.3. Self- and Peer-assessment

Self- and peer-assessment are combined when students are assessing peers but the self is also included as a member of the group and must be assessed. This combination fosters reflection on one's own learning process and one's learning activities compared to the other members in the group or class.

Because of the disadvantages of peer assessment stated above, almost all studies found on combinations of assessment forms were practically oriented and sought more proof on validity, inter-rater (or inter-peer) reliability, and positive involvement of students in the process.

In a study described by Burnett and Cavaye (1980), fifth-year medical students assessed their peers as part of the examination. They also were asked to assess their own performance. Peer-assessment highly correlated with the final grade \( r = 0.99 \) and staff-assessment \( r = 0.93 \), and that self-assessments highly correlated with the results of peer-assessments \( r = 0.99 \). This is a replicated finding (Birenbaum & Dochy, 1996; Falchikov, 1991; McDowell, 1995). As a consequence, friendship ratings should not be taken as too large a problem. Nevertheless, we have experienced in our own research on peer assessment that the problem lies more in the weakest students who over-rate themselves and are not able to judge their peers correctly. Such scores often behave as statistical outliers. In our investigations, we therefore exclude the highest and the lowest peer assessment scores for each individual in order to calculate the mean scores.

Strachan and Wilcox (1996) describe a peer- and self-assessment strategy in a third-year microclimatology course. Thirty students were asked to form groups of three persons to do a seminar presentation. The students were informed that they would undergo a peer-assessment and that they could develop their own criteria. After this 'brainstorming', a Group Project Peer- and Self-assessment sheet was established. It was also a self-assessment sheet because every student had to include himself or herself in the assessment. Each member of each group completed the sheet after handing in their papers. The students also were asked to give written feedback on this way of assessment. Some students preferred written evaluative comments to number ratings and some students felt that this way of assessment was not necessary because you always have a certain responsibility in a group. Strachan and Wilcox recommend, however, that it is important to give the student an active role in the development of assessment criteria. The process is thereby equally important as the product for the quality of learning.

Warkentin et al. (1995) investigated self- and peer-assessment in a study with 83 undergraduate educational psychology students. Warkentin et al.
hypothesised that students taking tests using individual and group assessments would perform better on a posttest based on educational psychology course concepts than students who took the traditional tests (individual examinations). The effects on student knowledge structure representations were examined. The results indicated that there were no significant differences between the two groups on achievement and knowledge structure. Warkentin et al. (1995), however, found that the reactions to the assessment procedure which they used were overwhelmingly positive. The students did like the group assessment and thought that it contributed to their learning through this process as they discussed and debated test items.

Sambell and McDowell (1998) studied six cases which included peer-and/or self-assessment. Students were generally positive towards involvement in the assessment process. Students’ awareness that self- and peer-assessment helped them to develop important skills (e.g. problem solving) was high.

Under- and over-marking in self- and peer-assessment were investigated by Falchikov (1991). The process of working together on a small-group project was assessed by the group members, namely, seven developmental psychology students. In the study, the development of a self/peer group process assessment checklist was carried out and was designed to compare the assessments of task and maintenance functions (Falchikov, 1991). Students worked on a piece of coursework. The checklist contained 16 task functions and eight maintenance functions. This list was developed with the co-operation of the students; this enabled them to become familiar with the assessment checklist. After finishing the coursework, the students had to rate their peers and themselves on the checklist. They rated the level of activity (high, medium, low) to which each group member including self had carried out the 16 functions (group activities). The results showed that there was no consistency in over- or under-marking when comparing self-ratings with peer-ratings. There was also a high level of agreement between peers. Falchikov states that “the process of working in a group is a simple and effective way of moving towards a solution to the problem of evaluation of group work” (p. 15).

