Using feedback requests to actively involve assessees in peer assessment: effects on the assessor's feedback content and assessee's agreement with feedback

Peer-reviewed author version

Voet, Michiel; GIELEN, Mario; Boelens, Ruth & De Wever, Bram (2018) Using feedback requests to actively involve assessees in peer assessment: effects on the assessor's feedback content and assessee's agreement with feedback. In: EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION, 33(1), p. 145-164.

DOI: 10.1007/s10212-017-0345-x
Handle: http://hdl.handle.net/1942/26546

Authors

Michiel Voet: http://www.tecolab.ugent.be/michiel.php

Mario Gielen http://www.tecolab.ugent.be/mario.php

Ruth Boelens http://www.tecolab.ugent.be/ruth.php

Bram De Wever: http://www.tecolab.ugent.be/bram.php

Using feedback requests to actively involve assessees in peer assessment: Effects on the assessor's feedback content and assessee's agreement with feedback.

ABSTRACT

Criticizing the common approach of supporting peer assessment through providing assessors with an explication of assessment criteria, recent insights on peer assessment call for support focussing on assessees, who often assume a passive role of receivers of feedback. Feedback requests, which require assessees to formulate their specific needs for feedback, have therefore been put forward as an alternative to supporting peer assessment, even though there is little known about their exact impact on feedback. Operationalizing effective feedback as feedback that (1) elaborates on the evaluation, and (2) to which the receiver is agreeable, the present study examines how these two variables are affected by feedback requests, compared to an explanation of assessment criteria in the form of a content checklist. Situated against the backdrop of a writing task for 125 first-year students in an educational studies program at university, the study uses a 2x2 factorial design that resulted in four conditions: a control, feedback request, content checklist, and combination condition. The results underline the importance of taking message length into account when studying the effects of support for peer assessment. Although feedback requests did not have an impact on the raw number of elaborations, the proportion of informative elaborations within feedback messages was significantly higher in conditions that used a feedback request. In other words, it appears that the feedback request stimulated students to write more focused messages. The use of a feedback request did, however, not have a significant effect on agreement with feedback.

1. INTRODUCTION

In education, assessment has been traditionally regarded as a responsibility of teachers, as their expertise and experience seems to make them the most qualified to inform students about how well they are doing (Hattie & Timperley, 2007; Ozogul & Sullivan, 2009). However, recent research has argued that there are good reasons to also involve students' peers in the assessment process. After all, one of the main goals of education is to create self-regulated learners, who must learn to evaluate performance against a given set of standards (Nicol & Macfarlane-Dick, 2006). More specifically, research has shown that peer assessment draws on several higher order thinking skills, such as critical thinking, problem solving, and decision-making, and may therefore contribute to the development of these skills (King, 2002).

Furthermore, peer assessment transfers part of the ownership of the assessment process to students, and, in this way, makes them feel more responsible for their own learning, which can in turn increase their motivation and engagement in class. (Ng, 2016; Ozogul & Sullivan, 2009). Finally, research has shown that, given enough time and support, peers can offer an evaluation that is of equal reliability and validity to that of a teacher (Topping, 2009). As, a result, scholarly interest in peer assessment has been growing steadily.

Looking at how peer assessment may impact students' learning, previous work has found that assessment is most effective in learning environments that enable students to seek, receive, and act on feedback (Price, Handley, Millar, & O'Donovan, 2010). In such learning environments, the focus shifts from 'assessment of learning' to 'assessment for learning' (Gielen & De Wever, 2015a). The latter contributes to learning by providing an evaluation of present performance, as well as the opportunity to act on this evaluation and further improve performance (Sadler, 1989). In this kind of setting, peer feedback seems to be most effective in improving performance when its content offers an elaboration of the evaluation (Butler, 1987; Walker, 2014), and is experienced as useful by the assessee (Anseel, Lievens, & Schollaert, 2009; Harks, Rakoczt, Hattie, Besser, & Klieme, 2014). In what follows, the concepts of elaboration and usefulness of feedback are therefore further explained.

Making a distinction between *verification*s, which merely indicate how well certain criteria are achieved, and *elaborations*, which further explain the evaluation, the literature on feedback content suggests that the latter kind of statements is particularly valuable to improving students' performance (Narciss, 2008). Looking further into the nature of elaborations, it appears that these statements can be either *informative*, giving more information about the reasoning behind particular judgements, or *suggestive*, providing specific directions with regard to performance improvement (Gielen & De Wever, 2015b). According to previous studies, both informative and suggestive elaborations may contribute to an increased performance on behalf of the assessee, through the particular information that these forms of elaborations provide with regard to the assessment (Butler, 1987; Walker, 2014).

Next to the content of feedback, research suggests that it is also important to consider how feedback is dealt with after assessees have received it (Walker, 2014). According to several studies, the effect of feedback on performance is mediated by assessees' satisfaction with the feedback, and especially their perceptions of its *usefulness* (Anseel et al., 2009; Harks et al., 2014). Regarding the question as to how usefulness of feedback can then be operationalized, the work by Strijbos, Narciss and Dünnebier (2010) suggests that students' perceptions of feedback's usefulness are indistinguishable from their *agreement* with this feedback. In relation to this, other studies have noted that students' agreement with feedback

is to a large extent determined by the characteristics of the feedback source, and its perceived expertise in particular (Ilgen, Fisher, & Taylor, 1979). As such, some have found that students are generally inclined to place less trust in assessments provided by their peers than those by the teacher (e.g. Kaufman & Schunn, 2011; Planas Lladó et al., 2014).

Unfortunately, it seems that not all students are able to provide elaborate and useful feedback, due to various reasons, such as limited competence, judgement bias, or low social skills (Cheng, Liang, & Tsai, 2015). Research therefore suggests that providing additional support is necessary for optimizing students' feedback, and might even be an essential precondition for peer assessment (Poverjuc, Brooks, & Wray, 2012). The most common approach consists of providing support to the assessor, through *explications of assessment criteria* in the form of checklists, inventories, response grids, marking schedules, or rubrics (Falchikov, 1995; Gielen & De Wever, 2015c; Panadero & Jonsson, 2013; K. Topping, 1998). By clarifying and exemplifying the expected performance, these tools guide assessors in providing elaborate and useful feedback (Topping, 1998).

