Studies in Higher Education

The use of self-, peer and co-assessment in higher education: A review

F. Dochy [a] , M. Segers [b] & D. Sluijsmans [c]

[a] University of Leuven, Belgium

[b] University of Maastricht, The Netherlands

[c] Open University of the Netherlands, The Netherlands

PLEASE SCROLL DOWN FOR ARTICLE

The Use of Self-, Peer and Co-assessment in Higher Education: a review

F. DOCHY
University of Leuven, Belgium

M. SEGERS
University of Maastricht, The Netherlands

D. SLUIJSMANS
Open University of the Netherlands, The Netherlands

ABSTRACT   *The growing demand for lifelong learners and reflective practitioners has stimulated a re-evaluation of the relationship between learning and its assessment, and has influenced to a large extent the development of new assessment forms such as self-, peer, and co-assessment. Three questions are discussed: (1) what are the main findings from research on new assessment forms such as self-, peer and co-assessment; (2) in what way can the results be brought together; and (3) what guidelines for educational practitioners can be derived from this body of knowledge? A review of literature, based on the analysis of 63 studies, suggests that the use of a combination of different new assessment forms encourages students to become more responsible and reflective. The article concludes with some guidelines for practitioners.*

Introduction

A New Era of Assessment

Alternatives in assessment have received much attention in the last decade and several forms of assessment have been introduced in higher education (see Birenbaum & Dochy, 1996). In Europe, as well as in the USA and Australasia, leading experts are claiming that the era of testing has changed in recent years into an era of assessment (Birenbaum, 1996). The era of testing can be characterised by a complete separation of instruction and testing activities, by a measurement that was passively undergone by the students, by measurement of knowledge of decontextualised subject matter that was unrelated to the student's experiences, and by measuring products solely in the form of a single total score (Wolf *et al.*, 1991). The assessment era promotes integration of assessment and instruction, seeing the student as an active person who shares responsibility, reflects, collaborates and conducts a continuous dialogue with the teacher. Assessment is then characterised by a pluralistic approach and by the use of interesting real-life (i.e. authentic) tasks (Segers, 1996). Moreover, new ideas have been developed concerning the main function of assessment. Assessment procedures are seen not only as serving as tools for crediting students with recognised certificates but also as

valuable for the monitoring of students' progress and to direct them, if needed, to remedial learning activities. Additionally, there is a strong support for representing assessment as a tool for learning (Arter, 1997; Dochy & McDowell, 1997). The view that the assessment of students' achievements is solely something which happens at the end of a process of learning is no longer tenable.

Changing Goals of Higher Education

For many years, the main goal of higher education has been to make students knowledgeable within a certain domain. Building a basic knowledge store was the core issue. Recent developments within society, such as the increasing production of new scientific knowledge and the use of modern communication technology, have encouraged us to implement new methods that are in line with these developments (Dochy & McDowell, 1997). These new methods, such as case-based and problem-based learning, are directed towards producing highly knowledgeable individuals, but do also stress problem-solving skills, professional skills and authentic learning, i.e. learning in real-life contexts:

> ... successful functioning in this era demands an adaptable, thinking, autonomous person, who is a self-regulated learner, capable of communicating and co-operating with others. The specific competencies that are required of such a person include
> (a) cognitive competencies such as problem solving, critical thinking, formulating questions, searching for relevant information, making informed judgements, efficient use of information, conducting observations, investigations, inventing and creating new things, analysing data, presenting data communicatively, oral and written expression;
> (b) meta-cognitive competencies such as self-reflection and self-evaluation;
> (c) social competencies such as leading discussions and conversations, persuading, co-operating, working in groups, etc. and
> (d) affective dispositions such as for instance perseverance, internal motivation, responsibility, self-efficacy, independence, flexibility, or coping with frustrating situations. (Birenbaum, 1996, p. 4)

Assessing the attainment of such goals will surely mean a change in our current assessment practices (Birenbaum & Dochy, 1996). Hence assessment will have to go beyond measuring the reproduction of knowledge.

It is not clear whether this change in perspective regarding assessment has been forced by the rising pressure of the labour market on education or whether the drives towards the change were generated within higher education itself. It is, however, clear that the main goal of higher education has moved towards supporting students to develop into 'reflective practitioners' who are able to reflect critically upon their own professional practice (Schön, 1987; Falchikov & Boud, 1989; Kwan & Leung, 1996). Students taking up positions in modern organisations need to be able to analyse information, to improve their problem-solving skills and communication and to reflect on their own role in the learning process. People increasingly have to be able to acquire knowledge independently and use this body of organised knowledge in order to solve unforeseen problems. As a consequence, higher education should contribute to the education of students as lifelong learners.

An increasing need for lifelong learning in modern society (Sambell & McDowell, 1997) will enhance the need for learning throughout one's entire working life. In such an era, traditional testing methods do not fit well with such goals as lifelong learning, reflective

thinking, being critical, the capacity to evaluate oneself, and problem-solving (Dochy & Moerkerke, 1997).

If the basics of higher education have to be directed towards lifelong learning, the approach to assessment will have to be in harmony. Research has shown that the nature of assessment tasks influences the approaches which students adopt to learning. Traditional assessment approaches can have effects contrary to those desired (Beckwith, 1991).

Research Questions

This article concentrates on one particular aspect of assessment, namely, assessments in which students play a role as assessors. The literature review focuses on forms of self-, peer and co-assessment from the point of view of their applicability in higher education. The following research questions are addressed.

(1) What are the main findings from research on self-, peer co-assessment?
(2) In what way can these results be brought together?
(3) What guidelines can be derived for educational practitioners from this body of knowledge?

Method

Selection of Studies

In order to answer these questions a literature search was conducted. The databases of the Educational Resources Information Center (ERIC) (1987–98) and of Current Contents on Disk (1996–98) were searched. The keywords used were self-assessment, peer assessment and co-assessment. Through the so-called 'snowball method' the references in all the aforementioned materials were checked for other studies.

The Internet was searched with the Alta Vista search engine. The following criteria were set to determine whether literature would be included in our study.

1. 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.
2. The studies had to be related to students in higher education. Thus, studies dealing with peer assessment of university personnel were excluded.

In total 63 articles were selected for further analysis.

Method of Analysis

A narrative review of the literature is used. This form of conventional literature review implies a careful reading of separate studies and integration of their findings (Slavin, 1986). A statistical meta-analysis could not be undertaken because only one of the selected studies included a control group and an experimental group.

