



Peer Assessment Training in Teacher Education: Effects on performance and perceptions

Dominique M. A. Sluijsmans , Saskia Brand-Gruwel & Jeroen J. G. van Merriënboer

To cite this article: Dominique M. A. Sluijsmans , Saskia Brand-Gruwel & Jeroen J. G. van Merriënboer (2002) Peer Assessment Training in Teacher Education: Effects on performance and perceptions, *Assessment & Evaluation in Higher Education*, 27:5, 443-454, DOI: [10.1080/0260293022000009311](https://doi.org/10.1080/0260293022000009311)

To link to this article: <https://doi.org/10.1080/0260293022000009311>



Published online: 27 May 2010.



Submit your article to this journal [↗](#)



Article views: 830



Citing articles: 97 [View citing articles](#) [↗](#)

Peer Assessment Training in Teacher Education: effects on performance and perceptions

DOMINIQUE M. A. SLUIJSMANS, SASKIA BRAND-GRUWEL & JEROEN J. G. VAN MERRIËNBOER, *Open University of the Netherlands, Heerlen, The Netherlands*

ABSTRACT *The implementation of peer assessment receives much attention in teacher education. This paper reports the effects of peer assessment training on the performance of student teachers. Ninety-three student teachers were randomly assigned to control groups and experimental groups. The experimental groups were trained in defining performance criteria, giving feedback and writing assessment reports. This was done through peer assessment tasks that were embedded in a redesigned course. Analyses of data derived from peer assessment reports written by the students showed that the experimental groups surpassed the control groups in the quality of the assessment skill. As a result of the training, students from the experimental groups also scored significantly higher grades for the end products of the course than students from the control groups. The results of the questionnaire showed that all students were significantly more satisfied with the redesigned course.*

Peer assessment practices have been applied in many institutions for over 50 years. Although Kane and Lawler in their publication of 1978 address the failure of recognition of the use of peer assessment, many studies since then have proven the importance to involve students in the assessment process through peer assessment (e.g., Arter, 1996; Boud *et al.*, 1999; Cutler & Price, 1995; Dochy & McDowell, 1997; Fallows & Chandramohan, 2001). There is an immense number of studies illustrating how peer assessment practices can be applied in curricula in both formative and summative ways (see Dochy *et al.*, 1999; Falchikov & Goldfinch, 2001; Sluijsmans *et al.*, 1999). Besides the analysis of peer–tutor correlations, which are reported in the majority of the studies, some of these studies also evaluate the perceptions and feelings of students regarding the process of peer assessment. These evaluations draw a diverse picture. On the one hand students regard peer assessment as useful for their own learning process. Orsmond *et al.*

(1996) found that students enjoyed carrying out the peer assessment and considered that it was beneficial to their learning. Keaten and Richardson (1992) also affirmed that peer assessment fosters an appreciation for internal awards and interpersonal relationships in the classroom.

On the other hand, students found that criticising their friends was difficult. Hanrahan and Isaacs (2001) mention the hostility that students show towards peer assessment in their university courses. Some quotations of students illustrate this reaction: “You don’t want to mark a fellow student too harshly”, or “I feel uncomfortable evaluating another student’s paper”. Investigators actually found increased opposition to peer assessment after student exposure to it (Rushton *et al.*, 1993). The part of giving grades to each other especially concerns students and is seen as “risky and unfair” (Kwan & Leung, 1996). Besides these findings, students also doubt the objectivity of peer assessment and claim to have no training in such assessment practices (Cheng & Warren, 1997; Sluijsmans *et al.*, 2001). This last point, the lack of instruction, is the central issue in this paper. The aim of the presented study is to examine how students can be involved in assessment and what the effects are of structural involvement on students performance and perceptions.