Recently, however, there has arisen some criticism with regard to peer feedback support that focusses solely on the assessor (Kollar & Fischer, 2010). Arguing that peer assessment represents an inherently collaborative activity (Falchikov & Goldfinch, 2000), scholars have called out to involve the assessee more closely in the feedback process, so as to establish an actual feedback dialogue (Nicol, 2010). To be more specific, it has been argued that the assessee could be further involved in the feedback process through *feedback requests* (Gielen & De Wever, 2015a), which allow the assessee to specify particular questions or components on which feedback is required (Gibbs & Simpson, 2004; Gielen, Tops, Dochy, Onghena, & Smeets, 2010; Nicol & Macfarlane-Dick, 2006). Similar to the explications of assessment criteria described above, this would then stimulate assessors to provide more elaborate and useful feedback.

As of yet, however, there is little information on the effect that a more active involvement of the assessee might have on the peer assessment process. In particular, seeing that explications of assessment criteria and feedback requests both provide directions to the assessor's review work, the question arises whether one approach might lead to better outcomes, or whether they might complement one another. In line with the description of effective feedback that is outlined above, these outcomes can be understood in terms of the feedback content provided by the assessor, and the assessee's agreement with feedback.

2. RESEARCH QUESTIONS

In short, the aim of the present study is to investigate how a more active involvement of the assessee in the feedback process, in the form of a feedback request, might influence the content of the assessor's feedback, as well as the assessee's agreement with the feedback. The effects of the feedback request are investigated in relation to those of an explication of assessment criteria for the assessor, which is commonly used by research that aims to support peer assessment (Falchikov, 1995; Gielen & De Wever, 2015c; Panadero & Jonsson, 2013; K. Topping, 1998). As such, the research questions are:

- How does support incorporating a feedback request, explication of assessment criteria, or a combination of both impact the content of the *assessor*'s feedback?
- How does support incorporating a feedback request, explication of assessment criteria, or a combination of both impact the *assessee*'s agreement with feedback?

3. DESIGN AND METHODS

In order to investigate the effects of a feedback request, as compared to an explication of assessment criteria, on the effectiveness of peer feedback, a quasi-experimental study was designed within the context of higher (university) education. This section offers more information on the task used for the study, the conditions, and the data analysis.

3.1. Task

As part of an educational studies program, 125 first-year students were divided over 27 groups, with three to five students per group, and given an assignment on writing a research abstract. As the literature stresses that multiple occasions of practice are required to develop students' ability to provide feedback, the present study consisted of three cycles, lasting nine weeks in total, or three weeks for each cycle. At the start of each cycle, groups were given one unpublished research paper for each student, from which the abstract had been removed. Each individual student then had to: (1) read their assigned paper and write a draft version of an abstract, (2) read the abstract of a peer and provide feedback, and (3) evaluate the feedback on their own abstract, and revise their abstract into a final version. The peer feedback was not reciprocal, but students were providing feedback to the same peer during all three cycles. Students carried out all of the work within the university's online learning management system, and could work on the task at any time, as long as the cycle lasted. Outside of the task, each student followed the same educational studies course together with all of the other participants.

To provide a set of standards for carrying out the peer feedback, all students were instructed to provide their feedback using the template by Gielen and De Wever (2015a). This

template asks the assessor to provide feedback and suggestions with regard to 7 criteria: aims, problem statement, methodology, results, conclusion, limitations, and general judgement. After having received the feedback, the assessee is asked to complete the template by adding, for each criterion, an evaluation of the assessor's comments. An example of a completed peer feedback template can be found in Figure 1.

criteria	**feedback** *(provided by assessor)*	**suggestions** *(provided by assessor)*	**reception** *(provided by assesse)*
1. aims	You formulated everything clearly. This is much better compared to your draft version of the previous article.	/	/
2. problem statement	The research questions are present and are described clearly.	If you could formulate the questions in your own words, it might be easier to better understand the text. If you literally copy the questions, it's best to also mention the sources.	I copied the questions because I was afraid that my interpretation would not be 100% correct. Saying that, I did change a few things: I have downsized the questions section to reduce the length of my abstract.
3. methodology	Almost complete. But it lacks information on how a wiki works? What do you need or what is expected from you?	I would include a brief explanation of the wiki. The information about methodology should be more concise.	It was quite difficult to explain how the wiki works, since my abstract is already quite long. Honestly, it does not seem necessary as the study mainly deals with providing and receiving peer feedback.

Figure 1. Example of a completed feedback template (first 3 criteria).

3.2. Conditions

Students were randomly assigned to groups and groups were randomly divided over four conditions. During cycle 1, all conditions had to carry out the feedback process using only the peer feedback template. This made it possible to establish a baseline, which would allow to evaluate the effect of the conditions against students' work without additional support. During cycle 2 and 3, students still carried out the peer assessment using the feedback template, but, depending on the condition, some student groups received additional support. The intervention was spread out over 2 cycles, so that students would be able to get accu£ stomed to working with the additional support during cycle 2, before carrying out their final

peer assessment task during cycle 3. A 2 x 2 factorial design was used to test the effects of two types of support: a feedback request, and an explication of assessment criteria in the form of a content checklist. Table 1 provides an overview of the activities within each condition during cycle 2 and 3.

Table 1

Overview of the activities within each condition during cycle 2 and 3

		condition			
activity	**actor**	control	feedback request	content checklist	combination
1. writing text	assessee	x	x	x	x
2. peer feedback request	assessee		x		x
3. preparing content checklist	assessor			x	x
4. providing feedback	assessor	x	x	x	x
5. revising text	assessee	x	x	x	x
6. evaluating peer feedback	assessee	x	x	x	x

(1) In the *control* condition, students did not receive additional support. (2) In the *feedback request* condition, *assessees* were required to complete a feedback request form, which the assessor then had to take into account when formulating feedback. This form enabled assessees to specify questions with regard to each of the 7 criteria in the feedback template. Apart from the form, assessees did not receive instructions on how to formulate a feedback request, but were instead left free to complete these forms as they best saw fit. Figure 2 provides an example of a completed feedback request.

criteria	**remarks, questions, issues?**
1. aims	- Should I also include something about the specific focus of this study?
2. problem statement	- /
3. methodology	- I wonder if I have sufficiently explained this part. How should I make it more concise, because I believe that it is rather long right now. - Is the structure of this part adequate?