3.4. Self- and Peer-assessment Related to Co-assessment

In the prior sections, the use of self-assessment, peer-assessment and a combination of these two forms was described. One step closer to the current mainstream in traditional educational practice is the assessment procedure in which the tutor plays a significant role in the process. The term ‘co-assessment’ refers to this situation.
Co-assessment, the participation of the students with the staff in the assessment process, is a way of providing an opportunity for students to assess themselves while allowing the staff to maintain the necessary control over the final assessments (Hall, 1995). Synonyms for co-assessment are ‘collaborative assessment’ and ‘cooperative assessment’.

Co-assessment can be used for summative purposes, while self- and peer-assessment are used in a formative way. Somervell (1993) sees collaborative assessment as a teaching and learning process in which the student and instructor meet to clarify objectives and standards. In this case, the student is not necessarily responsible for the assessment, but the student collaborates in the process of determining what will be assessed and perhaps by whom. Pain, Bull and Brna (1996) argue that the term ‘collaborative assessment’ can be applied to an assessor and an assessee working together to form a mutual understanding of the student’s knowledge. It is a true collaboration in so far as both parties work on the shared goal of providing a mutually agreed assessment of the student’s knowledge. This entails both parties negotiating details of the assessment and discussing any misunderstandings that exist. This is consistent with the less confrontational approach to assessment that we seek to adopt while stressing the need to develop an ongoing relationship between the assessor and assessee.

Co-assessment is often related to forms of self- and peer-assessment. In a single study, we found a combination of self- and co-assessment. In Hall’s study (1995), the students and staff set the criteria. The process involved a double-sided face-sheet. On the back of this sheet, the students had the opportunity to give their own self-assessment of the piece of work and then hand it to the staff member. The staff member used the outside of the sheet to record his or her assessment of the student’s work. Then, the staff member turned it over to see whether or not students had chosen to offer their own assessment on the other side. The findings showed that generally the staff member’s grade was higher than the student’s grade. The experiment which Hall described identifies three purposes of co-assessment to assist student teachers in making role changes from being students to being teachers; to provide insights into the assessment process which may be of use to them in assessing their own students; and to provide a skill-development step towards self-assessment.

Many other studies combine self-, peer- and co-assessment in some way. In the studies that we found, there is a majority of positive experiences for assessing essays and assignments and, to a smaller extent, assessing presentations.

Falchikov (1986) and Stefani (1992, 1994) also described studies with collaborative self- and peer-assessment. The study of Falchikov aimed to
implement and evaluate a method of collaborative self- and peer-assessment. First the tutors set criteria which were ranked in terms of their relative importance. Then students set criteria and tutor-student criteria comparisons were made. An essay marking schedule was drawn up. Students marked their own essays and then each group member and the tutor marked the essays. Self-, peer- and tutor-marks were compared. Results show that collaborative and self-assessment does appear to be comparable to traditional tutor methods of assessment, while collaborative and peer-assessment corresponds less well with either tutor- or self-grading. Stefani (1992) carried out an experiment in collaborative self- and peer-assessment involving a first-year undergraduate biochemistry laboratory practical experiment. The students themselves defined the marking schedule for a scientific report. The results show that students have realistic perceptions of their own abilities and can make rational judgements about the achievements of their peers. Many tutors expressed their fears in handing over the assessment to the student. Concerning the evaluation of the learning benefits, almost every student said that the scheme made them think more, learn more and was challenging.

Freeman (1995) conducted a peer-assessment experiment with 210 final-year undergraduate business students who were divided into 41 teams, with each team completing two of the four assessable tasks. The presentation, one of the two tasks, was chosen by staff to experiment with a peer-assessment worth 25% of the overall grade. In the first week of semester, each student was given the presentation marking and feedback sheet with 22 items, with eight items related to the content and 14 related to the presentation, weighted 60% and 40%, respectively. In the results of the team presentations rated by staff and peers, Freeman found that the quality of the presentations was very high. There was no statistically significant difference between the average staff ratings and average peer ratings. However, students tended to under-mark the good presentations and over-mark the poor presentations.

