The nature, reception, and use of online peer feedback in higher education

J. van der Pol [a,*], B.A.M. van den Berg [a], W.F. Admiraal [b], P.R.J. Simons [a]

[a] IVLOS Institute of Education, Utrecht University, P.O. Box 80127, 3508 TC Utrecht, The Netherlands
[b] ILO Graduate School of Teaching and Learning, University of Amsterdam, P.O. Box 19268, 1000 GG Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

This article focuses on the use of online interactive peer feedback in higher education and identifies the successful uptake of feedback as an important aspect. We investigate the link between the nature of students' feedback, the way it is evaluated by the receiver, and its consecutive use for the revision of students' products. Two separate studies were conducted to investigate the link between these three variables across different educational contexts and tools. Both studies showed a significant relationship between feedback containing concrete suggestions and a successful uptake of the feedback. Regarding the different tools that were used, these concrete suggestions were more often produced in the Annotation system than in the Blackboard discussion forum, the latter showing more evaluative forms of feedback. We also found significant relationships between elements of both the nature and the reception of feedback on the one hand, and the use of this feedback by the receiver on the other hand.

1. Online formative peer assessment in higher education: relating the nature, reception, and use of feedback

In higher education, the concept of *peer assessment* is increasingly common (Falchikov & Boud, 1989; Topping, 1998; Falchikov & Goldfinch, 2000; Topping, 2005). Peer assessment is to be understood as an educational arrangement in which students comment on the quality of their fellow students' work, for formative or summative purposes (Topping, 1998; Dochy, Segers, & Sluijsmans, 1999). There are several reasons for this growing popularity, one of which is rather pragmatic. University curricula today show an increased emphasis on the learning of complex skills such as writing, inquiring and problem-solving. This means that students produce more complex work, such as reports, articles and project presentations. As the available teacher time in most institutions does not increase, the teacher is urged to find new ways to provide feedback. A practical benefit of implementing peer assessment is that the feedback comes in much larger quantities than the teacher could ever provide alone, and becomes available much sooner. A more pedagogical reason for implementing peer assessment is that it resembles professional practice. Providing and receiving feedback from work colleagues is a common learning activity in many professional practices (Billet, 2002; Eraut, 2004). As such, the concept peer assessment fits in with recent developments in university teaching, such as collaborative learning and writing, and real-life task performance (see, for example, Van Weert & Pilot, 2003).

In recent years, the process of peer assessment is increasingly being facilitated by the use of electronic learning environments, such as Blackboard or WebCT.[1] Online peer assessment can simplify the logistics of peer assessment considerably, by reducing the complications of copying and distributing papers, turnaround time and keeping records (McCormack & Jones,

---

* Corresponding author. Tel.: +31 6 18135660; fax: +31 30 2532741.
 E-mail address: j.vanderpol@uu.nl (J. van der Pol).
[1] See www.blackboard.com.

1998; Bhalero & Ward, 2001; Tannacito & Tuzi, 2002). Compared to face to face peer assessment, implementing peer assessment online can also support its pedagogical aspects. As mentioned by Gehringer (2001) and Trahasch (2004), it allows higher degrees of interactivity between students and offers teachers better possibilities to monitor and guide this interactive process. In addition, peer feedback that is exchanged online may result more often in the revision of students' products than face to face feedback, as Hewitt (2000) and Tuzi (2004) demonstrated.

Despite these potential advantages of online peer assessment, its success is not unconditional. As shown by Van den Berg, Admiraal, and Pilot (2006a), of peer assessment must be adequately organized in order to produce feedback of sufficient quality. Regarding the quality of feedback, especially the formative function of assessment may be endangered when students (partially) take over the teacher's role of providing feedback. Because of the potential as well as the challenge for deepening students' learning, this study examines the formative use of online peer feedback in higher education. In order to motivate our specific research questions within this focus we will first attempt to provide some additional insight into the general process of peer assessment, both face to face and online.

1.1. Online peer assessment in higher education

Falchikov (1986) describes peer assessment as a method in which students engage in reflective criticism of the products of other students and provide them with feedback, using previously defined criteria. This process may consist of a single activity or involve a series of instances in which students supply feedback on increasingly polished versions of each other's work. The learning effect of this process is twofold and related to both the providing and the receiving of feedback. Regarding the providing of feedback, Dochy et al. (1999) and Topping (2003) emphasize that by assessing the work of fellow students, students also learn to evaluate their own work. Although *producing* peer feedback may indeed result in important learning outcomes, we think that *receiving* peer feedback should have a considerable profit as well in order to account for the time and effort that is required to engage in the process of exchanging peer feedback. Regarding the reception of feedback, Flower, Hayes, Carey, Schriver, and Stratman (1986) stress that novice writers often do not succeed in reviewing their own work well and can benefit from the reviewing of their products by peers when learning to write. Focusing on the learning effects of providing and receiving feedback places peer assessment in the realm of collaborative learning and can be viewed as such (Falchikov, 2001). This view is supported by Shekary and Tahririan (2006), who state that peer assessment resembles any other form of collaborative learning in the sense that it offers students the potential to develop new knowledge and understanding. Taking this perspective allows us to develop a better understanding of the formative effects of peer assessment by using the existing insights into the nature of collaborative learning. For instance, one can identify that peer assessment is often more limited than other forms of collaborative learning (or as Saunders calls it 'co-responding') in the sense that it generally offers a lower degree of interactivity. This affects students' possibilities for interactive meaning making and collaborative knowledge construction (Saunders, 1989).

When looking for studies on peer assessment from a collaborative learning perspective, however, not many studies can be found. Apart from a study by Orsmond, Merry, and Reiling (1996), none of the studies mentioned in the review of Dochy et al. (1999) report learning effects of receiving peer feedback. Instead, research on peer assessment often addresses the reliability and validity of students' markings and their appraisal (e.g. Bhalero & Ward, 2001; Sitthiworachart & Joy, 2004). In an elaborative review study on the use of peer assessment in higher education, Topping, Smith, Swanson, and Elliot (2000) underline the fact that many of the studies on peer assessment fail to investigate the relationship between peer feedback and its learning effects for the receiver as expressed in the revision of his or her product. Only two studies in Topping et al. examine the effect of peer feedback on the improvement of students' writing products (Chaudron, 1983; Jacobs & Zhang, 1989). Although focused on the formative effects of peer assessment, both studies relate these effects to the origin of the feedback (peer or teacher generated) but not to its characteristics or to the interactive process through which the exchange of formative feedback leads to these effects. The same goes for the studies of Hewitt (2000) and Tuzi (2004), which do investigate learning effects, but focus on its relationship with the form of peer feedback (being face to face or online). Similarly, a recent review study by Hattie and Timperley (2007) concludes that still surprisingly few studies have systematically investigated the impact of peer feedback on learning and achievement.

