

Student involvement in performance assessment: A research project

Dominique Sluijsmans & George Moerkerke, Otec, Open University of the Netherlands

© 1999

Abstract

Introduction

Starting phase: motivation and context

Phase I: analysis of the complex skill of assessment and design training peer-assessment

Phase II: experiment I: training peer-assessment

Phase III: experiment II: training self-assessment and peer-assessment

Closing phase: generalizability of the training, recommendations, guidelines, limitations, possibilities

Discussion

References

Abstract

To develop the skills and competencies required in professional organisations, students have to reflect on their own behaviour and that of their peers. In the past, a shift could be observed from a static view to learning and intelligence to questions about how differences in complex skills of students could be described and trained. The assumption that self- and peer-assessment strategies are important in order to solve complex problems is widely acknowledged in education in so far, but giving students opportunities to carry out self- and peer-assessments is not enough: explicit attention in training students in these strategies is still the missing link. Students seldom get training and practice in the development of assessment strategies. Analysis of 62 studies showed that self- and peer-assessment can be effective tools to develop competencies needed in the working field (Sluijsmans, Dochy & Moerkerke, 1999). The literature shows that students miss the training in self- and peer- assessment. Other studies (Arter & Spandel, 1992; Perkins, 1986) also indicated that students need instruction in self- and peer-assessment, in order to make reliable judgements. Self- and peer-assessment are complex skills which need to be trained. The main purpose of the current research project is to design tools which can be helpful in training learners to become professionals in the skill to self-assess and peer-assess.

Introduction

The main purpose of the research project is to design tools which can be helpful in training learners to become better problem solvers. The training to be developed in this project is focused on the development of the complex skill to self-assess and the complex skill to peer-assess. Self-assessment is defined as the learners taking responsibility for monitoring and making judgements about aspects of their own learning. It requires learners to think critically about what they are learning, to identify appropriate standards of performance and to apply them to their own work. Peer-assessment involves the responsibility to make critical judgements about the learning of peers and applying the standards to the work of these peers. Self-assessment and peer-assessment can be regarded as complex skills because they can be conceived of as a set of constituent skills or as a skills hierarchy (Van Merriënboer, 1997). This does not entail that the whole skill is merely the sum of its constituents; it is a highly integrated hierarchy for which the constituent skills are controlled by higher level strategies (expert-strategies). Training in self- and peer-assessment can be supportive in achieving higher order performances. Self- and peer-assessment calls upon an active participation of the student in their performance. This indicates that self- and peer-assessment can be useful in the development of a wide range of skills.

The basic research question in the project is: How to support teachers and learners who are active participants in performance-based assessments by means of self- and peer-assessment?

Research will be conducted in the following phases:

starting phase: theoretical basis (motivation and context)

phase I: analysis of the complex skill of assessment and design training peer-assessment

phase II: experiment I: training peer-assessment

phase III: experiment II: training self-assessment and peer-assessment

closing phase: generalizability of the training, recommendations, guidelines, limitations, possibilities.

This article is structured according to these phases.

Starting phase: motivation and context

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. 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- and peer-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.

To empower the theoretical basis for the research project a literature review was conducted. In this literature review 62 studies were analysed. The studies were classified into self-, peer- and co-assessment. For this paper the results of the self-assessment are the most interesting ones (For the complete article we refer to Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999)). 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. Griffiee (1995) also investigated the question 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. Another explanation for the fact that there was no difference between the self-assessments of the three classes was the teacher intervention during the year.

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; Griffiee, 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, Sams, and Hunt (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.

Beside the literature review, also a pilot study was organised to gather practical evidence for the purpose of the research project. In the pilot-study, conducted in 1998, 27 university students subscribed to a fourth year course in educational sciences and practicals using problem based learning. Two groups ($n_1=13$; $n_2=14$) worked together for four periods (blocks) of six weeks. Assessment criteria were decided upon for the collaborative work resulting in the reports and for the performance of students. Based on four criteria a peer-assessment took place. After the peer-assessment, students were asked to write down their experiences with this kind of evaluation. The most positive aspects were that the students had the opportunity to give their opinion about the contribution of each peer-student in the group. In this method there were more persons who made a judgement. The students think that their scores could be helpful to the tutor. The involvement in the assessment was regarded as fair, although the majority of the students doubted the reliability of the method. Some students were stimulated to think critically about their own

learning behaviour since they also had to give themselves a score. The peer-assessment was not only product-evaluation but also process-evaluation.