Results

Main Findings from Research on Self-, Peer and Co-assessment in Higher Education

We present the results related to the first and second research questions in the following way.

In four separate sections, the different combinations of self-, peer, and co-assessment are treated. For many of the studies reviewed, peer and self-assessment marks are compared with teacher marks in order to indicate the accuracy of peer or self-assessment. If it is not clearly indicated that there is a working together between assessor and assessees, we present these studies under the heading of peer assessment and/or self-assessment.

Within the sections, we maintain a similar structure. First, definitions are given. Then, the main findings are presented. Finally, the implications for practice are presented.

Self-assessment

Definition.   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. It is 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.

Main findings.   The literature on self-assessment concerns six main topics: the influence of different abilities of students on the accuracy of self-assessment, the time effect, the accuracy of self-assessment in relation to teacher assessment, the effect of self-assessment, methods for self-assessment and the content of the self-assessment.

The influence of different abilities.   Boud & Falchikov (1989) analysed studies published between 1932 and 1988, which investigated student self-ratings compared to the ratings of students by teachers, and reported the overrating and the underrating of students. They related these findings to the different abilities of students. Their finding was that good students tended to underrate themselves and that weaker students overrated themselves. Students in higher-level classes could better predict their performance than students in lower-level classes.

Time effect.   Griffee (1995) investigated the question of whether there was a difference in student self-assessment between first year, second year, and third year classes in a university department. The general conclusion was that there was no difference between the years. All students tended to rate themselves lower at the beginning of the academic year and higher as the semester progressed. As the semester progressed students gained more confidence in their ability to perform. The teacher 'intervention' during the year may account for the fact that there is no difference between the self-assessments of the three classes. Several studies confirm that the ability of students to rate themselves improves in the light of feedback or development over time (Boud & Falchikov, 1989; Griffee, 1995; Birenbaum & Dochy, 1996). 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.

Accuracy.   Longhurst & Norton (1997) designed a study to investigate how accurately 67 second year psychology students would be able to assess their own essays, and also to ascertain whether or not the students understood what taking a deep approach in their essays actually meant. Student grades were compared with tutor grades. The students were asked to rate themselves, in relation to one specified essay, on tutor-specified criteria which were designed to measure a deep approach, the expected grade and their level of motivation. The

tutor did not see the self-assessments since the self-assessment sheet was removed from each essay. The tutors also marked the essays against a set of deep processing criteria. Results show that the tutor grade for the essay correlated highly with the set of deep processing criteria (r between 0.69 and 0.88). There was also a positive correlation between students' and tutors' grades ($r = 0.43$). The results further indicated that, overall, students were 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.

Zoller & Ben-Chaim (1997) investigated the self-assessment ability of 71 biology majors enrolled in a 4-year college programme, with respect to Higher Order Cognitive Skills (HOCS) and their confidence in self-assessing. A specially designed self-assessment question-naire consisted of HOCS-related questions, interdisciplinary questions (oriented towards science, technology, environment and society) and Likert-type items related to students' confidence. Students assessed their knowledge and understanding on this questionnaire. Results indicated that the students evaluated themselves as quite knowledgeable. The results further showed that 75% of the students thought they were capable in self-assessing and peer assessing. Zoller & Ben-Chaim found a discrepancy between the students' assessments and the teachers' assessments, which they explained in terms of the lack of integration between assessment and learning in contemporary science teaching.

Effect.   In research conducted by Hassmèn *et al.* (1996), 128 women learned the correct answers on a specific task by either performing or observing. Participants took either a performance test 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 will obtain a higher percentage of correct responses during learning trials on a test than those who learn without self-assessments.

This is also illustrated in a study reporting successful language learning. McNamara & Deane (1995) designed a variety of activities that foster self-assessment. Three of them were: writing letters to the teacher, keeping a daily language learning log, and preparing an English portfolio. These activities were shown to help students to identify their strengths and weaknesses in English, to document their progress and to identify effective language learning strategies and materials. The students also became aware of the language learning contexts that worked best for them and were able to establish goals for future independent learning.

Methods for self-assessment.   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 which a student felt were his or her best or strongest. A second approach was to apply a Likert scale to a group of designated abilities ('in comparison to others of the same age, my art ability is: excellent; above average; average; below average; poor'). The third approach was, for each ability, to provide different examples of the ability's applications, on which individuals rated their performance level from high to low, and subsequently these were summed to obtain a total score. The self-assessment forms Harrington described were seen as cheaper and less time-intrusive than traditional ways of assessing students (Nevo, 1995).

An electronic interactive advice system for self-assessment was provided by Gentle (1994). The aim of this system was to see how accurately students were able to assess their own work without the involvement of their supervisor. The system was based on question and answer screens for 38 skills. These skills were arranged in four sections:

(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 was described 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 showed that students were able to assess themselves to within five percentage points. Students become more aware of the quality of their own work. They could predict their own mark and, while they were doing this, they reflected on their behaviour. Since the student reflected more often than once on his or her work, a higher standard of outcomes ensured. According to Gentle, the system is less time-consuming than conventional self-assessment: the supervisor has a minor part in the assessment.

Another instrument was used by Anderson & Freiberg (1995). These authors used an audiotape self-assessment instrument for student teachers to reflect on their teaching. This instrument—called the Low Inference Self-assessment Measure (LISAM)—was developed to enable student teachers to 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 & Freiberg gave three reasons why the LISAM was practical and effective:

(1) the use of LISAM made student teachers more independent, provided feedback and stimulated them to reflect on their own teaching;
(2) student teachers could practice LISAM immediately; and
(3) the LISAM records of teaching behaviours were observable and hence provided a basis for considering whether change was desirable.

Boud (1992, 1995) developed a self-assessment schedule in order to provide a comprehensive and analytical record of learning in situations in which students had substantial responsibility for what they did. The main guidance was a handout which suggested the headings students might use. The headings were goals, criteria, evidence, judgements and further action.

Adams & King (1995) identified activities that might develop self-assessment skills. A framework helped students to develop skills in self-assessment. Adams & King identified three levels of activity. At the first level students worked on understanding the assessment process. Students performed activities such as: discussing good and bad characteristics of sample work, discussing what was required in an assessment, and critically reviewing the literature. At the second level students worked to identify important criteria for assessment.

At the third level students worked towards playing an active part in identifying and agreeing assessment criteria and being able to assess peers and themselves competently.