Critical evaluation of the performances of peers is especially important in the teacher-training context. Conducting a peer assessment is considered as a specific skill in the Dutch vocational training profile of primary school teachers (LPC, 1995). Unfortunately, there are no explicit training programmes available that provide teachers in teacher training colleges with guidelines how to train their student teachers in these assessment skills (Tillema *et al.*, 2000). The literature on peer assessment particularly focuses on the importance of negotiating about performance criteria (Falchikov, 1995; Mehrens *et al.*, 1998; Orsmond, 1996 *et al.*, 2000), but that is only one of the skills required for conducting reliable assessments. To make an overview of the important skills, a literature review and expert interviews were conducted (Sluijsmans & Van Merriënboer, 2000). This resulted in a peer assessment model in which three main skills are taken into account. These skills are: (1) defining assessment criteria—thinking about what is required and referring to the product or process; (2) judging the performance of a peer reflecting upon and identifying the strengths and weaknesses in a peer’s product and writing an assessment report; and (3) providing feedback for future learning—giving constructive feedback about the product of a peer.

A training programme for peer assessment should be based on these skills (Van Merriënboer, 1997). The design of such a training program results in a number of peer assessment tasks, which should be embedded in an existing course. After all, the course content determines the object of the peer assessment. Regarding the skills in the model, it is a prerequisite that the teachers of the course agree on the course objectives and the assessment criteria, before they can guide students in their skills. If it turns out that an existing course contains course objectives that are ambiguous, it is necessary to discuss these course objectives with the teachers of the course. These discussions can lead to a thorough redesign to make it suitable for training in assessment skills.

The peer assessment tasks can be regarded as the learning exercises in which the assessment skills are practised. An example of a peer assessment task is the negotiation about assessment criteria for a specific product or process (e.g., Arter, 1996; Dochy & McDowell, 1997; Mehrens *et al.*, 1998; Stainer, 1997). A hypothesis is that understanding the criteria of a specific product in a certain subject matter domain and analysing the work of peers can lead to an improved awareness of the quality of one’s own work (Falchikov, 1995; Freeman, 1995; Mehrens *et al.*, 1998).

The aim of this study is to determine the effect of a course-embedded peer assessment training on: (1) students' assessment skill; (2) task performance in the domain of the course; and (3) perceptions regarding a redefined course. Because of the learning purpose and the research goals, the peer assessments were *not* used as part of the final examination. The students, however, received the study points for the course only if they conducted the peer assessments seriously and handed in their assessment reports for the research purpose.

Method

Participants

The sample consisted of 93 second-year students of a teacher training college in the Netherlands (19 male, 74 female) with an average age of 20.7 years ($SD = 1.6$). The students were randomly distributed amongst four educational groups, of which two were experimental groups ($n = 50$) and two were control groups ($n = 43$). Four teachers of the teacher training college participated in the study. Each teacher was responsible for a certain subject matter domain in the course. These domains were Art, Dutch Language, Music and Pedagogy.

Materials

Course. The second-year course "Designing Creative Lessons" was selected. This course has been part of the curriculum for several years. Because the course objectives were not revised for years and teachers had developed multiple perspectives on what the content should be, the teachers who were jointly responsible for this course first redefined the course objectives, by decomposing the skill of designing a creative lesson. This resulted in a concept map with a number of constituent skills. For the domains Art, Dutch Language and Music, four one-hour study tasks were defined, based on the constituent skills. In the study tasks the students of both the control groups and the experimental groups were supported in the content-related skill "designing creative lessons". The Pedagogy teacher designed four one-hour study tasks that integrated the tasks of the domains Art, Dutch Language and Music.

Peer assessment training. Four peer assessment tasks of one hour each were designed for the two experimental groups. These tasks were embedded in the pedagogy domain, but were closely related to the study tasks concerning designing creative lessons. The training focused on the three main constituent skills of the peer assessment model. The didactics used in the training were discussion and elaboration.

In Task 1, students were introduced to the meaning of peer assessment and the product that they were going to peer assess at the end of the course. This product was a video of a creative learning lesson taught by themselves. After this introduction students watched a creative learning lesson on video and discussed and elaborated on the fragments in which creativity was applied. This resulted in a first rough draft of the criteria that are required for a creative lesson.

In Task 2, the skill "defining criteria" was addressed. Examples of valid and invalid criteria were presented. Students then further elaborated on the rough criteria for designing a creative lesson they formulated in the first task. This exercise resulted in a list of 15 criteria that was accepted by the whole group of students and the teacher.

