FORMATIVE PEER ASSESSMENT IN A CSCL ENVIRONMENT: A CASE STUDY

Running head: FORMATIVE PEER ASSESSMENT IN A CSCL ENVIRONMENT

Formative Peer Assessment in a CSCL Environment: A Case Study

Frans J. Prins, Dominique M. A. Sluijsmans, Paul A. Kirschner, and J. W. Strijbos

OTEC (Educational Technology Expertise Center), Open University of the Netherlands

Final draft, July 27, 2004

Correspondence concerning this paper should be addressed to Frans Prins, Open University of the Netherlands, Educational Technology Expertise Centre, P.O. Box 2960 6401 DL Heerlen, The Netherlands, voice: ++31-45-5762292, fax: ++31-45-5762802, e-mail: frans.prins@ou.nl

Abstract

In this case study our aim was to gain more insight in the possibilities of qualitative formative peer assessment in a computer supported collaborative learning (CSCL) environment. An approach was chosen in which peer assessment was operationalised in assessment assignments and assessment tools that were embedded in the course material. The course concerned a higher education case-based virtual seminar, in which students were asked to conduct research and write a report in small multidisciplinary teams. The assessment assignments contained the discussion of assessment criteria, the assessment of a group report of a fellow group, and writing an assessment report. A list of feedback rules was one of the assessment tools. A qualitative oriented study was conducted, focussing on the attitude of students towards peer assessment and practical use of peer assessment assignments and tools. Results showed that students' attitude towards peer assessment was positive and that assessment assignments had added value. However, not all students fulfilled all assessment assignments. Recommendations for implementation of peer assessment in CSCL environments as well as suggestions for future research are discussed.

Formative peer assessment in a CSCL environment: a case study

Current educational practice in higher education stresses the development of lifelong learning skills. Among such lifelong learning skills is the ability to provide valuable feedback and suggestions for performance improvement to another person or a group. In spite of the apparent need to develop efficient and effective feedback skills, they are in general not explicitly included in most constructivist curriculae. One potentially valuable approach to train these skills is the use of *peer assessment*. We consider peer assessment as the process whereby groups of individuals rate their peers (Falchikov, 1995), or, more specifically, as an arrangement for learners and/or workers to consider and specify the level, value or quality of a product or performance of other equal-status learners and/or workers (Topping, 2003). In essence, peer assessment can be a specific form of collaborative learning and in the past decade collaborative learning has become a popular pedagogical approach in higher education (see Kirschner, Martens, & Strijbos, 2004).

A course that includes peer assessment may contain multiple learning goals (Sluijsmans, 2002; Sluijsmans & Prins, 2004), and may thus have multiple beneficial effects. We label the improvement of *content related* performance as the first order learning goal and the *acquisition of peer assessment skills* as a higher order learning goal. Recently, many researchers have concluded that the implementation of peer assessment in the curriculum was beneficial for one or both levels of learning goals. For instance, Cutler and Price (1995), Freeman (1995), Horgan, Bol, and Hacker (1997), and Sluijsmans, Brand-Gruwel and Van Merriënboer (2002) report an increase in the quality of learning due to peer assessment tasks, which represents the first order learning goal. With respect to the higher order learning goal, several studies report that peer assessment tasks encouraged students to reflect more on their own behaviour and/or performance (Anderson & Freiberg, 1995; Gentle, 1994; Longhurst & Norton, 1997; Sobral, 1997), they exposed students to the skills of critical reflection and analysis (Birenbaum, 1996; Sambell & McDowell, 1998), they resulted in the development of students' peer assessment skills (Sluijsmans, Brand-Gruwel & Van Merriënboer, 2002), and they increased students' confidence in their ability to perform according to specified criteria (e.g., Cutler & Price, 1995), as well as their awareness of the quality of their own work (e.g., Anderson & Freiberg, 1995; Gentle, 1994).

The aforementioned studies clearly indicate potential benefits of peer assessment, although the mechanism and conditions (or constraints) by which they may appear are still open for debate. Furthermore, the empirical support for the positive impact of peer assessment is based on studies that were conducted in face-to-face settings. Nowadays, however, computer-mediated communication (CMC) and Virtual Learning Environments (VLEs) – such as Blackboard® – allow for the implementation of peer assessment in a distance education setting, in which students are involved in learning activities that are independent of time and place. This raises the question whether it is possible to transfer the ideas of peer assessment to distance education. Will peer assessment be equally effective, will students appreciate peer assessment in distance education, and does a distance education setting cause extra or specific problems? We therefore conducted a case study to examine the effects of the implementation of peer assessment in a computer-supported collaborative learning (CSCL) environment. We specifically focus on students' perception and attitude concerning peer assessment, on students' discourse in the VLE concerning peer assessment, and on difficulties that may occur when peer assessment is implemented in a CSCL environment. Before we will describe the case study, we elaborate on our perspective of formative peer assessment, on peer assessment as a specific form of collaborative learning, and on peer assessment in CSCL environments.

Formative peer assessment

Many assessment approaches that are used in higher education are purely summative, and thus they aim to determine success or failure only after a student's performance (Topping, Smith, Swanson, & Elliot, 2000). By contrast, formative assessment intends to help students identify their strengths and weaknesses and guide students towards the achievement of learning goals during the learning process (e.g., Boud, 1990; Dierick & Dochy, 2001; Topping, 2003). According to Topping et al. (2000), "formative assessment seems likely to be most helpful if it yields rich and detailed qualitative feedback information about strengths and weaknesses, not merely a mark or a grade." (p. 150). In many current peer assessment practices, however, a quantitative or scoring-based approach is chosen. In this approach, peer assessment focuses mainly on peer ranking (for example each student ranks others from best to worst on one or more factors), peer nomination (for example each group member nominates the highest performing member of the group on several factors), or peer rating (for example each group member rates all other members on a set of performance or personal characteristics, using one or several kinds of rating scales). Nominations, rankings and ratings however, have been found to create quite strong adverse reactions (Hanrahan & Isaacs, 2001; Kwan & Leung, 1996; Rushton, Ramsey, & Rada, 1993). Above all, these methods though seem not to prevent a rating bias as reported by Pond, Ul-Haq and Wade (1995) who distinguish four origins of bias: over-marking ('friendship marking'), lack of differentiation within groups (collusive marking), individuals who dominate a group get the highest mark (decibel marking) and students failing to contribute benefiting from group marks (parasite marking). Thus, for our case study, we chose an approach in which students were asked to write an assessment report with qualitative feedback concerning a group product of fellow students.

In order to integrate peer assessment in courses – in face-to-face or distance education settings – several peer assessment assignments can be designed that are based on the peer assessment skill. Sluijsmans and Van Merriënboer (2000; Sluijsmans, 2002) analysed the peer assessment skill in the domain of teacher education and identified three important sub skills: *defining the assessment criteria* (thinking about what is required and referring to a product or process), *providing feedback* (such as constructive feedback about the product of contribution to group performance), and *writing a structured assessment report* (making the reflection explicit). The peer assessment assignments must be closely related to the regular course material and preferably integrated in the course assignments. For instance, students could be assigned to discuss performance criteria or, more specifically, they could be asked to create a *performance scoring rubric*. A performance scoring rubric contains predetermined standards that can be used to judge the level of achievement attained by individual students, group members or other groups by comparing their performance to these predetermined standards. By designing a rubric, students can play a crucial role in defining these standards, thereby increasing their involvement in the task. Moreover, the design of a rubric makes students aware of what is expected during the course and thus forces students to orient themselves on course demands. Orientation is an important metacognitive skill and has impact on the quality of learning task behaviour (Veenman, 1993; Prins, 2002). After students discussed assessment criteria, they could be asked to reflect on draft versions of other individual students, group members or peer groups (see for example Higgins, Hartley, & Skelton, 2002), and write their ideas down in an assessment report, which refers to the second and third skill, respectively. To make it a real interaction, student groups could also be asked to reply to their fellow group about the provided feedback. This will emphasise the importance of the role of being the assessee in peer assessment. We think that the

assessee should have the opportunity to communicate with the assessor and express whether the feedback is understood, whether the feedback is accepted, and how the feedback is processed during revision of the draft. In our opinion, the role of the assessee should receive more attention in research on peer assessment.