The idea of self-assessment for use in portfolios was described by Keith (1996). She suggested self-assessment assignments which asked students to report on their own learning. Assignments included sharing preconceptions about teaching and learning; comparing goals; creating a community of learners; generating student explanations and improving communication; group quizzes; challenging thinking dispositions; post-test evaluations; and collaborative assessing. The roots of all the described assignments lay in collaborative learning. The assignments encouraged students to feel responsible for their own learning. Keith found that the most influential variable for effective learning was the amount of meaningful energy the students put in.

Content. At the content level, it is striking that self-assessments are mostly used formatively (to foster skills and abilities) (Birenbaum & Dochy, 1996). For example, students of the Alverno College in Milwaukee have to develop problem-solving as one of eight abilities in order to graduate (Loacker & Jensen, 1988). At the heart of the educational process at Alverno stands assessment. The college sees it as a natural part of encouraging, directing and providing for development of abilities. Since self-assessment has been integrated into the students' problem-solving processes faculty have found that students show increasing understanding of interrelationships of ability, content, and context. Students take responsibility for their learning in a dynamic, continuing process. They gradually internalise their practice of problem-solving and their ability to self-assess.

Implications for practice. Overall, it can be concluded that research reports positive findings concerning the use of self-assessment in educational practice. Students who engage in self-assessment tend to score most highly 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, a higher standard of outcomes, responsibility for one's own learning and increasing understanding of problem-solving. The accuracy of the self-assessment improves over time. This accuracy is enhanced when teachers give feedback on students' self-assessment.

Adams & King's (1995) three levels for introducing self-assessment were noted earlier, and were found helpful in stimulating students' capacity to self-assess. Finally, a student's motivation influences the accuracy of self-assessment (Longhurst & Norton, 1997). Motivation can be enhanced by creating an educational setting where self-assessment is an inherent part of the learning process (Zoller & Ben-Chaim, 1997). This implies a learning environment in which the student's learning and not the teacher's teaching is central. A variety of instruments is available, including Likert scales, ability listings, written tests, portfolios, audiotape assessments and electronic interactive systems.

Peer Assessment

Definition. Falchikov (1995) defines peer assessment as the process through which groups of individuals rate their peers. This exercise may or may not entail previous discussion or agreement over criteria. It may involve the use of rating instruments or checklists which have been designed by others before the peer assessment exercise, or designed by the user group to meet its particular needs.

Somervell (1993) indicates that at one end of the spectrum peer assessment may involve feedback of a qualitative nature or, at the other, may involve students in marking. The assessment may be formative or summative and could form part of a larger scheme through

which peer feedback is given prior to self-assessment by the recipient of the feedback. Somervell stresses that peer assessment is not only a grading procedure, it is part of a learning process through which skills are developed. Peer assessment can be seen as a part of the self-assessment process and as informing self-assessment. The students have an opportunity to observe their peers throughout the learning process and often have a more detailed knowledge of the work of others than do their teachers. Keaten *et al.* (1993) report that peer assessment is a practice that can foster high levels of responsibility among students, requiring that the students be fair and accurate with the judgments they make regarding their peers.

Peer evaluation is also an alternative term to peer assessment (Weaver & Cotrell, 1986). Peer evaluation 'emphasises skills, encourages involvement, focuses on learning, establishes a reference, promotes excellence, provides increased feedback, fosters attendance, and teaches responsibility' (Weaver & Cotrell 1986, p. 25). Different forms of assessment are distinguished by Kane & Lawler (1978):

- *peer ranking*, which consists of having each group member rank all of the others from best to worst on one or more factors;
- *peer nomination*, which consists of having each member of the group nominate the member who is perceived to be the highest in the group on a particular characteristic or dimension of performance; and
- *peer rating*, which consists of having each group member rate each other group member on a given set of performance or personal characteristics, using any one of several kinds of rating scale.

Main findings. The studies of peer assessment focus on four different aspects: validity, fairness, accuracy and effects.

Validity. Dancer & 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. It is sometimes important to determine a individual's contribution to a group project, and hence training is of importance. Many studies of reliability appear actually to be studies of validity since they compare peer assessments with assessments made by professionals rather than with those of other peers or the same peers over time. Topping (1998) reviewed 31 studies and concluded that the majority of these studies (18) showed an acceptably high validity and reliability in a variety of fields and only seven studies found the validity or reliability to be unacceptably low.

Fairness. A study by Conway *et al.* (1993) indicated that students found group projects more interesting than traditional methods of teaching. Since the fairness of the (traditional mode of) 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. Second, 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 could be awarded individual marks, reflecting personal effort, for group projects. Conway *et al.* adopted the procedures suggested by Goldfinch & Raeside (1990) and simplified these. The results, using this method of calculating an individual weighting factor, showed that students felt that the peer assessment was a good method and sufficiently fair. Students felt that they should play a part in the assessment in order to make assessment results more objective.

In some studies the perceptions of students regarding innovative assessment and the

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 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.

Accuracy. The studies investigating the accuracy of peer assessment do not show consistent results. Two studies report on the peer assessment of essays and presentations. Oldfield & 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. Results show high correlations between student marks and lecturers' marks for individual essays and presentations. Fry (1990) describes a study in which the tutor introduced peer marking. The tutor first marked the scripts of the students and then handed them over to the students. The tutor asked the students to mark each others' work according to a marking scheme. The agreement between the tutor marks and the students' marks was generally very high. Fry's findings are confirmed by Rushton, *et al.* (1993), who developed a computerised peer assessment tool. Thirty-two computer science undergraduates were asked to write an essay on the viability of peer assessment. They typed their essays on the subject of peer assessment into the computer system. The class was split into groups of three or four students. Each group member used a peer assessment 'window' to mark the others' work. Contrary to expectations, the marks awarded by the peers were remarkably similar to those awarded by the tutors, suggesting that peer and teacher assessments were equally reliable.