The more negative aspects of the peer-assessment were that the contribution of the peers differed every block and that now a score had to be given that enclosed the average contribution of four blocks. Many students indicated that giving only a score is too simple and ineffective. There was no feedback moment. The criteria appeared to be difficult to interpret. One student introduced the idea to give a certain weight to the criteria, to indicate which criteria is important and which is of less importance. The peer-assessment was not sufficiently introduced. Some students experienced it as difficult and felt uncomfortable, because they had no prior experience in peer- and self-assessing. In the current situation at the university, the students think that the possibilities for this implementation in other courses are scarcely.

Experience from the 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 the way it is conducted in this study doesn't prevent 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, Ul-Haq & Wade, 1995).

The results of the study confirmed that adequate training in self- and peer-assessment strategies is priority. This training has to be embedded in the course domain, in order to ingrate the training and the instruction. An improvement could be that not only processes are evaluated in a peer-assessment but also the products. In the current study only the process was subject for assessment. A major improvement in future experiments would be that students are provided with a possibility to give feedback in function of following learning processes. The pilot study took place in a learning environment in which the development of professional/problem solving skills was central.

Phase I: analysis of the complex skill of assessment and design training peer-assessment

How can effective self-assessment and effective peer-assessment be identified? To answer this question the skill to assess has to be analysed. To analyse the complex skill of assessment, methods and techniques have to be carefully selected (Van Merriënboer, 1997). Merriënboer describes the following functions of skill decomposition:

Table 1: Functions of skill decomposition

Function	Description
identification	to identify the constituent skills that make up the complex cognitive skill;
description	to provide a clear description of each of the constituent skills involved
classification:	to classify each of the constituent skills involved as skills that must be selected for training;
sequencing:	to build a macro-level sequence for practising the selected constituent skills.

The first function of identification is important to analyse the skills of assessment and to develop an expert- and novice- model. To design a training which is focused on the growth in the skill to assess, it is important to identify specific novice-behaviour and specific expert-behaviour in assessment.

To have a complete picture of the idea of novice and expert, a short description of the novice- expert-theory may be helpful. Glaser (1984) and Larkin (1980) for example studied expert and novice problem solvers. They found that the distinction between them lies in the nature and amount of their organised knowledge. What experts are able to do is to look at a problem in various ways- to adopt new ways rapidly, and to discard ways rapidly. For any given problem area an expert calls upon a body of highly organised knowledge and skills which suggests a number of different mental sets. The expert is then able to try them out, one after another, if need be, in order to arrive at the desired solution (Gagné, 1988). Chi, Glaser and Farr (1988) described the following seven characteristics of expert performance:

- experts excel mainly in their own domains;
- experts perceive large meaningful patterns in their domain;
- experts are faster than novices at performing the skills of their domain, and they quickly solve problems with little error;
- experts have superior short-term and long-term memory;
- experts see and represent a problem in their domain at a deeper level than novices;
- experts spend a great deal of time analysing a problem qualitatively;
- experts have strong self-monitoring skills.

Glaser (1990) identified the four following constructs for dimensions of assessment, that has been generated by research on expert-novice-differences:

- structured principled knowledge: in assessment, an expert rapidly accesses the underlying meaningful patterns inherent in these structures, the novice will perceive primarily the surface features of the problem;
- proceduralised knowledge: in assessment, experts and novices may be equally competent at recalling a principle or a rule, but novices less frequently recognise where such knowledge applies or how to implement it;
- effective problem representation: in an early phase in the problem solving process, experts qualitative assess the nature of a problem and build a mental model from which they can organise the problem space, the novice quickly generate a superficial model that drives their performance;
- self-regulatory skills: experts use self-monitoring skills that control their performance. These skills are less available in novices in early stages of performance.

As stated, the novice- and expert-behaviour in assessment has to be determined. The phases in the instructional design of performance assessment in presented in Table 2 can be divided in two types of assessment-skills: the skill to design and the skill to rate. In both of them an novice- and an expert-behaviour can be characterised. The following matrix illustrates this:

	EXPERT	NOVICE
DESIGN	A	B
RATING	D	C

Ad A:

Systematic assessments have a clear purpose, are based on explicit criteria, rely on appropriate exercises, and include precise performance rating procedures. Stiggins (1987) distinguishes four phases in the design of performance assessment. The subphases can be regarded as tasks that need to be conducted in order to design sound assessments. In Table 2 the phases are illustrated.