The studies that do investigate the interactive process of formative peer assessment in more depth (see, for example, Nicol & Macfarlane-Dick, 2006; Shekary & Tahririan, 2006) confront us with a rather straightforward but essential condition: in order to achieve the potential of online peer assessment, a successful *uptake* of each other's comments seems to be essential. Uptake not only refers to an accurate understanding of the provided feedback, but also to its subsequent use in changing students' texts. What makes students process the peer feedback they received? Although there are studies that explore the role of the receiver during the feedback process (e.g. Prins, Sluijsmans, & Kirschner, 2006, who draw attention to the responsibility of the receiver to shape the feedback dialogue, among other things by asking for clarifications), the connection between the nature of peer feedback and its uptake by the receiver is still unclear. The effectiveness of peer feedback in terms of a successful uptake hinges at least partly on the quality of the feedback that students provide. As providing feedback to one's peers (without possessing a high level of expertise) is known to be difficult for students (Dochy et al., 1999; Topping et al., 2000), a sufficient quality of feedback to ensure its successful uptake is by no means guaranteed.

In order to shed more light on the factors that influence a successful uptake of online peer feedback, this study aims to examine the relationships between the nature of feedback, its reception by the receiver, and its consecutive use in the revision of students' texts. More specifically, our research questions are:

1. In what way is the nature of peer feedback related to its use for the revision of texts?
2. In what way is the nature of peer feedback related to its reception by the receiver?
3. In what way is the reception of peer feedback related to its use for the revision of texts?

   In order to investigate these relationships in different educational contexts in higher education, two separate studies were conducted: one in a professional Health Care course and one in academic Educational Science course. In addition, we implemented two different tools to facilitate the process of providing peer feedback in the Educational Science course, leading to a fourth research question:

4. In what ways do different tools for peer feedback evoke differences in the nature, reception and use of feedback?

2. Study 1

2.1. Method

2.1.1. Data collection

Data for the first study were collected over a period of 6 months at the Health Care Education study at the College of Arnhem Nijmegen in the Netherlands. The 27 participating students worked individually on several assignments including internship reports, essays and reflection reports. These documents were part of students' portfolios demonstrating their mastery of a range of competencies. The peer feedback process was organized in groups of four to ten students, and the feedback was aimed at each other's portfolio products. There were no structured procedures for how, how much, when, or where to provide peer feedback. A consequence was that students did not receive feedback on all their products and did not produce a revised version of all products. In this study, we limit our analysis to the products on which students received feedback and of which a revised version could be found, as only these products allow us to investigate the relationship between feedback and revision. This analysis of the relationship between feedback and revision will be studied at the level of individual feedback comments on a specific product. Thus, the revised products we have analyzed contain multiple feedback comments, some of which will have been used to revise the text and some of which have not. This selection resulted in a data set of 392 feedback comments on 78 products, created within four groups of students.

An electronic learning environment named the Virtual Learning Community (VLC)[2] was used to facilitate the process of providing peer feedback (see Fig. 1). This VLC showed students' documents on screen along with the provided feedback (displayed in the lower part of the screen), but lacked the possibility to link feedback to a specific section in the document. Neither did this VLC provide the opportunity to respond to each other's feedback. We used this system as it was already being implemented widely at the College of Arnhem Nijmegen, which chose the system for its compatibility with their principles of competency-based education.

2.1.2. Measures

Event sampling (see Shiffman & Stone, 1998) was used to determine coding units, meaning that each time the topic of communication changed, a new coding unit started. Two researchers were trained for approximately two hours, after which independent segmentation of 21 messages (12% of the data) resulted in an agreement of 98.3% from the perspective of both coders, which is well above the threshold of 80% (Riffe, Lacy, & Fico, 1998). Distinguishing topics is something that was already done very much by students themselves by using keywords or lay-out features to signal a division of separate points. Both coders divided the 21 messages into 58 feedback units (topics). Next, all feedback units were coded according to the three variables *nature of feedback*, *reception of feedback* and *use for revision*.

The *nature of feedback* was measured by means of a coding scheme of Van den Berg, Admiraal, and Pilot (2006b), developed for qualitative analysis of peer feedback on texts of university students. In this scheme, peer feedback was coded from two major perspectives: feedback function and feedback aspect. It is based on Flower et al. (1986) for the feedback functions, and on Steehouder, Jansen, Maat, Staak, and van de Woudstra (1992) for the feedback aspects. According to Flower et al., feedback on writing fulfils four functions, namely analysis (what does the text say, how is it put together), evaluation (does the text meet the requirements), explanation (why does the text go wrong), and revision (suggesting concrete changes to improve the text). The term 'feedback aspect' refers to the subject of feedback: content, structure or style (Steehouder et al., 1992). 'Content' refers to the relevance of the information, the clarity of the problem, argumentation, and explanation of concepts. 'Structure' refers to the inner consistency, for example, the relationship between the main problem and the subdivided research questions, the explanation and the conclusion. 'Style' refers to the outer form of the text, for example, the language, grammar, spelling and lay-out. Altogether, there are four feedback functions (analysis, evaluation, explanation and revision, see Table 1) and three feedback aspects (content, structure and style, see Table 2). After segmenting the data into individual comments with a single feedback function, Van den Berg et al. (2006b) report an inter-rater reliability of the coding instrument of .85 (Cohen's κ) for feedback function, $n = 88$ comments, and .93 (Cohen's κ) for feedback aspect, $n = 79$ comments. Following the coding scheme of Van den Berg et al., the data on feedback were coded in terms of functions and aspects.

---

[2] More information on www.citowoz.nl.

Fig. 1. Print screen of a student's product along with the provided feedback in the VLC, with the following elements: 1 – indicator that this is the portfolio part of the learning environment, 2 – display of the original product, and 3 – the received feedback comments on the product, with date and author.