Figure 2. Example of a completed feedback request (first 3 criteria).

(3) In the *content checklist* condition, *assessors* had to create a content checklist explicating the 7 criteria that were provided to students, before providing feedback. In order to do so, they were required to read through the research paper given to the assessee, and select and

categorize the content that they thought was essential. For an example of a completed content checklist, see Figure 3. (4) Finally, the *combination* condition incorporated both the feedback request and content checklist. Assessors were instructed to create the checklist first and to take both the checklist and the feedback request into account when formulating their feedback.

Criteria	relevant content for each criterion
1. aims	- the goal of this paper is not so much to promote the use of commercial video games in education in se, as to understand, explain and predict changes in teachers' behavior in view of adopting these tools. - The study contributes to an established body of research that has examined general reasons for playing video games, the play behavior of teachers and teachers-in-training and teachers' acceptance of educational computer games. - In this paper, a model-based approach to teachers' beliefs is presented and evaluated, based on the understanding that teachers are faced with many variables that interact with each other to either facilitate or discourage the acceptance of technology.
2. problem statement	- The present study focuses on the factors that influence the acceptance of commercial video games as learning tools in the classroom. - When discussing teachers in relation to digital game-based learning, the focus is often on what they perceive as potential barriers to the implementation of games in their own practice. - …to measure the concerns of the teachers regarding the difficulty of using games in their practice
3. methodology	- …the focus is on teachers in practice. - Secondary schools were contacted based on their denomination (i.e. community/subsidized public schools, and subsidized private schools), type of education (general, technical, and vocational) and geographical distribution. - The teachers could fill in the questionnaires using the medium of their choice. This way, 505 teachers could be involved. - The questionnaire consisted of three parts, examining demographic information, teacher related variables, and the constructs of the research model.

Figure 3. Example of a completed content checklist (first 3 criteria).

3.3. Data analysis

In order to analyse the content of the peer feedback, and assesees' agreement with this feedback, a subsample of 16 out of 27 groups (4 randomly selected out of each condition),

with 79 students in total, was selected. Selecting this subsample significantly reduced the workload for the analysis, while the dataset remained representative of the whole group's performance, and sufficient for carrying out the required analyses. To investigate the evolution in this subsample's performance throughout the intervention, the analysis focused on the feedback templates that were completed during the first cycle (i.e. the baseline, during which none of the conditions received support other than the feedback template) and the third cycle (i.e. the second cycle during which the additional support differed depending on conditions). These completed feedback templates contained both the assessor's feedback and assessee's evaluation of this feedback. Due to one student failing to provide feedback during the feedback cycle, 157 completed feedback templates were available for analysis.

Following the procedure of Strijbos, Martens, Prins, and Jochems (2006), data were segmented before the actual coding was carried out. The completed feedback templates were segmented based on sentences, as sentences, or parts of compound sentences, often contain a single concept, statement, or expression (Strijbos et al., 2006). This also allowed a more fine-grained analysis compared to approaches that use broader units of analysis, such as thematic units, or paragraphs (Neuendorf, 2002). The segmentation procedure resulted in 4428 segments of feedback content, and 1202 segments covering students' evaluation of feedback. The average number of segments in assessors' feedback messages was 28.20 (SD=10.09), while that in assessees' evaluations of feedback was 7.66 (SD=1,5). In other words, assessees' evaluations of feedback were a good deal shorter compared to the original feedback messages that the assessors composed.

The content of the feedback messages was analysed using a coding scheme based on the work by Gielen and De Wever (2015b), which can be found in appendix 1. This coding scheme categorizes segments of feedback messages as either (1) verifications, (2) elaborations, or (3) neutral statements (not focussing on the quality of the work). Given the particular importance of elaborations in feedback (e.g. Walker, 2014; Webb & Mastergeorge, 2003), it further distinguishes between (2a) informative elaborations and (2b) suggestive elaborations.The coding scheme also takes into account assessees' (4) evaluation of feedback, by checking whether they (4a) agree, (4b) partially agree, or (4c) disagree with the feedback.

The coding was carried out by a coder who followed a four-hour training given by the second author. During this training, each category within the coding scheme was explained using various example segments. To check interrater reliability for feedback content, the coder and second author independently coded a dataset consisting of 1506 segments of feedback content (which was part of a previous study). The results indicate good interrater reliability with a Cohen's K of .91 for feedback content in general (i.e. 3 categories: verifications, elaborations, and general statements), and a Cohen's K for elaboration type (i.e.

2 categories: informative elaborations, and suggestive elaborations) that was also .91. The interrater reliablity for students' evaluations of feedback was checked by a third coder, who indepently coded 1166 segments of evaluations. Cohen's Kappa for students' agreement with feedback (i.e. 4 categories: agreement, partial agreement, disagreement, no evaluation) was .92, again indicating good interrater reliability.

The data are hierarchically structured, as there are multiple performances, one during each of the two cycles (level 1), for each student (level 2), who is in turn nested within a particular group (level 3). MLwiN 2.10 was therefore used to analyze the data, as this program allows to estimate multilevel models that, compared to single-level analyses, provide more precise estimates of the data by allowing variance at each level (Hox, 1998). Given that the data were gathered at two different times, a growth curve modeling approach was used to investigate the impact of the different forms of support on both feedback content and agreement with feedback. This approach allows to investigate whether evolutions in performance from cycle 1 to cycle 3 are in fact significant, and whether the different conditions significantly influence this evolution.