Peer assessment can be considered as a complex skill in which students often need to be supported or trained (Sluijsmans, 2002). In general, support or training is necessary for each sub skill because it cannot be assumed beforehand that students are experienced in peer assessment practices (Sluijsmans, 2002). Examples of supporting tools are templates for writing an assessment report, a protocol for negotiation about assessment criteria, a work out example of a performance scoring rubric, and a list of feedback rules. Some of these tools were used in our case study and will be described in more detail in the method section.

Peer assessment as a specific form of collaborative learning

Peer assessment incorporates many features of collaborative learning. Collaborative learning refers to an instructional approach in which students work together in small groups toward a common goal (Dillenbourg, 1999). Strijbos, Martens and Jochems (2004) illustrated that collaborative learning can be regarded as specific form of group-based learning. A collaborative learning environment essentially entails student-student interaction and the pedagogy by which they do so depends on the type of learning objective, the task and the extend to which the collaboration is pre-structured in advance (a high level of pre-structuring versus a low level of pre-structuring). Regardless of the educational setting, peer assessment thrives on interaction and thus it is a form of collaborative learning at its very core. Determining and negotiating criteria, as well as assessing a product by a fellow student or group and providing constructive feedback, amplify that peer assessment is a specific pedagogical approach of collaborative learning. Moreover, both mechanisms prone to any form of group-based learning occur in peer assessment as well.

Most approaches to group-based learning rely on two central mechanisms: *individual accountability* and *positive interdependence*. Individual accountability refers to the extent to which group members are held individually accountable for the jobs, tasks or duties, central to group performance or group efficiency. It was introduced by Slavin (1980) to counter the 'free-rider effect', i.e. some students would deliberately not invest any (or little) effort into group performance. Thus, individual accountability implies specifying individual responsibility, something someone can be held accountable for. Peer assessment makes students individually responsible for an active contribution to group discussions that focus on establishing a shared set of criteria (Sluijsmans, 2002). In addition, when it is part of the group's task to ensure that every group member has learned something, it is in the interest of every group member to spend time providing feedback to their peers (Slavin, 1989). Positive interdependence refers to the extend that the performance of a single group member depends on the performance of all other members. Johnson (1981) implemented it to foster group cohesion and a heightened sense of 'belonging' to a group. It can be achieved through the task, resources, goals, rewards, roles or the environment (Brush, 1998). Although positive interdependence can have a strong influence on the level of cohesion, establishing such cohesion depends also on familiarity and mutual trust. Especially trust is important prerequisite for peer assessment, as most students tend to be hesitant to assess their peers and regard assessment as the exclusive realm of the teacher (Cheng & Warren, 1997; Sambell, McDowell & Brown, 1997; Sluijsmans, 2002). In peer assessment, positive interdependence can be enhanced through role interdependence by assigning the roles of assessor and assessee to individual students. Peer assessment may be performed on the *individual*

level (e.g., a student rates some other students), the *intra-group* level (e.g., each student rates the performance or contribution to shared product by all other group members) or on the *inter-group* level (e.g., one or more students from a group rate the performance or product of another group). In our opinion, it remains essential in any peer assessment format that a student or group (or student representative of a group) indicates explicitly whether – and to what extend – they used the feedback provided by the individual assessor(s) or (representatives of) the assessing group.

In all, it is apparent that peer assessment through the use of positive interdependence and individual accountability can enhance a student's sense of task ownership (see Kirschner, 2002) and stimulate involvement in his/her learning. However, assessment – let alone peer assessment – has not been a focus of collaborative learning approaches. Most assessment techniques still rely on individual quizzes, group grades or a combination of the individual and group level achievement on quizzes (Slavin, 1995). Clearly, peer assessment can be an asset to regular group-based learning approaches and decrease the emphasis on individual performance.

Peer assessment in computer-mediated learning environments

Simultaneously with the increased interest in collaborative learning pedagogies, higher education has implemented computer-mediated communication (CMC) technology on a large scale to enhance student-student and student-staff interaction. Although the opportunity for interaction exists in so-called Virtual Learning Environments (VLE) such as Blackboard®, providing students with communication technology does not automatically result in collaboration – let alone learning (Guzdial & Turns, 2000; Lehtinen, Nurmela, & Salo, 2001). Although De Graaff, De Laat and Scheltinga (2004) illustrate that no direct relationship exists between the technological tool and the pedagogical orientation it affords, it is apparent that current higher education practice relies primarily on what De Graaff et al. refer to as 'guided learning', i.e. the lecturer/teacher owns and controls the task. Peer assessment provides an opportunity to decrease the emphasis on teacher ownership and increase student involvement by transferring ownership to the students.

As it appears to be difficult to incorporate the contributions of individual students in the actual assessment in face-to-face settings, opportunities for doing so are enhanced by the use of new communication media. For example e-mail or a discussion forum not only forces students to make their thoughts explicit but it also provides a powerful repository for the teacher (and to the students!) to take individual contributions into account. Irrespective of these benefits, researchers in the field of CMC and computer-supported collaborative learning (CSCL) in particular, however, are starting to take a reserved position towards popular views on collaboration and technology, which can be summarised by Salomon's (2000) claim: what is technologically possible, is not always educationally desirable.

Computer systems should be supportive of the needs of students and not all CMC tools provide equal opportunities for interaction (Chin & Carroll, 2000). Course designers should not be lured in thinking that students use technological support as they intended (Martens, 1998); in fact "whether the opportunities are actually taken and whether taking them upgrades performance and leaves some desired cognitive residue, is less dependent on the technology and far more on other factors" (Salomon, 1992, p. 63). Therefore, like any educational approach, CSCL has to provide a sound pedagogical context to support students' learning – and in the case of peer assessment the students should be provided with appropriate support tools.

Similar to face-to-face group-based learning practices, assessment has not been a focus in CSCL research and practice. Gradually it is acknowledged that CSCL is not merely a matter of changing the technology, but it requires redesigning the learning environment – including the

assessment! (see Strijbos, Kirschner & Martens, 2004). Nevertheless, innovative assessment such as the use of peer assessment has surfaced in CSCL practices. A recent example of peer assessment during CSCL at the K-12 level is provided by Chan and Van Aalst (2004). Students were asked to select the best contribution in a KnowledgeForum ® threaded discussion forum, but the criteria by which the students performed their evaluation were set by the teacher in advance. In addition, students were not explicitly trained to apply these criteria. Moreover, studies of peer assessment in CSCL environments in distance education with adult learners are often limited to a quantitative approach whereby students give scores to peers on a list of criteria (Topping, 2003). Such peer assessment practices appear to be reliable, but students often express their preference for teacher-based assessment and sometimes even reject peer assessment (Rushton, Ramsey & Rada, 1993). In our study, we addressed these issue regarding peer assessment in CSCL and distance education by providing students with explicit instructions and a set of tools to guide and support the peer assessment. In the next section the design of our learning environment will be discussed in more detail.

