Designing student peer assessment in higher education:

analysis of written and oral peer feedback

There is an increasing amount of attention in higher education for the concept of student peer assessment (PA), which is to be understood as an educational arrangement in which students assess the quality of their fellow students' work and provide one another with feedback (Dochy *et al*. 1999). This development is in line with other recent developments in university teaching, such as collaborative learning and writing, and real-life task performance (see, for example, Van Weert & Pilot, 2003).

There have been many studies on the results of PA in relation to grading and student satisfaction, as well as on effective organization of PA in higher education (Topping, 1998, and Van den Berg *et al*., in press). However, little has been said as to the content of students' feedback in educational designs employing PA. More insight into the nature of the feedback would indicate more clearly how students could support one another and what kind of assistance teachers should preferably provide. For example, teachers facing adequate feedback on style and appeal, but nothing on textual coherence or content, will

1

know where to direct their assistance. Additionally, we hoped our study would reveal more about the relation between content of peer feedback and design features of PA.

Peer assessment and collaborative writing

Peer assessment can, to a certain extent, be understood as a type of collaborative learning (Falchikov, 2001). Compared to collaborative, co-operative, or peer learning- in all of which students produce a collective outcome, share knowledge, and learn from the collaboration- PA is generally more limited, requiring students simply to assess one another's work by means of relevant criteria, and to provide feedback.

Peer feedback seems to be most valuable in the collaborative writing process. Based on a typology of Stodolsky (1984), Saunders (1989) distinguished five types of collaborative writing, as summarized in Table 1.

= Table 1 =

2

Saunders states that the type of collaborative task influences the content of peer feedback as well as the learning outcomes. When 'co-writing', students go through the entire writing process together (planning, composing, revising and editing). This process also includes elements of selection: what will be our topic, how shall we structure it? As it involves all phases of the writing process, the interaction between co-writers may be more extensive than that of 'co-publishers', which in turn may be more extensive than the interaction of co-responders and co-editors. When operating as co-publishers, students plan their collective text together and divide their labor into separate, individual writing tasks, upon completion of which the group pieces together the collective end product. In a 'co-responding' collaboration, students plan and write individually, but are required to assist one another in the reviewing process. The interaction will be extensive when students discuss their interpretation of the text and negotiate about measures for text improval. But the interaction may also be restricted to a simple pattern of (unilateral) comments and responses, in which feedback is indeed provided and accepted (or not, as the case may be), but not discussed. Finally, co-editing constitutes the least intensive form of collaborative writing. When co-editing,

students plan, write and review individually, only being required to assist one another in correcting the last version of the text, at the levels of grammar and layout. The interaction, if any, will be superficial.

Most designs of PA, such as those described by Topping, 1998, and Dochy *et al.*, 1999, have students engage in peer feedback of the co-responding mode, presumably as it is relatively easily organized in a classroom setting, with the students reviewing the work of their fellow students.

Flower *et al.* (1986) stress the importance of reviewing in learning to write. Essentially, reviewing aims to evaluate the text thoroughly, in order to adequately define problems and take proper measures. However, with evaluation not properly processed, reviewing will not yield a higher quality of text. Lacking a clear understanding of standards, many novice writers do not succeed in properly reviewing their own work.

Flower *et al.* (1986) argue that particularly novice writers have yet to learn how to systematically perform the four steps of reviewing: to analyze (what does the text say, how is it put together), to evaluate (does the text meet the requirements), to explain (why does it go wrong), and to revise (what changes are necessary to make the text

4

meet the requirements). In our opinion, especially the co-responding type of PA may be helpful in this learning process. It may be expected that students, as a result of applying the assessment criteria to the work of fellow students, also learn to evaluate their own work. As added practical benefits, peer feedback comes in much larger amounts than teachers could ever provide on their own and becomes available much sooner.

Based on a list of organization characteristics from Topping (1998), found in reported systems of PA in higher education, presented in Table 2, we set up PA in several ways, in order to examine design factors geared to support the quality of feedback. The term 'peer feedback' is used for the feedback students give one another when assessing one another's performance.

=Table 2=

Designing peer assessment

PA designs were developed and implemented in the History curriculum of Utrecht University. At the time, in 1999, the curriculum

5

was a four-year programme that may best be compared to a combined bachelor and master programme, then the standard university curriculum in The Netherlands. The designs were implemented in seven courses, distributed over the entire programme, covering different levels of competency and different types of writing product.

