



Universiteit
Leiden
The Netherlands

Immediate and delayed effects of a modelling example on the application of principles of good feedback practice: A quasi-experimental study

Blankenstein, F.M. van; Trutescu, G.; Rijst, R.M. van der; Saab, N.

Citation

Blankenstein, F. M. van, Trutescu, G., Rijst, R. M. van der, & Saab, N. (2019). Immediate and delayed effects of a modelling example on the application of principles of good feedback practice: A quasi-experimental study. *Instructional Science*, 47(3), 299-318.
doi:10.1007/s11251-019-09482-5

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](#)
Downloaded from: <https://hdl.handle.net/1887/83343>

Note: To cite this publication please use the final published version (if applicable).



Immediate and delayed effects of a modeling example on the application of principles of good feedback practice: a quasi-experimental study

Floris M. van Blankenstein¹ · Gabriela-Olivia Truțescu¹ · Roeland M. van der Rijst¹ · Nadira Saab¹

Received: 31 January 2018 / Accepted: 11 February 2019 / Published online: 28 February 2019
© The Author(s) 2019

Abstract

The learning benefits of peer assessment and providing peer feedback have been widely reported. However, it is still not understood which learning activities most facilitate the acquisition of feedback skills. This study aimed to compare the effect of a modeling example, i.e., a model that demonstrated how to give feedback, on the acquisition of feedback skills. The participants were second-year bachelor students in pedagogical sciences ($N=111$). They were assigned randomly to a *practice condition*, in which they practiced giving feedback on oral presentations, or a *modeling example condition*, in which a teacher demonstrated how to give feedback on a good and a bad presentation. Students then gave feedback to a presenter in a video (direct feedback measure). One week later, they gave each other peer feedback on oral presentations (delayed feedback measure). On the direct feedback measure, students in the modeling example condition used assessment criteria more often in their feedback, and produced significantly more overall feedback, and significantly more positive and negative judgments than students in the practice condition. There was no significant difference in the amount of elaboration and feed-forward between the two conditions. On the delayed feedback measure, there were no significant differences between the two experimental conditions. The results suggest that, at least in the short term, a modeling example can stimulate the use of assessment criteria and judgments in feedback. The results and implications for future research and practice are discussed.

Keywords Feedback · Formative assessment · Peer assessment · Peer feedback · Modeling examples · Observational learning

✉ Floris M. van Blankenstein
f.m.van.blankenstein@iclon.leidenuniv.nl; f.m.van_blankenstein@lumc.nl

¹ Leiden University Graduate School of Teaching/Leiden University Medical Center, P.O. Box 905, 2300 AX Leiden, The Netherlands

Introduction

Peer assessment has been implemented and studied widely in higher education, both as a learning tool and an assessment tool (Topping 1998; Van den Berg et al. 2006a; Van Zundert et al. 2010). Several studies have shown positive effects of training and experience on students' peer feedback skills (Van Zundert et al. 2010). However, little is known about which specific instructional methods contribute to these learning effects. Modeling examples may be an effective instructional method when students have little experience with giving feedback. Therefore, we tested the effect of a modeling example on undergraduate students' feedback skills.

Peer assessment can be described as “a process whereby students evaluate, or are evaluated by their peers” (Van Zundert et al. 2010, p. 270). It is implemented for educational purposes in many different ways, ranging from peer marking or grading (grading the work of peers, Sadler and Good 2006) to more elaborate peer assessment with narrative feedback and dialogue between students (Panadero 2016). From a pedagogical perspective, peer assessment can be used to help students develop a range of professional and generic skills, such as feedback skills, collaboration skills, social skills, and communication skills (Sadler and Good 2006; Topping 1998; Van Popta et al. 2017). Therefore, peer assessment is implemented in universities to prepare students for future professions (Van der Pol et al. 2008).

The role of feedback in peer assessment

Giving and receiving feedback plays an important role in peer assessment. The role of feedback in education has been studied extensively (for literature reviews see, e.g., Evans 2013; Hattie and Timperley 2007; Shute 2008). A central question in these studies is what makes feedback effective.

Nicol and Macfarlane-Dick (2006) have derived seven principles of good feedback from literature about self-regulated learning and formative assessment. First, feedback should clarify clear goals and criteria, which relates to what Hattie and Timperley (2007) call '*feed-up*' (goals that need to be reached). Second, feedback should encourage self-assessment, which can be accomplished by self-monitoring performance in relation to the goals and criteria. Third, feedback should provide high-quality information, which Nicol and Macfarlane-Dick define as information that helps the receiver to reduce the discrepancy between what he or she intended to achieve and what were the results of these actions. In addition, several types of feedback have been suggested to specify 'high quality information'. Based on Kulhavy and Stock (1989), Shute (2008) proposed that effective feedback contains two types of information: verification and elaboration. A verification is a judgment of whether an answer is correct or not, whereas an elaboration is a more detailed explanation of why the answer is correct or incorrect. There seem to be various synonyms for the term 'elaboration' in research literature about feedback, such as 'justification' (see Gielen et al. 2010), or 'explanation' (see Van den Berg et al. 2006b). In addition, several authors argue that feedback should contain suggestions for improvement (e.g. Hattie and Timperley 2007; Gielen et al. 2010), which can also take the form of a thought-provoking question.