The results of a study by Orsmond *et al.* (1996) are less positive regarding accuracy of peer assessment. They described an experiment in peer assessment for a first year undergraduate animal physiology poster assignment. Thirty-nine pairs of students completed a poster assignment, having been 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 in the other laboratory against five criteria. Each criterion had a graded scale ranging from 0 to 4. The result of the peer assessment was that each poster was given a total mark, with a possible maximum of 20. After the students marked the posters the tutor also marked the posters without seeing the marks the students had given. Orsmond *et al.* found that there was 18% agreement between students and tutor, with 56% of the students overmarking and 26% of the students undermarking. The correlation between students and tutors was 0.54. These results are in line with the findings of Stefani (1992). Stefani reported 14% complete agreement between students and tutor, 58% of students overmarking and 28% undermarking. Peer marks (both in the case of under- and overmarking), however, did differ by less than 10 percentage points from the tutor. In the Orsmond *et al.* (1996) study, the students filled in a questionnaire which showed that 76% of them thought that 'the peer assessment had make them think more, and work in a more structured way' (p. 243).

Effects. Different positive effects are reported. Orsmond *et al.* (1996) found that students did enjoy carrying out the peer assessment and that it was beneficial to their learning. Keaten & Richardson (1993) also affirmed that peer assessment fostered an appreciation for the individuals' performance within the group and interpersonal relationships in the classroom. Cheng & Warren (1997) conducted a research in the English Department of the Hong Kong

Polytechnic University to gauge the 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 in a questionnaire with four items. The results of the questionnaire show that the students were mostly positive towards the peer assessment, but that only a few students thought that beginning students were able to conduct the assessment in a fair and responsible manner. (The same had been found earlier by Falchikov & Boud, 1989.) Further, the students were not entirely confident in their ability to assess their peers. There was, however, a positive shift over time in both attitudes and confidence.

Williams (1992) found that the vast majority of students saw benefits in peer assessment. Benefits were seen in three main categories: in comparison of approaches, in comparison of standards and in exchange of information. However, students found criticising their friends to be difficult (see also Strachan & Wilcox, 1996). Students also found peer assessment to be difficult or undesirable when guidelines for evaluation were not established first. The two major findings in the study of Williams (1992) were that students liked to have more say in how they approached their learning and its assessment and that students needed guidance and training in new role behaviours before this could actually happen. Cheng & Warren concluded from their study that there was a need for students to be given systematic and comprehensive training in how they were to assess their peers and how to establish criteria.

Implications for practice.   Experience from peer assessment indicates that peer assessment can be valuable as a formative assessment method and hence as a part of the learning process. Students become more involved, both in the learning and in the assessment process. They find peer assessment sufficiently fair and accurate. However, the following can also be observed during peer assessment (Pond *et al.*, 1995): friendship marking, resulting in overmarking; 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. These problems can be prevented by combining peer assessment with self-assessment or co-assessment. This may be why the majority of studies investigate these combinations of assessment forms.

Self- and Peer Assessment

Definition.   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 the student's own learning process and learning activities compared to those of the other members in the group or class.

Main findings.   Because of the above-mentioned disadvantages of peer assessment, almost all studies of combinations of assessment approaches were practically oriented and were seeking ways in which to achieve fairness, accuracy and the positive effects of self- and peer assessment on learning and students' satisfaction.

Fairness.   Cutler & Price (1995) described an investigation in which presentations and seminars, built into each of the 3 years of a geography programme, were peer assessed against a set of criteria. Self-appraisal forms were also a part of the assessment procedure. The results of the peer assessment showed that the majority of the students were happy and confident in being assessed by their peers. Half of the students felt that their assessments of their peers were accurate. A third of the students thought that they had improved in confidence, organisation of materials and their use of voice.

Accuracy. Most studies report positive results regarding the accuracy of self- and peer assessment. The studies have been conducted in courses from different disciplines (medicine, psychology, law). In a study described by Burnett & Cavaye (1980) fifth year medical students had to assess their peers as part of the examination. They also were asked to assess their own performance. Results show that the peer assessment highly correlated with the final grade ($r = 0.99$) and the staff assessment ($r = 0.93$) and that the self-assessments highly correlated with the results of the peer assessments ($r = 0.99$) [1]. Other studies have come up with similar findings (Falchikov, 1991; McDowell, 1995; Birenbaum & Dochy, 1996). These studies suggest that friendship rating should not be seen as a major problem.

Falchikov (1991) investigated under- and overmarking in a study which also included self- and peer assessment. The process of working together in a small group project was assessed by the group members, seven developmental psychology students. In the study a self/peer group process assessment checklist was developed to compare the assessments of task and maintenance functions in respect of a piece of coursework. The checklist contained 16 task functions (related to the task performance) and eight maintenance functions (related to the cooperation within the group). This list was developed with cooperation of the students, enabling the students to become familiar with its content. 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) which each group member including self had contributed to the 16 functions (group activities). The results showed that there was no tendency to over-or undermark when self-ratings were compared with peer ratings. There was a high level of agreement between peers. The students also were asked to give written feedback on this mode of assessment. Some preferred written evaluative comments to number ratings, and some students felt that this approach to assessment was not necessary because in a group one always has a certain responsibility. Strachan & Wilcox (1996) recommend, however, that it is important to give the student an active role in the development of assessment criteria. The process is thus just as important for the quality of learning as is the product.

Oldfield & Macalpine (1995) described an approach to self-assessment in achievable steps—first, comparisons of contributions to group activities excluding self; then including self; and finally a self-assessment of individual work. The students first made an individual assessment of the activities of all the group members (viewed as contributions to the groups' achievements). To train in the skills of self-assessment, students also had to do this for their own group. The same procedure took place within the group: students first assessed the group members and then their own contributions. Oldfield & Macalpine found that this assessment procedure strengthened the confidence of students to assess the work of others and that of themselves.

Effects. A set of studies describe the effect of peer and self-assessment. There seems to be a striking difference between the results based on expectations and perceptions of students and those based on test scores.

Warkentin *et al.* (1995) investigated self- and peer assessment in a study with 83 undergraduate educational psychology students. They hypothesised that students taking tests using individual and group assessments would perform significantly better on a post-test of educational psychology course concepts than students who took the traditional tests (individual examinations). The effects on student learning were examined. The results indicated that there were no significant differences between the two groups on achievement and knowledge structure. Warkentin *et al.* (1995) found, however, that the reactions to the self- and peer-assessment procedure they used were overwhelmingly positive. The students did like the

group assessment and thought it contributed to their learning through this process as they discussed and debated test items.

Sambell & McDowell (1997) studied six cases which included peer and/or self-assessment, and found that students were generally positive towards an involvement in the assessment process. Students' awareness was high that self- and peer assessment had helped them to develop important skills, such as problem-solving. 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 liked self- and peer assessment. This finding cannot be interpreted, however, as indicating that the students were totally committed to traditional forms of assessment. The possibility of being able to redraft assignments after tutor feedback was viewed more favourably.