At the first meeting of every course, the students were informed about the objectives of PA, and the applicable procedure. As a basic method, we adopted Bean's elaboration of the concept 'advice-centred feedback' (Bean, 1996). Students were asked to exchange their drafts, and assess them according to the same criteria the teacher would use for the final versions. Students were to write their findings on a standardised assessment form, at the end of which they were asked to reflect on their notes and to formulate at least three recommendations to the writer. After receiving PA, one could rewrite one's draft. The teacher monitored the whole process by receiving copies of the draft versions and the written feedback reports, only providing feedback after the students had received peer feedback.

We have developed designs for:

❑ A first-year course in which students take their first steps in learning how to report on historical research (course 1);

6

❑ Two second-year courses. In one of them (course 2), students plan, carry out and report on a limited piece of historical research. The other one (course 3), has students perform a more extensive historical study;

❑ A third-year course, in which the students have to write a biography of an important historian (course 4);

❑ A third/fourth-year specialisation course, in which students learn to write a newspaper article under a strict deadline (course 5), and

❑ Two third-/fourth-year specialization courses, one an introduction to cultural education which has the students write an exhibition review (course 6), the other one requiring them to summarize and discuss literature in the form of an article (course 7). The designs are summarised in Table 3.

== Table 3==

We clustered Topping's variables in four groups and varied ten of them. Seven were not varied, for practical or pedagogical reasons.

Cluster 1 (variables 1-6) relates to PA as an assessment instrument. We varied the required length and completeness of the writing assignments, and the relation between peer and teacher assessment. Courses 1, 2, 3 and 5 featured supplementary PA in the sense that students provided a second source of feedback, next to the teacher's feedback. In courses 1 and 2, the teacher also gave written and/ or oral feedback on the draft version which was to be peer assessed, without grading it. In courses 3 and 5, the teachers graded the draft, stating the arguments for their mark in the assessment form. The teachers were asked always to have the students exchange their comments first and, in courses 3 and 5, to hand out teachers' assessment forms and grades only after peer assessment had taken place.

In courses 4, 6 and 7, PA served as the only formative assessment, coming before the teacher's end of course assessment. In none of our designs PA was meant to be substitutional.

Cluster 2 (variables 7-9) concerns the mode of interaction. First, interaction can be directed one-way or mutual, the latter meaning that assessors and assessees switch roles. One-way assessment means that the assessor is to be assessed by students other than the one(s) he

himself has to assess. In our study, courses 1 and 7 were designed with one-way directionality. Course 7 required written feedback only. Next, the assessment outcomes may be presented in public, or in a smaller feedback group. Course 1 had the students presenting their oral feedback at a plenary session. The assessment can take place partly or entirely outside the classroom, and with or without face-to-face contact. In all courses, students performed the written part of the PA outside the classroom. This includes reading the draft, making notes and completing an assessment form. All oral feedback was provided face-to-face, in the classroom.

Cluster 3 (10-13) relates to the composition of the feedback group. We varied the size of feedback groups from two students (course 3) to three (course 5) or four (courses 2, 4 and 6). In courses 1 and 7, both having one-way assessment, every student assessed the work of two others. Except for courses 2 and 6, the teacher generally grouped students at random. In course 2, the teacher formed feedback groups from students working on related subjects. In course 6, the students had already formed groups of their own, and we saw no reason to change this. In course 4, the assessors had to reach consensus on their feedback, before communicating it to the assessees. In courses 5 and

6, all students studied the same subjects and material, the subjects of the other courses being similar or non-related.

Cluster 4 (14-17) relates to external factors, such as requirements and rewards. In all courses, participation in the PA procedure was mandatory. The quality of the PA was only rewarded in course 5, where students could receive up to 2.5% of their final course grade upon having assessed the written feedback.

Method

Subjects and data collection

Our study involved nine teachers from the History Department of the University of Utrecht and 131 students from the History programme of the Faculty of Arts. Courses 2 and 4 were randomly divided into two PA groups (a and b, both with the same PA design, 2a and 2b having different teachers). Neither teachers nor students had had any previous experience with student PA.

In order to check the participation in PA, we observed classroom activities, gathered all products and peer feedback, and administered

10

two student questionnaires. The first questionnaire was to be completed directly after the students had provided their oral feedback (so before having received the credits for their final version). The items related to the students' time investment in assessing their peers, the practicability of the PA procedures, the usefulness of the received feedback, the workability of the assessment form and the progress of their writing. The second questionnaire was administered at the end of each course, before the students received the credits for the final exam. In terms of reliability and validity, the quality of both questionnaires was satisfactory. Classroom observations covered the entire process, that is the implementation of the PA procedures, their introduction by the teacher and students responses, forming of feedback groups, exchange of written products and written feedback, participation in oral feedback, plenary discussion after oral feedback, and interaction between students and students and teacher.