Table 2: Expert-design and -development of assessments

1. Clarify the reasons for assessment: how are the results to be used?
specify decisions to be made on the basis of assessment: e.g. individual diagnosis, group needs assessment, grading, selection, evaluation. The given context determines which decision best describe the intended use;
specify decisionmakers by identifying the persons who will use the results to make the decisions;
specify use to be made of results: rank examinees or determine mastery;
describe students to be assessed: level, number, characteristics.
2. Clarify performance to be evaluated: what is the type of performance to be observed?
specify the content or skill typology for the assessment;
select the type of performance: e.g. evaluation of process of behaviour as it occurs, product. The question is for example if steps must occur in specific order or not;
list performance criteria: the dimensions of examinee performance (observable behaviour or products). The dimension of performance is to be specified in a definition and a performance continuum.
3. Design exercises
select form of exercises: the natural availability of dependable evidence and the seriousness of the decision are two factors that have to be taken into account in this subphase;
determine the obtrusiveness of the assessment key difference between obtrusive and unobtrusive is observed in the motivation of the examinee and the test anxiety;
determine the amount of evidence.
4. Design performance rating plan

determine the type of score needed: overall index of performance (holistic rating) or detailed presentation of criteria (analytic rating);

determine who is to rate performance;

clarify score recording method to be used: checklist, portfolio.

Beside the criteria for the performance assessment itself, also the whole design prior to the actual assessment have to be determined. In each phase of this design, the importance of the development of assessment strategies has to be pointed out.

Ad B:

The assessment in the traditional approach was often regarded as a task which was supposed to be carried out at the end of a course. As is stated in Table 2, the design of the performance assessment has to be started at the beginning of a course. Novice-behaviour in design can thus be described in terms of missing several steps in the design procedure.

Ad C:

Novice-behaviour in rating is based on "rating-errors" or the naive strategies learners exhibit when they are novices in using self- and peer assessment. In the following paragraph these errors are described and related to the experiences of naive behaviour in peer-assessments. Rating processes are subject to a variety of measurement errors. Hogarth (1981) states that the literature shows a depressing picture of human judgmental ability. When raters make judgements about a particular performance, a number of rating errors can be identified:

- personal differences among raters in their standards and their rating styles (Coffman, 1971; De Groot, 1975). Raters may differ in their severity or leniency. Some raters consistently tend to give high grades (lenient raters), while others (severe raters) consistently tend to give low grades (see also Lunz, Wright & Linacre, 1990);
- raters differ in the extent to which they distribute grades throughout the score scale. Some raters tend to distribute scores closely around their average; others will spread scores much more widely. In other words, some raters avoid giving extreme grades while others prefer to use them;
- the halo effect: the term for the tendency of human raters to base distinctive aspects of the rating on an overall impression of a single dominating aspect. This may mean that raters cannot differentiate among distinct aspects of one product or procedure (Borman, 1975);
- the significant effect: refers to the fact that raters may have different opinions about the rating tasks. According to Voss and Post (1990), this problem was not so much related to the deviating views of an individual, but rather to the deviating opinions of groups of individuals. Voss and Post believed that in particular in the assessment of less 'tangible' skills, objectivity was decreased significantly due to divergence of views among raters of different schools;
- evaluation policy: judges differ also in the ways they employ criteria (Sadler, 1983). Every assessor has his own evaluation policy. Some act: the performance must achieve a minimum qualifying level on a number of criteria. Other judges act conjunctively: the performance is excellent on one criteria and weak in the rest of the criteria. Thirdly, one could judge compensatory, poor showings on some criteria can be balanced by high performance on other criteria.

In the following Table these errors are directly coupled with novice peer-assessment behaviour.

Table 3 Errors in rating and peer-assessment

Errors in rating	Occurrence in peer-assessment
leniency	over-marking peers
friendship marking	over-marking peers
severity	under-marking peers
collusive marking	lack of differentiation within groups
decibel marking	individuals dominate groups and get the highest marks
parasite marking	students fail to contribute but benefit from group marks
the halo effect	students find one criteria the most important one and this influences the objectivity towards the other criteria
the significant effect	the students have different views on what a good performance is

evaluation policy

each student has his own interpretation of the importance/meaning of the established assessment criteria

Ad D:

De Groot (1975) stated that a testing and decision procedure is, by definition, 'completely objective' if the processing of all the pertinent behavioural data from a subject up to the final decision is carried out or could be taken over by a mechanised program. Since humans are imperfect machines, complete objectivity is not possible with ratings but should be striven for. What is required is a satisfactory degree of intersubjective agreement. This can be said to exist if there are sufficiently valid grounds to assume that the rater is guided by a system of reasonably constant criteria.

Expert-raters construct sound guidelines, consisting of instruments like model answers, product scales, checklists and rating scales, and advice for scoring procedures. These guidelines prevent raters from making errors like the halo effect and the sequence effect. The use of multiple raters also can eliminate accidental errors and personal strategies.