Implications for practice.   The development of criteria through active cooperation between teacher and students seems to be a critical success factor for self- and peer-assessment (Falchikov, 1991; Strachan & Wilcox, 1996), as is the development of a series of instructions for students to set criteria for themselves. A third critical success factor is congruence between the mode of group activity and the evaluation of the group work (Falchikov, 1991). Oldfield & Macalpine (1995) suggest a stepwise approach to group project assessment starting with peer assessment from the other groups' work, moving on to peer assessment of their own group's work, and ending with self-assessment.

Until now there has been little empirical evidence of the effect of self- and peer assessment on students' performance on traditional individual examinations which measures their knowledge base. However, most studies report positive expectations and perceptions of students regarding the contribution of self- and peer assessments to their learning.

Self- and Peer Assessment Related to Co-assessment

In previous sections the use of self-assessment, peer assessment, and the use of a combination of these two forms have been discussed. One step closer to traditional assessment practice are assessment procedures in which the tutor plays a significant role in the process.

Definition.   Co-assessment, the participation of students and staff in the assessment process, is a way of providing an opportunity for students to assess themselves whilst allowing the staff to maintain the necessary control over the final assessments (Hall, 1995).

Co-assessment can be used for summative purposes whereas self- and peer assessment tend to be 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 *et al.* (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 over the details of the assessment and discussing any misunderstandings that exist, and is consistent with the less confrontational approach to assessment increasingly being sought, through the development of an ongoing relationship between the assessor and assessee. Synonyms for co-assessment are collaborative assessment and cooperative assessment.

Main findings.   Co-assessment is often related to forms of self- and peer assessment. We found a single study which combined self- and co-assessment (Hall, 1995), in which the students and staff jointly set the criteria. In Hall's account three purposes of co-assessment are identified. One is to assist the student teachers in making the role-change from being student to being a teacher, a second is to provide insights into the assessment process which may be of use to them in assessing their own students, and a third is to provide a skill-development step towards self-assessment. The process involved a double-sided cover-sheet for an assignment. On the back of this sheet the students had the opportunity to give their own self-assessment and then hand it to the staff member. The staff member then used the other side of the sheet to record his or her assessment of the student's work. Then the staff member turned the sheet over to see whether or not the student had chosen to offer their own assessment on the reverse. The findings were that generally the staff member's grade was higher than the student's grade.

Orpen's (1982) study combined co- and peer assessment. Twenty-one students in an organisational behaviour course and 21 students in a political philosophy course were required to write an essay, having been 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 following criteria: coverage of the relevant material, coherence and strength of the underlying argument, and fluency and clarity of expression. Results show that 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 course-end examinations.

A number of studies deal with combinations of self-, peer and co-assessment. Falchikov (1986) aimed to implement and evaluate a method of collaborative self- and peer assessment. First, the tutors set criteria and ranked these criteria in terms of their relative importance. Then students set criteria and comparisons were made between the criteria generated by the tutors and the students. 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 did appear to be comparable with traditional tutor methods of assessment, whereas collaborative and peer assessment corresponded less well with either tutor or self grading. Stefani (1992) carried out an experiment in collaborative self- and peer assessment of a first year undergrad-uate biochemistry laboratory practical experiment. The students themselves defined the marking schedule for a scientific report. The results showed that students had a realistic perception of their own abilities and could make rational judgements on 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.

Horgan *et al.* (1997) used the predictions of grades, actual grades, peer reviews, and reflective essays on self-assessment of undergraduate teacher education students to analyse the relationships between self-, peer and co-assessment. The students were trained in self-assessment. The students had to complete three multiple-choice examinations, of which the third was a cumulative final. Students had to predict their grade and after the examination they had to reflect on their performance. The students also had to undertake a written analysis of a case study, which was self-assessed and then reviewed by three peers and the instructor against five criteria. A third part of the assessment procedure Horgan *et al.* (1997) described was an oral case analysis as part of a group. These individual presentations were

also reviewed by peers. The final part was an essay reflecting on the self-assessment activities. Results of the assessments described here 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, whereas weaker students used feedback to find the errors.

Two studies (Freeman, 1995; Kwan & Leung, 1996) present findings of tutor and peer assessment. Kwan & Leung (1996) report the results of tutor and peer group assessment of student performance of 96 students in a simulation exercise on hotel personnel training. The group was divided into five tutorial sub-groups. Then students were paired and each student had to conduct a training session with the partner in front of 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 than that reported by Falchikov (1986) and Stefani (1994). The weaker level of agreement may be due to student unfamiliarity with self-assessment (it was their first attempt), and to the fact that the students had made no contribution to the identification of the criteria, and to the lack of any negotiation between tutor and students in understanding the criteria.

Freeman (1995) conducted a peer assessment experiment with 210 final year undergraduate business students. Students were divided into 41 teams, and each team had to complete two of the four assessable tasks. The presentation, one of the two tasks, was chosen by staff as a vehicle for experimenting with a peer assessment worth 25% of the overall grade of the students. In the first week of the semester each student was given the presentation marking and feedback sheet with 22 items, eight items related to the content and 14 related to the presentation, weighted 60% and 40% respectively. Freeman found that the quality of the presentations was rated very highly by staff and peers. There was no significant difference between the average staff ratings and average peer ratings. Students, however, tended to undermark the good presentations and overmark the poor presentations.

Implications for Practice

The findings indicate that the use of the combination of self-, peer, and co-assessments is effective. The results regarding accuracy (Orpen, 1982; Stefani, 1992; Falchikov, 1986; Horgan *et al.*, 1997) indicate that self- and peer assessment can be used for summative purposes as a part of the co-assessment, by giving the tutor the power to express the final decision about a process or a product. In this way the traditional assessment, where 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 the student becomes teacher, this role-change provides him or her with insights into the assessment process.

The studies regarding accuracy show the importance of the setting of criteria, jointly by peers and teacher or by the students independently (Stefani, 1992; Falchikov, 1986; Kwan & Leung, 1996). Horgan *et al.* (1997) additionally stress the effect of time and training. In contrast to Warkentin *et al.* (1995), who find no relation between self- and peer assessment marks and test performances, Orpen (1982) shows a correlation between peer and co-assessment marks and performances in course-end examinations. Probably there are differ-

ences between both settings in the way that the criteria were developed and in the extent and the mode of training of the students.