Longhurst and Norton (1997) designed a study to investigate how accurately 67 second-year psychology students would be able to assess their own essays and thereby ascertain whether or not they understand what taking a deep approach in their essays actually means. Student grades were compared with tutor grades. The students were asked to rate themselves on tutor-specified criteria which were designed to measure a deep approach, essay grade and level of motivation in completing one specified essay. The tutor didn’t see these self-assessments because the self-assessment sheet was removed from each essay. The tutors also marked the essays on the deep-processing criteria. The tutor grade for the
essay highly correlated with each of the five criteria. There also was a positive correlation between student and tutor grades ($r = 0.43$). Overall, students were very accurate in grading their own essays, but less accurate in assessing their own deep processing. Less motivated and weaker students appeared to be less clear on understanding the individual criteria.

Oldfield and Macalpine (1995) investigated the competence of students in making assessments. The peer assessment was designed in steps from individual tasks to group assignments. Each task was assessed by the peer group and compared with the assessment of the lecturer. High correlations were evident between student marks and lecturers marks for individual essays and presentations. The self-assessment which Oldfield and Macalpine describe is also an approach in achievable steps, with the first being a comparison of contributions to group activities excluding self, then including self, and finally a self-assessment of individual work. The students first make a peer assessment of all of the groups’ achievements. To train self-assessment skill, students also had to do this for their own group. The same procedure takes place within the group: assessing the group members and then assessing one's own contribution. Orpen and Macalpine found that this assessment procedure strengthens the confidence of students to assess the work of others and of themselves.

Orpen (1982) describes a study with 21 students in an organisational behaviour course and 21 students in a political philosophy course. These students had to write an essay. The students were informed that “their papers would be marked by five lecturers later in the year, and that their final grade would be the average of the marks they received from their fellow-students and from the lecturers” (p. 568). The marks were given according to the criteria of (1) coverage of the relevant material, (2) coherence and strength of the underlying argument, and (3) fluency and clarity of expression. There was no difference between the lecturers and students in their average marks, in the variation of their marks, in the extent to which their marks agreed with each other, and in the relationship between their marks and the writer’s performance in end-of-course examinations.

Rushton et al. (1993) developed a computer-based peer assessment tool. A group of 32 computer science undergraduates was asked to write an essay on the viability of peer assessment. They typed their essays on the subject of peer assessment into the system. The class was split into groups of three or four students. Each group member used the peer-assessment window to mark each other's work. Contrary to expectations, the marks awarded by peers were remarkably similar to those awarded by tutors; so, peer and teacher assessment seem to be equally reliable forms of assessment.
Zoller and Ben-Chaim (1997) investigated the self-assessment ability of 71 biology majors enrolled in a four-year college program, with respect to higher-order cognitive skills and their confidence in self-assessing. A specially-designed self-assessment questionnaire consisted of interdisciplinary science-technology-environment-society (STES) questions and Likert-type questions involving students’ confidence. Students assessed their knowledge and understanding on this questionnaire. Students evaluated themselves as quite knowledgeable. The results further showed that 75% of the students thought that they were capable in self-assessing and peer-assessing. Zoller and Ben-Chaim found a discrepancy between the student assessment and the teacher assessment, and they accounted for this in terms of the lack of integration between assessment and learning in contemporary science teaching.

Kwan and Leung (1996) investigated tutor- and peer-group assessment of the performance of 96 students in a simulation exercise on hotel personnel training. The group was divided into five tutorial groups. Then students were paired and each student conducted a training session with the partner to an audience. The performance of each student was assessed by the tutor and the peers according to a checklist. Results show that there was some agreement between tutor and peer group markings, but somewhat less agreement than that reported by Falchikov (1986) and Stefani (1994). Arguments for this finding are that students aren’t capable of assessing it because this was the first time that they did it. Secondly, students made no contribution in identifying the criteria, and there was no negotiation between tutor and students in understanding the criteria.