Aim of the case study

It has been illustrated that peer assessment, collaborative learning and CSCL are closely related. Effective collaboration – of which peer assessment is a specific form – requires fruitful interaction and students must be individual accountable and positive interdependence between the group members has to be stimulated. Whereas the systematic integration of peer assessment support has proven to be effective in face-to-face environments with students age 18-22 (Sluijsmans, 2002), our knowledge about peer assessment assignments and procedures for CSCL – as well as the support that *can* and *has* to be provided – is still very limited. As the experience with the support of peer assessment skills and qualitative peer assessment is limited and reports that may help to design such support is rare, the integration of peer assessment support in distance courses is a challenging task. We therefore conducted a pilot case study in which we investigated the applicability as well as the design of instructional support and tools for peer assessment in CSCL environments.

The pilot case study was conducted in a distributed case-based CSCL-course at the Open University of the Netherlands. The study focused on inter-group peer assessment with the group product as the subject of the assessment. For the summative assessment of this product a scoring rubric was developed. In addition, a rubric was developed to assess each group members' contribution to the online collaborative process. To explore the possible additional value of peer assessment support in a CSCL environment, several peer assessment assignments (e.g., discuss assessment criteria, construct an assessment form, write an assessment report of your peer groups' product) and assessment tools (e.g., examples of products, templates for assessment forms, feedback rules, and scoring rubrics) were integrated in the learning process. Two assessment conditions were researched: a rich assessment condition in which assessment tools were provided, and a bare assessment condition without tools. This way we were able to examine the impact of the tools as well as the impact of the assignments on assessment products. In this exploratory study, the number of participating groups was limited – which is not unusual in CSCL research (see Stahl, 2002) – and the focus was mainly qualitative. Data was gathered during the whole course from both the students and the tutors. Our research questions were the following: (1) what are students' attitudes towards peer assessment and towards evaluating others in a CSCL environment and how do students perceive their own assessment skill? (2) What are the effects of the assessment assignments and tools on communication behaviour and

assessment products? (3) What does the implementation of peer assessment in a CSCL environment require?

Method

Participants

Participants in this study were 27 university students (14 female, 13 male) attending the European Virtual Seminar (EVS). The students were from five different European countries: two students from Belgium, two from Germany, eight from The Netherlands, seven from Poland, and eight from Spain. Each of the students had expertise in one of the following areas: environmental sciences, marketing, geography, economics, law, anthropology, geology, sociology, chemistry or engineering. Six multidisciplinary student groups were formed, consisting of three to six students. Two students dropped out of the course.

Design and Procedure

When participants enrolled in the EVS course, they could indicate their preference for one of the four available cases in the EVS (agriculture, water management, energy, or spatial planning). Six multidisciplinary groups of four to six students were formed, based on their nationality, discipline, and case preference. There were three conditions, a *rich* assessment condition, a *bare* assessment condition, and a *control* condition. In each condition, two student groups of four to six students participated. The students in the *rich* assessment condition (group 1 and 6, $n = 7$) were (1) encouraged to discuss performance criteria, (2) had to write a self-assessment report on the first draft of their report, (3) had to write an assessment report on the first draft of the report of a fellow student group, (4) had their first draft of the report assessed by a fellow student group, and (5) had to respond to the assessment report of this fellow student group. Moreover, in the rich assessment condition student groups were also provided with the assessment tools to support the assessment process. In the *bare* assessment condition (group 2 and 5, $n = 8$), students received the same assessment assignments as students in the rich assessment condition except for the assignment on discussing performance criteria. Also, they did not receive assessment tools. Groups in the *control* condition (group 3 and 4, $n = 12$) received neither assessment assignments nor assessment tools.

After the general mini-course, the groups in the assessment conditions attended a mini-course on peer assessment. Information on assessment tools was given only in the mini-course of the rich assessment condition. At the end of the mini-course, students were asked to complete the questionnaire concerning their attitude towards peer assessment, perception of their assessment skills, and attitude towards evaluating others. The student groups in the control condition did not attend a mini-course on peer assessment and did not complete a questionnaire. Next, the student groups worked on writing the group report. Students in the rich assessment condition received all three types of assessment assignments, whereas students in the bare assessment condition only received assessment assignments of the second and third type. When the first draft was handed in, the groups in both assessment conditions exchanged their reports and were asked to assess the first draft of the report of a fellow group. The groups in the control condition received feedback from their tutor. When the final draft of the group report was handed in, the students in the assessment conditions once again completed the questionnaire concerning their attitude towards peer assessment. Moreover, they evaluated the quality of the peer assessment mini-course and the assessment assignments via the second questionnaire with self-report open–ended questions.

Finally, tutors marked the final drafts of the group reports and the students' participation in the collaborative learning process using the scoring rubrics. The marks were discussed during

a meeting of tutors, researchers and designers of the EVS course. Moreover, during this meeting the whole EVS course, including the peer assessment conditions, was evaluated.

Materials

The European Virtual Seminar (EVS). In this course, four cases were designed concerning sustainable development and enlargement of the European Union. The topics of the cases were agricultural policy, integrated water management, energy technology, and spatial planning and policy. The case description contained background information, a general assignment, sources and links. For instance, the general assignment concerning agriculture was: "The aim of the assignment is to find out the impact of the enlargement process on sustainability in agriculture in Poland." The student groups were asked to conduct research based on their case and to write a group report in which they operationalised the main concepts of their case, integrate the different disciplinary views on the problem described in the case, and provide recommendations to the European Union for a policy change concerning the topic of their case. Three main phases of the writing process were distinguished, namely: (1) writing a research proposal, (2) writing a first draft of the report, and (3) revising the first draft into a final draft of the report. For each phase, the student groups received specific assignments.

The tutors in the EVS course fulfilled the role of coach rather than expert on the research topic. They were specifically instructed to keep track of group participation and take initiative when students were about to drop out. Furthermore, the tutors decided whether the research proposal of their group was of sufficient quality and whether the group could continue with the research and the second writing phase or if they had to revise or rewrite the proposal. Finally, the tutor was responsible for the marking after the third phase of the writing process.

The student groups collaborated in Blackboard 5®, a virtual learning environment (VLE). General information concerning cases, assignments, performance criteria, planning, and deadlines were available for all groups during the course in the *course documents* and in the *course information*, two specific sections in theVLE. For group communication, group specific *discussion boards* and *chat facilities* were used. The assignments were posted in the group discussion board according to a timetable that was available in the course information section in the VLE. When an assignment was posted, a new forum was started, with one or more discussion threads. Students were asked to use these threads to discuss matters concerning the assignment. In this way, groups were invited to structure their discussions. The students could communicate with their tutor by using a special discussion thread called 'Communication with the tutor'. By attending three distributed mini-courses at the beginning of the course, group members were able to become familiar with Blackboard®, small group skills, and communication skills.

Mini-course on peer assessment. In a mini-course on peer assessment, basic information about peer assessment was provided. It took students approximately half an hour to read and understand it. The mini-course dealt with a description of peer assessment, its purposes, and the learning goals of the assessment assignments.