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

Where application of self-assessment and peer assessment has been mostly used for formative purposes, combinations of these forms with co-assessment do appear to have worked out well for summative assessments. There are various possibilities, in the combination of different approaches, ranging from using the peer assessments as a contribution of, say, 25% to the overall score, to using peer assessments as a correction score for tutor assessment. Developments in this area do clearly open up the possibility of assessing skills and abilities in circumstances in which higher education has traditionally had problems: they may also help to keep down the costs of assessing. If peer and co-assessment indeed is a valid, fair and useful method for assessing essays and assignments, it may well become more widespread in the near future.

Improving the Quality of Learning

Overall, it does appear that self-, peer and co-assessment do improve different aspects of the quality of learning of students. Apart from the fact that the research in this field suggests that these assessment methods do fit better with more problem-based and authentic learning contexts and are mostly valid and reliable methods (Topping, 1998), we tried to summarise the concrete effect. Searching for an answer to the third research question, we detected eight positive effects of self-, peer, and/or co-assessment which arise from our body of research.

1. Increased student confidence in the ability to perform (Orpen, 1982; Cutler & Price, 1995; Griffee, 1995).
2. The increased awareness of the quality of the student's own work (Gentle, 1994; Anderson & Freiberg, 1995; McNamara & Dean, 1995).
3. Increased student reflections on their own behaviour and/or performance (Gentle, 1994; Anderson & Freiberg, 1995; McNamara & Dean, 1995; Longhurst & Norton, 1997; Sobral, 1997).
4. Increased student performance on assessments, increased quality of the learning output (Loacker & Jensen, 1988; Stefani, 1992; Cutler & Price, 1995; Freeman, 1995; Warkentin *et al.*, 1995; Orsmond *et al.*, 1996; Hassmen, 1997; Horgan, *et al.*, 1997; Martens & Dochy, 1997; Sambell & McDowell, 1997).
5. Effectiveness of approaches to learning (McNamara & Deane, 1995).
6. Taking responsibility for learning; the independence of students (Loacker & Jensen, 1988; Keaten *et al.*, 1993; Anderson & Freiberg, 1995).
7. Increased student satisfaction (Williams, 1992; Conway *et al.*, 1993; Boud, 1995; Cutler & Price, 1995; Warkentin *et al.*, 1995; Orsmond *et al.*, 1996; Peters, 1996; Cheng & Warren, 1997; Sambell & McDowell, 1997).
8. Ameliorated learning climate (Keaten & Richardson, 1993).

Guidelines for educational practitioners derived from this body of knowledge.

Boud (1989), Boud & Knight (1994) and Falchikov & Boud (1989) elaborate on the issue of the impact of assessment on learning. Falchikov & Boud (1989) stated in their discussion:

Although we have focused on student-teacher agreement over rating, we must not

be distracted by the search to maximise 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, the main reason why these forms of assessment need to be integrated in curricula in higher education in their impact on the learning process. Boud (1995, p. 41) stresses that an important concept that links assessment with the quality of learning is that of consequential validity. This 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 students have in being able to assess themselves. 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 & Knight, 1994). The relationship between reflection and self-assessment is also pointed out by Sobral (1997). Self-assessment of self-directed learning supports reflection and learning partnerships and is facilitated by discussions and exercises. Longhurst & Norton (1997) state that self-assessment is clearly an important part of helping students to improve their own learning, as it focuses students' attention on the metacognitive aspects of their learning and teaches them to be more effective at monitoring their own performance.

A second issue is the relationship between higher education and professional life. Boud (1990) recognised the gap between what is required of students in higher education and what happens in real life. 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 & King (1995) 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 deal with weaknesses as appropriate.

The following guidelines have emerged from our study.

1. Training in the skill to self-assess or to peer assess has to be provided in order to obtain an optimal impact on the learning process, at least for beginning students. The first assessments which involve students as assessors should perhaps be implemented with groups of third and fourth year university students.
2. Self-assessment takes time, and sometimes support for students will be necessary during the self-assessment.
3. Self-assessment can be used fairly easily for formative purposes. Students should learn to see this as a tool for learning.
4. The habit of academics to do the teaching and all the marking is hard to change, and it seems likely that a staff development programme will be needed if the approaches discussed in this article are to be implemented widely.
5. In peer assessment, criteria should be determined beforehand. Experiences show that it works well if these criteria are determined jointly by staff and students.
6. Peer assessment criteria should be presented in operational terms with which all students are familiar. Students can play a role in the process of operationalisation.
7. Peer assessment can be used as a tool for summative assessment, in combination with other assessment instruments. It can lead to a student profile or a peer assessment factor, i.e. a correction factor calculated from the peer assessment scores that adjusts a preceding

group score for a collaborative product. Peer assessment measures should not be used as the sole indicator in a summative assessment.

Concluding Observations

There is much evidence which supports the view that students' contributions to assessment can be consistent with the assessment of staff, and of other students. There is also empirical evidence that the students perceive positive effects. Involving students in assessment is perceived as being valid, reliable, fair and as contributing to a growth in competence. Some issues, however, need further research.

First, in our search we only found two studies referring to the measures of students' performances on end-of-course individual examinations, measuring student's knowledge representations (Orpen, 1982; Warkentin *et al.*, 1995). A number of studies refer to the assessment of essays, group projects and presentations. Although they may play an important role in formative assessment, end-of-course examinations are widely used for summative reasons. It can be expected that by reflecting on their own thinking and performing and that of peers, students will perform better on end-of-course examinations measuring their knowledge representations and skills. However, there is as yet little evidence that this is the case.

Second, more studies investigating the generalisability of peer and co-assessment, and studies scrutinising the different effects of different scoring methods and the use of different scales in peer and co-assessment could give information on how to improve the quality of these assessments.

Third, most of the studies that we reviewed do not elaborate on the characteristics of the learning environment in which the study took place. A comparative study of the effects of the use of self-, peer, and co-assessment in different educational settings could give valuable information on the educational conditions for the effective use of these methods of assessment. One context in which these methods could be particularly useful is the problem-based learning environment. Peer and co-assessment are inherent aspects of working on problems within small tutorial groups. To what extent do these tutorial groups influence the accuracy and the effects of these methods of assessment in comparison with, for example, a merely teacher-centred educational setting with only a few project-based courses? Do students in these different settings have different perceptions and needs with respect to the effective use of self-, peer and co-assessment?