Prior to model estimation, the data were checked for outliners, based on the rule of thumb of z = 3 (see Osborne, 2004). The results revealed a maximum of 3 outliers (out of 157 cases) per variable. Upon closer inspection, all outliers turned out to be legitimate cases, and were therefore not removed from the dataset (Orr, Sacket, & Dubois, 1991). Model estimation then began with the estimation of two null models: one for feedback content, and the other for students' agreement with feedback. Based on these models, two growth models were estimated for the dependent variables of feedback content, and students' agreement with feedback. In both cases, the four conditions were added as independent variables, with the control condition as the reference. In these growth curve models, the intercept represents the control condition's performance during cycle 1, while the variables representing each of the other conditions refer to their respective differential performances during cycle 1. A separate variable, 'growth', represents the evolution from cycle 1 to cycle 3 for the control condition, while interaction effects between this variable and the other conditions expresses these conditions' differential growth from cycle 1 to cycle 3.

Finally, since message length was found to vary considerably between conditions (see Figure 4), the number of segments within the assessors' feedback messages was added as a control variable in the growth model for feedback content, while the number of segments within assessees' evaluations of the feedback was added as a control variable into the growth model for agreement with feedback.

4. RESULTS

This section first examines the conditions' impact on assessors' feedback content. In particular, it examines differences between effects when the length of the feedback messages is first left out of the equation, and then taken into account. In the second part of this section, the focus shifts toward the conditions' effects on assessees' agreement with the feedback they received, in relation to the respective lengths of their evaluations.

4.1. The content of the assessor's feedback

To examine conditions' effect on the evolution of feedback content from cycle 1 to cycle 3, a multivariate multilevel growth curve model was estimated, with three dependent variables: verifications, informative elaborations, and suggestive elaborations. Table 2 presents the estimates of the null model, which was estimated first. The results indicate that, for cycle 1 and cycle 3 combined, and across all conditions, the average feedback message has 11.43 segments coded as verifications, 8.03 as informative elaborations, and 7.96 as suggestive elaborations. In addition, Table 2 shows that for all dependent variables, variance was significant at the student level, but not at the group level, meaning that, in this case, not taking into account the nesting of students in groups would not have yielded different estimates.

Table 2

Multivariate multilevel null model of conditions' effects on feedback content

	dependent variables		
	verifications	informative elaborations	suggestive elaborations
fixed part			
intercept	11.43 (0.36)***	8.03 (0.63)***	7.96 (0.51)***
random part			
intercept group σ^2_f	0.52 (0.81)	2.73 (2.37)	2.43 (1.5)
intercept student σ^2_v	2.201 (1.69)	10.38 (3.49)**	4.78 (1.63)**
intercept measurement σ^2_u	11.41 (1.82)***	15.62 (2.50)***	7.42 (1.19)***

Note. Parameter estimates with standard deviations between brackets. * Significant at .05; ** significant at .01; *** significant at .001

Following this, a growth curve model with conditions as the independent variable was estimated. A likelihood ratio test comparing this model's deviance (2503.85) to that of the null model (2601.05) indicates that this more elaborate model is a significant improvement over the null model (X^2=3097.2, df=1, p<.001). The output of the growth curve model can be found in Table 3, while Figure 4 presents a graphical overview of this model's estimates.

Looking first at the results of cycle 1, during which none of the conditions received additional support, Table 3 indicates that there are no significant differences between the number of verifications and informative elaborations across conditions. Contrary to what

would be normally expected, there is, however, a significantly lower number of suggestive elaborations in the combination condition during cycle 1. The reason for this difference is unclear, as students were randomly grouped into conditions, and all students followed the same course in educational studies outside of the task. This significant difference is, however, not problematic, as the growth curve model takes initial differences between conditions into account.

Table 3

Multivariate multilevel growth curve model of conditions' effects on feedback content, without message length as control variable

	dependent variables		
	verifications	informative elaborations	suggestive elaborations
fixed part			
intercept	10.15 (0.74)***	7.4 (1.16)***	9.75 (0.82)***
feedback request	1.6 (1.04)	-1 (1.63)	-1.85 (1.16)
content checklist	1.75 (1.04)	-0.8 (1.63)	-1.2 (1.16)
combination	1.9 (1.06)	0.33 (1.65)	-2.9 (1.75)*
growth (from cycle 1 to cycle 3)	2.8 (0.97)**	2.75 (1.16)*	0.25 (0.843)
growth x feedback request	-3.48 (1.38)*	-0.07 (1.66)	-1.06 (1.2)
growth x content checklist	-2.6 (1.37)	-1.55 (1.64)	-1.45 (1.19)
growth x combination	-5.54 (1.39)***	-1.28 (1.66)	-1.04 (1.2)
random part			
intercept group σ^2_f	0 (0)	1.86 (1.67)	0.29 (0.7)
growth group σ^2_f	0 (0)	0 (0)	0 (0)
intercept student $\sigma 2_v$	10.85 (1.73)***	17.34 (2.98)***	12.02 (2.01)***
growth student $\sigma 2_v$	18.83 (3.01)***	26.75 (4.28)***	14.14 (2.26)***

Note. Parameter estimates with standard deviations between brackets. * Significant at .05; ** significant at .01; *** significant at .001

Most important, Table 3 provides more information on the growth in verifications and elaborations in feedback messages from cycle 1 to 3. The results show a significant growth in verifications within the control condition. In the other conditions, the number of verifications during cycle 3, as compared to cycle 1, is lower compared to that in the control condition, although this is only significant in the feedback request and combination conditions. Furthermore, there is a significant growth of informative elaborations across all conditions, with no significant differences between conditions. In contrast, the growth of suggestive elaborations is not significant in any of the four conditions.

Most important, Figure 4 shows that the number of segments within a feedback message, as well as its growth from cycle 1 to cycle 3, differs across conditions. As it is possible

that differences in message length may mask the conditions' effects, the number of segments within a feedback message was added as a control variable into the model.