Assessment assignments. During the EVS course, three types of assessment assignments were provided to the students in the assessment conditions, namely: (1) discussing assessment criteria, (2) assessing a draft of their own report and the report of a fellow group, and (3) writing a reply to the fellow group indicating how the report was revised according to the assessment from the fellow group. Each assignment consisted of an introduction, a task, a desired output and a deadline.

The first type of assessment assignment, discussing assessment criteria, was given prior to the assignment to write a research proposal and the assignment to write the first draft of the report. Student were provided the following instruction:

> Take a close look at the criteria for the assessment and marking of group reports (see course information, tool 5.1). Check the criteria systematically. Is each criterion explicit enough? Are all criteria of equal importance? Which criteria should be added, which should be deleted? Write down your group's adjusted criteria list. Use the discussion thread called 'criteria' to discuss and improve the criteria.

Following the discussion of the criteria for group reports, students were asked to make an assessment form based on their adjusted criteria list.

The second assessment assignment, assessing the draft of their own report and that of a fellow group, was provided right after the first draft of the group report was written. Two tutors with groups that worked on the same case exchanged their group reports. Students had to do this assignment individually, and, consequently, each student group received more than one assessment report. The following instruction was provided:

> Determine the strong aspects as well as the aspects that could use some improvement of the report of the fellow group by completing your group's assessment form. Send the completed peer-assessment reports to your tutor using the discussion thread 'Communication with tutor'.

The final assessment assignment, sending a reply to the fellow group, was provided after the third phase of the writing process when the final draft of the report was ready. The instruction given was the following:

> Take a close look at the completed assessment reports of the fellow group, provided by your tutor. Adjust your group's report and send the final draft to your tutor. Your group as well as the fellow group have invested time and effort in assessing a report of another group. Therefore, let the fellow group know how your group processed their comments, by writing a short reply to the fellow group. Send this reply to your tutor (by using the discussion thread Communication with tutor).

Assessment tools. The assessment tools were designed to support students while they were working on the assessment assignments. Tools for support during the first type of assessment assignments, that is, discussing assessment criteria, were the following: (1) an outline for writing a research proposal, (2) examples of last year reports' tables of content, (3) the *scoring rubric* that tutors used for the summative assessment (see Appendix 1), and (4) a *template* for making an assessment form. Both the outline and the scoring rubric were designed by domain experts in collaboration with the first two authors of this article. A scoring rubric has several components, including one or more dimensions for rating performance, definitions and examples illustrating the attribute(s) being measured and a rating scale for each dimension. Thus, the rubric was also used for the assessment and marking of the group reports. The outline and the scoring rubric were available in the course documents. However, providing these tools linked to the particular assessment assignment gave the tools much more emphasis, and it was expected that students would be more willing to use them. With the 'template for making an assessment form', students were expected to be challenged to formulate their own criteria explicitly and to make an assessment form based on their own adjusted criteria list.

Information on *feedback rules* (see Appendix 2) was available during the second assessment assignment, that is, assessing a draft of their own report and the report of a fellow

group, to support students. Basically, by providing these feedback rules, students are challenged to give constructive feedback, be specific, and be the owner of the feedback.

Questionnaires. The first questionnaire concerned student's attitude towards peer assessment and consisted of Likert-scale type questions and open-ended questions. The Likert-scale type questions were statements about (1) students' attitude towards peer assessment (7 items), (2) their perception of their own assessment skills (8 items), and (3) students' attitude towards evaluating other students (8 items). Students were asked to decide how much they agreed with each statement according to their beliefs and experiences. Their answers could vary from 1 (strongly disagree) to 6 (strongly agree). An example of a question about attitude towards peer assessment: *I believe that novices in peer assessment are able to assess peers in a responsible manner.* An example of a question about perception of one's own assessment skills: *I am confident that I can give feedback to peers.* And finally, an example of a question about attitude towards evaluating others: *I often compare myself with other people.*

The second questionnaire consisted of open-ended questions concerning the value of the mini-course on peer assessment, the clarity of the goals of the assessment assignments, the clarity of the assessment assignments, the perception of learning how to assess, the perception of how comfortable students felt when making the assessment, the value of the feedback students received from their fellow students, the way students responded to the assessment report of their fellow students, and the positive and negative aspects of peer assessment during the course. An example of an open question is: "Did you feel comfortable in making assessments concerning a report of students with whom you did not meet? Please explain."

Scoring rubrics. For the assessment of the joint report and the group process, two scoring rubrics were designed. The scoring rubric for the product counted for 70% of the end mark (10 criteria, see Appendix 1). The scoring rubric for the group process counted for 30%. Criteria for the group process were 1) planning research, 2) planning individual tasks, 3) cooperation within the group, 4) cooperation via the internet, 5) participation, and 6) incorporate comments.

Data-analysis

Given the exploratory nature of the study and the limited sample size, the results of both the closed and open-ended questions of the pre- and post-test questionnaires were analysed qualitatively. Also students' discourse via the discussion boards and chat facilities were examined. In general, the number of messages posted in the discussion threads concerning the content, the collaboration process, and the assessment assignments and tools were reported. Messages and chat sessions were analysed qualitatively as far as they concerned the assessment assignments and assessment tools. We focussed on the effectiveness of the assessment assignments and tools, that is, we searched for messages in discussion threads and remarks in chat sessions referring to the assessment assignments and tools, and for cues in the electronic discourse indicating whether students comprehended the assignments, whether they were using the tools adequately, and how students reached consensus about assessment criteria.

The assessment reports of the fellow student groups were analysed for use of the criteria that were provided, use of new criteria created by the group or individuals, and use of positive and negative feedback statements (Sluijsmans, 2002). We also examined whether the provided feedback rules were used.

Finally, tutors were asked about their experiences with the peer assessment assignments and the scoring rubric as a means for the final marking of the group report.

Results

Attitude towards peer assessment

The pre-test questionnaire about students' attitude towards peer assessment was completed by five out of fifteen students and the post-test questionnaire by seven out of fifteen students in the rich and bare assessment condition. In spite of the limited response, reliabilities were acceptable for the scale *attitude towards peer assessment* (Cronbach's alphas were .88 and .49 for pre-test and post-test, respectively, 7 items) and the scale *perception of one's own assessment skills* (Cronbach's alphas were .80 and .68 for pre-test and post-test, respectively, 8 items). For the scale *attitude towards evaluating others* Cronbach's alphas for pre-test and post-test were below .40.

The answers students gave provided general and valuable information on students' attitude towards peer assessment and assessment skills. The questionnaire results indicated that initially, as well as after the EVS-course, students had a positive attitude towards peer assessment: the majority of the questions concerning the attitude towards peer assessment, towards assessment skills, and towards others were scored 4 or higher for the pre-test (69 % of the questions) as well as for the post-tests (76 %).

Analysis of students' discourse concerning assessment assignments and tools

The student groups could use discussion boards and chat facilities in Blackboard® to communicate about the learning tasks and the assessment assignments and tools. Since the assessment assignments and tools were the main focus of our study, we analysed the discourse of the groups that were involved in assessment tasks (group 1, 2, 5, and 6) in a qualitative way. Groups varied substantially concerning the number of messages posted by students in the discussion threads and the number of chat sessions (see Table 1). The vast majority of the discourse concerned the content of the case study and the group report and the collaboration process. Discussion threads were often used to arrange a chat session. Only a few messages referred to assessment assignments and tools. Surprisingly, no messages in the discussion threads concerned off task communication. Students used chat to exchange information about their social life and about the characteristics of their country. Chat was also frequently used for discussions about the course content. It should be noted that groups were could also use e-mail facilities for communication, which was unfortunately unobservable for research.