The fourth principle of good feedback that Nicol and Macfarlane-Dick suggest is that students should be able to give meaning to the feedback, for instance by engaging in dialogue with the provider of the feedback. Fifth, feedback should encourage motivation and self-confidence. Sixth, feedback should contain information that closes a gap between desired and current performance. Hattie and Timperley (2007) use a similar definition of good feedback and call information about current performance '*feed-back.*' Information on how to close the gap (analogous to suggestions for improvement) is what they call '*feed-forward.*' As a seventh and final principle of good feedback, Nicol and Macfarlane-Dick argue that teachers should use information that they gather from their feedback to improve their teaching.

Feedback literature does not only describe principles of effective feedback, but also aspects of less effective feedback. More specifically, Kluger and DeNisi (1996) discovered that feedback is less effective when it focuses on personal aspects of the receiver than when it focusses on task-related aspects. These findings formed the basis of their Feedback Intervention Theory (FIT), which postulates a hierarchy of feedback effectiveness. In this hierarchy, feedback aimed at the person is considered the least effective and feedback aimed at task details is regarded the most effective. However, in a more recent synthesis of feedback literature, Narciss (2008) concludes that feedback can be aimed at more than only task details. More specifically, she described that feedback can be aimed at five aspects: task constraints, conceptual knowledge about the task, knowledge on how to perform the task, detected errors or metacognitive strategies.

The learning benefits of peer assessment have been shown in several studies. These studies have focused on both the provider and the receiver of peer feedback. Several studies have shown learning benefits for the assessor (Cho and MacArthur 2011; Li et al. 2010; Lu and Law 2012; Lundstrom and Baker 2009; Van Popta et al. 2017). For instance, Cho and MacArthur (2011) conducted an experiment in which students either reviewed papers written by peers, merely read papers of peers, or read other papers. Students in the reviewing condition had to rate their peers' work and explain their marks with comments. Subsequently, these students wrote better papers than students in the other two conditions.

A possible explanation behind this learning benefit may be that providing feedback triggers *generative learning*, which occurs when learners actively try to give meaning to new information that they are learning (Fiorella and Mayer 2016). Wittrock (1989) proposed that generative learning occurs when learners actively construct relations between (a) different elements of new information that they learn, or (b) new information and their own prior knowledge. This requires motivation to invest the necessary cognitive effort, attention to relevant aspects of the learning task and active attempts to relate prior knowledge to new information. Fiorella and Mayer propose that teaching is one form of generative learning, because teaching requires selecting relevant information, explaining this information coherently to others and elaborating on those explanations by including one's own existing knowledge in the explanations. It can be argued that providing peer feedback triggers the same mechanisms, i.e. selecting (based on feed-up) relevant aspects of the peer's work that need improvement, explaining why and how these aspects need improvement and elaborating on those explanations by incorporating one's own knowledge in the feedback.

For the receiver, peer assessment has learning benefits when he or she is able to use the feedback to act upon (Topping 1998). Zhang et al. (2017) showed that peer feedback was related to a large degree with changes that students made in their papers. Van der Pol et al. (2008) related different aspects of peer feedback (among which judgments, elaborations and feed-forward) to students' choice to use peer feedback. In one study, they found a significant positive correlation between the amount of received feed-forward and the number

of revisions that the receiver made. In a second study, they also found a significant positive relationship between the amount of received judgments and the number of revisions made by the receiver. Further analyses revealed that the judgments in the second study were more objective and content-oriented than the judgments in the first study. Thus, it seems that students can use judgments as implicit feed-forward to improve their work.

Preparing students for giving feedback

Although the learning benefits of peer assessment thus seem to be clear, they may not occur automatically if students are not prepared for giving peer feedback. Research shows that the quality of peer feedback increases when students are trained and more experienced in peer assessment (Van Zundert et al. 2010). Various training methods for giving (peer) feedback have been developed and studied. In general, these training methods include a variety of instructional scaffolds, such as learning new theory about feedback (Voerman et al. 2015), studying worked examples of feedback (Alqassab et al. 2018; Sluijsmans et al. 2002), observing models which demonstrate how to give feedback (Van Steendam et al. 2010), following a structured format for giving feedback (Gielen and De Wever 2015), defining assessment criteria (Sluijsmans et al. 2002), following instructions to use received feedback actively (Gielen et al. 2010; Wichmann et al. 2018), and receiving expert feedback on feedback provided to others (Voerman et al. 2015).

Some studies have tested the combined effect of various instructional scaffolds in a holistic way. Several of these studies found positive effects of peer feedback training on students' feedback skills (Alqassab et al. 2018; Sluijsmans et al. 2002; Voerman et al. 2015). For instance, Sluijsmans et al. (2002) developed a training program in which undergraduate students defined assessment criteria, gave peer feedback, studied examples of expert feedback, and learned how to write a feedback report. The researchers found that students who followed the training, used assessment criteria more often in their peer feedback and gave more constructive feedback than students who did not follow the training. Voerman et al. (2015) designed a training program for teachers that was based on four principles: learning new theory about feedback, observing demonstrations of good feedback, practicing giving feedback, and receiving individual coaching on how to give feedback. Teachers who followed this training program gave more positive feedback and more specific feedback after the training than before the training.