Data on the written feedback were gathered by means of the completed standardized assessment forms and by taping classroom sessions in which students provided feedback.

Instruments

To examine both the written and oral peer feedback, we analyzed the data from two perspectives: the function of the feedback and the aspect referred to. The reliability of both coding systems was satisfactory in all courses (with Cohen's $\kappa \geq .85$ for function and $\kappa \geq .93$ for aspect).

Feedback functions

Based on Flower *et al*. (1986) and Roossink (1990), we coded the feedback in relation to its several functions, of which we distinguish: Analysis, Evaluation, Explanation, and Revision. These are product-oriented functions (referring directly to the product to be assessed). 'Analysis' was used for comments aimed at understanding the text. 'Evaluation' was used for all quality statements, whether explicit or implicit. Arguments supporting the evaluation were coded 'Explanation'. Suggested measures for improvement were coded 'Revision'. Next, we distinguish Orientation and Method. These are process-oriented, aiming as they do to structure the discussion, or to discuss the writing process. Comments not fitting one of these categories were coded 'Not applicable'.

12

Feedback aspects

By the term 'aspect' we refer to the subject of feedback, distinguishing between content, structure, and style of the writing (see Steehouder *et al*., 1992). 'Content' refers to relevance of information, clarity of the problem, argumentation, and explanation of concepts. 'Structure' refers to the inner consistency of a text, e.g. the relation of the main problem to specified research questions, explanation, and the conclusion. 'Style' refers to the 'outer' form of the text, which includes use of language, grammar, spelling and layout. Feedback simultaneously aimed at more than one aspect was coded 'Ambiguous'. Comments on the PA procedure were coded as 'Procedure', feedback on subjects not from the text 'Non applicable'.

Results

Students' participation in PA activities

The outcome of students' participating in PA is summarized in Table 4. Generally, procedures were followed as scheduled. Around 80% of

the students handed in their drafts on time, received feedback from at least one other student, and assessed the work of at least one peer. Only some two thirds received the number of comments prescribed in the PA design. Some participants of course 2 did not receive any peer feedback at all, despite the teacher's monitoring efforts. Their peers only incidentally attended classes, neither handed in a draft on time, nor complied with any PA activity, for reasons unknown to us. In course 7 none of the draft versions were exchanged on time. As it turned out the assignment to prepare a PowerPoint presentation in the field of art took more of students' time than was estimated in the course design, so it superseded the peer feedback activities.

=== Table 4 ==

Written and oral feedback

Generally, most written feedback was product-oriented, students concentrating on simply evaluating the product, not engaging themselves much in asking questions, explaining their judgments, or proposing revisions. The oral feedback was less product-oriented,

14

about one third of it being aimed at structuring the discourse, or discussing subject matter which was related to but not written in the text. In their product-oriented oral feedback, students were less focused on evaluation and engaged more in all four required activities: analysis, evaluation, explanation and revision. Feedback on the writing process was hardly provided, although more in oral than in written form. On the whole, students' written and oral feedback addressed the aspects of content and style, but not much on structure.

In order to decide which combination of design features yields the best quality of feedback, we combined the designs and their peer feedback in a meta-matrix (cf. Miles and Huberman, 1994), with columns for the seven designs, and rows for the functions and aspects of both written and oral feedback. Table 5 presents the frequencies in which the different functions and aspects figure in the written and oral feedback.

== Table 5==

Written feedback

Cross-course analysis of the written feedback shows significant differences between the seven designs in the frequency of function (χ^2= 143; df =48; p ≤ .001 for all feedback functions, χ^2= 77; df =24; p ≤ .001 for product-oriented feedback only). Written feedback in courses 5 and 6 concentrated on analysis and revision, students asking more detailed questions and providing more suggestions for text improval than in the other courses. This may be explained by both these designs requiring students to write about the same subject while having studied the same materials. As a likely result of this design feature (summarized in Table 3 under 11-Ability), it may be that, in assessing one another's text, students focused on differences with their own writing product, feeling more need to ask the writer why things were written the way they were. Another consequence of this feature may have been that these students felt a strong need for communicating their individual solutions to problems experienced during writing.