Orsmond et al. (1996) describe an experiment in peer-assessment for a first-year undergraduate animal physiology poster assignment. The 39 pairs of students completed a poster assignment. Students were informed about the poster requirements. At the end of the 12-week lecture course, the students were divided between two laboratories. Later, the students of each laboratory were asked to mark all posters of students in the other laboratory on five criteria. Each criterion had a grade of 0–4. Each poster was marked and a peer assessment grade for each criterion was calculated based on the maximum number of ticks (the modal value). After the students marked the posters, the tutor also marked the posters without seeing the marks which students had given. Orsmond et al. (1996) found that there was 18% agreement between students and tutor, with 56% of students over-marking and 26% of students under-marking. The correlation was 0.54. The students also filled in a questionnaire which showed that 76% of the students thought that “the peer assessment had made them think more, and work in a more structured way” (p. 243).
Fry (1990) describes a study in which the tutor introduced peer marking. The tutor first marked the students’ scripts and then handed them back to the students. The tutor asked the students to mark each other’s work (peer assessment) according to a marking scheme. The agreement between the tutor’s marks and the students’ marks were generally very good. Fry further found five strong aspects of peer marking, as discussed in a later section.

The relationships among self-, peer- and co-assessments were examined by Horgan, Bol and Hacken (1997). They used the predictions of grades, actual grades, peer reviews, and reflective essays on self-assessment of undergraduate teacher education students to analyse these relationships. The students were trained in self-assessment. The students completed three multiple-choice examinations, for which the third was a cumulative final. Students predicted their grade and, after the examination, they reflected on their performance. The students also did a written analysis of a case study which was self-assessed and reviewed by three peers and the instructor based on five criteria. A third part in the assessment procedure described by Horgan et al. (1997) was an oral case analysis as part of a group. These presentations were also reviewed by peers. The final part was an essay about reflection on the self-assessment activities. Results of the assessments described above showed (1) agreement across assessors, (2) little consistency of self-assessment across tasks, (3) improvement in accuracy over the semester, (4) increased accuracy with increased performance, and (5) that better students used self-assessments to guide work, while weaker students used feedback to find the errors.

The literature reviewed above illustrates how effective self-, peer- and co-assessments are when used in combination. Self- and peer-assessment can be used for summative purposes as part of the co-assessment by giving the tutor the power to make the final decision about a process or a product. In this way, the traditional assessment, in which the tutor makes an autonomous decision, is not comparable with co-assessment. The combination of self-, peer- and co-assessment makes tutors and students work together in a constructive way and, as a result, they come to higher levels of understanding by negotiation. When students become teachers, this role-changing provides them with insights into the assessment process.

Several studies indicate that the marks given by the tutors and those given by the students are highly correlated (Freeman, 1995; Fry, 1990; Longhurst & Norton, 1997; Oldfield & Macalpine, 1995; Orpen, 1982). Only a few researchers found low correlations between student and tutor marks (Kwan & Leung, 1996; Rushton et al., 1993). Arguments for these findings were that the assessment and the learning weren’t sufficiently integrated, or that
students are not capable of assessing themselves when they first have to.

It can be concluded that the use of self-, peer- and collaborative-assessment is important to remove the student/tutor barrier, to develop enterprising competencies in students, and leads to greater motivation and 'deeper' learning (Somervell, 1993).

Where applications of self-assessment and peer-assessment were mostly used for formative purposes, combinations of these forms with co-assessment work out well for summative assessments. Various applications show various possibilities, ranging from using the peer-assessment as a contribution (e.g. 25%) to the overall score, to using peer-assessment as a correction score for tutor-assessment.

Developments in this area clearly open possibilities for assessing skills and abilities in areas in higher education where traditionally there were problems in assessing or in the costs for assessing. If peer- and co-assessment indeed is a valid, fair and useful method for assessing essays and assignments, it could become a widespread assessment method in the near future.