Below, the results are reported for the three main phases of the writing process separately.

Writing the research proposal. Both groups in the rich assessment condition used the group discussion board to discuss the research proposal. Moreover, they followed the instruction to discuss the research focus, the proposal, and the tools in separate threads. A substantial part of the messages concerned the content of the proposal (9 of 16 messages in group 1; 14 of 35 in group 6). Moreover, messages about the content of the research proposal consisted of much more text than messages about other topics. The students made few and short superficial remarks about the tables of contents examples that were provided. For instance, in group 1 a student stated: "For me example 3 looks good, it is, I agree with B. A little extensive but it looks like a logical path to work it through like that." A fellow student replied: "Yes, example 3 looks good for me too." Group 1 also used the chat facilities to discuss these examples, again in a superficial way:

M. >   And as I said I think the third one is good
T. >   Why the third, I thought the second one is a good one?

T. >   It is in my opinion not so important which one we choose, but I think it is important that we agree on the focus and divide tasks, so that we can start writing

M. >   The second one was short

T. >   Yes, that is why I liked it. Why did you like number 3?

M. >   It is more detailed

B.>   The third is a good proposal

T.>   So do you think we can provide content for those details?

M.>   We can also create one ourselves

T.>   That I think is more complicated, two like the third, let us agree on the third

In addition, the messages on planning ("I think it is very important to plan our work") and on the focus of the report (e.g., "I would like to focus on economic and ecological impact") were short and superficial. The structure of the research proposal of group 1 showed that this group used the tool 'outline for writing a research proposal'. Unexpectedly, the research proposal of group 6 did not show any use of this tool. This group came up with a proposal that was structured differently. In sum, the groups in the rich assessment condition noticed the tools but rapidly turned to researching and discussing the content of the case study and the report and failed to use some of the tools effectively.

The discussions about the research proposals of the groups in the bare assessment condition were less extensive. In group 2 only six messages were posted, in group 5 one message was posted about the focus, and 10 about the proposal. Interestingly, in both groups in the bare assessment condition, the tutor played an important part in the discussion. The tutor of group 2 more or less determined the focus, while the tutor of group 5 directed the group to the outline for writing a research proposal that was available in the course information section of the VLE.

Writing the first draft. In this second phase of the writing process, both groups in the rich assessment condition were assigned to discuss the provided criteria for writing a report and to create an assessment form. Unexpectedly, both groups did not discuss the provided criteria via their discussion board or chat facilities, and they also did not hand in an assessment form. One of the students of group 6 stated in a message s/he posted in a discussion thread: "We haven't made any adjusted criteria. (...) We should reshape the criteria into our own? I think they are pretty clear. What should we change? Did anybody think about this already, or made their own personal criteria yet?" None of his fellow students answered this message.

Revising the first draft. In the third phase of the writing process students were asked to assess the group report of a fellow group and to revise their own report by processing the feedback provided by their fellow group. Similar to the second phase, the discussions about the assessment assignments and tools, as well as about the group report by a fellow group, were very limited. Messages in this phase concerned practical matters, like "Where can I find the first draft of the other group?" Apparently, students fulfilled the assignment to assess the report of a fellow group individually without engaging in a discussion with their group members by means of a discussion thread or chat.

Assessment reports and reply

Students from three of the four groups in the assessment conditions carried out the peer assessment, although in different ways. Group 2 performed no peer assessment. Below we describe for each group the assessment reports in terms of the use of the criteria that were

provided, use of new criteria created by the group or individuals, and use of positive and negative feedback statements.

One student of group 1 (rich assessment condition) assessed the first draft of the fellow group 2 (bare assessment condition) and wrote a short assessment report, containing three statements. Unexpectedly, neither the criteria provided for writing a group report (see Appendix 1) nor the template for an assessment form was used for the assessment. The assessment contained three implicit new criteria, which were not explicitly labelled in the assessment report. The new criteria concerned the data sources used by group 2, the many details in the report, and the suggested solutions for Polish agriculture. All statements in the assessment report were negative. However, the assessor used 'I' statements and was the owner of the feedback.

One student of group 5 (bare assessment condition) assessed the first draft of the report of group 6 (rich assessment condition). Two criteria from the provided list were used, although not explicitly labelled (i.e., *quality of language used* and *relation problem definition-analysis-solution*). Three new criteria were mentioned, that is, *use of Figures, length of chapters,* and *the way topics were discussed*. Seven comments of this assessment report were negative statements, one was positive.

Four students of group 6 performed the assessment of the report of the fellow group 5. Three of them used their own criteria for the assessment, which showed some overlap with the criteria from the provided list. All three somehow used the criterion *problem definition-analysis-solution,* two mentioned *quality of language used,* and *integration of disciplinary contributions* was mentioned once. The three students came up with a total of nine new criteria that concerned content (e.g., correctness of information), *structure* of the report, or *layout.* Sometimes the feedback dealt with details (e.g., "Sometimes the writer uses the term 'closed lines', other times 'closed system', or 'closure of circuits'"). Recommendations were quite general (e.g., "add new information"). The first student made 1 positive and 5 negative statements, the second 1 positive and 4 negative, and the third 3 positive and 8 negative. The fourth student, in contrast, used all of the provided criteria as well as the assessment form template, which resulted in a much broader assessment. This report contained 3 positive and 4 negative statements. The student considered the last criterion of the provided list, that is, *creativity,* as a subjective criterion and thus, he did not assess the report on creativity.

One of the students of group 5 sent a reply to the four assessment reports the group received. He summarized the comments his/her group received from group 6 and s/he described whether s/he agreed with the opinion of the assessors. Basically, s/he agreed with many of the remarks and explained how s/he revised the report. For instance, at the end of his/her reply, s/he stated: "I tried to adjust every chapter to make the story more logical. Also I added a chapter concerning policies of different European countries."

Open questions on peer assessment

Seven students (five from the rich assessment condition and two from the bare assessment condition, together 47% of all students in either assessment condition) completed the post-test questionnaire concerning their experience with the assessment mini-course and assessment assignments. Their answers to the open questions gave us an indication on how they perceived the assessment material.

Mini-course. In general, the students evaluated the mini-course on peer assessment moderately positive, but some students also stated that it is difficult to actually perform the peer assessment. For instance, on the question whether the mini-course was valuable, E.M. said: "Yes it was. In peer assessment you try to value the members of the other group and the final project

of all groups, but it is much too difficult to make this true."

Assessment assignments. The goals of the assessment assignments and the assignments itself were considered to be clear. However, again students added that actually executing the assignments is a challenge. For instance, T.S. stated: "The goals were clear. What was not clear was how exactly it was supposed to be done." Two students mentioned time pressure as an obstacle to complete the assignment. For instance, J.S. wrote: "Because of time pressure we did not have enough time and energy to construct our own criteria for assessment."

Perception of learning. Four out of seven students indicated that they learned to assess products of fellow students. As T.S. stated:

> I did [learn]. I don't assess my fellow students very often. At my university we
> work rather individually. One of the challenges for me during this course was to
> work in a team. The assessments are part of working in a team.

Three students stated that that they did not learn to assess. One of them did not perform the peer assessment at all, and the second explained that s/he really tried to do so but that s/he was not an expert.