Table 1
Feedback functions with examples

Feedback function	Example
Analysis	*Is this not already a very constructivist viewpoint, advocating situated learning and an authentic context?*
Evaluation	*This is a very restricted way of testing, for one, and I would like to see the constructivist part incorporated - I don't see it*
Explanation	*This is good because in cognitivism you should inform the learner of the learning objectives.*
Revision	*I think that according to the theory of sociocultural learning, you should consider not only the teacher, but also other students.*

Table 2
Feedback aspects with examples

Feedback aspect	Example
Content	*This is also an important element of the sociocultural perspective*
Structure	*Nice overview of the topics you described. Now everything is connected to each other. Well done*
Style	*Using 'we' is not very APA-style*

The *reception of feedback* was measured in two different ways: in terms of the 'importance' of a feedback comment as indicated by the receiver (going from 1 = not important to 4 = very important) and, in cases where the receiver – in spite of the lack of functionality for this in the VLC[3] – also managed to respond to the feedback, in terms of his or her 'agreement' with it (these reactions were coded by the researchers as 0 = do not agree, 1 = partly agree, and 2 = completely agree). For the latter, an inter-rater reliability of .74 (Cohen's κ) was established by comparing the independent scores of two researchers on 35 feedback comments on 8 products.

In order to measure the *use of feedback* for revision of the text, all feedback comments were first sorted to the document concerned. Second, anti-plagiarism software was used to systematically identify all changes in the original and revised ver-

---

[3] Some of the students inserted personal reactions to the received feedback in their revised documents (indicated with a different font colour). We will elaborate on this phenomenon in 'Results' section.

sions of the documents. Generally, this software compares versions of the same document and identifies regions where differences can be found. These flagged regions then supported our manual process of looking for revisions that corresponded to a certain feedback comment. Thus, each separate feedback comment was scored as being processed and incorporated in the revised version of the text in some way (score 1), or as not leading to a change in the document at all (score 0). The inter-rater reliability is .77 (Cohen's κ), based on 35 scores of two researchers referring to 8 student products.

2.1.3. Analysis

2

Data on the three variables are available at the level of topic within a feedback message. We call this the feedback unit. Analyses of the relationship between the three variables *nature of feedback*, *reception of feedback* and *use of feedback* were performed at the level of this feedback unit. The scores in the feedback units were analyzed using binary logistic regression tests (stepwise), with use of feedback as the dependent variable, and with the feedback functions and feedback aspects, respectively, as independent variables. Pearson's correlations were used to examine the relationship between the reception of feedback and the revision of texts, and between the nature of feedback and the reception of feedback.

2.2. Results

In the VLC, 392 feedback units were produced by the four groups of students. Table 3 shows the percentage of feedback units with the particular feedback function and feedback aspect, the percentage of feedback units that was evaluated by the receiver on importance and agreement, as well as the mean scores for both aspects, and the percentage of feedback units that resulted in revision of the texts.

A remarkable finding is that students responded to 4% of the feedback comments, although the system did not provide the opportunity to respond to the received feedback. These 14 responses were created by students who found an alternative way to respond to the provided feedback, by including their reactions in the revised versions of their products. Although these reactions were delivered in an unconventional way, their content was very similar to the reactions found in the second study.

2.2.1. Relationship between the nature of feedback and revision of products

The results with respect to the nature of feedback (function and aspect) are summarized in Table 4. With respect to the feedback function, we found only one significant relationship with the revisions. The more the feedback included recommendations for revision, the more it resulted in the revision of texts. The *B* coefficients can be interpreted as the chance that a certain feedback comment can be tied to a corresponding revision in the text, given the presence of a particular feedback aspect or function. With respect to the feedback aspect, we found both content and style to be significantly related to revision. The more students focused in their feedback on the content and style of the written work of their peers, the more their peers revised it.

2.2.2. Relationship between the reception of feedback and the use of feedback

The results with respect to the relationship between the reception of feedback and the revision of the texts are summarized in Table 5. We only found one significant (positive) correlation between the appreciation of importance and the revision of texts: the more feedback was appreciated as important, the more feedback was used to change the written work. It must be noted that the absence of a significant relationship between agreement and both appreciation and use for revision

Table 3
Descriptive statistics for nature, reception, and use of feedback ($N = 392$)

Variable	%				Total
Nature of feedback					
Feedback functions					
Analysis	40				157
Explanation	30				118
Evaluation	61				237
Revision	52				202
Feedback aspects					
Content	71				279
Structure	3				12
Style	25				96
		Mean	SD	%	Total
Reception of feedback					
Importance of feedback as appreciated by the receiver (1 = low, 4 = high)		2.50	.88	33	131
Agreement of the receiver with the received feedback as coded by the researcher (0 = low, 2 = high)		.64	.63	4	14
		%		Total	
Use of feedback					
Feedback comments that led to a revision in the text		49		192	

Note. The sum of the percentage per variable can be more than 100% as more than one feedback function or aspect can be assigned to one feedback unit.

Table 4
Binary logistic regressions with feedback functions and aspects as independent variables and the use of feedback as dependent variable

Variable	N	B	SE	df	p
Feedback functions					
Analysis	157	.38	.25	1	ns
Explanation	118	.44	.27	1	ns
Evaluation	237	.06	.27	1	ns
Revision	202	.72	.23	1	.001
Feedback aspects					
Content	279	.95	.36	1	.009
Structure	12	−.82	.66	1	ns
Style	96	1.12	.38	1	.003

Note. ns means not significant.

Table 5
Pearson's correlations between evaluation, agreement, and use of feedback

	Appreciation of importance	Agreement receiver	Used for revision of text
Appreciation of importance		ns $n = 10$	.50[*] $n = 126$
Agreement receiver			ns $n = 13$
Used for revision			

Note. ns means not significant and [*]$p \leqslant .05$.

may have resulted from the small number of feedback comments for which the agreement of the receiver could be coded (due to constraints of the tool).

2.2.3. Relationship between the nature of feedback and the reception of feedback

In the correlations between the nature of feedback (its functions and aspects) and the reception of feedback (evaluation of importance and agreement of the receiver) we did not find any significant results.

2.2.4. Summary of the main findings

In sum, we found significant positive relations between concrete suggestions for revisions and aspects of content and style. Moreover, the more feedback was appreciated as being important, the more students used it to revise their texts.