4. Effects of Using Self-, Peer- and Co-assessment on the Quality of the Learning Environment

In order to find an answer to the second research question, regarding the effect of using self-, peer- and co-assessment on the learning environment, we can refer to several previously-discussed studies which showed quantitative effects. Studies such as Hassmén, Sams & Hunt (1996) and Martens and Dochy (1997) show that students who use self-assessment procedures get better scores on a final test. However, such studies mostly involved assessments as support devices embedded in the learning materials and use a traditional final test.

Some studies reviewed in earlier sections of this article stress the agreement between self- or peer-marks and teacher-marks (e.g. Falchikov & Boud, 1989; Stefani, 1992, 1994). However, these findings don't give us much information about the impact which the form of assessment has on the learning process. Falchikov and Boud (1989) stated that “although we have focused on student-teacher agreement over-rating, we must not be distracted by the search to maximize congruence at all costs. Self-assessment can be a valuable learning activity, even in the absence of significant agreement between student and teacher, and can provide potent feedback to the student about both learning and educational and professional standards” (p. 427). Above all, this is the main reason why these forms of
assessment need to be integrated into curricula in higher education.

The studies in this literature review are more concerned with integrating learning and instruction, which means not only using assessment instruments to improve the learning process, but also adapting the assessment instruments (including the final examination) to the main goals of higher education. In this case, we look at effects of new assessment forms on the quality of the learning process and the creation of a learning environment. Therefore, the studies primarily in the fifth category of our classification (studies about the relationship between assessment and learning/reflection) were analysed.

An important concept that links assessment with the quality of learning is that of consequential validity (Boud, 1995, p. 41), which refers to the effects of assessment on learning and other educational matters. Assessment procedures of high consequential validity should be developed. Encouraging deep approaches to learning is one aspect which can be explored in considering consequences. Another is the impact which assessment has on the competencies and skills that students have in being able to assess themselves.

The importance of developing students into reflective practitioners is already pointed out. Boud (1992, 1995) developed a self-assessment schedule to provide a comprehensive and analytical record of learning in situations where students have substantial responsibility for what they do. The main guidance is a handout which suggests the headings (goals, criteria, evidence, judgements and further actions) that students might use. Self-assessment schedules are effective tools to use in enabling students to bring together a wide range of their learning, to reflect on their achievements and to examine the implications for further learning (Boud, 1992; Boud & Knights, 1994). Boud (1990) further recognised the gap between what is required of students in higher education and what happens in real life. He stresses the need for examination of assessment practices to see if they are compatible with the goals in higher education as described in the introduction of this article. The two main goals of student assessment in his view are improvement of the quality of learning (e.g. by reflection) and the need to certify. Boud sees self- and peer-assessment as fundamental to learning. Existing assessment practices might be more defensible if they could bear some relationship to the ways in which academic and other professional work is assessed in actual working environments and the situation in which knowledge is used.

Adams and King (1995) investigated the perceptions of self-assessment held by different student groups and the skills required for self-assessing. They also recognised that employment at a professional level usually
requires specialist knowledge. An important part of this knowledge is the ability to have a continual knowledge of one’s own capabilities and to be able to update weaknesses as appropriate. Adults, for example, like to learn and are capable of considerable amounts of high-quality learning of their own. Adams and King identified activities that can develop self-assessment skills. A framework helps students to develop self-assessment skills to be competent at self-assessment. Adams and King identified three levels. At the first level, students work on having an idea about the assessment process. Students perform activities such as discussing good and bad characteristics of sample work, discussing what is required in an assessment, critical reviews on literature, etc. At the second level, students work on identifying important criteria for assessment. At the third level, students work towards playing an active part in identifying and agreeing on assessment criteria and being able to assess peers and themselves competently.

The relation between reflection and self-assessment is also pointed out by Sobral (1997). Self-assessment of self-directed learning support reflection and learning partnerships and are facilitated by discussions and exercises. Therefore, Longhurst and Norton (1997) claim that self-assessment is clearly an important part of helping students to improve their own learning, as it focuses students’ attention of the meta-cognitive aspects of their learning and teaches them to be more effective at monitoring their own performance.