Discussing the purpose and guidelines for giving constructive feedback was the central topic in Task 3. In the peer assessment model, this is the skill “provide feedback for future learning”. The teacher first asked the students what their ideas were about feedback and criticism. After a short discussion, the teacher presented an expert assessment report to the students, which was written by two experts on creative learning regarding the video lesson that was analysed in Task 1. Students discussed the good examples of constructive feedback in the expert assessment report. These were that the feedback was clear and unambiguous. The teachers started with positive comments and provided suggestions. They also posed questions. At the end of the task, students had to give each other feedback on some aspects of their own work. The output of this task was a list of criteria for constructive feedback. Examples of criteria were that the feedback has to be specific, direct, accurate, achievable, practicable and comprehensible to the peer.

In Task 4, the students were trained in the third main skill of the peer assessment model, namely “judge the performance of a peer”. In this final task, the three prior tasks were integrated. To confront the students with ways in which an assessment report can be written, they analysed the expert assessment report and discussed the structure that was applied by the experts. They also discussed the language used in the assessment, for example the use of naive words, like ‘nice’. It was concluded that novice assessors frequently use words such as ‘nice’ and ‘good’ instead of more substantiate words, due to a lack of content and assessment expertise. Experts use more words regarding content and even sometimes avoid words like ‘nice’ and ‘good’.

Based on the findings of the discussion, students defined a peer assessment form. A summary of this form is presented in the Appendix.

Peer assessment form. At the end of the course, students from both groups had to assess the videotapes of the creative lessons of three peer groups on a peer assessment form. The peer assessment form defined by the students from the experimental groups in the fourth peer assessment task, consisted of four pages. A summary of this form is presented in the Appendix. The students from the control groups also used this form.

Rating form. To analyse the quality of the peer assessments that were written by the students, a rating form was developed. In this rating form seven variables were included. These variables were based on an expert peer assessment that was used in the peer assessment training. One variable was related to the criteria (“use of criteria”), four variables were related to giving feedback (“positive comments”, “negative comments”, “constructive comments”, and “posed questions”), and another two variables were related to judging the performance of a peer by writing an assessment report (“naive word use”, and “structure”). Each student could score a maximum of 15 points on the first variable “use of criteria”, because the students defined 15 criteria in the first peer assessment task. The four variables for “giving feedback” were measured by counting the number of comments and questions. The variable “naive word use” was measured by counting words such as ‘nice’, ‘good’, ‘excellent’, and ‘fine’. Four items in the rating form were included to measure the variable “structure”. On these four items a maximum score of 13 points could be gained.

Three independent research assistants scored the peer assessment forms with the rating form. For each variable the interrater reliabilities were calculated. These reliabilities were acceptable for all variables (Cohen’s Kappa > 0.95).

Examinations. To measure an effect of the peer assessment training on the task performance of students, the marks on the end products given by the teacher were analysed. These end products were a collection of assignments, a group report, an individual report and a creative product for each domain. The teachers marked all products. The average of the four marks was calculated for each student. The score could range from 0 to 100.

Student questionnaire. Before and after the course, the students filled out a questionnaire about their perceptions on instruction and assessment. Seventy-two items were divided among 13 variables. Five variables were related to instruction, four variables were related to vision on instruction and assessment, and another four were related to the role of the student in assessment. The students had to score the items on a 5-point Likert scale, varying from “I totally disagree” to “I totally agree”. The pre-test was carried out to investigate the students’ perceptions on prior courses that were comparable to the course on “Designing Creative Lessons”. These courses were not designed in a skill-based way. The post-test concerned students’ perceptions after the redesigned course. The clusters, variables, number of items, reliability coefficients and example items of the 13 variables are presented in Table 1.

Design and Procedure

The study was set up according to a pre-test/post-test control group design. Before the start of the course, all students filled out the questionnaire. During the course, all students worked in subgroups of five or six students and each student prepared his or her lesson that was taped on video and the subject of the peer assessment. In between classes, the whole group of students worked independently on assignments.

During the course the students of the experimental groups performed the peer assessment tasks. Instead of these tasks the students of the control groups attended four extra hours in the pedagogic domain. During these hours they had the opportunity to elaborate on certain aspects of creative learning. At the end of the course, a peer assessment session was organised for each group, in which the video lessons of each subgroup were shown (four video lessons in each group).