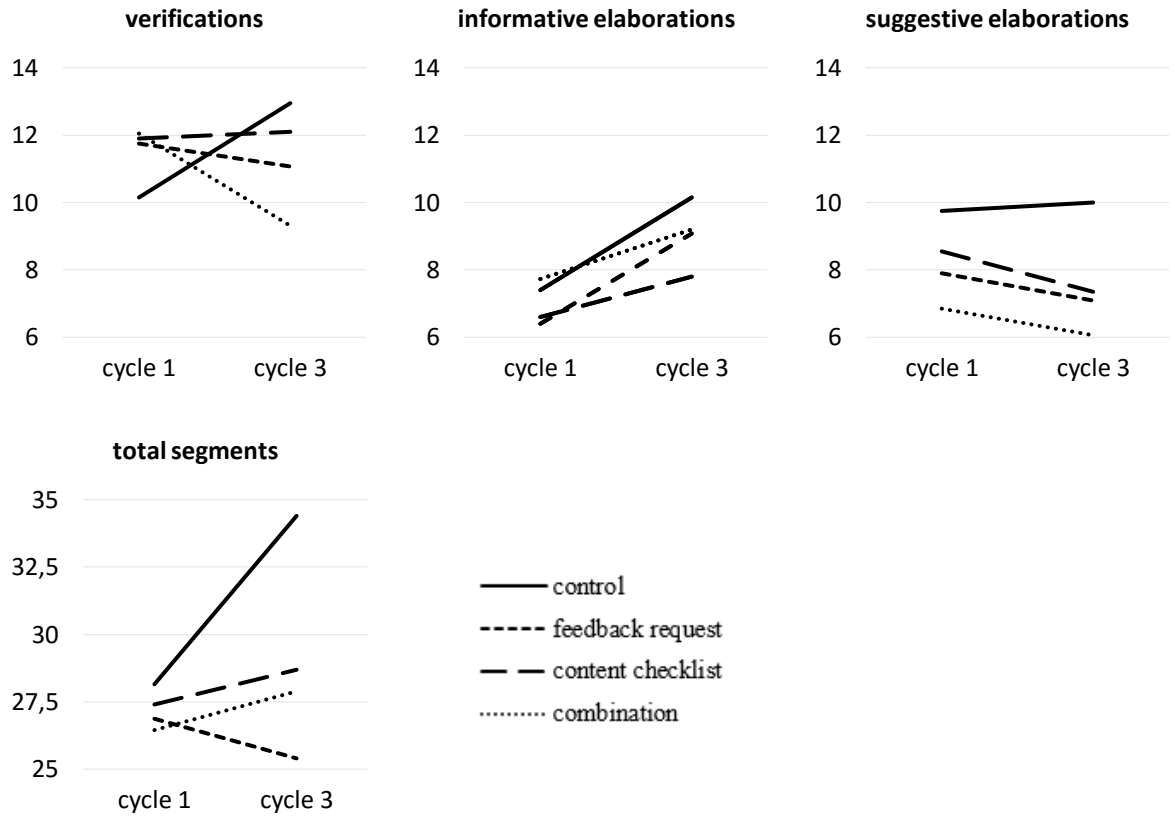

Figure 4. Feedback content during cycle 1 and cycle 3, without taking message length into account (based on estimates in Table 2, and a separate multilevel growth curve model for total segments).

Similar to before, the difference in deviance between the model with segments within a feedback message as a control variable (1899.73), and the model without (2503.85), suggests that the more elaborate model is a significant improvement (X^2=604.12, df=1, p<.001). Table 4 contains the output of the model controlling for the number of segments, and Figure 5 presents the graphs of this model's parameter estimates. Together, they paint a different picture of the conditions' impact on feedback content.

Starting with the results of the baseline measure of cycle 1, the results shows that, in addition to the significantly lower number of suggestive elaborations in the combination condition, the proportion of verifications in the feedback request, content checklist, and combination condition is significantly higher than that in the control condition. As noted before, while these initial differences between conditions are taken into account by the

growth model, and therefore do not complicate the analysis, there is unfortunately no clear-cut explanation for these differences.

Table 4

Multivariate multilevel growth curve model of conditions' effects on feedback content, with message length as control variable

	dependent variables		
	verifications	informative elaborations	suggestive elaborations
fixed part			
intercept	10.17 (0.59)**	7.42 (0.75)***	9.76 (0.78)**
feedback request	2.07 (0.83)*	-0.26 (1.06)	-1.45 (1.1)
content checklist	1.96 (0.83)*	-0.47 (1.06)	-1.02 (1.1)
combination	2.25 (0.84)**	0.89 (1.07)	-2.62 (1.11)*
growth (from cycle 1 to cycle 3)	1.08 (0.67)	0.01 (0.67)	-1.23 (0.75)
growth x feedback request	-2.18 (0.95)*	2.02 (0.93)*	0.11 (105)
growth x content checklist	-1.22 (0.94)	0.64 (0.93)	-0.27 (1.04)
growth x combination	-3.41 (0.96)***	2.11 (0.95)*	0.79 (1.07)
segments (centered around mean)	0.28 (0.02)***	0.44 (0.02)***	0.24 (0.02)***
random part			
intercept group σ^2_f	0 (0)	1.09 (0.65)	0.92 (0.58)
growth group σ^2_f	0 (0)	0 (0)	0 (0)
intercept student σ^2_v	6.88 (1.1)***	5.84 (1)***	7.5 (1.26)***
growth student σ^2_v	8.69 (1.39)***	8.41 (1.35)***	10.78 (1.72)***

Note. Parameter estimates with standard deviations between brackets. * Significant at .05; ** significant at .01; *** significant at .001

As the results on the growth from cycle 1 to cycle 3, which are also presented by Table 4, indicate, the number of segments has a significant effect on the number of verifications, informative elaborations, and suggestive elaborations. In other words, message length does need to be taken into account to get an accurate overview of each condition's effects. Looking, then, at these effects, it appears that the feedback request and combination condition have a significant effect on the growth of verifications and informative elaborations. Compared to the control condition and content checklist conditions, where growth was found to be not significant, the growth score for verifications was significantly lower, while that for informative elaborations was significantly higher in the feedback request and combination conditions. In other words, these results show that, in conditions that did not offer a feedback request, the proportion of informative elaborations remained the same throughout the intervention. In contrast, this proportion significantly increased in conditions with a feedback request, while that of verifications significantly decreased. Finally, and similar to before, the

results show that there was no significant evolution of suggestive elaborations in any of the conditions.