In some studies, the perceptions of students towards innovative assessment and its impact on learning are investigated. Sambell et al. (1997), for example, investigated the perceptions of students towards different aspects of innovative assessment. When discussing innovative assessment, many students believed that success more fairly depended on consistent application and hard work, not on a last-minute burst of effort or sheer luck. Many students felt that openness and clarity were fundamental requirements of a fair and valid assessment system. Students were very positive about the effects of alternative assessment on their learning.

A small-scale study of the views of a group of newly-enrolled Open University students in London resulted in a mixed response to alternative methods of assessment (Peters, 1996). The majority of the students disagreed with self- and peer-assessment. This finding, however, did not mean that the students were totally committed to traditional forms of assessment. The possibility of being able to re-draft assignments after tutor feedback was viewed more favourably. Williams (1992) found that the majority of students (90%) see benefits in peer assessment. Benefits are seen in three main categories: in comparison of approaches; in comparison of standards; and in exchange of information. However, students found that
criticising their friends was difficult (see also Strachan & Wilcox, 1996). Students also found peer-assessment difficult or undesirable when guidelines for evaluation are not established first. The two major findings in the study of Williams (1992) were that (1) students like to have more say in how they approach their learning and its assessment and (2) students need guidance and training in new role behaviours before this can actually happen. Ormond et al. (1996) found that students enjoyed carrying out the peer assessment and considered that it was beneficial to their learning. Keaten and Richardson (1992) also affirmed that peer assessment fosters an appreciation for internal awards and interpersonal relationships in the classroom. Cheng and Warren (1997) conducted research in the English department of the Hong Kong Polytechnic University to gauge students' attitudes prior to and after a peer-assessment. The students and the teacher assessed each group seminar and oral presentation. Before and after the peer-assessment, the students filled out a questionnaire with four items. The results of the questionnaire show that students were mostly positive towards the peer assessment, but that a few students thought that beginning students were able to conduct the assessment in a fair and responsible manner. The same result was reported by Falchikov and Boud (1989). Further, the students were not entirely confident in their ability to assess their peers. However, there was a positive shift overall in both attitudes and confidence. Finally, Cheng and Warren concluded that there is a need to give students systematic and comprehensive training in how they can assess their peers and how to establish criteria (see also Williams, 1992).

Overall, self-, peer- and co-assessment seem to improve different aspects of the quality of the learning environment and the learning of students. However, training in the skill to self-assess or to peer-assess is needed for an optimal impact. The prior section stresses the use of self-, peer- and co-assessment as tools for learning (i.e. ways to develop reflective practitioners in higher education).

5. DISCUSSION

Self-, peer- and co-assessment in higher education can be used in different ways and can be reliable and valid to a large extent. Assessment as a tool for learning has considerable impact on the students’ learning and development into reflective practitioners. In this final part of this article, the main strengths and weaknesses of self-, peer- and co-assessment that arise from the studies are discussed.
According to McDowell (1995), the main strengths are that (1) there is a development of evaluative and critical abilities, (2) there are opportunities for skill development, (3) knowledge is more integrated and (4) students collaborate, are motivated and are satisfied. The weaker points of innovative assessment lie sometimes in organisational issues, such as the occurrence of cheating, stress and time constraints. At last, there could be a mismatch between learning and marks without feedback. Self-assessment improves the independence in students’ learning, responsibility for decision making related to assignments, proactivity, and creativity in taking charge of their own work (Klenowski, 1995). Constraints on student self-assessment were lack of time, the paucity of professional development and support for student self-evaluation, and the change process itself. Adams and King (1995) remind us of a real problem of perception. The idea that teachers do the teaching and marking is hard to change. Also there is the possibility that students take advantage of their role and that they become strategic in their approach to their studies. Once students experience the self-assessment schedule described by Boud (1992, 1995), they see its value. The majority of students were initially supportive and became enthusiastic having been through the process of constructing a schedule (Boud, 1992, p. 191). There is much more concern about the self-grading aspect than there is about the qualitative assessment. If it is not possible to demonstrate that students can produce marks which are acceptable for teachers, the self-assessment should be restricted to a purely learning role and as a skill to be developed (Boud, 1989). Until now, we looked at self-assessment as a formative tool. Moreover, it should be clear that students have to know the criteria clearly and that peer assessment can be time consuming.