The ultimate purpose of the research project is to design tools which can be helpful in training students in higher education to become professionals in assessment strategies. A strategy is simply defined as a sequence of straightforward steps. Complex or expert strategies involve multiple decision points because alternative courses of action are possible because of a complex context (Sadler, 1983). Strategies are behaviours of a student that are intended to influence how he or she processes information (Mayer, 1988). Assessment strategies need to be explicitly taught and cannot be expected to develop without intervention (Birenbaum, 1996). Training students to be assessors represents an excellent instructional strategy (Stiggins, 1987).

The training has to be focused on design-tasks and on rating tasks, based on the findings illustrated in the prior sections. The key theme in self and peer-assessment is the involvement of students in the practices of performance assessment. The training has to support students in demonstrating better performances. Stiggins (1987) stated: "once students internalise performance criteria and see how those criteria come into play in their own and each other's performance, students often become better performers" (p.38). The term performance is chosen, since the tasks that the students have to work on during a course, can be described in terms of a task performance (This task performance can be perceived in many ways, for example essays, oral assessments, paper-and-pencil-assessments, process-analyses, group-products, work-samples, peer-and self-ratings). This task performance is object of evaluation. In performance assessment, evaluations of student achievement are based on the professional judgement of the assessor (Stiggins, 1991). The keys to performance assessment quality include the development of (a) an appropriate method of sampling the desired behaviours or products and (b) a clearly articulated set of performance criteria to serve as the basis for evaluative judgements.

In Table 4 the functions of a system for performance assessment are shown (Moerkerke, 1996). In the current approach of assessment the student could be involved in several functions, e.g. in the construction of tasks (negotiation about task criteria), in the compilation of tasks for the test (negotiation about assessment criteria), the rating and the way feedback is provided. This involvement is an important aspect for the development of the skill to self- assess and the skill to peer-assess.

Table 4 Phases in the instructional design of performance assessment

Phases	Student
Construction of tasks (DESIGN)	x
Compilation of tasks for the performance (DESIGN)	x
Assessment/rating products or behaviour (DESIGN)	x
Providing feedback (RATING)	x
Deciding (RATING)	

The first three phases can be placed in the design-procedure, the two last phases in the rating-procedure. One of the future goals in higher education is to integrate the assessment and the instruction. The Table shows that this integration is accomplished by defining a performance assessment-system as a tool that is embedded in the design of the instruction. The phases could be regarded as systematic steps in the instructional design. The reason why the phase deciding is not marked is that students are not yet qualified to make summative, that means certifying, decisions in the current educational systems. According to the law, only the teaching staff is qualified to make summative decisions. In the research project, self- and peer-assessments are regarded as formative manners to improve the ability to assess.

Phase II: experiment I: training peer-assessment

In this section the design of the first experiment is explained. As the results of starting phase (literature review and pilot study) and phase I (phase I: analysis of the complex skill of assessment and design training peer-assessment) indicate, self-assessment and peer-assessment are used as means to implement features of student centred, competence-based curricula, like the enlargement of student responsibility or

the gradual development towards a reflective practitioner. Peer-assessment is used to help students to develop skills of assessment, deepen ones understanding of the process of assessment, method, or systematic problem solving, develop skills of group work and facilitate the development of reflective learning. Furthermore, information from self-assessment and peer assessment sometimes contributes to formal assessment programs (Brown, Bull & Pendlebury, 1997).

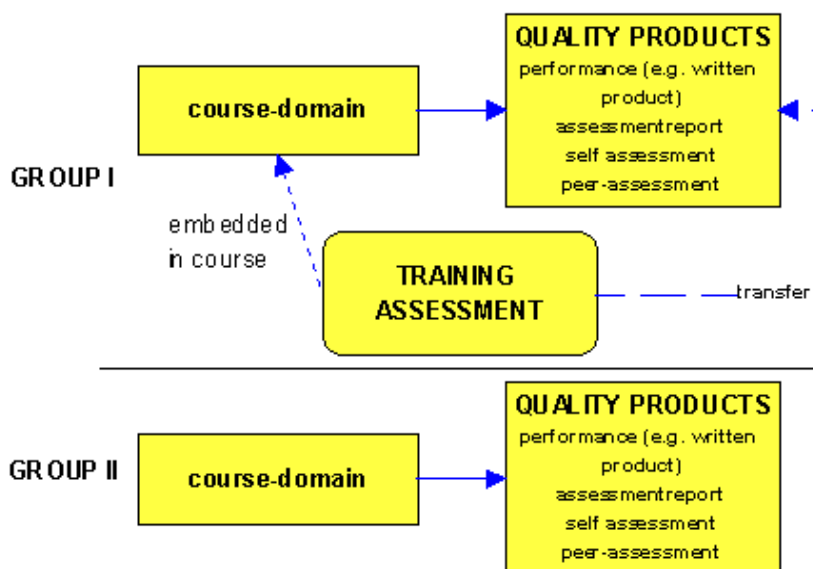
Table 1 Typology of objectives of higher education and related methods for assessment and supporting professional development