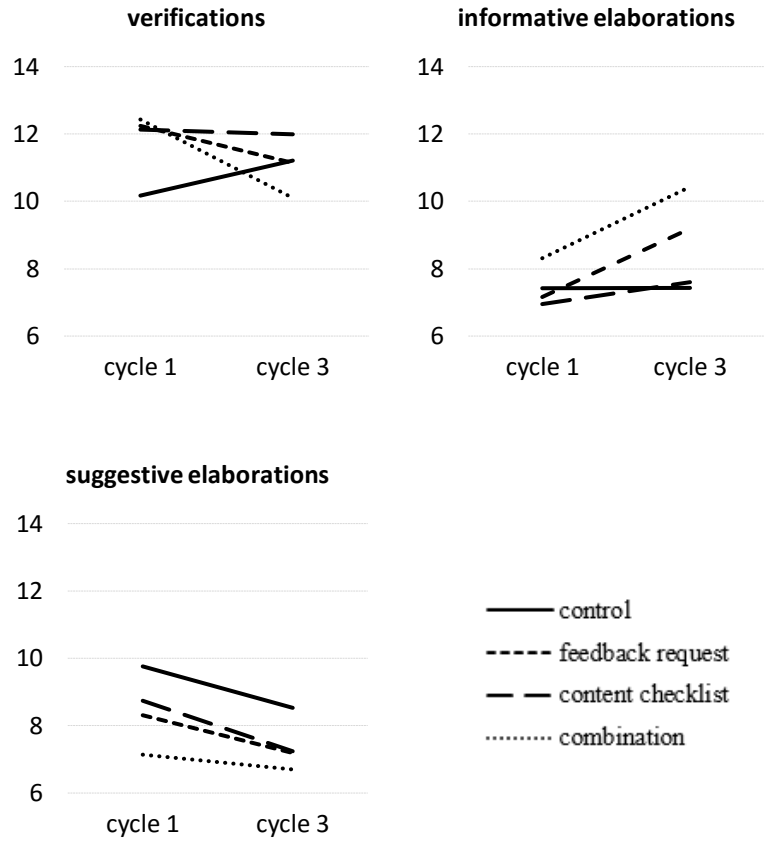

Figure 5. Feedback content during cycle 1 and cycle 3,
with message length taken into account (based on estimates in Table 3).

4.2. The assessee's agreement with feedback

Similar to the analysis of feedback content, a multivariate multilevel model was estimated to determine conditions' effect on agreement with feedback. The model considered the following three independent variables: agreement, partial agreement and disagreement with feedback.

Table 5 gives an overview of the estimates for the null model, which indicate that, across cycle 1 and cycle 3, and all four conditions, assessees on average expressed 4.72 agreements, 0.81 partial agreements, and 0.93 disagreements with the feedback that they received. Furthermore, there is significant variance for all dependent variables at student level, but not at group level.

Table 5

Multivariate multilevel null model of conditions' effects on agreement with feedback

	dependent variables		
	agreement	partial agreement	disagreement
fixed part			
intercept	4.72 (0.17)***	0.81 (0.12)***	0.93 (0.13)***
random part			
group level σ^2_f	0	0.11 (0.08)	0.13 (0.09)
student level σ^2_v	0.96 (0.42)*	0.29 (0.13)*	0.25 (0.15)
measurement level σ^2_u	2.6 (0.42)**	0.74 (0.12)**	0.95 (0.15)**

Note. Parameter estimates with standard deviations between brackets. * Significant at .05; ** significant at .01; *** significant at .001

Table 6

Multivariate multilevel growth curve model of conditions' effects on agreement with feedback, with evaluation length as control variable

	dependent variables		
	agreement	partial agreement	disagreement
fixed part			
intercept	4.43 (0.51)	0.9 (0.29)	1.09 (0.21)**
feedback request	-0.12 (0.72)	0.08 (0.41)	-0.27 (0.3)
content checklist	-0.09 (0.73)	0.23 (0.41)	-0.15 (0.3)
combination	0.24 (0.73)	-0.18 (0.41)	0.06 (0.3)
growth (from cycle 1 to cycle 3)	0.77 (0.57)	0.1 (0.29)	-0.07 (0.28)
growth x feedback request	-0.29 (0.81)	-0.64 (0.41)	0.02 (0.39)
growth x content checklist	0.37 (0.81)	-0.51 (0.41)	-0.05 (0.39)
growth x combination	-0.98 (0.82)	-0.25 (0.41)	-0.21 (0.4)
segments (centered around mean)	0.37 (0.09)***	0.09 (0.05)	0.34 (0.06)***
random part			
intercept group σ^2_f	0.53 (0.35)	0.09 (0.12)	0 (0)
growth group σ^2_f	0.6 (0.41)	0.07 (0.13)	0 (0)
intercept student σ^2_v	2.6 (0.45)***	1.22 (0.22)***	0.9 (0.14)***
growth student σ^2_v	3.46 (0.59)***	1.29 (0.23)***	1.5 (0.24)***

Note. Parameter estimates with standard deviations between brackets. * Significant at .05; ** significant at .01; *** significant at .001

When estimating the multilevel growth curve model, the number of segments within each evaluation was immediately added as a control variable, so as to focus on students' relative agreement and disagreement with feedback, in relation to the length of the evaluations they wrote. A comparison of this model's deviance (1383.03) to that of the null model (1530.2) shows that the full model is a significant improvement over the null model (X^2=147.17, df=1, p<.001). The model is presented in Table 6, and a visualization of its estimates in Figure 6.