The peers were instructed to write a qualitative peer assessment with regard to the content of the video lesson. The experimental groups were free to use the output of the peer assessment tasks. For the peer assessment, the students from the control group had to use the regular course materials from the study tasks. Each student wrote three peer assessments, because in each group there were three other subgroups to assess. After the course, all students filled out the same questionnaire as in the pre-test.

Data Analyses

Three independent research assistants analysed the 279 peer assessment forms (93 students who completed three assessments). These research assistants had prior experience of the use of rating forms. One-way analyses of variance were applied to identify differences between the control groups and experimental groups on the seven variables of the rating form. One-way analyses of variance were also applied to identify differences between the control and experimental groups on the product performances. Means and standard deviations were calculated for the 13 variables of the student questionnaire for the control and experimental groups. The scores of each variable were

TABLE 1. Clusters, variables, number of items, reliability coefficients and example items of the 13 variables of the student questionnaire

| Variable | # | α | Example items |
|---|----|----------|--|
| Cluster: Instruction | | | |
| Satisfaction classes | 5 | 0.65 | The study tasks evoked interesting discussions |
| Transparency classes | 4 | 0.68 | The course objectives were comprehensible |
| Learning access level | 4 | 0.61 | I felt that I could distinguish main issues from side issues |
| Practical relevance | 3 | 0.71 | The study tasks are practically oriented |
| Quality of the instruction | 7 | 0.88 | The goals of the study tasks were explained very clearly |
| Cluster: Vision on instruction and assessment | | | |
| Relation instruction and assessment | 4 | 0.72 | The study tasks and the assessment were interrelated |
| Fear for assessment | 3 | 0.58 | I'm usually very nervous before taking an exam |
| Obrusiveness assessments | 5 | 0.76 | The questions on an exam have to be public to students before the exam is taken |
| Overall vision on assessment | 2 | 0.71 | I support the way I am assessed |
| Cluster: Role of student in assessment | | | |
| Involvement in assessment | 8 | 0.45 | I think that students should be more involved in the development of assessment criteria |
| Group behaviour | 5 | 0.65 | I don't think like it when students don't make an individual contribution to a group product |
| Collaborative learning | 3 | 0.71 | I prefer to elaborate on problems with my peers |
| Assessment skill | 18 | 0.87 | I'm able to analyse a product of a peer |

TABLE 2. Means and standard deviations of the experimental and control groups on the peer assessment forms at the post-test

| Variable | Experimental groups | | Control groups | |
|-------------------------|---------------------|-----------|----------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Use of criteria** | 9.62 | 2.34 | 6.68 | 2.14 |
| Positive comments | 6.55 | 2.11 | 6.13 | 3.14 |
| Negative comments | 1.47 | 1.06 | 1.32 | 0.88 |
| Constructive comments* | 2.56 | 1.58 | 1.26 | .97 |
| Posed questions | 0.14 | 0.21 | 0.13 | 0.24 |
| Naive word use* | 1.42 | 0.78 | 1.92 | 0.87 |
| Structure ^{b*} | 6.67 | 1.46 | 6.21 | 1.47 |

* $p < 0.01$.

^a maximum score = 15.

^b maximum score = 13.

analysed with a 2 (Groups) \times 2 (Time of Testing) analysis of variance with repeated measures on the last factor.

Results

To what extent does the assessment skill of students receiving the training in peer assessment differ from those students who did not receive training?

Table 2 presents the means and standard deviations of the seven variables that were measured with the rating form for the experimental and control groups. The analyses reveal significant differences between the groups on four of the seven variables. The experimental groups applied more criteria (variable “use of criteria”, $F(1,85) = 36.10$, $MSE = 5.07$, $p < 0.01$) and gave more constructive feedback (variable “constructive comments”, $F(1,85) = 20.13$, $MSE = 1.77$, $p < 0.01$). They also use fewer naive words (variable “naive word use”, $F(1,85) = 7.57$, $MSE = 0.68$, $p < 0.01$). Finally, the experimental groups scored better on the variable “structure” ($F(1,85) = 7.89$, $MSE = 1.98$, $p < 0.01$).