Objectives of higher education	Method for assessment	Method for supporting professional development
Subject matter expertise Acquiring subject matter expertise and skill directly related the core competencies of the domain	Multiple choice tests Essay examinations Subject-oriented self testing	Curriculum based multiple choice tests
Professional skills Learning to solve problems which are relevant to the tasks and jobs in the domain Securing communication skills that provide access to the knowledge network of others	Individual and/ or groups assignments eliciting systematic problem solving behaviour which results in authentic products or presentations Self assessment Peer assessment Teacher assessment	Development Centre for standardised assessment of skill Skills Management System which collects data from self, peer and teacher assessment Portfolio which collect products, testimonials, etc.
Skills regulating of professional development Developing meta-cognitions conducive to locating paths leading to new knowledge and its use Procuring skills that regulate motivation and affections related to self, peers, employers, etc.	Standardised inventories aiming at learning styles, study behaviour, interests in the professional domain, etc., Systematic reflection and appraisal by student self on long term development Systematic reflection and appraisal by tutor, peers, clients on long term development	Personal reports on progress on development on subject matter expertise and professional skills Learning contracts Career and Development planning interview

In order to get a structured view on the use of assessment methods in curricula, Moerkerke (1998) linked different aspects of competence to (a) methods for assessment and (b) methods for monitoring professional development. In Table 1 a similar framework for the analysis of assessment needs is given. A typology for objectives in higher education was derived from the work of Kessels (1997) who breaks down professional competence towards three aspects.

In the first column the facets of expertise are taken as an organising principle for classifying assessment methods. In the second column optimal assessment methods are mentioned. These assessment methods evaluate the quality of knowledge, performance or attitude at a certain time and place during the study program. In the third column methods which are optimal for supporting development during the study program are mentioned. Theories on independent learning and research on the effects of embedded informal tests in subject-oriented courses predict positive effects of self testing on learning. Moerkerke and Dochy (1998) revealed that students profit most from so-called progress tests, which provide regular feedback on learning. Although acquiring subject matter expertise is an important objective of higher education, the core of competency-based curricula lies in the development of skills which enables students to work on problems in teams. The shaded part of the table emphasises that in the research project the accent will be on the development of problem solving skills. That means that the selected course domain involves activities that contribute to the development of these skills.

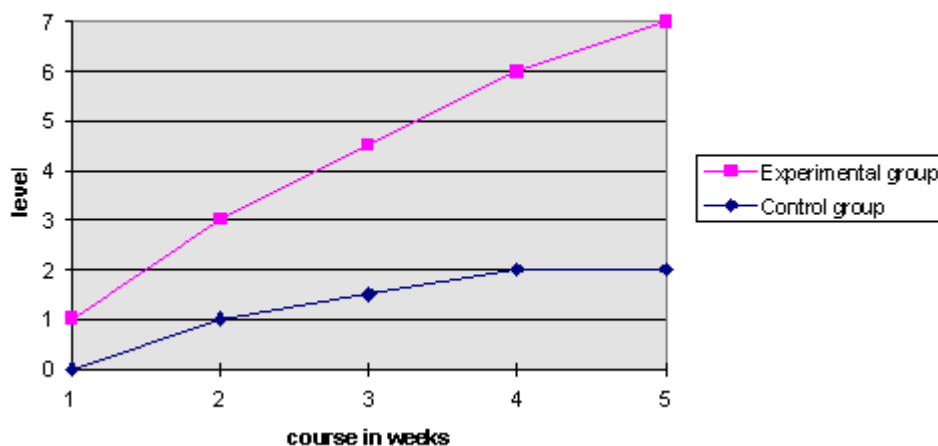
Figure 1: Design first experiment



At the beginning of the course students will undergo a pre-test. This probably will be a report of a peer-assessment of a similar product that has to be supplied at the end of the course and additionally a questionnaire, checklist or observation based on the empirical model which represents the naive strategies. One group will be trained in peer-assessment strategies. This training (the independent variable) is embedded in the course. The control group just take the course without the training. The students in both groups supply four products (the dependent variables): the actual performance (for example an essay), a self assessment of this product, a peer-assessment of the product of a peer/peers and an extensive assessment report similar to the pre-test.