As to the aspect of written feedback, we also found significant differences per course (χ^2= 116; df =40; p ≤ .001 for all feedback, χ^2= 86; df =24; p ≤ .001 for product-oriented feedback only). Feedback concentrated more on content in courses 1, 4 and 6 than in the other

16

designs. We have no explanation for this. The fact that there was no feedback on structure in course 6 is striking. Presumably the students thought the list of topics from their teacher to be the prescribed order they had to follow in structuring their product. In courses 2 and 3, feedback was more concentrated on structure.

Regarding the relation between the function of written feedback and the feedback aspect, only three out of seven courses (1, 3 and 6) show significant relations (α = .05). In course 1, the evaluative comments were more directed on style. In courses 3 and 6 both the evaluative comments and the suggested measures for revision were more style-directed. Courses 1 and 6 show weak correlation (Cramérs C= .19, C= .22, respectively), while course 3 shows moderate correlation (.30).

Oral feedback

Like in written feedback, oral feedback shows significant differences in the frequency of function (χ^2= 351; df =42; p $\leq$ .001 for all feedback functions, χ^2= 93; df =21; p $\leq$ .001 for product-oriented feedback only). In course 1, the students concentrated strongly on evaluation and explanation, and hardly on revision. This can be a

result of the design, as the task instruction required the students to present their feedback in a plenary session, one directly after the other (see Table 3 under 8-Privacy). With the number of students involved, the schedule left little time between presentations, from which students presumably concluded that they were only supposed to unilaterally present a list of pros and cons, supported by arguments.

That feedback was also rarely aimed at revision in course 4a must be explained by another phenomenon altogether. The products handed in for PA in this course were frequently so incomplete, that it was difficult to determine what was exactly required to improve the text. The writers often said in advance they did intend to add essential missing parts, such as the conclusion or the introduction, but had not yet found any time to do so. Those products then left so much to be revised that it was difficult to find a starting point for giving feedback.

As to the aspect of oral feedback, we also found differences between courses (χ^2= 330; df =35; p ≤ .01 for all feedback, χ^2= 251; df =21; p ≤ .001 for product-oriented feedback only). More so than in the other designs, feedback was focused on content in courses 1, 2 and 4. Feedback in designs in which all students wrote about the same topic, and studied the same material (courses 5 and 6), was not more

18

directed at content. In course 2 all feedback was relatively more concentrated on structure, which is due to the instruction. Here, students assessed in two steps, the first one had them assessing the paper-outline, particularly evaluating the main question and its relation with sub-questions. In this early phase of writing, students provided more feedback on structure. A second assessment had them investigating a more complete text, namely a chapter (see Table 3 under 4-Product).

Courses 3, 5, 6 produced relatively large amounts of oral feedback on style. In course 5 this may have been due to the specific assignment in writing a newspaper article, which had the students strongly focusing on assessment from the point of view of a newspaper editor. The results of the oral feedback in course 3 must be considered a misrepresentation. As one of the pairs had only one product to discuss, the assessor used all available time with comments on style only.

Relating the oral feedback with their feedback aspects, three of the seven courses (2a, 4 and 6) show significant relations ($\alpha = .05$). In these courses students' analysis concentrated more on content, than on structure or style. In all courses, the correlation is weak (Cramérs C= ≤.19). These findings lead us to believe that there are no considerable

differences between the PA designs concerning the relationship of function and aspect, neither in written, nor in oral feedback.

Relation between written and oral feedback

As to feedback function, written feedback differed significantly from oral feedback in all courses, showing moderate correlation (Cramérs C between .21 and .42).

In their written feedback, students concentrated on evaluating the product, while oral feedback, even with the focus still on evaluation, contained more explanation and revision. Moreover, oral feedback included more non-product-oriented feedback.

In the feedback aspect, written and oral feedback also differed significantly in all courses (Cramérs C between .24 and .42). Students commented on structure more in their written than in their oral feedback. In their oral feedback students commented more on style.

Conclusions and discussion

We have described the written and oral peer feedback in seven PA designs and related our findings to the design features. Adequate
20

feedback addresses the writing product and the writing process as well. Regarding the writing product, feedback should fulfill four main feedback functions: analysis, evaluation, explanation and revision, and cover three main aspects: content, structure and style. Our main conclusions are as follows.