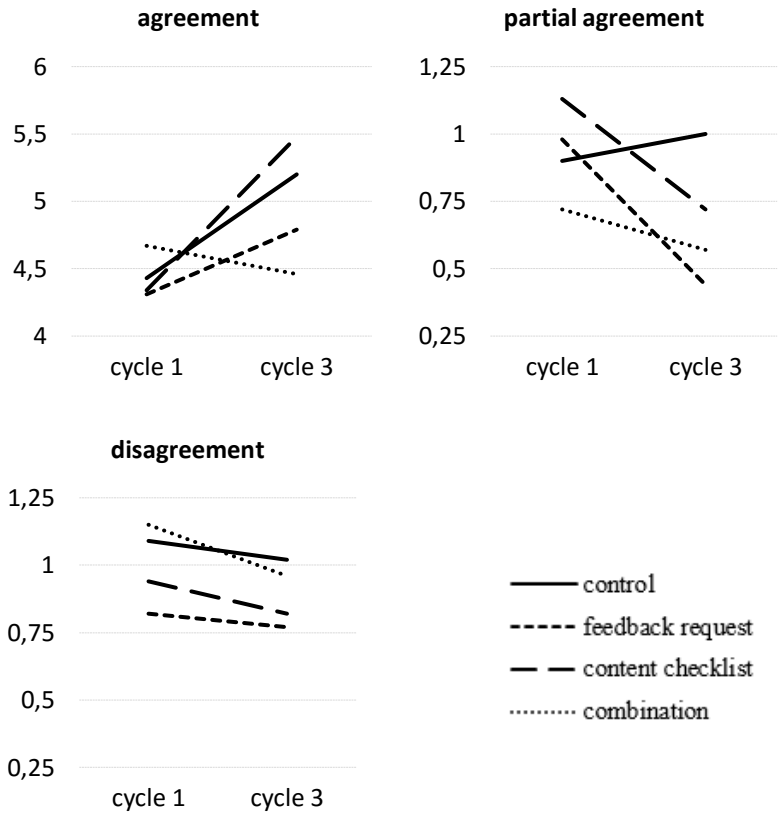

Figure 6. Agreement with feedback during cycle 1 and cycle 3 (based on estimates in Table 5).

Quite logically, the results show that the number of segments within assessees' evaluation of received feedback messages has a significant positive effect on the number of agreements and disagreements they express. However, the number of segments does not have a significant effect on the extent to which assessees partially agreed with the received feedback. In addition, the absence of significant results with regard to the conditions' effects, indicates that the evolution of assessees' agreement, partial agreement or disagreement with feedback from cycle 1 to cycle 3 does not vary across conditions.

5. DISCUSSION AND CONCLUSION

Building on previous work calling for a more active involvement of the assessee in the peer feedback process (Prins & Mainhard, 2009), the present study set out to investigate the impact of a feedback request form on the effectiveness of peer feedback. Contrary to traditional approaches that support peer feedback by providing assessors with an explication of assessment criteria (e.g. Falchikov, 1995; Panadero & Jonsson, 2013; Topping, 1998), feedback requests facilitate peer feedback by providing assessees with the opportunity to inform the assessor of their needs (Gibbs & Simpson, 2004; Gielen, Tops, Dochy, Onghena, & Smeets, 2010; Nicol & Macfarlane-Dick, 2006). To determine the differential impact of peer feedback requests, their effects were examined against an explication of assessment criteria in the form of a content checklist. In line with the present study's definition of effective feedback (see section '1. Introduction'), the effects on both feedback content and agreement with feedback are taken into account.

With regard to feedback content, the results underline the importance of taking message length into account when examining the effects of support for peer assessment. Comparing the feedback content from cycle 3 to that of cycle 1, without taking message length into account, the results suggest a growth in informative elaborations, without significant differences between conditions, together with non-significant growth in suggestive elaborations. In this case, the only difference between conditions appears to be related to the number of verifications, which increased in the control and content checklist condition, but decreased in the feedback request and combination conditions.

However, when the number of segments is entered into the equation, the results reveal an altogether different picture. Again comparing the feedback content from cycle 3 to that of cycle 1, but this time taking the number of segments into account, the results show a significantly higher proportion of informative elaborations, and significantly lower proportion of verifications in feedback messages within the feedback request and combination condition. In other words, although the raw growth of informative elaborations did not differ across conditions, it seems that support through a feedback request actually helped students to deliver more focused feedback messages, containing a higher proportion of informative elaborations per message. This, then, also helps to explain why assessors in these conditions used less verifications, which, contrary to elaborations, merely express a judgement without further explaining it.

The finding that a feedback request promoted more focused feedback messages, with a higher proportion of informative elaborations, provides evidence for Webb's (1991) argument that tools such as a feedback request can motivate and guide assessors to provide more "responsive" feedback. Likewise, it complements and confirms the conclusion by Gielen, Tops, et al. (2010), who, on finding that student groups using a feedback request made more progress in learning, proposed that: "A possible explanation for this is that assessors may provide more useful feedback when informed of the assessee's questions and doubts beforehand (p. 157)." Still, it appears that the positive effect of a feedback request on feedback content is limited to verifications and informative elaborations, as there were no significant differences between conditions with regard to the growth of suggestive elaborations.

Moving on to assessees' agreement with feedback, the absence of significant differences between conditions suggests that the use of a feedback request did not have an impact on the assessee's reception of feedback. Although this may sound contrary to expectations, this finding is actually in line with that of Gielen, Tops, et al. (2010). According to this previous study, implementation of a feedback request did not have an influence on the overall percentage of assessees who rated feedback as helpful, as compared to student groups without feedback request. The most likely explanation for this finding is that students' perceptions with regard to usefulness of feedback is to a large extent determined by their perceptions of the feedback source's expertise (Ilgen et al., 1979). To be more specific, several studies have found that students are generally inclined to regard their peers' feedback as less useful compared to that of the teacher (e.g. Kaufman & Schunn, 2011; Planas Lladó et al., 2014). Seeing that, although the feedback requests used in the present study helped to provide responsive feedback, the actual feedback source remained unchanged, this might explain why no significant influence of the feedback requests on agreement with feedback was found. Another possible explanation may be that, even though the work of Strijbos et al. (2010) suggests that usefulness of feedback and agreement with feedback are indistinguishable from one another, the measure used by the present study may not be the best approach to operationalizing students' reception of feedback. In relation to this, others have proposed different conceptualizations of students' reception of feedback, stressing the extent to which students feel personally addressed or are more inclined to apply the feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991).