Other studies have tested the effect of independent instructional scaffolds more specifically, using experimental or quasi-experimental designs (Gielen and De Wever 2015; Gielen et al. 2010; Peters et al. 2018; Van Steendam et al. 2010). For instance, Gielen and De Wever (2015) varied the level of structure in instructions for giving feedback (no structure, basic structure, or elaborate structure). Students who used an elaborate or basic structure provided significantly more elaborative feedback than students who used no structure. In addition, the elaborate instructions led to significantly more negative judgments and more feedback focused on assessment criteria.

Experimental and quasi-experimental studies are relatively uncommon in feedback research in higher education. Evans (2013) reviewed a large body of research literature on assessment feedback in higher education and revealed that the majority of the studies were case studies. Only 12.6% of these studies used an experimental design and 5.3% of the studies used quasi-experimental designs. Evans did not advocate the design of more experimental studies; she rather argued that small-scale case studies are helpful in developing

both theory and practice, and that randomized, experimental designs are difficult to perform in educational settings. In contrast, Van Zundert et al. (2010) concluded, after reviewing a large body of research to peer assessment, that more experimental and quasi-experimental studies are needed to test the effects of independent variables more specifically. The authors criticized the lack of experimental studies that clearly describe their methods and the outcomes of these methods. Therefore, there seems to be no consensus on which research methodology should be used to investigate peer assessment, or assessment and feedback in general. A mixture of quantitative, experimental studies and qualitative case studies might be preferable over choosing one specific research design.

Using modeling examples to prepare students for giving feedback

As described above, one way of scaffolding the acquisition of feedback skills is by demonstrating how to give feedback. In the educational research literature, such models have also been described as modeling examples. *Modeling examples* are adult or peer models that demonstrate how to perform a certain task (Van Gog and Rummel 2010). Modeling examples can either be *mastery models* that demonstrate how to perform a task without making any errors, or *copying models* that make errors and correct those errors while performing the task (Schunk and Hanson 1985; Schunk et al. 1987).

Like other types of examples, such as worked examples, modeling examples may help learners to acquire new skills more efficiently. Modeling examples differ from worked examples in the sense that the former demonstrate how to perform a task, whereas the latter are fully worked-out, written examples that explain step by step how a problem should be solved (McLaren et al. 2016; Van Gog and Rummel 2010). For novice learners, worked examples have proven to be a more effective learning strategy than unguided problem solving (McLaren et al. 2016; Nievelein et al. 2013). This positive learning effect has been attributed to a reduction in cognitive load on working memory resulting from worked examples. More specifically, worked examples cause a reduction in *extraneous cognitive load*, that is, cognitive load that is caused by an ineffective way of presenting learning tasks (Paas et al. 2003; van Merriënboer and Sweller 2005).

Modeling examples may also reduce cognitive load on working memory and facilitate the acquisition of, for instance, feedback skills (Van Steendam et al. 2010). There is some evidence that observing a modeling example that demonstrates how to give an oral presentation has a more positive impact on oral presentation skills than merely practicing oral presentation skills (De Grez et al. 2014). However, there is little evidence for the hypothesis that modeling examples also improve peer feedback skills.

In one empirical study by Van Steendam et al. (2010), students either observed a video of students modeling a text revision strategy or practiced this revision strategy in pairs. Subsequently, all students were instructed to apply the text revision strategy on a new letter that contained several errors. Based on previous work by Zimmermann and Kitsantas (2002, also see Zimmerman 2013), this phase was called the 'emulation phase.' Students were thus instructed to emulate the text revision strategy. Moreover, they did this either alone or in pairs. The feedback that students wrote down during this emulation phase was collected and scored by the researchers. After one week, students reviewed a new letter and again wrote down their feedback. This feedback was also collected and scored. The investigators expected that observing the student model, followed by emulating the text revision strategy in pairs, would cause the lowest cognitive load. Therefore, observing the

student model should lead to the highest quality of feedback, whereas practicing the revision strategy plus emulating the strategy individually would cause the highest cognitive load and therefore lead to the lowest quality of feedback. These hypotheses were partially confirmed. As expected, students who observed the student model detected more errors than students who practiced the revision strategy right away. However, the former group did not suggest more revisions than the latter group. As expected, observation followed by emulation in pairs seemed to be more effective than observation plus individual emulation. However, contrary to the expectations, practice seemed to be more effective when it was followed by individual emulation than when it was followed by emulation in pairs. Practice plus emulation in pairs actually led to the lowest quality of feedback. Therefore, the authors concluded that emulation in pairs was not effective unless it was preceded by observing the student model. Possibly, collaborative emulation after practice caused too much distraction or cognitive overload, whereas individual emulation after the practice phase led to more efficient time-on-task.