Does following the training lead to better performances? All students handed in several products to ground their development in the skill to design creative lessons. Four marks were given to each student: one for the collection of assignments students carried out during and between the study tasks, one for the group report about the process of making the video lesson, a mark for the individual report about the design of creative lessons, and one average mark for the creative products for each domain. The average of the four marks was calculated. The score of the experimental groups was 72.77 ($SD = 6.72$) compared to an average score of the control groups of 68.42 ($SD = 5.27$). The difference between both groups was significant ($F(1,83) = 10.59$, $MSE = 37.45$, $p < 0.01$). The students from the experimental groups thus performed better on the skill of designing creative lessons than students from the control groups. Further analyses show that the significant difference between the two groups is caused by the significant effects in the performance of two of the four products, namely the collection of assignments and the group report. The students from the experimental groups scored an average of 72.34 ($SD = 6.98$) on the assignments against an average of 67.63 ($SD = 6.75$) of the students from the control groups, which is a significant effect ($F(1,83) = 9.84$, $MSE = 47.34$, $p < 0.01$). The average score on the group report gained

TABLE 3. Means and standard deviations of the experimental and control groups' students questionnaire results at the pre-test and post-test on a 5-point Likert-scale

| | | Experimental groups | | Control groups | |
|---|-----------|---------------------|-----------|----------------|-----------|
| | | pre-test | post-test | pre-test | post-test |
| Cluster: Instruction | | | | | |
| Satisfaction classes | <i>M</i> | 3.10 | 3.99 | 2.99 | 3.85 |
| | <i>SD</i> | 0.51 | 0.36 | 0.53 | 0.44 |
| Transparency classes | <i>M</i> | 3.15 | 3.99 | 3.28 | 3.88 |
| | <i>SD</i> | 0.63 | 0.34 | 0.54 | 0.58 |
| Learning access level | <i>M</i> | 3.45 | 3.92 | 3.49 | 3.79 |
| | <i>SD</i> | 0.81 | 0.41 | 0.80 | 0.51 |
| Practical relevance | <i>M</i> | 3.52 | 4.14 | 3.57 | 4.01 |
| | <i>SD</i> | 0.84 | 0.39 | 0.69 | 0.68 |
| Quality of the instruction | <i>M</i> | 2.81 | 3.63 | 2.91 | 3.61 |
| | <i>SD</i> | 0.75 | 0.55 | 0.50 | 0.67 |
| Cluster: Vision on instruction and assessment | | | | | |
| Relation instruction and assessment | <i>M</i> | 2.18 | 3.89 | 2.37 | 3.85 |
| | <i>SD</i> | 0.72 | 0.42 | 0.59 | 0.74 |
| Fear for assessment | <i>M</i> | 2.70 | 2.28 | 2.88 | 2.14 |
| | <i>SD</i> | 0.91 | 1.25 | 0.99 | 0.88 |
| Obtrusiveness assessment | <i>M</i> | 2.41 | 3.39 | 2.49 | 3.49 |
| | <i>SD</i> | 0.56 | 0.56 | 0.57 | 0.57 |
| Overall vision on assessment | <i>M</i> | 2.80 | 3.89 | 3.17 | 3.91 |
| | <i>SD</i> | 1.02 | 0.74 | 0.88 | 0.84 |
| Cluster: Role of student in assessment | | | | | |
| Involvement in assessment | <i>M</i> | 3.20 | 3.42 | 3.13 | 3.28 |
| | <i>SD</i> | 0.45 | 0.41 | 0.44 | 0.49 |
| Group behaviour | <i>M</i> | 4.17 | 4.27 | 3.93 | 4.04 |
| | <i>SD</i> | 0.45 | 0.51 | 0.54 | 0.60 |
| Collaborative learning | <i>M</i> | 3.80 | 3.92 | 3.98 | 3.87 |
| | <i>SD</i> | 0.60 | 0.59 | 0.44 | 0.74 |
| Assessment skill | <i>M</i> | 3.82 | 3.80 | 3.69 | 3.73 |
| | <i>SD</i> | 0.41 | 0.35 | 0.37 | 0.42 |

by the experimental groups was 74.38 ($SD = 11.09$), compared to an average of 63.66 for the control groups ($SD = 9.94$). This difference was also significant ($F(1,83) = 22.70$, $MSE = 111.87$, $p < 0.001$).

What are the perceptions of students regarding the redesigned course?