Comfort. Remarkably, five out of seven students felt really comfortable in conducting peer assessments regarding a report by students they did not meet. For instance, T.S. stated: "I think I found it easier to assess a person I did not meet or know. It is easier to write what you think, instead of feeling that you are held back by personal relationships." One student who did not feel comfortable explained that s/he was not aware of the objectives of the other group when they wrote their piece. Therefore, s/he did not know whether his/her help was useful.

Reply. After the students in one of the assessment conditions received the assessment of their group report by their fellow group, they were asked to respond. Three students stated that they did not respond, and the other four indicated that they felt comfortable responding to the assessment by their fellow group. For instance, M.B. said: "Definitely. Many comments I agreed with and I was able to reply that I would change them."

Tools. Two students in the rich assessment condition did not use the available tools. One replied s/he did not do so because his/her group did not carry out the peer assessment, and the other stated that there was no more time available at that moment. The other three students in the rich assessment condition that completed the post-test considered the tools to be valuable, although they still reported that they experienced trouble using them during the execution of the assessment assignments.

Negative aspects of peer assessment. The negative aspects of peer assessment that were mentioned in the post-test considered the organisation of the peer assessment and the feedback students received from their fellow group. T.S. suggested making one document about peer assessment including all information such as the mini-course, the criteria, and the assessment form, and put this in a logical and easily accessible place, that is, in the course documents. J.S. stated that the most negative aspect of peer assessment was "… the remark about the incoherence of the report". J.A. mentioned "… the inexperience in this kind of work for most students" and M.B. said about the negative aspect of peer assessment: "The first shock on a tough comment. I will have that all my life (…) so [I] better get used to it".

Positive aspects of peer assessment. The students mentioned some positive aspects of peer assessment. For instance, T.S. stated: "The person assessing you has had the same experience and knows what the problems in writing a text like this are. He or she has gone through the same process and can therefore give better feedback." Peers could also function as an audience, as M.B. stated: "I too easily assume people will understand data. I should write more elaborate and structured."

End marks based on scoring rubrics for product and process

The product counted for 70% of the end mark, the quality of the group process for 30%. Each tutor scored the performance of their group, using the scoring rubrics, on the product ($M = 44,5$; max = 70; range between 22 and 57) and the process ($M=20,2$; max = 30; range between 9 and 27). Most tutors provided feedback together with the scores.

Exit-meeting with tutors: use of peer assessment and scoring rubrics

During the exit meeting with the tutors and course developers at the end of the EVS-course, possible reasons were discussed for the fact that not all assessment assignments were performed by all students in the assessment conditions. According to the tutors, the main reason concerned the lack of awareness of the added value of peer assessment. "Students focussed on content, and were less interested in conducting peer assessment", one tutor stated. Obviously, students did not realise that the assessment assignments were closely related to the content. Some tutors and students had some difficulties understanding the assessment assignments right away. For instance, a tutor said: "It took me more time to realise what it meant to do peer assessment."

The scoring rubric (see Appendix 1) appeared to be very helpful for most of the tutors. They valued the fact that all group reports were administered and scored in a consistent way for all student groups. The two tutors who developed the scoring rubric appreciated that it was now made more explicit what was expected of students. One tutor expressed his/her difficulties with the rubric, because s/he could not differentiate between individual students and discovered that this scoring rubric resulted to a much lower mark than would be the case if he used his own marking scheme.

Discussion and conclusions

In the present case study our aim was to gain more insight in the possibilities of qualitative formative peer assessment in a CSCL environment. An approach was chosen in which peer assessment was operationalised in assessment assignments and assessment tools that were embedded in the course material. Our leading research questions were the following: (1) what are students' attitudes towards peer assessment and towards evaluating others in a CSCL environment and how do students perceive their own assessment skill? (2) What are the effects of the assessment assignments and tools on communication behaviour and assessment products? (3) What does the implementation of peer assessment in a CSCL environment require?

The answer to the first research question is predominantly based on the results of the questionnaires. These results showed that in general students had a positive attitude towards the use of peer assessment. Moreover, students who were actively involved in the peer assessment assignments felt positive about the clarity of the assignments and the added value of peer assessment in general.

The assessment assignments and tools affected students' behaviour and assessment products, although not exactly the way as expected. It appeared that some students in the assessment conditions skipped some assessment assignments, particularly the assignment to discuss the assessment criteria and to design of their own assessment form, and they did not use all the assessment tools. Discussions of students who discussed the assessment assignments and tools in discussion threads and chat, were short and rather superficial. Possibly, instructional prompts to discuss assessment criteria, as we did in the EVS course, is not sufficient to involve students in an effective way in discussing and negotiating assessment criteria, especially when it concerns students who have limited experience with peer assessment. For future implementation

of peer assessment we may consider specific support for each sub skill. For instance, a protocol can be provided with the necessary steps to be taken for an effective discussion of assessment criteria.

The quality of the assessment reports was rather low, with negative statements rather than positive, probably because the majority of the students did not apply the provided criteria and feedback rules. It is therefore no surprise that students reported that they sometimes had difficulties receiving critical feedback. However, they also indicated that the peer feedback was valuable for the revision of their report. Only one student wrote a reply to the assessors to indicate whether s/he agreed with the comments and how s/he processed the peer feedback and recommendations. We value a reply as an important part of the peer assessment procedure and stress that this needs more attention in future research and implementation concerning peer assessment. Not only should the rules for *providing* formative feedback be given to students, they also need more support or explicit rules for *receiving* and *accepting* feedback and *coping* with feedback.

The answers to our first and second research question reveal at least two difficulties that educational designers and teachers may encounter when they implement peer assessment in CSCL environments: the risk of a limited participation of students in peer assessment assignments and the risk of rather low quality of assessment products. We have some recommendations to overcome these difficulties. First, for a successful implementation of peer assessment in CSCL environments, we recommend to reconsider the role of the tutor. In a VLE like Blackboard 5®, tutors hardly have any control over the learning behaviour and communication activities of the students. Consequently, students have the opportunity to skip assignments, become a free rider, or keep silent for weeks. In face-to-face environments, students acquire collaboration skills by engaging in tasks defined and supervised by the teacher. In CSCL environments, in contrast, the tutor's presence is often limited to management and minor feedback functions (McLoughlin & Luca, 2002). Also for the EVS course we made the choice to let tutors fulfil the role of coach. In fact, tutors were specifically instructed to keep track of group participation and take initiative only when students were about to drop out. They had, in other words, a minor role in the group process. However, integration of peer assessment activities in online learning may require trained e-moderators (Salmon, 2002). These e-moderators are supposed to be more active during group communication by prompting and encouraging students to reflect on their group activities in interaction with team members. This reflection is essential for successful peer assessment. E-moderation by tutors could increase the number of postings but also the quality of the messages. In addition, attention must be given to the underlying ideas and belief that tutors and students have about learning, assessment, and peer assessment in particular. Greenbank (2003) stresses the importance of values of tutors and students and the influence of these values on their attitude towards new modes of assessment. Greenbank therefore advocates an analysis of these values before introducing collaborative tasks and 'new' assessments, such as peer assessment activities.

Second, we recommend letting student's performance on the assessment tasks be part of the marking. This makes it more difficult for students to skip assessment assignments. In our view, marking should involve peer assessment skills when a particular course aims at the achievement of higher order learning goals like the acquisition of feedback skills. It should be noted, however, that the acquisition of feedback skills was not an explicit goal of the EVS course that we examined in this case study.