The hypothesis is that the products supplied by the experimental group are of a significantly higher quality than the products supplied by the control group, because of a transfer effect of the training (see figure II).

Figure II: The growth in assessment skills

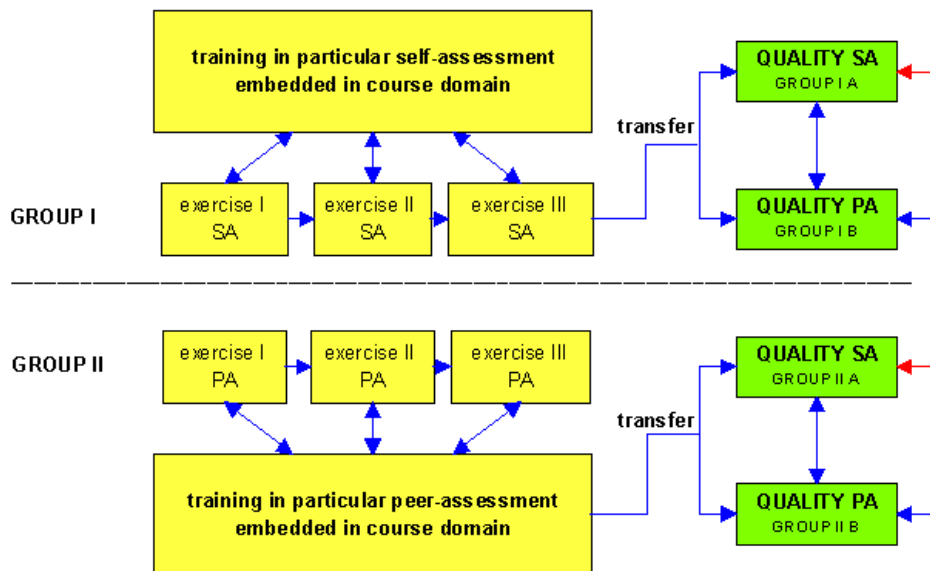


The growth of the control group is accounted for the content of the course-domain which aimed at the development of problem solving skills.

Phase III: experiment II - training self-assessment and peer-assessment

Another important question in the research project is the way students should be trained in assessment-techniques (see Figure III).

Figure III: Design second experiment



One group is trained in particular self-assessment strategies and the other group is trained in particular peer-assessment strategies. These strategies will be trained in a specific course domain.

At the end of the course and training each group will be separated in two groups (A and B). The A-groups will only carry out a self-assessment and the B-groups will only carry out a peer-assessment.

The hypothesis is that when students have followed a training in peer-assessment, this leads to higher quality self-assessments (GROUP II A perform better than GROUP I A). When students are familiar with peer-assessment strategies, the transfer of these strategies will be higher to the self-assessment than when a reversed procedure is followed.

Closing phase: generalizability of the training, recommendations, guidelines, limitations, possibilities

In the closing phase of the research project the following two questions will be explored:

- how can teachers be facilitated using the training for peer-assessment?
- what is the generalizability of the method?

Discussion

The research project aims at the development of models and an instructional training and experimental research. Specific factors and considerations need to be taken into account:

- the subject domain in which the experiments will take place need to be chosen with care, as well as the educational domain (e.g., higher vocational education or education at a university level);
- the type of education;
- the characteristics of the learning environment in which studies take place must be determined (competency-based);
- the explication of a large set of research variables and processes is needed in order to be able to compare and integrate the results of the studies.

To begin a process of instructional design in which the self- and peer-assessments strategies will be trained, it is important to reflect on the environment in which these training can achieve an optimal effect. A clear definition of the context in which self- and peer-assessment strategies are trained is required. The "new" assessment approach, the instructional approach, is based on cognitive learning theory. Learning involves the student's active construction of schemata in order to understand materials and processes. The instructional approach is called the assessment culture, which strongly emphasises integration of instruction, learning and assessment. De Corte (1990) refers to the design of powerful learning environments. They are characterised by the view that learning means actively constructing knowledge and skills on the basis of prior knowledge, embedded in contexts that are authentic and offer ample opportunities for social interaction. Since the goals as well as the methods of instruction are oriented towards more complex curricular objectives, it is necessary for assessment practices to increasingly use various kinds of performance assessments in which students have to interpret, analyse and evaluate problems and explain their arguments. It can be suggested that the assessment culture fits in well with a learning environment based on the principles of problem based learning. Problem based learning approaches for example organise study around key professional problems rather than traditional disciplinary knowledge. With staff support and access to appropriate study materials, students plan their

own learning to address problems with which they are confronted (Boud & Feletti, 1991). The development of assessment skills is an important feature in these problem-based learning approach. Schön (1983, 1987) introduced the idea of the reflective practitioner, the ability of practitioners to monitor what they do as they are doing it. Boud (1990) stresses that assessment practices in higher education have to be compatible with the curricular goals. One of these goals is that students take responsibility for their own learning and can self-evaluate their own learning at different phases in the instruction process.