3. Study 2

3.1. Method

3.1.1. Data collection

The data for the second study were gathered during a 3-month Educational Science course at Utrecht University. In this course, 38 students in groups of three or four had to collaboratively create a set of course materials for high-school students. These course materials encompassed a theoretical chapter, a chapter with assignments, and a report accounting for the educational choices that were made. Students chose their own topics, as well as their target groups. The teacher provided some guidelines on how to design educational materials and a set of criteria according to which they would be evaluated (both in the peer feedback round as in the final teacher evaluation). Students of each group provided feedback on the product of one other group. The teacher assigned different roles to the students in commenting on each other's product, along with different criteria for the feedback. The feedback was provided on a concept version of the materials over a fixed one-week period. After this, the students had 1 week to revise their chapters and hand them in to receive their final grades.

In this second study we introduced a new tool. Besides students' regular online learning environment (Blackboard), we used a system for anchored discussion called the Annotation system[4]. When used for the collaborative processing of texts, anchored discussion has been found to produce to-the-point feedback in a highly task-oriented learning environment (Van der Pol, Admiraal, & Simons, 2006). Moreover, Trahasch (2004) has earlier suggested incorporating the possibility of annotation in tools for online peer feedback in order to afford more specific feedback. The importance of detailed feedback for learning has been underlined by Gibbs and Simpson (2004) and Webb and Mastergeorge (2003). In addition, Fiehn et al. (2003) found shared annotations, with students anchoring their evaluations in the artefacts created by their peers, to be particularly suitable for peer

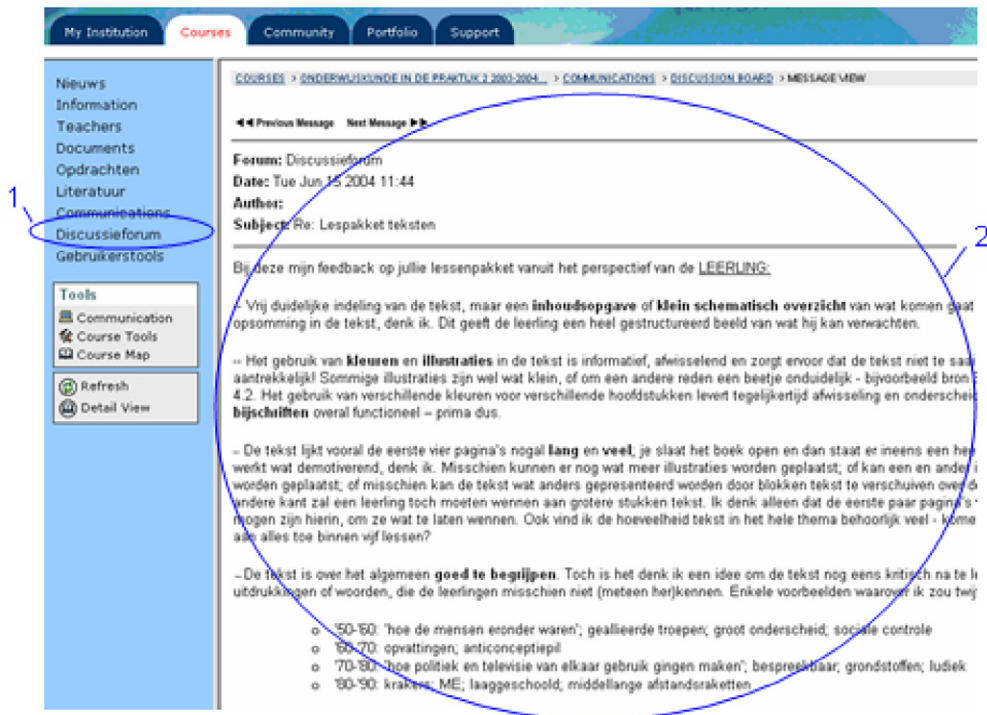---

[4] See www.annotatiesysteem.nl.

Fig. 2. Print screen of students' feedback in Blackboard, with the following elements: 1 – indicator that this is the threaded discussion part of the learning environment and 2 – display of a message with the different feedback comments of a particular reviewer on a single product, with date and author.

assessment scenarios. In order to investigate these claims and check the affordances of normal threaded discussion versus anchored discussion for the exchange of peer feedback, students were randomly placed in either the discussion board of Blackboard to write their feedback and upload their documents (see Fig. 2), or in the Annotation system developed by Van der Pol et al. (2006), see Fig. 3. The main functional differences between these tools are that the Annotation system displays both the document and discussion on screen and allows students to anchor their comments to specific sections of the document, whereas Blackboard does not, presenting a regular forum discussion with the possibility to attach products to individual reactions.

3.2. Measures

In order to code all data according to the three variables *nature of feedback*, *reception of feedback* and *revision of texts*, we used the same procedures as described in study 1. Event sampling was used in order to determine the coding unit. This means that if the topic of the communication changed, a new coding unit started. The *nature of feedback* was measured in terms of the four feedback functions (analysis, evaluation, explanation and revision) and three feedback aspects (content, structure and style of the students' writing). The *reception of feedback* was measured in terms of the 'usefulness'[5] of a comment as indicated by the receiver (on a 5-point scale, ranging from 1 = low to 5 = high) and the receivers' agreement with the provided feedback expressed in a reaction (coded by the researcher as 0 = do not agree, 1 = partly agree and 2 = completely agree). Appreciating the usefulness of the received feedback was done both online and offline for the groups in the Annotation system, whereas for the groups using Blackboard these evaluations were only collected on paper. In both cases the evaluations were collected after the exchange of peer feedback and the revision of texts. Students' appreciation of the received feedback in the Annotation system was directly visible online for all students, whereas the results of the evaluations in Blackboard collected offline were not available for other students. This difference could not be prevented and results from the different functionalities of the two systems. As we expect that the visibility of the evaluations may influence their contents (related social factors), we will not directly compare the evaluations found in Blackboard with those found in the Annotation system. In order to measure the use of the feedback in the revision of text, again the same procedure was followed as described in study 1, with a score of 1 for a feedback unit leading to a revision in the text in some way and '0' for feedback not leading to a change in the document at all.