Research question

Considering the scarce empirical evidence for the effectiveness of modeling examples in peer feedback training, this study aimed at answering the following research question: what is the effect of observing a modeling example in feedback training on the acquisition of feedback skills among undergraduate students? Students gave peer feedback on oral presentations, a learning activity that was fairly new to them. More specifically, learning how to present in front of an audience and how to give feedback on presentations had not yet been taught in the curriculum. Therefore, observing a modeling example was expected to be more effective for acquiring feedback skills than practicing giving feedback. Students who observed a modeling example were expected to generate more feedback than students who practiced giving feedback, both on a direct and a delayed measure of their feedback skills.

Method

We chose a quasi-experimental research design to divide a group of students in two groups: one in which they observed a teacher who modeled how to give feedback and one in which they practiced themselves how to give feedback. Subsequently, all students provided feedback at two moments: directly after observing the modeling example or practicing how to give feedback, and after one week, when they gave each other peer feedback. On both moments, students wrote down their feedback, so it could be collected for further analysis. The feedback that was collected directly after observing or practicing, served as a direct measure of students' feedback skills. The peer feedback that was collected one week later served as a delayed measure to evaluate if the experimental treatment had a long-term effect on students' feedback skills.

Participants, context, and design

The study was carried out at a research university in the Netherlands and was approved by the Leiden University Graduate School of Teaching Research Ethics Committee. 121 second-year bachelor students in pedagogical sciences were asked to participate in the study; 118 students agreed ($M=20.25$ years, $SD=1.42$; 3 male, 108 female; 111 valid cases). In the second year of the bachelor program the students learned several academic skills, such as conducting research in teams, scientific writing, and presenting in front of an audience. In seminar groups of about 15 students, they developed their academic skills through various assignments. One of those assignments (not part of this study) was to give a presentation about one of the master's programs that the institute offered after the bachelor program. Students were free to choose which program they wanted to present. The goal of this assignment was to raise consciousness about potential prospective master's programs among students, and to let them practice their oral presentation and feedback skills. The assignment was also a preparation for a presentation that students gave at the end of the academic year. This presentation was one of the final assignments of a research project that students conducted throughout the year.

In agreement with the coordinator of the course, two lectures were developed to prepare students for their presentations. The learning goal of the first lecture was how to present in front of an audience. The learning goal of the second lecture was how to give feedback on oral presentations. Students could use the input of both lectures to prepare their presentation.

Materials

Feedback forms

Two basic formats were developed for the feedback form, one for training (*training form*) and one for data collection (*data collection form*). Inspired by earlier research (De Grez and Valcke 2013; De Grez et al. 2014), the training form contained two main categories: "presentation," i.e., feedback aimed at the content and structure of the presentation itself, and "presenter," i.e., feedback aimed at the and how the presentation was delivered. "Presentation" was specified with four assessment criteria: "introduction," "structure," "conclusion," and "PowerPoint." "Presenter" was also specified with four assessment criteria: "vocal delivery," "enthusiasm," "interaction with audience," and "body language." In addition, the training form contained three columns: one for feed-up, one for feed-back, and one for feed-forward. This structure was inspired by an earlier study by Gielen and De Wever (2015). The feed-up column contained the eight criteria plus short explanations of these criteria. These explanations had also been given during the lecture on public speaking before the feedback training. An example of such an explanation was "Mentions or repeats the take-home message." In the feed-back column, students could write down observations and feed-back based on those observations. In the feed-forward column, they could write suggestions for improvement. The training form can be found in the Appendix.

The data collection form contained two columns, one for feed-up and one for feed-back and feed-forward combined. Feed-up and feed-forward was combined in one

column to give students as much autonomy as possible, and to influence them as little as possible in addition to the experimental treatment, while writing their feedback. In the feed-up column, students could write down which criteria they would pay attention to. In the feed-back/feed-forward column, they could freely write down their feed-back and feed-forward, based on the self-generated criteria.

Videos

Two videos that were found online were used for the experimental treatment. These videos were created specifically for training purposes to identify errors and good presentation techniques. One video showed an actress delivering a good presentation and the other video showed the same actress delivering a bad presentation. The video of the good presentation lasted 6:59 min (<https://youtu.be/Dv1HuXTWE5c>) and the video of the bad presentation lasted 4:50 min (<https://youtu.be/ATfY8dvbuFg>). The good presentation was of good quality according to the assessment criteria presented in the feedback lecture, e.g., the presenter made good eye contact. In the bad presentation, she did not meet the assessment criteria, e.g., she did not make eye contact. A third video that was found online was used for the data collection. This video showed a professor who gave a lecture about behavioral change that lasted 15:28 min (<https://youtu.be/jI2Mmh3v3Fc>).

Feedback lecture

The experimental intervention took place during a lecture on giving feedback. The first author (FvB) wrote a script for and designed the PowerPoint slides of this lecture. The script was as follows: The teacher started by explaining the context and goal of the lecture. Students were reminded about the presentations they would give to each other in one week's time and the lecture on presentation skills they had attended earlier. The goal of the lecture was that students should be able to explain how they could give good feedback on oral presentations. Then, the teacher explained some principles of feedback based on the definitions of feed-up, feed-back, and feed-forward provided by Hattie and Timperley (2007). Feed-up was explained as goals or criteria that form the basis of feed-back ("Which goals or criteria do I need to meet?"). The teacher then showed the assessment criteria for oral presentation skills on a PowerPoint slide. These assessment criteria were the same as those on the training form: introduction, structure, conclusion, PowerPoint, interaction with audience, enthusiasm, vocal delivery, and body language. Feed-back was explained as information about performance with regard to feed-up ("How am I performing with regard to the goals or criteria?"). The teacher explained that feed-back should contain references to specific observations and add (subjective) interpretations of those observations. Feed-forward was explained as steps to be taken to fulfil the goals or criteria (even) better. The teacher instructed the students to phrase feedback in a positive way, so not only to mention what could be improved, but also to mention what went well. Then, the teacher explained that he would now show videos of a good and a bad presentation. This is the point at which the experimental intervention took place.