One could state that the research context in which self- en peer-assessment is trained, has to meet the criteria of problem-based learning environments. In order to make the assessment practices congruent with instructional principles and practices the following criteria for the research context can be formulated:

- essential for the research context is that students learn by analysing and solving problems which are representative of the problems which they will have to apply their professional knowledge in the future;
- in the research context students are confronted with novel problems, asking them to transfer their knowledge and skills and to demonstrate the understanding of the influence of contextual factors on problem analysis as well as problem solving;
- in the research context, the problem analysis assessment tasks ask students to argue their ideas on the basis of various relevant perspectives;
- in the research context, the assessment of the application of knowledge of the course-domain are important aspects. Therefore, the performance assessment require examinees to apply their knowledge to commonly occurring and important problem-solving situations. Because a sufficient level of domain-specific knowledge is a determinant of productive problem solving, assessments measuring the coherence of students' knowledge base serve at least feedback function. Dochy (1992) defines knowledge profiles as "a plotting as a graph of raw or standardised scores of a group or individual on certain parameters" (p. 143). These indicate strengths and weaknesses in the student's knowledge base. Research has shown that such knowledge profiles can be seen as basic determinants of higher order achievement and can accurately identify specific deficits that contribute significantly to low achievement (Dochy, 1994; Dochy, Valcke & Wagemans, 1991; Letteri, 1980; Letteri & Kuntz, 1982);
- in the research context, the performance assessment asks for more than the knowledge of separate concepts. The assessment of integrative knowledge, requiring the integration of relevant ideas and concepts, is stressed. Since real life problems are mostly multidimensional, and as such integrate different disciplines within one field of study, assessment focuses on problems with this integrative characteristic.

During the project it is inevitable that some problems will occur in the design of the experimentation, the definition of the control group, gathering sound data and make general statements. This on the other hand makes the project dynamic and innovative. The major goal is to provide educational institutions tools which support them to develop their students into professional life-long learners. 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