---

[5] Please note that this is a different aspect of the reception of feedback than the 'importance', which was measured in the first study (using a 4-point scale on that occasion), caused by the different functionalities of the tools.
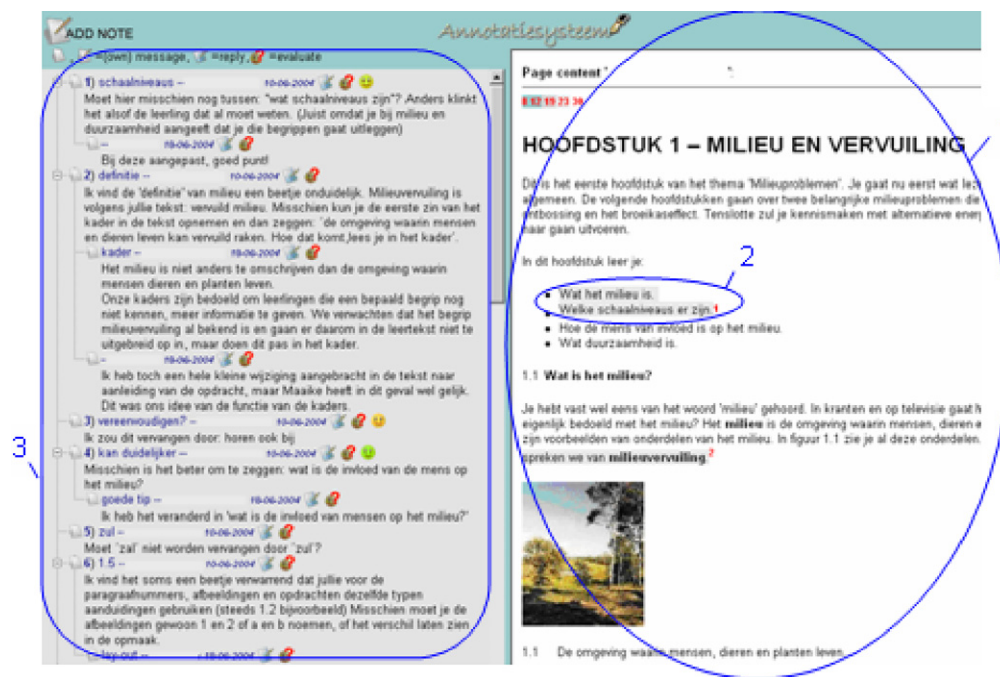
Fig. 3. Print screen of students' feedback in the Annotation system with the following elements: 1 – the original product, with two selected passages ('anchors') in the text to indicate the object of different feedback comments (3).

3.2.1. Analysis

As in study 1, data on the three variables are available at the level of topic within a feedback message. We call this the feedback unit. Analyses of the relationship between the three variables *nature of feedback*, *reception of feedback* and *use of feedback* were performed at the level of feedback unit. Again, the scores were analyzed using binary logistic regression tests (stepwise), with use of feedback as the dependent variable and with the feedback functions and feedback aspects, respectively, as independent variables. Because we will see an influence of 'tool' on the presence of several feedback functions, we included it as a covariate in this regression analysis. Pearson's correlations were used to examine the relationship between the reception of feedback and the revision of texts, and between the nature of feedback and the reception of feedback. Because in the Annotation system the evaluation of usefulness is linked to entire messages, we disaggregated this evaluation score in cases where messages contained multiple topics, attributing the same score to the several feedback units within the message. In this way, a total of 160 original evaluation scores resulted in 192 disaggregated scores. This means that the statistical analyses using the evaluation scores from the Annotation system may overestimate the results. χ^2-tests and Mann–Whitney U-tests were performed in order to examine the relationship between tool (Annotation system or Blackboard) and nature of feedback, reception of feedback and revisions of texts.

3.3. Results

In both tools (Annotation system and Blackboard), 335 feedback units were produced by the six groups of students (four groups in the Annotation system and two in Blackboard). Table 6 shows the proportion of feedback units with the particular feedback function and feedback aspect, the proportion of feedback units that were appreciated on usefulness and on agreement, and the proportion of feedback units that resulted in revision of the texts.

Table 6 shows an obvious difference with the results found in study 1. Now, students provided an appreciation of about 70% for the received feedback comments and a reaction to about 34% (for a more detailed description of the differences between the two systems, see Table 7). The latter is a direct result of the more interactive nature of the tools used in this study that, in contrast to the VLC in study 1, did allow students to react to the feedback they received.

3.3.1. Relationship between the nature of feedback and use of feedback

The results with respect to the nature of feedback (function and aspect) are summarized in Table 7. Again, the *B* coefficients can be interpreted as the chance that a certain feedback comment can be tied to a corresponding revision in the text, given the presence of a particular feedback aspect or function. Regarding the feedback function, we found three significant relationships with the revision of the texts: the more students included an analysis of an issue, an evaluation, or a recom-

Table 6
Descriptive statistics for nature, reception, and use of feedback ($N = 335$)

Variable	%			Total
Nature of feedback				
Feedback functions				
Analysis				87
	26			
Explanation				140
	42			
Evaluation				235
	70			
Revision				167
	50			
Feedback aspects				
Content				168
	50			
Structure				19
	6			
Style				147
	44			
	Mean	SD	%	Total
Reception of feedback				
Importance of feedback as appreciated by the receiver (1 = low, 5 = high)	3.56	1.30	70	234
Agreement of the receiver with the received feedback as coded by the researcher (0 = low, 2 = high)	1.39	.85	34	115
	%			Total
Use of feedback				
Feedback comments that led to a revision	55			184

Note. The sum of proportion per variable can be more than 100% as more than one feedback function or aspect can be assigned to one feedback unit.

Table 7
Binary logistic regressions with feedback functions and aspects together with tool as independent variables, and the use of feedback as dependent variable

Variable	N	B	SE	df	p
Feedback functions					
Analysis	87	1.18	.33	1	$\leqslant$.001
Explanation	140	.02	.28	1	ns
Evaluation	235	.75	.35	1	.032
Revision	167	1.20	.25	1	$\leqslant$.001
Feedback aspects					
Content	168	.96	.34	1	.005
Structure	19	−.19	.51	1	ns
Style	147	.70	.34	1	.042

Note. ns means not significant.

mendation for revision in their feedback, the more students changed their written work. With respect to the feedback aspect, we found two significant relationships: the more students focused in their feedback on the content and style of the written work of their peers, the more their peers revised it.

3.3.2. Relationship between the reception of feedback and the use of feedback

The results with respect to the relationship between the reception of feedback and the use of feedback are summarized in Table 8. We found two significant (positive) correlations. The first correlation refers to the relationship between the two variables that measure the reception of feedback: the appreciation of usefulness and the agreement with feedback. The more feedback was perceived as useful, the more the receiver agreed with the provider of feedback. This correlation may point to a partial overlap in both concepts. The second correlation refers to the relationship between the level of agreement and the revision of texts: the more a student agreed with the provider of feedback, the more feedback was used to change the written work.