In one version of the script (henceforth called the *modeling example condition*), the teacher explained that he would now show videos of a good and a bad presentation. He also explained that he would give feedback, first on the presentation and then on the presenter. Students observed the video of the bad presentation first, followed by the good

presentation. After each video, the teacher presented feed-back and feed-forward, first on the presentation and then on the presenter. This feed-back and feed-forward was spelled out in the script for the lecture, and was based on the assessment criteria on the training form. Students were not instructed to give feed-back and feed-forward at this moment.

In another version of the script (called the *practice condition*), the teacher again explained that he would show videos of a good and a bad presentation. However, this time the teacher instructed students to note observations during the video, and to write down feed-back and feed-forward after the presentation. He also instructed students to use the assessment criteria on the training form as feed-up. Students observed the videos and wrote down their feedback after each video, first on the presentation and then on the presenter.

The remainder of the script was the same for both versions of the lecture. The teacher instructed students to observe the third video and write down feed-up (criteria they would pay attention to), feed-back and feed-forward on the data collection form. Only the first 10 min and 24 s were to be shown to the students. Students were instructed to write down their feedback while watching the video and received up to 5 more minutes after the video to complete their feedback. To prevent confounding effects additional to the experimental treatment, they were left free to choose what feed-up, feed-back, and feed-forward they wrote down. The data collection forms were collected and served as the direct feedback measure.

The script indicated how much time the teacher should spend on each of the sections of the lecture and which PowerPoint slide belonged to which section. Apart from small differences in the PowerPoint slides designed for the experimental intervention, the PowerPoint slides of both lectures were similar.

Procedure

Table 1 shows an overview of the research procedure. During a seminar group meeting, all students were briefed about the study via an informed consent letter and could sign this letter voluntarily if they wanted to participate in the study. Data from students who did not sign the letter were excluded from the analyses. Students were also asked to report the grade point average (GPA) of their secondary school final exams (in the Netherlands, secondary school is from age 12 to 18). One week later, they attended a lecture of 1.5 h on presentation skills. In this lecture, they received practical tips for planning the introduction, structure, and conclusion of a presentation. They also received tips for vocal delivery, body language (e.g., mimics and hand gestures), how to interact with the audience by asking questions, and how to use PowerPoint. The teacher showed several videos of presentations

Table 1 Procedure of the study

Week 1	Week 3	Week 5 (T1)	Week 6 (T2)
Informed consent	Lecture: oral presentations	Lecture: giving feedback a. Modeling examples condition b. Practice condition	Peer feedback

T1 = time of first data gathering

T2 = time of second data gathering

to illustrate the tips and students discussed in pairs how they could draw the attention of the audience at the beginning of a presentation.

Three weeks after this lecture, students attended the lecture on giving feedback. This lecture took 1.5 h. Half of the students ($N=53$) attended a version of the lecture in which they observed examples of how to give feedback (modeling example condition). At the same time, the other half of the students ($N=55$) attended a version of the lecture in a different room in which they practiced giving feed-back and feed-forward (practice condition). The lectures were given by two external teachers who were not involved in the rest of the course. The modeling example lecture was given by the first author of this study (FvB). The practice lecture was given by the third author of this study (RvdR). Before the lectures, both teachers discussed the script in detail, including the quality of the feedback that should be given by the first teacher. They then prepared the lecture using the script and the PowerPoint slides. Students in the modeling example condition received only the data collection form. Students in the practice condition received the training form and the data collection form.

One week after the lecture, all students gave their presentations in their own seminar groups. They gave the presentations in subgroups of two students and each presentation lasted 5 min. The seminar teachers were informed about the study and each of them received the data collection forms to hand out to their students. The teachers were instructed to let students fill in one form after each presentation. They were also instructed to ask students to write down criteria (feed-up) on the form before each presentation. During the presentations, students could write down feed-back and feed-forward for the presenters. After each presentation, students had 5 min to finish writing and present their feed-back and feed-forward to the presenters. The teachers collected the data collection forms and handed these into the coordinator of the course, who forwarded the forms to the first author. These forms served as the delayed feedback measure.

Coding the feedback

The written feedback on the data collection form was transcribed verbatim to a digital text format. Based on the method developed by Mayer (1985) and used by Van Blankenstein et al. (2013), the second author parsed the sentences of this feedback into *idea units*. An idea unit “expresses one action, event or a state and corresponds to a single verb clause” (Mayer 1985, p. 71). Each sentence can include one or multiple idea units. For example, in a sentence that contains one verb, there is one unit: “She speaks calmly.” In sentences with two (or more) verbs, there will be two (or more) units: “Questions are understood correctly” (first unit) “but cannot be answered correctly” (second unit).