In Table 3 the means and standard deviations of the student questionnaire are given. The scores of each variable were analysed according to a 2 (Groups) \times 2 (Time of Testing) analysis of variance with repeated measures on the last factor. There were highly significant main effects for the factor Time of Testing. This is the case for 11 of the 13 variables. All students—experimental and control group students together—were more positive in the post-test than in the pre-test about the classes ($F(1,68) = 65.82$, $MSE = 13.74$, $p < 0.001$), and the transparency of the study tasks ($F(1,68) = 60.42$, $MSE = 13.83$, $p < 0.001$). Students found the redesigned course more attuned to their learning access level ($F(1,70) = 74.38$, $MSE = 21.14$, $p < 0.001$), and were more positive about the practical relevance of the course ($F(1,70) = 38.19$, $MSE = 12.76$, $p < 0.001$), and the quality of the instruction ($F(1,68) = 37.90$, $MSE = 9.74$, $p < 0.001$). They were less afraid about the assessment ($F(1,70) = 72.59$, $MSE = 53.95$, $p < 0.001$), and regarded the assessment in the redesigned course as more obtrusive (F

(1,69) = 184.99, $MSE = 37.36$, $p < 0.001$). Students felt more involved in the assessment ($F(1,69) = 17.57$, $MSE = 3.13$, $p < 0.01$). The relation between instruction and assessment became more apparent for the students ($F(1,70) = 64.01$, $MSE = 29.04$, $p < 0.001$). Students' overall vision on assessment changed positively ($F(1,69) = 45.02$, $MSE = 32.99$, $p < 0.001$).

On the factor Groups two significant main effects were found. One on the variable "involvement in assessment" ($F(1,69) = 4.66$, $MSE = 1.29$, $p < 0.05$), and one on the variable "group behaviour" ($F(1,69) = 8.34$, $MSE = 3.09$, $p < 0.01$). In general the experimental groups outscored the control groups.

The interaction Time of Testing \times Groups yielded no significant effects. The peer assessment training had no effect on students' perceptions.

Conclusion and Discussion

This study was carried out to investigate the effects of a peer assessment training embedded in a course on Designing Creative Lessons, on the development of the peer assessment skill. A second research question was whether this training had effects on the performance of students. A third question was how students perceived the redesigned course.

Regarding the first question, the analyses of the qualitative peer assessment reports clearly revealed that the training had the expected positive effects on the development of the peer assessment skill. The experimental groups were more likely to use the criteria and to give more constructive comments than the student teachers from the control groups. The students who received training also scored higher on structure and used fewer naive words. In spite of these positive results, student teachers cannot be regarded as expert assessors after this training. The experimental groups only applied nine of the 15 possible criteria. Also, two constructive comments per assessment report indicates small progress. Therefore longer training periods seem to be recommendable.

On the second research question, whether the peer assessment training led to an improved performance in the content domain, an even more important result was found: a positive effect of the peer assessment training on the performance in the content domain. The student teachers from the experimental groups outperformed the students from the control groups. Besides the overall significant difference, it was also found that this effect occurred thanks to significant results on two of the four end products. A more profound training may lead to an effect on the performance of all the end products.

The results of the analysis of the questionnaire data relating to students' self-perceptions regarding the redesigned course were positive (research question 3: *What are the perceptions of students regarding the redesigned course?*). The student teachers indicate a positive change in their view on several aspects of assessment, instruction and the role of the teacher. Assessment and instruction were more aligned in the redesigned course. Test anxiety decreased and the student teachers felt more involved in the assessment procedure.

A surprising result was that there is no effect on the variable "assessment skill", because this is precisely what the students were trained in. An explanation for this odd effect may be caused by the design of the study. A problem in the design of the presented study is that the pre-test/post-test comparison results of the student questionnaire may be an inaccurate measurement of the impact of the redesigned course, because the students may have limited knowledge about assessing at the beginning of the course that prevents them from accurately self-assessing baseline opinions (Sprangers, 1989).

By the end of the course, their new understanding of peer assessment and the skills involved may have an impact on the responses on their self-evaluation.