3.3.3. Relationship between the nature of feedback and the reception of feedback

In the correlations between the nature of feedback (its functions and aspects) and the reception of feedback (appreciation of usefulness and agreement by the receiver), we did not find any significant results.

Table 8
Pearson's correlations between evaluation, agreement, and use of feedback

	Appreciation of usefulness	Agreement receiver	Used for revision of text
Appreciation of usefulness		.32[*] $n = 77$	ns $n = 234$
Agreement receiver			.33[*] $n = 115$
Used for revision			

Note. ns means not significant. [*]$p \leqslant 0.05$.

Table 9
Differences between the annotation system and Blackboard

Variable	Annotation system ($N = 200$), %	Blackboard ($N = 135$), %	χ^2	df	p
Nature of feedback					
Feedback function					
Analysis	27	24	ns		
Explanation	41	44	ns		
Evaluation	64	80	10.5	1	.001
Revision	55	42	5.3	1	.014
Feedback aspects					
Content	51	49	ns		
Structure	5	7	ns		
Style	43	46	ns		
	Mean	Mean	Z		
Reception of feedback					
Usefulness of feedback as appreciated by the receiver (1 = low, 5 = high)	3.7	2.9	–3.36		<.001
	$n = 192$	$n = 42$			
Agreement with feedback as observed by the researcher (0 = low, 2 = high)	1.43	1.36	ns		
	$n = 51$	$n = 64$			
Use of feedback					
Feedback comments that led to a revision (%)	56	54	ns		

Note. The sum of proportion per variable can be more than 100% as more than one feedback function or aspect can be assigned to one feedback unit. ns means not significant.

3.3.4. Differences between tools

A total of 335 feedback units were produced (200 by the four groups of students in the Annotation system and 135 by the two groups of students in Blackboard). An interesting detail when comparing the use of both tools, is that one student in Blackboard (with the regular, non-anchored discussion) spontaneously anchored his feedback (consisting of 8 feedback units) in the corresponding text by using the review function of Word and attaching this reviewed document to the discussion forum. This may tell us something about students' need for annotation possibilities and the natural suitability of this functionality for the task of exchanging peer feedback. The results of χ^2-tests and nonparametric *t*-tests (Mann–Whitney *U*) that check for differences across tools are summarized in Table 9.

In Table 9, we see four significant differences between the Annotation system and Blackboard. First, the percentage of feedback units with an evaluative feedback function was lower in the Annotation system than in Blackboard ($\chi^2 = 10.5$; df = 1; $p = .001$). Second, the proportion of feedback units with suggestions for revisions was higher in the Annotation system than in Blackboard ($\chi^2 = 5.3$; df = 1; $p = .014$). Third, we see a significant difference in the average scores on the scale of usefulness. In the Annotation system, students evaluated the received feedback as more useful than in Blackboard ($Z = -3.36$; $p \leqslant .001$). This difference, however, is likely to be caused by the fact that the online evaluations in the Annotation system were visible to the others students (making them part of a social process), which was not the case in Blackboard. Fourth, we see that in the Annotation system a larger amount of feedback units was evaluated, which can also be traced back to the different ways of collecting students' evaluations (online vs. on paper).

3.3.5. Summary of the main findings

In sum, we found significant relationships between feedback with the function of analysis, evaluation, and proposing revisions and feedback on both content and style of a text on the one hand, and whether or not a text had been revised on the other hand. Moreover, the more students agreed with the received feedback, the more the feedback was evaluated as useful; and the more students agreed with the provided feedback, the more they revised the corresponding parts of their texts. Finally, in the Annotation system students showed less evaluative feedback and more feedback with suggestions for revisions than in Blackboard.

4. General discussion

4.1. Comparison of results of the two studies

4.1.1. The nature of feedback

Our first research question concerned the way the nature of peer feedback is related to its use for the revision of texts. With regard to the *functions* of feedback, the main result that was found with both studies is that students' feedback in which they suggest concrete revisions (feedback function 'revision') is positively related to revision of the corresponding parts of students' texts. This is understandable, because these concrete suggestions for revisions give the receivers the most direct lead for a potential change in their text. These results are in line with those of a study by Tuzi (2004), who found that concrete and specific feedback resulted in a higher degree of revision of students' products. Furthermore, the second study (the Educational Science course) also showed a positive relationship between the feedback functions 'analysis' and 'evaluation' and the revision of students' written texts. This difference may have resulted from the different contexts (domains, tasks and participants) of the two studies.

In order to investigate these differences, we conducted a post-hoc analysis on the nature of the analysis and evaluation comments in both studies. From this, it appeared that the difference in domain and the *type* of texts that were peer reviewed could have been responsible for producing slightly different kinds of analytical and evaluative feedback comments, leading to different results. The analytical feedback comments seemed to come in two different forms. While analytical comments normally serve to deepen a subject by asking for elaboration, they can also function as an indirect suggestion for revision. In the first study (the Health Care course), students' products that were part of their personal portfolio were very diverse, and consisted mainly of personal reflection (e.g. on their internships). Here, peers served mainly as an 'interested fellow-student" in their feedback, typically producing more subjective comments like: *Why do you think that you can work more on quality of care in the DDZ than in other organisations*? In the second study (the Educational Science course), however, students not only commented on products that were far less personal (proposed book chapters for secondary education), but they were also instructed to review from a professional point of view (e.g. taking the role of a teacher or a student). Here, analytical comments were often more task-directed, being objective questions for elaboration such as identifying missing or ambiguous information. Thus in these comments, the implicit suggestions for revision were more specific and straightforward. Some typical examples of analytical comments in this study are: *Do you intend to let the discussion of sources take place within the whole group?*, *The core does not contain much text, is that a deliberate choice?*, and *Have you deliberately put the pictures at the end of the chapter and not in between its text*? Our post-hoc analysis of the evaluative feedback comments showed a similar effect. Evaluative feedback comments also were also shown to be of a different nature in both studies, being more subjective and opinion-oriented in study 1 and more objective and content-oriented in study 2. Again, this difference can be traced back to the existence of many controversial issues in the Health Care course that transcended into the students' internship reports. For example, the raised question of to what extent one can restrict the range of freedom of psychologically challenged persons 'for their own good' when treating them is useful for raising discussion, but offers less concrete or easy leads for the revision of students' texts. These analyses may imply that using peer feedback for certain more objective tasks may yield higher gains than when using them for more subjective tasks, at least when it concerns the number of revisions that is made (and discarding possible opinion-developing effects). In order to shed more light on the exact relationship between peer feedback and its learning effects across different tasks, we suggest further research on possible subdivisions for the feedback functions as defined by Van den Berg et al. (2006) in line with the post-hoc analysis presented above.