Based on a pilot study, the second author wrote a codebook with instructions for coding the data. She used this codebook to code part of the data from the pilot study, adjusted the codes and instructions based on these new coding experiences, and discussed these adjustments with the first author. This process was repeated several times to fine-tune the codes and coding instructions. This resulted in a final codebook with the following coding structure: “assessment criteria,” “judgments (positive or negative),” “feed-forward,” “elaborations,” and “other.” “Other” referred to unclear or ambiguous statements and double remarks. These idea units were excluded from the analysis.

The coding structure is outlined in further detail in Table 2. Idea units were coded as a positive judgment if they suggested a positive aspect of the presentation. Signaling words

Table 2 Coding scheme of the qualitative feedback data

Codes and subcodes	Explanation	Examples
Assessment criteria	Idea units that refer to assessment criteria for oral presentation skills.	“Enthusiastic;” “no distracting gestures or movements;” “no coherent story;” “little eye contact;” “not too much text on the slide;” “figure can be explained better;” “clear voice”
Vocal delivery	These codes were based on De Grez and Valcke (2013) and the criteria that students learned during the feedback lecture	
Enthusiasm		
Interaction with audience (including eye-contact)		
Supporting media (e.g. PowerPoint)		
Body language		
Introduction		
Structure		
Conclusion		
Illustrations/examples		
Judgments		
Positive judgments	Idea units that refer to positive or negative judgments	Positive judgments: “clear,” “logical,” “self-assured,” “enthusiastic,” “knew a lot about the subject,” “not too much text on the slide,” “no distracting gestures or movements”
Negative judgments		Negative judgments: “hard to follow;” “minimal,” “slow,” “not clear;” “no coherent story;” “without explanation;” “fittle eye contact;” “techniques were not discussed”
Feed-forward	Idea units that refer to suggestions for improvement. Suggestions could be to stop or reducing a certain behavior, start or increasing the frequency of a certain behavior, or continue to perform a certain behavior. Suggestions were expressed in a direct way or in a more indirect way, e.g., by asking a question	“Show less text on the slides;” “try to take more pauses;” “keep doing this;” “try to start with a clearer introduction;” “perhaps a figure?”
Elaborations	Idea units that express a reason or consequence of another idea unit. Conditional statements (“If ... then...”) were also coded as elaborations.	“It was easy to follow the presentation, <i>because you spoke clearly.</i> ” “The presentation was easy to follow <i>if you already knew something about the subject.</i> ” (Parts in <i>italics</i> were coded as elaboration)

for positive judgments were, e.g., “clear,” “energetic,” or “not too much text on the slide.” Idea units were coded as a negative judgment if they indicated a negative aspect of the presentation. Examples of signaling words for negative judgments were: “little eye contact,” “not spoken fluently,” or “there is too much...” Idea units that expressed suggestions to perform a new behavior, continue a certain behavior, or stop or reduce a certain behavior were coded as feed-forward. Signaling words for feed-forward were, for instance: “Perhaps you can...,” “You could have...,” “Could you add...?” etc. Idea units were coded as “elaboration” if they expressed a cause for or a consequence of a judgment or feed-forward. Signaling words for elaborations were “Because...,” “caused by...,” “If ... then...,” and “as a consequence...”

Idea units could be assigned multiple codes. For instance, the statement “The figure could be explained better” was coded as “illustration” (one of the assessment criteria) and “feed-forward.” The statement “It was easy to follow the PowerPoint presentation” was coded as “supporting media” and “positive judgment.” Codes that referred to assessment criteria were coded once per student. Codes that referred to positive and negative judgments, elaborations, and feed-forward were coded as many times as they were given by each student. This created both categorical data (assessment criteria: mentioned or not) and continuous data (judgments, elaborations, and feed-forward: mentioned n times).

The second author used the final codebook to code the data. To check the inter-rater reliability, the first author coded a random sample of seven students (6.31%). Consistent with the coding procedure of the second author, codes that referred to assessment criteria were coded only once per student. The other codes were coded as many times as they occurred. Cohen’s kappa (κ) was used to estimate the inter-rater reliability of the categorical data (assessment criteria) and intra-class correlation (ICC) coefficients were used to estimate the inter-rater reliability of the continuous data (judgments, elaborations, and feed-forward), resulting in a κ value of 0.610 and an ICC of 0.964 ($p < 0.0001$).

Quantitative analyses

An independent t test with high school GPA as the dependent variable and lecture (modeling examples/examples plus practice) as the independent variable was conducted to estimate whether or not there were a priori differences in academic performance between the two groups. The t -test showed no significant difference for high school GPA between students in the modeling examples and practice condition, ($t(1, 89) = 0.44, p = 0.66$).

As explained in the procedure, feedback data were collected at two time points, i.e., a direct and a delayed feedback measure. We henceforth call these two measures T1 and T2. The context at T1 differed from that at T2. At T1, students gave feedback to a person in a video. At T2, they gave feedback to each other on real-life presentations. Therefore, T2 can be seen as a transfer test that measured the extent to which the effect of the experimental treatment transferred to a real-life situation one week later.