Reflecting on the results, some recommendations for future research can be given. First, the training period has to be extended considerably, to attain full mastery of the assessment skill. Second, we have to be aware of the finding that using more criteria does not necessarily mean that the experimental groups apply the criteria *adequately and correctly*. Further analyses of the assessment reports by content experts is recommended. Third, it is important to investigate peer assessment in different subject matter domains to test the generalisability of training. Peer assessment is a skill that can have a life beyond the task for which it is employed in this study (James, 2000). Peer assessment can be used in comparable tasks within other domains. The question is: can students assess products in other subject matter domains after a training in one particular domain, or do they need additional training in order to be able to assess in other domains?

Fourth, a suggestion is to begin training students in their first year of teacher training. Observation during informal evaluation of the peer assessment procedure with students showed that students are conservative and conditioned in their attitude towards teachers and assessment. They still feel that the teacher is the expert and the only objective assessor. When students are trained to rely on their own judgement and that of their peers, they may develop a belief that a teacher is first of all a coach, who supports and adjusts the decisions that students make. The effects of ongoing involvement in assessment and instruction on the development of these beliefs need to be further examined.

A final issue that deserves further elaboration is the relationship between content knowledge and assessing (see Mehrens *et al.*, 1998). Peer assessing is a complex skill that cannot be demonstrated outside a particular subject matter domain. It can be hypothesised that students who are novices in a certain domain are also less capable of assessing. One student stressed this issue in his evaluation: "I have problems with assessing a product of a peer that is of a higher level than my own product. Then I realize that I don't have enough domain knowledge to criticize it." This can also occur the other way round: the more domain expertise a student has, the more capable he or she is of evaluating the work of peers. An interesting question for further research would be to what extent domain expertise influences the development of assessment skills.

In summary, it can be concluded that students can be trained in assessment skills and that such training positively affects the performance level of student teachers. A thorough redesign of courses, whereby student teachers become more involved in assessment and instruction, leads to more satisfaction among students. Although student questionnaires should not be considered as determinative, the results of this study at least show that the benefits of peer assessment may only be realised after serious effort is made to incorporate it into the everyday teaching practices in a way which is positive, non-threatening and attractive to students. As Boud *et al.* (1999) state: "It is unrealistic to discuss assessment in isolation from curriculum content and teaching strategies" (p. 424).

Notes on Contributors

DOMINIQUE SLUIJSMANS is an educational technologist at the Educational Technology Expertise Center of the Open University of the Netherlands. Her main interests are student involvement in assessment, teacher education and curriculum design.

Correspondence: Dominique Sluijsmans, Open University of the Netherlands, Educa-

tional Technology Expertise Center, P. O. Box 2960 6401 DL Heerlen, The Netherlands, Tel: + 31-45-5762178, Fax: + 31-45-5762802. E-mail: dominique.sluijsmans@ou.nl.

SASKIA BRAND-GRUWEL is an educational technologist at the Educational Technology Expertise Center of the Open University of the Netherlands. Her areas of specialisation are comprehensive reading, higher order skills and competency-based education.

JEROEN VAN MERRIËNBOER is a full professor at the Educational Technology Expertise Center of the Open University of the Netherlands, where he also heads the research program. He specialises in Instructional Design (ID) for complex learning, computer-based learning environments, and intelligent performance support for ID.