Finally, the use of self-, peer, and co-assessment is consistent with the need of society for lifelong learners who reflect continuously on their behaviour and the learning processes they experience (Moerkerke, 1996). Studies which investigate the long-term effects of these methods of assessment would be very valuable.

Correspondence: Prof. Filip Dochy, University of Leuven, Department of Instructional Sciences, Vesaliusstraat 2, 3000 Leuven, Belgium; e-mail: Filip.Dochy@Ped.Kuleuven.ac.be

NOTE

[1] Nevertheless, we have experienced that, in our own research on peer assessment, the problem lies more in the weakest students who overrate themselves and are not able to judge the peers accurately. Scores from such students are often statistical outliers. In our own investigations, we therefore exclude the highest and the lowest peer assessment score for each individual in order to calculate the mean scores.

REFERENCES

ADAMS, C. & KING, K. (1995) Towards a framework for student self-assessment, *Innovations in Education and Training International*, 32, pp. 336–343.

ANDERSON, J.B. & FREIBERG, H.J. (1995) Using self-assessment as a reflective tool to enhance the student teaching experience, *Teacher Education Quarterly*, 22, pp. 77–91.

ARTER, J. (1997) Using assessment as a tool for learning, in: R. BLUM & J. ARTER (Eds) *Student Performance Assessment in an Era of Restructuring* pp. 1–6 (Alexandria, VA; Association for Supervision and Curriculum Development).

BECKWITH, J.B. (1991) Approaches to learning, their context and relationship to assessment performance, *Higher Education*, 22, pp. 17–30.

BIRENBAUM, M. (1996) Assessment 2000: towards a pluralistic approach to assessment, in: M. BIRENBAUM & F. DOCHY (Eds) *Alternatives in Assessment of Achievement, Learning Processes and Prior Knowledge*, pp. 3–31 (Boston, MA, Kluwer Academic).

BIRENBAUM, M. & DOCHY, F. (Eds)(1996) *Alternatives in Assessment of Achievement, Learning Processes and Prior Knowledge* (Boston, MA, Kluwer Academic).

BOUD, D. (1989) The role of self-assessment in student grading, *Assessment and Evaluation in Higher Education*, 14, pp. 20–30.

BOUD, D. (1990) Assessment and the promotion of academic values, *Studies in Higher Education*, 15, pp. 101–111.

BOUD, D. (1992) The use of self-assessment schedules in negotiated learning, *Studies in Higher Education*, 17, pp. 185–200.

BOUD, D. (1995) *Enhancing Learning through Self-assessment* (London and Philadelphia, Kogan Page).

BOUD, D. & FALCHIKOV, N. (1989) Quantitative studies of self-assessment in higher education: a critical analysis of findings, *Higher Education*, 18, pp. 529–549.

BOUD, D. & KNIGHT, P. (1994) Designing courses to promote reflective practice, *Research and Development in Higher Education*, 16, pp. 229–234.

BURNETT, W. & CAVAYE, G. (1980) Peer assessment by fifth year students of surgery, *Assessment in Higher Education*, 5, pp. 273–278.

CHENG, W. & WARREN, M. (1997) Having second thoughts: student perceptions before and after a peer assessment exercise, *Studies in Higher Education*, 22, pp. 233–239.

CONWAY, R., KEMBER, D., SIVAN, A. & WU, M. (1993) Peer assessment of an individual's contribution to a group project, *Assessment and Evaluation in Higher Education*, 18, pp. 45–56.

CUTLER, H. & PRICE, J. (1995) The development of skills through peer assessment, in: A. EDWARDS & P. KNIGHT (Eds) *Assessing Competence in Higher Education*, pp. 150–159 (Birmingham, Staff and Educational Development Association).

DANCER, W.T. & DANCER, J. (1992) Peer rating in higher education, *Journal of Education for Business*, 67, pp. 306–309.

DOCHY, F.J.R.C. & McDOWELL, L. (1997) Assessment as a tool for learning, *Studies in Educational Evaluation*, 23, pp. 279–298.

DOCHY, F. & MOERKERKE, G. (1997) The present, the past and the future of achievement testing and performance assessment, *International Journal of Educational Research*, 27, pp. 415–432.

FALCHIKOV, N. (1986) Product comparisons and process benefits of collaborative peer group and self-assessments, *Assessment and Evaluation in Higher Education*, 11, pp. 146–166.

FALCHIKOV, N. (1991) Group process analysis in: S. BROWN & P. DOVE (Eds) *Self-and Peer Assessment*, pp. 15–27 (Birmingham, Standing Conference on Educational Development, Paper 63).

FALCHIKOV, N. (1995) Peer feedback marking: developing peer assessment, *Innovations in Education and Training International*, 32, pp. 175–187.

FALCHIKOV, N. & BOUD, D. (1989) Student self-assessment in higher education: a meta-analysis, *Review of Educational Research*, 59, pp. 395–430.

FREEMAN, M. (1995) Peer assessment by groups of group work, *Assessment and Evaluation in Higher Education*, 20, pp. 289–300.

FRY, S.A. (1990) Implementation and evaluation of peer marking in higher education, *Assessment and Evaluation in Higher Education*, 15, pp. 177–189.

GENTLE, C.R. (1994) Thesys: an expert system for assessing undergraduate projects, in: M. THOMAS, T. SECHREST & N. ESTES (Eds) *Deciding our Future: technological imperatives for education*, pp. 1158–1160 (Austin, TX, University of Texas).

GOLDFINCH, J. & RAESIDE, R. (1990) Development of a peer assessment technique for obtaining individual marks on a group project, *Assessment and Evaluation in Higher Education*, 15, pp. 210–231.

GRIFFEE, D.T. (1995) *A Longitudinal Study of Student Feedback: self-assessment, course evaluation and teacher evaluation* (Birmingham, AL).

HALL, K. (1995) Co-assessment: participation of students with staff in the assessment process. A report of work in progress, paper given at the 2nd *European Electronic Conference on Assessment and Evaluation*, European Academic & Research Network (EARN); http://listserv.surfnet.nl/archives/earli-ae.html

HARRINGTON, T.F. (1995) *Assessment of Abilities* (Greensboro, NC; ERIC Clearinghouse on Counseling and Student Services).

HASSMEN, P., SAMS, M.R. & HUNT, D.P. (1996) Self-assessment responding and testing methods: effects on performers and observers, *Perceptual and Motor Skills*, 83, pp. 1091–1104.