Firstly, in all designs students' feedback focused on the evaluation of the product and neglected the writing process. In the view of Saunders (1989), who reminds us of the relationship between tasks and outcomes, it is hardly surprising to conclude that, generally, students' feedback was mainly evaluative in nature. After all, the task instruction was to assess someone else's writing product. An other explanation for students not addressing the writing process is, to our opinion, the fact that most of them were not used to receive process-oriented feedback from the teacher, so they had no model. If we had wanted students to comment not only on products, but also on the working process, they had to be trained to do so. However, they were only informed about the PA procedure and did not receive any training. An interesting issue for further study would be to develop tasks of PA designed to induce more process-oriented and more constructive feedback.

Secondly, there is a significant relationship between whether feedback is provided written or orally and the feedback functions it fulfills. As the written feedback was concentrated mainly on evaluative comments, students provided arguments and suggestions for text revision in their oral feedback. Similarly, written feedback concentrated more on content, oral feedback more on style. From this, we can conclude that, for the process of feedback in PA to yield adequate results, oral explanation of the written assessment is essential.

Thirdly, most feedback, written and oral, was directed at content and style, not at structure. When students were asked to assess at the end as well as at the start of the writing process, by assessing the rough outline of a paper or essay plan, more feedback was yielded on method and structure. To that end, we would redesign the task, so that the PA applies to 'earlier' draft versions which are still capable of further revision, or to products at various stages of the writing process. We summarize the design features, in terms of Topping's typology (see Table 2) supporting the quality of peer feedback:

1- PA at the start (paper-outline) as well as at the end (draft version) generates more feedback on the text's structure and writing process (Table 2, product).

2- When students present the oral part of PA in small feedback groups they engage more in analyzing and revising, than when they have to present the outcome of their assessment publicly (Table 2, privacy).

3- Combination of written and oral feedback is more profitable than written or oral feedback only. In their oral feedback, students interact to clarify the text and develop measures for revison. In their written feedback, students attend more on structure, while in oral feedback they focus more on style (Table 2, contact).

The results of our study lend support for our view that PA is not aimed at erasing the teacher from the learning process. On the contrary, as peer feedback tends to neglect the writing process and the structure of the text, the teacher has to fill up the gap, while designing tasks in order to evoke peer feedback on a deeper level. As an expert in the subject area, the teacher guides students not only in the subject matter,

but also in the reflection and adjustment of their own working process. Thus, the teacher helps students in learning from providing as well as receiving feedback, in order to apply their newly found knowledge to future writing. Next, the teacher faces new tasks, which include supervising the PA process and creating a class climate in which students feel safe in commenting on the performances of their fellow students.

References

Bean, J. C. (1996). *Engaging ideas: the professors' guide to integrating writing, critical thinking and active learning in the classroom.* San Francisco: Jossey-Bass Publishers.

Berg, B.A.M. van den, Admiraal, W.F., & Pilot, A. (in press). PA in university teaching: evaluating seven course designs. *Assessment and Evaluation in Higher Education.*

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*, 331-350.

Falchikov, N. (2001). *Learning together; peer tutoring in higher education.* London: RoutledgeFalmer.

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. *College Composition and Communication, 37,* 16-55.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook (2nd ed.).* Thousand Oaks, California: SAGE.

Saunders, W. M. (1989). Collaborative writing tasks and peer interaction. *International Journal of Educational Research, 13,* 101-112.

Roossink, H.J.(1990). Terugkoppelen in het natuurwetenschappelijk onderwijs, een model voor de docent.[*Feeding back in science education, a feedback model for the teacher]*. Doctoral dissertation, University of Twente, The Netherlands.

Sluijsmans, D.M.A. (2002). *Student involvement in assessment; the training of PA skills.* Doctoral dissertation Heerlen: Open Universiteit Nederland, The Nederlands.

Steehouder, M., Jansen, C., Maat, K., Staak, J. van de, & Woudstra, E. (1992). Leren communiceren [*Learning to communicate]*. Groningen, The Netherlands, Wolters-Noordhoff.

Stodolsky, S.S. (1984). Frameworks for studying instructional processes in peer work-groups. In P.L. Peterson, L.C.Wilkinson, & M.Hallinan (Eds.), *The social context of instruction: group organization and group processes.* Orlando: Academic Press Inc.

Topping, K. (1998). PA Between Students in Colleges and Universities. *Review of Educational Research, 68*, 249-276.

Weert, T.J.van, Pilot, A. (2003). Task-Based Team Learning with ICT, Design and Development of New Learning. *Education and Information Technologies, 8,* 195-214.