REFERENCES

- ARTER, J. (1996) Using assessment as a tool for learning, in: R. Blum & J. Arter (Eds) *Student performance assessment in an era of restructuring* (Alexandria, VA, Association for Supervision and Curriculum Development).
- BOUD, D., COHEN, R. & SAMPSON, J. (1999) Peer learning and assessment, *Assessment and Evaluation in Higher Education*, 24 (4), pp. 413-426.
- CHENG, W. & WARREN, M. (1997) Having second thoughts: student perceptions before and after a peer assessment exercise, *Studies in Higher Education*, 22 (2), pp. 233-239.
- CUTLER, H. & PRICE, J. (1995) The development of skills through peer assessment, in: A. EDWARDS & P. KNIGHT (Eds) *Assessing competence in higher education* (Birmingham, UK, Staff and Educational Development Series).
- DOCHY, F. J. R. C. & McDOWELL, L. (1997) Assessment as a tool for learning, *Studies in Educational Evaluation*, 23 (4), pp. 279-298.
- DOCHY, F., SEGERS, M. & SLUIJSMANS, D. (1999) The use of self-, peer-, and co-assessment in higher education: a review, *Studies in Higher Education*, 24 (3), pp. 331-350.
- FALCHIKOV, N. (1995) Peer feedback marking: developing peer assessment, *Innovations in Education and Training International*, 32 (2), pp. 175-187.
- FALCHIKOV, N. & GOLDFINCH, J. (2001) Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks, *Review of Educational Research*, 70 (3), pp. 287-322.
- FALLOWS, S. & CHANDRAMOHAN, B. (2001) Multiple approaches to assessment: reflections on use of tutor, peer and self-assessment, *Teaching in Higher Education*, 6 (2), pp. 229-246.
- FREEMAN, M. (1995) Peer assessment by groups of group work, *Assessment and Evaluation in Higher Education*, 20 (3), pp. 289-300.
- HANRAHAN, S. & ISAACS, G. (2001) Assessing self- and peer assessment: the students' views, *Higher Education Research and Development*, 20 (1), pp. 53-70.
- JAMES, P. (2000) A blueprint for skills assessment in higher education, *Assessment and Evaluation in Higher Education*, 25 (4), pp. 353-367.
- KEATEN, J. A. & RICHARDSON, M. E. (1992) A field investigation of peer assessment as part of the student group grading process. Paper presented at the Western Speech Communication Association Convention, Albuquerque, NM.
- 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 (3), pp. 205-214.
- LANDELIJKE PEDAGOGISCHE CENTRA (LPC) (1995) *Beroep in beweging. Beroepsprofiel leraar primair onderwijs [Profession in action. Vocational training profile for the primary school teacher]*. (Utrecht, Forum Vitaal Leraarschap).
- MEHRENS, W. A., POPHAM, W. J. & RYAN, J. M. (1998) How to prepare students for performance assessments, *Educational Measurement: Issues and Practice*, 17 (1), pp. 18-22.
- ORSMOND, P., MERRY, S. & REILING, K. (1996) The importance of marking criteria in the use of peer assessment, *Assessment and Evaluation in Higher Education*, 21 (3), pp. 239-249.
- ORSMOND, P., MERRY, S. & REILING, K. (2000) The use of student derived marking criteria in peer and self-assessment, *Assessment and Evaluation in Higher Education*, 25 (1), pp. 23-38.

- RUSHTON, C., RAMSEY, P. & RADA, R. (1993) Peer assessment in a collaborative hypermedia environment: a case study, *Journal of Computer-Based Instruction*, 20 (3), p. 75.
- SLUUSMANS, D., DOCHY, F. & MOERKERKE, G. (1999) Creating a learning environment by using self-peer- and co-assessment, *Learning Environments Research*, 1 (3), pp. 293–319.
- SLUUSMANS, D., MOERKERKE, G., DOCHY, F. & VAN MERRIËNBOER, J. J. G. (2001) Peer assessment in problem based learning, *Studies in Educational Evaluation*, 27 (2), pp. 153–173.
- SLUUSMANS, D. & VAN MERRIËNBOER, J. J. G. (2000) *A peer assessment model* (Heerlen, Open University of the Netherlands, Center for Educational Technology and Expertise).
- SPRANGERS, M. (1989) *Response shift and the retrospective pre-test* (SVO, The Hague).
- STAINER, L. (1997) Peer assessment and group work as vehicles for student empowerment: a module evaluation, *Journal of Geography in Higher Education*, 21 (1), pp. 95–98.
- TILLEMA, H. H., KESSELS, J. W. M. & MEIJERS, F. (2000) Competencies as building blocks for integrating assessment with instruction in vocational education: a case from the Netherlands, *Assessment and Evaluation in Higher Education*, 25 (3), pp. 265–278.
- VAN MERRIËNBOER, J. J. G. (1997) *Training complex cognitive skills* (Englewood Cliffs, NJ, Educational Technology Publications).

Appendix

Summary of the peer assessment form

PEER ASSESSMENT on CREATIVE TEACHING

Names of peer assessor _____

Name of peer assessee _____

Please write a peer assessment report on the following pages with regard to the video lesson you observed based on the personal notes you made during the observation.

General impression:

Criteria that I applied:

Criterion 1

Criterion 2

Criterion 3

Criterion ...

Feedback:

I have the following conclusions and recommendations:

END OF THE PEER ASSESSMENT