HORGAN, D.D., BOL, L. & HACKER, D. (1997) An examination of the relationships among self, peer, and instructor assessments, paper presented at the *European Association for Research on Learning and Instruction*, Athens, Greece.

KANE, J.S. & LAWLER III, E.E. (1978) Methods of peer assessment, *Psychological Bulletin*, 85, pp. 555–586.

KEATEN, J.A. & RICHARDSON, M.E. (1992) A field investigation of peer assessment as part of the student group grading process, paper presented at the *Western Speech Communication Association Convention*, Albuquerque, NM.

KEITH, S.Z. (1996) Self-assessment materials for use in portfolios, *Primus*, 6, pp. 178–192.

KWAN, K. & LEUNG, R. (1996) Tutor versus peer group assessment of student performance in a simulation training exercise, *Assessment and Evaluation in Higher Education*, 21, pp. 205–214.

LOACKER, G. & JENSEN, P. (1988) The power of performance in developing problem solving and self-assessment abilities, *Assessment and Evaluation in Higher Education*, 13, pp. 128–150.

LONGHURST, N. & NORTON, L.S. (1997) Self-assessment in coursework essays, *Studies in Educational Evaluation*, 23, pp. 319–330.

MARTENS, R. & DOCHY, F. (1997) Assessment and feedback as student support devices, *Studies in Educational Evaluation*, 23, pp. 257–275.

MCDOWELL, L. (1995) The impact of innovative assessment on student learning, *Innovations in Education and Training International*, 32, pp. 302–313.

MCNAMARA, M.J. & DEANE, D. (1995) Self-assessment activities: toward language autonomy in language learning, *Tesol*, 5, pp. 17–21.

MOERKERKE, G. (1996) *Assessment for Flexible Learning* (Utrecht, Lemma).

MOERKERKE, G. & DOCHY, F. (1997) Het toetsen van complexe vaardigheden [The assessment of complex skills], paper presented at the *Conference of the Dutch Educational Research Association*, University of Twente, Enschede, The Netherlands.

MOERKERKE, G. & TERLOUW, C. (1998) Herontwerp van toetsing [The redesign of testing], *THEMA* [Journal for higher education management], 16, pp. 19–25.

NEVO, D. (1995) *School-based Evaluation: a dialogue for school improvement* (Pergamon).

OLDFIELD, K.A. & MACALPINE, J.M.K. (1995) Peer and self-assessment at the tertiary level—an experiential report, *Assessment and Evaluation in Higher Education*, 20, pp. 125–132.

ORPEN, C. (1982) Student versus lecturer assessment of learning: a research note, *Higher Education*, 11, pp. 567–572.

ORSMOND, P., MERRY, S. & REILING, K. (1996) The importance of marking criteria in the use of peer assessment, *Assessment and Evaluation in Higher Education*, 21, pp. 239–249.

PAIN, H., BULL, S. & BRNA, P. (1996) A student model 'for its own sake'. Available electronically at http://cbl.leeds.ac.uk/~paul/papers/euroaiedpapers96/smpaper/smpaper.html

PETERS, M. (1996) Student attitudes to alternative forms of assessment and to openness, *Open Learning*, 11, pp. 48–50.

POND, K., UI-HAQ, R. & WADE, W. (1995) Peer review: a precursor to peer assessment, *Innovations in Education and Training International*, 32, pp. 314–323.

RUSHTON, C., RAMSEY, P. & RADA, R. (1993) Peer assessment in a collaborative hypermedia environment: a case study, *Journal of Computer-based Instruction*, 20, pp. 75–80.

SAMBELL, K. & MCDOWELL, L. (1997) The value of self- and peer assessment to the developing lifelong learner, in: C. Rust (Ed.) *Improving Student Learning—improving students as learners*, pp. 56–66 (Oxford, Oxford Centre for Staff and Learning Development)

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, pp. 349–371.

SCHÖN, D.A. (1987) *Educating the Reflective Practitioner: towards a new design for teaching and learning in the professions* (San Francisco, CA, Jossey-Bass).

SEGERS, M. (1996) Assessment in a problem-based economics curriculum, in: M. BIRENBAUM & F. DOCHY (Eds) *Alternatives in Assessment of Achievement, Learning Process and Prior Knowledge*, pp. 201–226. (Boston, Kluwer Academic).

SLAVIN, R. (1986) Best-evidence synthesis: an alternative to meta-analysis and traditional reviews, *Educational Researcher*, 15(9), pp. 5–11.

SOBRAL, D.T. (1997) Improving learning skills: a self-help group approach, *Higher Education*, 33, pp. 39–50.

SOMERVELL, H. (1993) Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment, *Assessment and Evaluation in Higher Education*, 18, pp. 221–233.

STEFANI, A.J. (1992) Comparison of collaborative, self, peer and tutor assessment in a biochemistry practical, *Biochemical Education*, 20, pp. 148–151.

STEFANI, L.A.J. (1994) Peer, self- and tutor assessment: relative reliabilities *Studies in Higher Education*, 19, pp. 69–75.

STRACHAN, I.B. & WILCOX, S. (1996) Peer and self-assessment of group work: developing an effective response to increased enrolment in a third-year course in microclimatology, *Journal of Geography in Higher Education*, 20, pp. 343–353.

TOPPING, K. (1998) Peer-assessment between students in colleges and universities, *Review of Educational Research*, 68, pp. 249–276.

WARKENTIN, R.W., GRIFFIN, M.M., QUINN, G.P. & GRIFFIN, B.W. (1995) An exploration of the effects of cooperative assessment on student knowledge structure, paper presented at the *Annual Meeting of the American Educational Research Association*, San Francisco, CA.

WEAVER II, R. & COTRELL, H.W. (1986) Peer evaluation: a case study, *Innovative Higher Education*, 11, pp. 25–39.

WILLIAMS, E. (1992) Student attitudes towards approaches to learning and assessment, *Assessment and Evaluation in Higher Education*, 17, pp. 45–58.

WOLF, D., BIXBY, J., GLENN, J. & GARDNER, H. (1991) To use their minds well: investigating new forms of student assessment, *Review of Research in Education*, 17, pp. 31–74.

ZOLLER, Z. & BEN-CHAIM, D. (1997) *Student self-assessment in HOCS science examinations: is it compatible with that of teachers?*, paper presented at the meeting of the *European Association for Research on Learning and Instruction*, Athens, Greece.