With regard to the *aspects* of feedback, both studies established a significant positive relationship between the aspects content and style and the revision of texts: more than half of all feedback remarks that focused on content or style were related to revision of students' texts (52% for content and 58% for style in the first study and 61% for content and 55% for style in the second study). In both studies we found only a small amount of feedback comments on the structure of students' texts, similar to findings by Van den Berg et al. (2006). Their explanation for this finding is that feedback on structure is difficult for students to formulate properly and also difficult to apply in revising texts, especially texts that are already in their final stages of production. Van den Berg et al. also suggest that the task of providing feedback on the structure of a text, which is also an important kind of feedback, may be better suited to the teacher.

Our second research question concerned the way the nature of peer feedback is related to its reception by the receiver. In neither of the two studies did any of the coded functions or aspects of students' feedback comments correlate significantly with the way it was received (in terms of the receivers' evaluation of importance or usefulness and the agreement with the provider of feedback). This would mean that there are other variables that play a role in how feedback is received that, for example, may be less bound to the general characteristics of feedback and more to the particular content and domain of the feedback.

4.1.2. The reception of feedback

Our third research question concerned the way the reception of peer feedback relates to its use for the revision of texts. With regard to the relationship between the reception of feedback and its use for the revision of texts, we found different results for students' appreciation of the received feedback. While students' appreciation of the *importance* of feedback correlated with the revision of texts (in study 1), their appreciation of the *usefulness* of feedback did not (in study 2). Unfortunately, as we have measured user's appreciation of usefulness and importance in two different contexts, we are unable to

attribute their different relationships with revision to either the difference between importance and usefulness, or to a difference in context. We suspect, however, that this difference may result from a different number of *positive* appreciations in both studies. Feedback messages that *solely* consist of compliments do not provide the receiver with many options for revision, but they are often appreciated very positively. Hence, the existence of these messages could blur the relationship between the given appreciation of certain feedback comments and their use for revision. After an additional post-hoc analysis we found that indeed the second study showed more feedback comments that consisted only of compliments (5.4% in the first and 19.1% in the second study), whereas in the second study these compliment often formed the introduction to a larger message, combining it with other feedback functions. Thus, the presence of feedback units that consisted only of compliments may have caused the potential correlation between the receivers' appreciation of the usefulness of feedback and its use for revision not to become apparent in our analysis in the second study. Further research may shed more light on this relationship by identifying the 'direction' of feedback comments (positive or negative), including more qualitative data collection of students' reasons for giving certain scores on importance or usefulness.

With regard to the relationships between the different variables grouped under 'reception of feedback', we found a significant relationship between the appreciation of usefulness and the agreement with the received feedback. In turn, the agreement with feedback showed a significant relationship with its use for revision. Although these relationships between appreciation, agreement, and revision were not found in study 1, this can be explained by the low number of expressed (dis)agreements. In this study, the tool did not offer students the possibility to react, but this limitation was circumvented by several students. The 13 reactions that were thus collected did show a correlation that was positive ($\rho = .26$, $p = .40$, $n = 13$), which can reasonably be expected to become significant with an increased number of reactions.

Thus, the way feedback is received seems to play an important role in the use of feedback, but the exact nature of this relationship needs further investigation. Although we have found several significant relationships, we lack a thorough understanding of how the processes of reception and uptake of feedback take place. Why students use particular feedback is still unclear: they might evaluate it as positive and therefore use it, they might be triggered to change something else, or they might receive feedback but revise their writing accordingly without knowledge of the feedback. For future research, the way feedback is processed should be studied more qualitatively in order to understand the reasons for revisions.

4.1.3. Tools

Our fourth research question concerned the way different tools for peer feedback evoke differences in the nature, reception and use of feedback. In study 2, we found two interesting differences between the Annotation system and Blackboard. When used side-by-side in the same course, Blackboard proved to elicit more evaluative feedback remarks, whereas the Annotation system produced more suggestions for revision. This is in line with the outcomes of a previous study on the collaborative processing of texts, that students' interaction was more critical in Blackboard and more constructive in an Annotation system (Van der Pol et al., 2006).

Furthermore, the unconventional and creative use by some students of both Blackboard and the VLC points to a lack of certain functionalities in these two systems that *are* offered in the Annotation system. Firstly, some students in Blackboard circumvented its constraints for annotation by attaching annotated Word documents to their discussion messages, showing students' preference for annotation. Secondly, several students in the VLC circumvented its constraints on interaction by including reactions to the received feedback in their revised documents, pointing towards students' desire for interactivity. As the Annotation system allows both annotation and interactivity, facilitating formative peer assessment with an interactive system for anchored discussion seems promising. Future experiments with systems for anchored discussion might be aimed at finding out how to increase the quality of interaction, e.g. by moderating or modelling, in order to generate helpful and inspiring feedback. An important question in this regard is also how and to what degree interaction in peer feedback (not only between reviewer and receiver, but also between reviewer and object, between different reviewers, and between students, products and teachers) can increase its effectiveness and efficiency.

4.1.4. Limitations of the study

Regarding the presented differences between tools, we must note that these are not very 'hard', because the main focus of this research was on investigating the relationships between the nature, reception, and use of feedback and not on the comparison of different tools for peer feedback. Although we have mentioned several interesting differences regarding the Annotation system and the Blackboard discussion board, more research is needed to elaborate on these differences between tools. Regarding the relationships between the nature, reception and use of feedback it is important to note that we can identify a correspondence between feedback and revisions, but we cannot be sure that the revision is a *consequence* of the feedback. It might be possible that the authors of a text were already planning to make certain revisions, regardless of the feedback they received. Self-report questionnaires or a more controlled research design as in quasi-experimental studies might be useful to examine these relationships in the future. When elaborating on the relationship between feedback and revision, it could also be relevant to distinguish different types of revision (for example distinguishing deep and more superficial revisions). For example, Annotation systems may promote specific rather than holistic feedback, which may have different consequences for its use in revising texts.