- Anderson, J.B. & Freiberg, H.J. (1995). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly*, 22, 77–91.
- Arter, J. (1996). 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, 17–30.
- Birenbaum, M. (1996). Assessment 2000: Towards a Pluralistic Approach to Assessment. In M. Birenbaum, & F.J.R.C. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 3-29). Boston: Kluwer Academic Press.
- Borman, W.C. (1975). Effects of instruction to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of applied psychology*, 60, 556-560.
- Boud, D. (1990). Assessment and the Promotion of Academic Values. *Studies in Higher Education*, 15(1), 101-111.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D. & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18, 529–549.
- Brown, G., Bull, J., & Pendlebury, M. (1997) *Assessing student learning in higher education*. London: Routledge.
- Chi, M. T. H., Glaser, R. & M. J. Farr (Eds.) (1988). *The nature of expertise*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Coffman, W.E. (1971). Essay examinations. In R.L. Thorndike (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education.
- De Corte, E. (1990). Learning with New Information Technologies in Schools: Perspectives from the Psychology of Learning and Instruction. *Journal-of-Computer-Assisted-Learning*, 6, 2, 69-87.
- De Groot, A.D. (1975). *Methodologie* [Methodology]. (9th ed.). 's-Gravenhage: Mouton.
- Dochy, F.J.R.C. (1992). *Assessment of prior knowledge as a determinant of future learning: the use of knowledge state tests and knowledge profiles*. Utrecht/London: Lemma B.V./Jessica Kingsley Publishers.
- Dochy, F.J.R.C. (1994). Investigating the use of knowledge profiles in a flexible learning environment: analyzing students' prior knowledge states. In S. Vosniadou, E. De Corte, & H. Mandl (Eds.). *Psychological and Educational Foundations of technology-Based Learning Environments*. NATO ASI Series F, Special Programme AET. Berlin, New York: Springer Verlag.
- Dochy, F., Valcke, M., & Wagemans, L. (1991). Learning economics in higher education: an investigation concerning the quality and impact of expertise. *Higher Education in Europe*, 4, 123 -136.
- Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59, 395–430.
- Gagné, R.M. (1988). Some reflections on thinking skills. *Instructional Science*, 17, 4, 387-390.
- 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: The University of Texas.
- Glaser, R. (1984). Education and Thinking: The Role of Knowledge. *American Psychologist*, 39, 2, 93-104.
- Glaser, R. (1990). *Assessment and Education: Access and Achievement*. CSE Technical Report 435, CRESST/University of Pittsburgh, Learning Research and Development Center.
- Griffiee, D.T. (1995). Criterion-referenced test construction and evaluation. In J.D. Browne & S.O. Yamashita (Eds.), *Language testing in Japan* (pp. 20–28). Tokyo, Japan: The Japan Association for Japan Language Testing.
- Harrington, T.F. (1995). *Assessment of abilities*. Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services.
- Hassmén, P., Sams, M.R. & Hunt, D.P. (1996). Self-assessment responding and testing methods: effects on performers and observers. *Perceptual and Motor Skills*, 83, 1091–1104.
- Hogarth, R.M. (1981). Beyond discrete biases: functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, 90, 197-217.
- Keith, S.Z. (1996). Self-assessment materials for use in portfolios. *Primus*, 6, 178–192.
- Kessels, J. (1997). *Learning, the corporate curriculum and knowledge productivity*. Paper presented at the Invited Seminar Knowledge Productivity: Concepts and Issues. November, 20-22 at Leiden University, The Netherlands.
- 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, 205–214.
- Larkin, J. (1980). Teaching problem solving in physics: The psychological laboratory and the practical classroom. In F. Reif & D. Tuma (Eds.), *Problem solving in education: Issues in teaching and research*. Hillsdale, NJ: Lawrence Erlbaum.
- Letteri, C.A. (1980). Cognitive profile: basic determinant of academic achievement. *The Journal of Educational Research*, 4, 195-198.
- Letteri, C.A., & Kuntz, S.W. (1982). *Cognitive profiles: examining self-planned learning and thinking styles*. Paper presented at the Annual American Educational Research Association Meeting, New York City, March 19-23.
- Lunz, M.E., Wright, B., & Linacre, M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 4.
- Mayer, R.E. (1988). Learning strategies: An Overview. In C.E. Weinstein & E.T. Goetz (Eds), *Learning and study strategies: issues in assessment, instruction and evaluation* (pp. 11-22). San Diego: Academic Press.
- McNamara, M.J. & Deane, D. (1995). Self-assessment activities: toward language autonomy in language learning. *TESOL Journal*, 5(1), 17–21.
- Moerkerke, G. (1996). *Assessment for flexible learning*. Phd. Thesis. Utrecht, The Netherlands: Lemma.

- Moerkerke, G. (1998). Toetsing van academische vaardigheden [Assessment of academic skills]. *Tijdschrift voor Hoger Onderwijs*, 16(3).
- Moerkerke, G. & Dochy, F. (1998). Effects of prior knowledge state assessment and progress assessment on study results in independent learning. *Studies in Educational Evaluation*, 24(2), 179-201.
- Nevo, D. (1995). *School-based evaluation: a dialogue for school improvement*. London: Pergamon Press.
- Perkins, D.N. (1986). Thinking frames: an integrative perspective on teaching cognitive skills. In J.B. Baron & R.S. Sternberg (Eds), *Teaching thinking skills: Theory and Practice* (pp. 41-61). New York: W.H. Freeman.
- Pond, K., Ul-Haq, R. & Wade, W. (1995). Peer review: a precursor to peer assessment. *Innovations in Education and Training International*, 32, 314–323.
- Ritts, V., Patterson, M.L., & Tubbs, M.E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research*, 62, 413-426.
- Sadler, D.R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54(1), 60-79.
- Sambell, K. & McDowell, L. (1998). 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, UK: Oxford Centre for Staff and Learning Development.
- Schön, D.A. (1983). *The reflective practitioner: How professionals Think in Action*. London; Temple Smith.
- Schön, D.A. (1987). *Educating the reflective practitioner: towards a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*. (in press)
- Stiggins, R.J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice*, 6, 33-41.
- Stiggins, R. (1991). Relevant Classroom Assessment Training for Teachers. *Educational Measurement: Issues and Practice*, 10, 1, 7-12.
- Van Merriënboer, J. (1997). *Training complex cognitive skills*. Englewood Cliffs, NJ: Educational Technology Publications.
- Voss, J.F., & Post, T.A. (1990). On the solving of ill-structured problems. In N. Frederiksen, R. Glaser, A. Lesgold & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 261-285). Hillsdale, NJ: Lawrence Erlbaum Associates.