Strengths in using self- and peer-assessment (see also Brown & Dove, 1991) are that it (1) can foster students’ feelings of ownership for students’ own learning, (2) can motivate students and encourage their active involvement in learning, (3) makes assessment a shared activity rather than a lone one (i.e. more objective), (4) promotes a genuine interchange of ideas, (5) leads to more directed and effective learning, (6) encourages students to become more autonomous in learning, (7) signals to students that their experiences are valued and their judgements are respected, (8) develops transferable personal skills, (9) produces a community of learning in which students feel that they have influence and involvement, (10) reduces the teacher’s workload (Rushton et al., 1993), and (11) makes students think more deeply, see how others tackle problems, pick up points and learn to criticise constructively.

Considering these strengths and weaknesses, it may be concluded that probably the most difficult aspect of self-, peer- and co-assessment is to
determine the criteria. Criteria are the basis of evaluating student progress; they identify the critical aspects of a performance or a product that describe in specific terms what is involved in meeting the learning outcomes. It is necessary for the concept that the criteria are presented in operational terms with which all participants are familiar. Criteria should include information about the area to be assessed, the aims to be pursued and the standards to be reached (Boud, 1995). Boud and Falchikov (1989) identified two elements in any assessment decision: the identification of criteria or standards to be applied to one’s work; and the making of judgements about the extent to which work meets these criteria. In self-assessment, students judge their own performance and products against their own assessment criteria (Falchikov, 1986). Students also have to be trained in self- and peer-assessment. They have to learn and understand their role in the assessment process. Assessment should only be used in a “summative assessment system whose outcome is not a grade or label but a profile of the student to which all who are able to speak about him can contribute what they know – and in which conflicting assessments are highlighted rather than ironed out” (p. 297). Involvement in learning, including assessment, is vital to effective learning, and the teacher is the key person to help students to develop this learner autonomy (Dickinson, 1998).

An important shift is that the future labour market will play a distinguished part in the way in which the curriculum and the goals are revised (Pilot, 1997). The form of assessment determines whether the student achieves the skills required for the working field. This working field will thus also influence the content of the assessments (Moerkerke & Terlouw, 1998). Curricula nowadays are becoming more competency-based. This redesign of the curriculum requires a redesign of the assessment in order to create a learning environment that fosters learning. Self-, peer- and co-assessment can discourage passive, reproductive forms of learning. By integrating these forms of assessment into the curriculum, students are likely to develop into competent persons and lifelong learners who reflect continuously on their behaviour and learning process (Moerkerke, 1996). These alternative forms of assessment should be a part of a process of change towards a student-centred learning environment. This change requires a shift in emphasis from the norm-referenced to the criterion-referenced testing, from purely summative to formative and summative assessment, from external to internal evaluation, and from the assessment of product to the assessment of process as well. In order to be successful, the following supporting factors seem to be necessary: pedagogical change; a shared value system between students and teachers; and an organisation-wide evaluation ethic.
REFERENCES


D. SLUISMANS AND G. MOERKERKE
Center for Educational Technology and Expertise
Open University of the Netherlands
PO Box 2960, 6401 DL Heerlen
The Netherlands

F. DOCHY
Center for Research on Instructional Technology and Teacher Training
Department of Educational Sciences
University of Leuven
Vesaliusstraat 2, 3000 Leuven
Belgium