Hewitt (2000) found that more specific feedback aimed at concrete writing issues resulted in a different kind of revision than feedback directed at general and abstract development of ideas.

4.1.5. A general perspective on online formative peer feedback

The potential beneficial effects of formative peer feedback lie in two activities: in *providing* feedback and in *receiving* it. While providing feedback is important in giving students a good idea of the criteria for the product and in developing a sense of confidence by seeing how their peers are performing, receiving feedback mostly serves to improve upon their products. As such, receiving good quality feedback also fulfils an important motivational function: investing time into providing feedback for others is only worthwhile if students also receive useful comments in return. Contrary to most research on peer feedback this study has focused on increasing the learning effect of receiving formative peer feedback, because these effects cannot be taken for granted as much as the effects of providing it. Learning effects of providing feedback will be accomplished relatively simply: as long as student invest time and effort into actively constructing content-oriented reactions, we can expect certain learning gains. The learning effects of receiving feedback, however, highly depend on its quality, which in its turn hinges on the expertise of the provider. It can thus be expected that asking students – who by definition are not experts on the subject matter they are supposed to appropriate – to effectively fulfil this role, poses great challenges.

This study has demonstrated that the uptake and use of peer-generated feedback by its receiver is not straightforward and does depend on certain characteristics of the feedback. In these studies 49% (study 1) and 55% (study 2) of all feedback comments were related to a revision of students' texts. The main characteristic that was positively related to revision of the corresponding parts of students' texts was the presence of concrete suggestions for revisions. Thus, a recommendation for teachers when implementing peer feedback could be to ask students to provide each other with concrete suggestions for revision or to provide modelling or moderating to this end. In order to further investigate the relation between feedback and its learning effects, we recommend more detailed research into the quality of feedback, such as its level of detail and its accuracy, as related to a successful uptake and learning effect. Important questions remain which characteristics of formative peer feedback can optimize its learning effects for both provider *and* receiver, which of these effects need facilitation to be achieved, and how this facilitation may be realized online.

References

Bhalero, A., & Ward, A. (2001). Towards electronically assisted peer assessment: A case study. *ALT-J, 9*(1), 26–37.

Billet, S. (2002). Critiquing workplace learning discourses: Participation and continuity at work. *Studies in the Education of Adults, 34*(1), 56–68.

Chaudron, C. (1983). Evaluating writing: Effects of feedback on revision. In *Paper presented at the 17th annual teachers of English to speakers of other languages TESOL convention.*

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.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative self and peer group assessments. *Assessment and Evaluation in Higher Education, 11*(2), 146–166.

Falchikov, N., & Boud, D. (1989). Students self-assessment in Higher Education: A meta analysis. *Review of Education Research, 59*(4), 395–430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Fiehn, T., Lauer, T., Lienhard, J., Ottmann, T., Trahasch, S., & Zupancic, B. (2003). *From lecture recording towards personalized collaborative learning.* In B. Wasson, S. Ludigsen, & U. Hoppe (Eds.), *Designing for change in networked learning environment. Proceedings of the international conference on computer support for collaborative learning* (pp. 471–475). London: Kluwer Academic Publishers.

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.

Gehringer, E. (2001). Electronic peer review and peer grading in computer-science courses. In H. Walker (Ed.), *Proceedings of the 32nd ACM Special Interest Group on Computer Science Education (SIGCSE) technical symposium on computer science education* (pp. 139–143). New York: ACM Press.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1,* 3–31.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hewitt, B. L. (2000). Characteristics of interactive oral and computer-mediated peer group talk and its influence on revision. *Computers and Composition, 17*(3), 265–288.

Jacobs, G., & Zhang, S. (1989). Peer feedback in second language writing instruction: Boon or bane? In *Paper presented at the annual meeting of the American educational research association.*

McCormack, C., & Jones, D. (1998). *Building a web-based education system.* New York: Wiley.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31,* 199–218.

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,* 239–249.

Prins, F., Sluijsmans, D., & Kirschner, P. (2006). Peer feedback exercises for general practitioners: Improving skills for professional development. In *Paper presented at the conference of the American educational research association.*

Riffe, D., Lacy, S., & Fico, F. (1998). *Analyzing media messages: Using quantitative content analysis in research.* Mahwah, NJ: Lawrence Erlbaum Associates.

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

Shekary, M., & Tahririan, M. H. (2006). Negotiation of meaning and noticing in text-based online chat. *The Modern Language Journal, 90*(4), 557–573.

Shiffman, S. M., & Stone, A. A. (1998). Ecological momentary assessment: A new tool for behavioural medicine research. In E. David, S. Krantz, & E. A. Baum (Eds.), *Technology and methods in behavioural medicine* (pp. 17–131). Mahwah, NJ: Erlbaum.

Sitthiworachart, J., & Joy, M. (2004). Using web-based peer assessment in fostering deep learning in computer programming. In *Paper presented at the international conference on education and information systems: Technologies and applications.*

Steehouder, M., Jansen, C., Maat, K., Staak, J., & van de Woudstra, E. (1992). *Leren communiceren (3^{e} herziene druk).* Groningen: Wolters-Noordhoff.

Tannacito, T., & Tuzi, F. (2002). A comparison of e-response: Two experiences; one conclusion. *Kairos: A Journal of Rhetoric, Technology, and Pedagogy, 7*(3), 1–21.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

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

Topping, K. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645.

Topping, K., Smith, E., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education, 25*(2), 149–169.

Trahasch, S. (2004). Towards a flexible peer assessment system. In Y. Akpinar (Ed.), *Proceedings of 5th international conference on information technology based higher education and training* (pp. 16–20). Istanbul: Institute of Electrical and Electronics Engineers (IEEE).

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition, 21*(2), 217–235.

Van den Berg, B. A. M., Admiraal, W. F., & Pilot, A. (2006a). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.

Van den Berg, B. A. M., Admiraal, W. F., & Pilot, A. (2006b). Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education, 11*, 135–147.

Van der Pol, J., Admiraal, W. F., & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *International Journal of Computer Supported Collaborative Learning, 1*(3), 339–357.

Van Weert, T. J., & Pilot, A. (2003). Task-based team learning with ICT: Design and development of new learning. *Education and Information Technologies, 8*(2), 195–214.

Webb, N. M., & Mastergeorge, A. M. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39*, 73–97.