Finally, the present study also has a number of limitations. A first important limitation is that, despite random assignment of students to conditions, the number of suggestive elaborations during cycle 1 was significantly lower in the combination condition than in the other conditions. It is still unclear what may have caused this difference in students' initial

approach to peer assessment, as none of the conditions had yet received additional support, and students followed the same course outside of the task. Even so, this unexpected finding did not affect the present study's outcomes, as growth curve models control for initial performance. A second limitation is that, although a check was carried out to make sure that students in the feedback request, content checklist, and combination condition, had used the additional support as instructed, the relative quality of students' feedback requests and content checklists was not analyzed. This was outside the scope of the present study, which mainly focuses on the outcomes of peer assessment (i.e. feedback content and agreement with feedback). Still, it would be interesting if future research could further analyze students' use of these tools, as this may reveal further differences within conditions. Finally, as was noted before, with regard to the absence of significant effects on students' agreement feedback, the question remains whether feedback requests have indeed no effect on students' reception of feedback, due to students' perceptions of their peers' expertise, or whether the present study's operationalization of this concept falls short of capturing these effects.

Notwithstanding these limitations, the present study does shed more light on the effects of a feedback request on the effectiveness of feedback in peer assessment. Its findings hold a number of implications for both practice and future research, which are discussed in the next section.

6. IMPLICATIONS

With regard to the application of peer assessment in educational practice, the present study shows that, although the use of a feedback request does not lead to a significant difference in the raw growth of informative elaborations in students' feedback messages, it does stimulate students to write more focused feedback messages, containing a higher proportion of informative elaborations. It thus appears that active involvement of the assessee in peer assessment, through the use of feedback requests, can help to improve feedback content.

Moving on to future research on peer assessment, the present study holds methodological as well as theoretical implications. On the methodological level, the findings underline the importance of taking message length into account when examining the effects of support for peer assessment on feedback content. As the findings of the present study demonstrate, this leads to a more nuanced picture of an intervention's impact. On the theoretical level, further research is necessary with regard to the conceptualization of students' reception of feedback. Although the present study focused on students' agreement with feedback, the concept may also be investigated in terms of students' perception of feedback quality, or their actual application of feedback. Future research that takes different

measures of students' reception of feedback into account, may therefore help to get a better theoretical understanding of this construct, as well as how it is impacted by the use of feedback requests.

7. REFERENCES

Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110*(1), 23–35.

Bangert-Drowns, R. L. ., Kulik, C.-C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238.

Butler, L. B. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *79*(4), 474–482.

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduates' writing peroformance during an online peer assessment activity. *The Internet and Higher Education*, *25*(1), 78–84.

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

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.

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

Gielen, M., & De Wever, B. (2015a). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on feedback quality and product improvement. *Computers & Education*, *88*(1), 370–386.

Gielen, M., & De Wever, B. (2015b). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*(1), 315–325.

Gielen, M., & De Wever, B. (2015c). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and feedback quality. *Journal of Computer Assisted Learning*, *31*(5), 435–449.

Gielen, S., Tops, L., Dochy, F., Onghena, P., & Smeets, S. (2010). A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal*, *36*(1), 143–162.

Harks, B., Rakoczt, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on

achievement, interest and self-evaluation: The role of the feedback's perceived usefulness. *Educational Psychology*, *34*(3), 269–290.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 88–112.

Hox, J. (1998). Multilevel Modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways: Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation e.V., University of Potsdam, March 12--14, 1997* (pp. 147–154). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371.

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*, *39*(3), 387–406.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, *41*(1), 33–39.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*(4), 344–348.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–143). Mahwah, NJ: Lawrence Erlbaum.

Neuendorf, K. A. (2002). *The content analysis guidebook*. London: Sage Publications.

Ng, E. M. W. (2016). Fostering pre-service teachers' self-regulated learning through self- and peer assessment of wiki projects. *Computers and Education*, *98*(1), 180–191.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback. *Studies in Higher Education*, *31*(2), 199–218.

Orr, J. O., Sacket, P. R., & Dubois, C. L. Z. (1991). Outlier detection and treatment in I/O psychology: A survey of research beliefs and an empirical illustration. *Personnel Psychology*, *44*, 473–486.

Osborne, J. (2004). The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1–8.

Ozogul, G., & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self and peer evaluators. *Educational Technology Research and Development*, *57*(3), 393–410.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144.

Planas Lladó, A., Soley, L. F., Fraguell Sansbelló, R. M., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., … Moreno, L. M. (2014). Student perceptions of peer assessment: an interdisciplinary study. *Assessment & Evaluation in Higher Education*, *39*(5), 592–610.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: A multiple case study. *Teaching in Higher Education*, *17*(4), 465–477.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, *35*(3), 277–289.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29–48.

Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*(4), 291–303.

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

Topping, K. J. (2009). Peer assessment. *Theory into Practice*, *48*(1), 20–27.

Walker, M. (2014). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education*, *40*(2), 232–247.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, *22*(5), 366–389.

APPENDIX 1

Table 7

Coding scheme for peer feedback content, and agreement with peer feedback

Category	Subcategory	Description	Example
Peer feedback content	Verification	Is the feedback sentence an evaluative statement expressed as a positive or negative remark with regard to the work?	*Your intro is well formulated!*
	Informative elaboration	Is the feedback sentence an informative statement, which gives more details about a previous evaluative statement, without providing suggestions for adapting the work?	*I like it because you used your own words.*
	Suggestive elaboration	Is the feedback sentence a suggestive statement, which gives more details about a previous evaluative statement, by providing directions for adapting the work?	*In your final version, you should integrate the limitations, which you can find on page 9.*
	Neutral	Is the feedback sentence a neutral statement, with none of the characteristics of a verification or elaboration?	*This week, I'm providing feedback on your second abstract.*
Peer feedback agreement	Agree	Does the assessee states that he agrees with the assessor?	*I believe your suggestion regarding lacking limitations is correct.*
	Partly agree	Does the assessee states that he only partly agrees with the assessor?	*I followed your advice on the limitation, but I believe that the number of participants should still be included.*
	Disagree	Does the assessee states that he totally does not agree with the assessor?	*I believe that my original problem statement was already clear enough.*