Third, the ratio between time available for the course and time needed for the assessment assignments has to be guarded. The EVS course contained several assessment assignments that

needed substantial investment of time and effort while the content-related assignments also happened to be very much time consuming. Observations of student participation in the groups, group discussions and chats, revealed that the students were very much content driven and that they regarded the peer assessment assignments as an extra investment. Maybe we demanded a bit too much from the students in the EVS-course. Students who do not have experience with peer assessment, like the students in our study, probably need more time to do the assessment assignments. For these students, a step-by-step implementation of peer assessment during the curriculum could be more effective, with one or two small assessment assignments at the beginning of a curriculum and an increasing amount and size of the assessment assignments in subsequent courses further on in the curriculum. For instance, in the first courses of a curriculum, discussing criteria could be supported and practiced, while the actual assessment of products of fellow students could be supported and practiced in subsequent courses. An integration of the peer assessment support with the content-related tasks is necessary. This has already proven to be realisable and effective in face-to-face courses (Sluijsmans, 2002).

Finally, effective instructional conditions should be created in which students can acquire collaboration skills just as purposefully and precisely as academic skills. We believe that collaboration skills and social skills are conditional for effective peer assessment in a CSCL environment. As one of the students in our case study phrased it: "The assessments are part of working in a team." When students are involved in peer assessment according to our perspective, they have to communicate in groups, communicate between groups, negotiate about criteria, exchange products in time, and so on. For these activities complex collaboration skills and social skills are needed. However, these skills do not magically appear when tasks are employed in which students are expected to collaborate. Providing an online mini course on collaboration skills, as we did in the EVS course, may not have been sufficient to establish a sufficient quality of students' collaboration skills. McLoughlin and Luca (2002) provide some design features for the development of collaboration skills in CSCL environments. In our opinion, it is important to focus more on collaboration skills and social skills in future research, since they may have predictive value for the development of peer assessment skills. Especially when groups are formed with students from different countries, who have different cultural backgrounds, this investment in social skill development seems essential.

In our case study, we did not specifically focus on the quantitative impact of peer assessment on learning gain. Instead we conducted a qualitative oriented case study on a smaller scale, focussing on attitude and practical use of peer assessment assignments and tools. To us, increasing student's involvement in assessment is also a step forward. In addition, we first need to know how students respond to assessment assignments and tools and what obstacles may be encountered before we can conduct a large scale quantitative study in which the impact of peer assessment on learning gain can be examined. This is, of course, a direction for further research. Preferably, a longitudinal perspective should then be taken because it takes much time to acquire peer assessment skills (Sluijsmans, 2002). In our current research regarding peer assessment in CSCL environments, we continue our work on the tackling the aforementioned challenges for the design and implementation of peer assessment. In a current study, for example, attention is given tot the development of team skills by stimulating students to negotiate about criteria for effective teamwork. We do acknowledge that our ideas about peer assessment require a big investment of students and tutors, especially in situations where students and tutors are completely dependent on distance communication. A *step-by-step* approach seems to be desirable, whereby we constantly bear our ultimate goal in mind: high student involvement in educational practice where learning, instruction and assessment are completely aligned.

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.

Birenbaum, M. (1996). Assessment 2000: Towards a Pluralistic Approach to Assessment. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). Boston, MA: Kluwer Academic Press.

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

Brush, T. A. (1998). Embedding cooperative learning into the design of integrated learning systems: Rationale and guidelines. *Educational Technology Research & Development*, *46*, 5-18.

Chan, C. K. K., & van Aalst, J. (2004). Learning, assessment and collaboration in computer-supported environments. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL: And implementing it in higher education* (pp. 87-112). Boston, MA: Kluwer Academic Publishers.

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

Chin, G., Jr., & Carroll, J. M. (2000). Articulating collaboration in a learning community. *Behaviour & Information Technology*, *19*, 233-245.

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, UK: Staff and Educational Development Series.

De Graaff, R., De Laat, M., & Scheltinga, H. (2004). CSCLware in practice: Goals, tasks and constraints. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL: And implementing it in higher education* (pp. 210-219). Boston, MA: Kluwer Academic Publishers.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, *27*, 307-329.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1-16). Amsterdam: Pergamon.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, *32*, 175-187.

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

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.

Greenbank, P. (2003). Collaboration in the assessment process: An initial evaluation of collaboration on an undergraduate business and management course. *Teaching in Higher Education*, *8*(3), 317-331.

Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, *9*, 437-469.

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer assessment: The students' views. *Higher Education Research and Development*, *20*, 53-70.

Horgan, D. D., Bol, L., & Hacker, D. (1997, August). *An examination of the relationships among self, peer, and instructor assessments.* Paper presented at the seventh meeting of the European Association for Research on Learning and Instruction, Athens, Greece.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, *27*, 1, 53-64.

Johnson, D. W. (1981). Student-student interaction: The neglected variable in education. *Educational Research*, *10*, 5-10.

Kirschner, P. A. (2002). Can we support CSCL? Educational, social and technological affordances for learning. In P. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL* (7-47). Heerlen: Open University of the Netherlands.

Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL: And implementing it in higher education* (pp. 3-30). Boston, MA: Kluwer Academic Publishers.

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.

Lehtinen, E., Nurmela, K., & Salo, A. (2001, August). Case-based learning in CSCL environments. In F. Fischer & H. Mandl (Chair), *Computer-mediated cooperative learning.* Paper presented in a symposium conducted at the ninth EARLI conference, Fribourg, Switzerland.

Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation*, *23*, 319-330.

Martens, R. L. (1998). *The use of embedded support devices in independent learning.* Doctoral dissertation. Utrecht: Lemma.

McLoughlin, C., & Luca, J. (2002). A learner-centered approach to developing team skills through webbased training and assessment. *British Journal of Educational Technology*, *33*(4), 571-582.

Pond, K., Ul-Haq, R., & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education and Training International*, *32*, 314-323.

Prins, F. J. (2002). *Search and see: The roles of metacognitive skillfulness and intellectual ability during novice inductive learning in a complex computer-simulated environment.* Unpublished doctoral dissertation, Leiden University, The Netherlands.

Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: A case study. *Journal of Computer-Based Instruction*, *20*(3), 75-80.

Salmon, G. (2002). Mirror, mirror, on my screen…Exploring online reflections. *British Journal of Educational Technology*, *33*, 4, 379-391.

Salomon, G. (1992). What does the design of effective CSCL require and how do we study its effects? *SIGCUE Outlook*, *21*(3), 62-68.

Salomon, G. (2000, June). *It's not just the tool, but the educational rationale that counts.* Keynote address at the 2000 ED-MEDIA Meeting, Montreal, Canada. Retrieved January 12, 2004, from: http://construct.haifa.ac.il/~gsalomon/edMedia2000.html

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, *23*, 391-402.

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*, 349–371.

Slavin, R. E. (1980). Cooperative learning in teams: State of the art. *Educational Psychologist*, *15*, 93-111.

Slavin, R. E. (1989). Research on cooperative learning: An international perspective. *Scandinavian Journal of Educational Research*, *33*, 231-243.

Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice* (2nd ed.). Boston: Allyn & Bacon.

Sluijsmans, D. M. A. (2002). *Student involvement in assessment: The training of peer assessment skills*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.

Sluijsmans, D., Brand-Gruwel, S., Van Merriënboer, J. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education, 27*, 5, 443-454.

Sluijsmans, D. M. A., & Prins, F. J. (2004). *A theoretical framework for integrating peer assessment in teacher education.* Manuscript submitted for publication.

Sluijsmans, D., & Van Merriënboer, J. J. G. (2000). *A peer assessment model*. Heerlen: Open University of the Netherlands.

Sobral, D. T. (1997). Improving learning skills: A self-help group approach. *Higher Education*, *33*, 39-50.

Stahl, G. (Ed). (2002). *Computer support for collaborative learning: Foundations for a CSCL community*. Mahwah, NJ: Lawrence Erlbaum Associates.

Strijbos, J, W, Kirschner, P. A., & Martens (Eds.) (2004). *Computer-supported collaborative learning: Vol 3. What we know about CSCL: And implementing it in higher education*. Boston, MA: Kluwer Academic Publishers.

Strijbos, J. W., Martens, R. L., & Jochems, W. M. G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, 42, 403-424.

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

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*, 149-169.

Veenman, M. V. J. (1993). *Intellectual ability and metacognitive skill: Determinants of discovery learning in computerized learning environment.* Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Table 1

Number of messages posted in the discussion threads by the groups in the rich and bare assessment condition, concerning content, collaboration process, assessment assignments and tools, and other.

	Content	Collaboration process	Assessment	Other	Total messages discussion threads	Total chat sessions
Group 1 (rich)	16	32	4	2	47	41
Group 2 (bare)	1	1	0	0	2	10
Group 5 (bare)	18	6	1	1	26	9
Group 6 (rich)	47	106	7	9	163	33

Note. Some messages contained content as well as process-related statements

Appendix 1

The performance scoring rubric in EVS, concerning content and process

Content-related criteria	Above Standard	At Standard	Below Standard	Attribute Points Earned
	10 - 9	**8 - 6**	**5 - 0**	
Sustainable development is made operational	Students give a definition used in their report and give practical tools to measure their solutions on this points	Students give a definition used in their report but do not give practical tools to measure their solutions on this points or vice versa	Students do not give a definition used in their report and do not give practical tools to measure their solutions on this points	**/10**
	5	**4 - 3**	**2 - 0**	
The different aspect of SD are used in coherence	The ecological, social and economic aspects of Sustainable Development are used in coherence and balance. Arguments are given for priority.	Not all aspect of sustainable development are used, but the one that are used, are in balance and coherent.	The different aspect of sustainable development are not in used coherently, nor balanced.	**/10**
	10 - 9	**8 - 6**	**6 - 0**	
Consistency of the content, awareness of lacuna	In the different chapters the same definitions are used, there is no overlap between the different chapters and the content of one chapter is not striking with another. Insight is given in lacuna in knowledge.	In the different chapters the same definitions are used, there is no overlap between the different chapters and the content of one chapter is not striking with another. Lacunas in knowledge are disguised.	Different definitions are used. Chapters are striking with one another. Lacunas in knowledge are disguised.	**/10**
	5	**4 - 3**	**2 – 0**	
Integration of disciplinary contributions	The different disciplines are integrated in each chapter and not only at the end.	The different disciplines are only integrated at the end of the report.	The different disciplines are not integrated.	**/5**

	10- 9	8 - 6	5 – 0	
Relation problem definition –analysis – solution	Scientific quality of report and logical forthcoming of conclusions and recommendations. Questions asked in the beginning are answered.	Scientific quality of report and logical forthcoming of conclusions and recommendations. Not all questions asked in the beginning are answered	Low scientific quality of report and conclusions and recommendations do not come from the chapters in the report. Not all questions asked in the beginning are answered	**/10**
	5	4 - 3	2 – 0	
Application of result	A target group is distinguished, is involved in the process and is ready to work on the applicability of the results	A target group is distinguished, is involved in the process. It is not clear in which way the target group will work further with the results.	A target group is distinguished but is not involved in the process.	**/5**
	5	4 - 3	2 – 0	
Quality of language used	The style of the different chapters is the same and the English used is of good quality	The style of the different chapters is different. The English used is of good quality	The style of the different chapters is different. The English used is of bad quality	**/5**
	5	4 - 3	2 – 0	
Creativity	The different knowledge is linked to each other in a creative way. The recommendations are provocative and sharp.	The different knowledge is linked to each other in a creative way.	No new insight is given because knowledge of different disciplines and sources are left apart.	**/5**
	10 - 9	7 –6	5 - 0	
Summary (separately!)	A 2 - 3 page summary is added, with: Background research, recommendations, target group, possible implementation route. The summary is sharp, and provocative.	Summary is lacking one of the four points mentioned or leaves room for interpretation.	Summary is lacking two or more of the four points mentioned and leaves room for interpretation. Or no summary is added at all.	**/10**

Process-related criteria	Above Standard	At Standard	Below Standard	Attribute Points Earned
	5	4 – 3	2 - 0	
Planning research	Not all work was done at the end, the spreading was reasonable	Most of the work was done at the end.	One or more of the deadlines have not been met.	/5
	5	4 - 3	2 - 0	
Planning individual tasks	The task division was clear and every student had a reasonable task.	The task division was clear but not every student had a reasonable task.	The task division was not clear.	/5
	5	4 - 3	2 - 0	
Cooperation within the group	Decisions where made together and every group member has a vote in the group	Decisions where made but not every group member has a vote in it.	The group did work as a group, so no common decisions where made.	/5
	5	4 - 3	2 - 0	
Cooperation via the internet	The internet is used for cooperation so decisions made are traceable	The internet is not always used for cooperation, not all decisions made are traceable	The group did not use internet for decisions.	/5
	5	4 - 3	2 - 0	
Participation	Each group member participated equally. Visit to internet, input in the report are equally	Not every group member participated equally. Visit to internet, input in the report differs, but stays within reasonable variety.	Not every group member participated equally. Visit to internet, input in the report differs strongly and caused problems in the group.	/5
	5	4 - 3	2 - 0	
Incorporate comments	The project team dealt with the comments, given by other teams, by the staff and the target group in	The project team dealt with the comments, given by other teams and by the staff in a way that is recognizable and that it fits with the rest of the report. Comments from the target	The project team did not dealt with the comments	/5

	a way that is recognizable and that it fits with the rest of the report.	group are left out of the report.		

Appendix 2
Characteristics of good feedback

Bregquist, W. H., & Phillips, S. R. (1975). Components of an effective faculty development
      program. *Journal of Higher Education*, *46*, 177-211.

- promotes reflection as part of a dialog between the giver and receiver of feedback. Both parties are involved in observing, thinking, reporting, and responding.
- focuses on observed behaviour rather than on the person. Refers to what an individual does rather than to what we think s/he is.
- is descriptive rather than judgmental. Avoiding judgmental language reduces the need for an individual to respond defensively.
- is specific rather than general.
- promotes reflection about strategies and the students' or observer's responses to a specific strategy.
- is directed toward behaviour that the receiver can change.
- considers the needs of both the receiver and giver of feedback.
- is solicited rather than imposed. Feedback is most useful when the receiver actively seeks feedback and is able to discuss it in a supportive environment.
- is well timed. In general, feedback is most useful at the earliest opportunity after the given behaviour.
- involves sharing information rather than giving advice, leaving the individual free to change in accordance with personal goals and needs.
- considers the amount of information the receiver can use rather than the amount the observer would like to give. Overloading an individual with feedback reduces the likelihood that the information will be used effectively.
- requires a supportive, confidential relationship built on trust, honesty, and genuine concern.