Because the contexts in which T1 and T2 took place differed, the data were not analyzed as repeated measures. Kolmogorov–Smirnov tests of normality were significant for all feedback types at T1 (assessment criteria: $D(80) = 0.112, p = 0.015$; positive judgments: $D(80) = 0.130, p = 0.002$; negative judgments: $D(80) = 0.205, p < 0.0001$; feed-forward: $D(80) = 0.236, p < 0.0001$, elaborations: $D(80) = 0.219, p < 0.0001$) and T2 (assessment criteria: $D(80) = 0.119, p < 0.0001$; positive judgments: $D(80) = 0.114, p < 0.0001$; negative judgments: $D(80) = 0.209, p < 0.0001$; feed-forward: $D(80) = 0.260,$

$p < 0.0001$; elaborations: $D(80) = 0.226$, $p < 0.0001$). Therefore, non-parametric tests were used to analyze the differences between the modeling example condition and the practice condition. Mann–Whitney tests were performed on assessment criteria and total amount of feedback (summed judgments, feed-forward, and elaborations) to evaluate whether students generated more feedback after the modeling example condition. Assessment criteria and total feedback were added separately to this analysis because they overlapped. To analyze differences between the two conditions on specific types of feedback, Mann–Whitney tests were performed on positive judgments, negative judgments, feed-forward, and elaborations. Exact tests (two-tailed) were used to analyze significant differences.

Results

Confirming our expectations, at T1 students in the modeling example condition ($N = 47$) generated significantly more feedback overall than students in the practice condition ($N = 49$, $U = 656.00$, $z = 3.638$, $p < 0.0001$, $Mdn = 13.00$ vs. $Mdn = 9.00$). In addition, the modeling example condition generated significantly more feedback referring to assessment criteria than the practice condition ($U = 413.00$, $z = 5.464$, $p < 0.0001$, $Mdn = 7.00$ vs. $Mdn = 5.00$). Focusing on the different types of feedback, at T1 the modeling example condition generated significantly more positive judgments ($U = 584.00$, $z = 4.187$, $p < 0.0001$, $Mdn = 7.00$ vs. $Mdn = 4.00$) and negative judgments ($U = 673.00$, $z = 3.591$, $p < 0.0001$, $Mdn = 2.00$ vs. $Mdn = 1.00$) than the practice condition. The amount of feed-forward ($U = 1079.00$, $z = 0.551$, $p = 0.585$) and number of elaborations ($U = 977.00$, $z = 1.316$, $p = 0.190$) did not differ significantly between the two conditions.

At T2, there were no significant differences between the modeling example condition ($N = 47$) and the practice condition ($N = 45$). This accounted for all types of feedback:

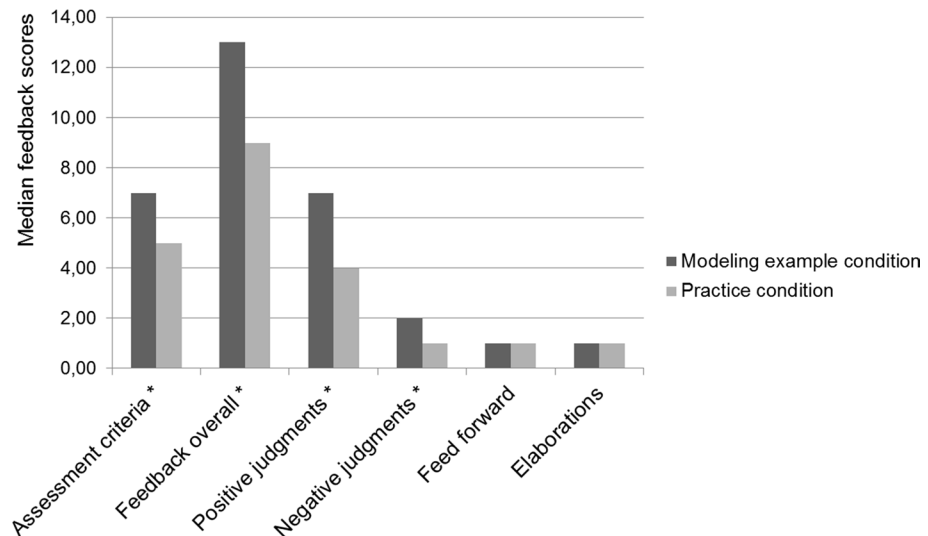


Fig. 1 Median feedback scores at immediate measure (T1). * $p < 0.0001$

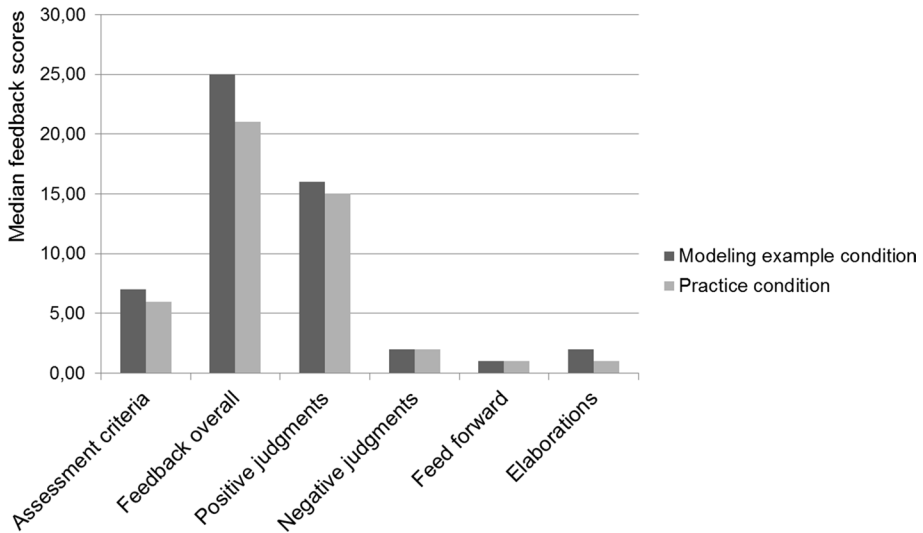


Fig. 2 Median feedback scores at delayed measure (T2)

assessment criteria ($U=827.00$, $z=1.832$, $p=0.067$), feedback overall ($U=810.00$, $z=1.931$, $p=0.053$), positive judgments ($U=876.50$, $z=1.417$, $p=0.158$), negative judgments ($U=956.00$, $z=0.804$, $p=0.424$), feed-forward ($U=851.00$, $z=1.658$, $p=0.098$), and elaborations ($U=1008.00$, $z=0.394$, $p=0.698$). Figures 1 and 2 show the median scores of the overall feedback and types of feedback at T1 and T2.

Discussion

This study examined the effect of observing modeling examples on the acquisition of feedback skills among undergraduate students. The research hypothesis was that students would provide more feedback after observing a modeling example than after practicing how to give feedback. We expected this outcome both on a direct and a delayed feedback measure. The hypothesis was confirmed only and partly on the direct feedback measure. Directly after the intervention, students generated more feedback in the modeling example condition than in the practice condition. Furthermore, students in the modeling example condition referred to the assessment criteria more often than students in the practice condition, a finding that is consistent with previous empirical studies in which students were trained to give feedback (Gielen and De Wever 2015; Sluijsmans et al. 2002). In the modeling example condition, students also expressed significantly more judgments, both positive and negative, than in the practice condition. However, students did not provide more elaborations or feed-forward in the modeling example condition. On the delayed feedback measure, there were no significant differences between the two experimental conditions, neither on the total amount of feedback, nor on any type of feedback.

These findings raise two questions. First, why were significant differences found only for references to assessment criteria and judgments, but not for elaborations and feed-forward on the direct feedback measure (T1)? Second, why were no significant differences found for references to assessment criteria and judgments on the delayed feedback measure

(T2)? To answer the first question, it is possible that students emulated the use of assessment criteria after observing the modeling example, who showed how to use these criteria to give feedback. Zimmermann (2013) has theorized that self-regulated learning can occur at four levels: observation, emulation, self-control and self-regulation. At the level of observation, students observe a model who demonstrates how to perform a certain task. At the level of emulation, students emulate the model's general behavior. Because the teacher in the modeling example condition explained the use of the criteria with examples of feedback, it seems plausible that emulation was the working mechanism behind the increased use of criteria. Furthermore, the teacher explicitly phrased judgments based on these criteria, which may also explain why students in the modeling example condition phrased more judgments than students in the practice condition. In the practice condition, the teacher showed the assessment criteria without illustrating how to use the criteria to provide feedback. Instead, students used the assessment criteria immediately to give feedback, making use of the training form (see Appendix).

Emulating elaborations and feed-forward may be more difficult, because this requires more creativity than emulating the use of assessment criteria and making judgments based on those criteria. This may explain why modeling had no effect on elaborations and feed-forward. Alternatively, looking at the low frequencies of elaborations and feed-forward, a floor effect also seems plausible. Students probably had too little time during the lecture to elaborate on their judgments and come up with suggestions for improvement.

Why did the intervention not yield significant long-term effects on the use of assessment criteria and judgments? The most probable reason for this is that the experimental intervention was not powerful enough to yield a long-term effect. In addition, the contexts in which feedback was given differed between T1 and T2. At T1, students gave feedback in a simulated setting, whereas at T2, they gave peer feedback in a real educational setting. The real setting at T2 may have caused reciprocity effects, which are known to occur in peer assessment (Panadero et al. 2013). Reciprocity effects can take the form of, for instance, *collusive marking* (giving high marks to group members) or *friendship marking* (giving high marks to friends, Carvalho 2013; Dochy et al. 1999; Pond et al. 1995). Students generated more positive judgments at T2 than at T1, which shows that they assessed their own peers less critically than the presenter in the video at T1, whom they did not know personally. A certain form of friendship bias may have occurred, i.e., students may have been more positive in their judgments because they knew each other and were reticent about giving each other negative feedback. Friendship bias can be reduced by using clear assessment criteria and making peer feedback anonymous. Anonymity may help assessors to focus more on the task rather than the person, thereby increasing the objectivity of the peer assessment. However, anonymity may also lead to inconsiderate and perhaps even disrespectful feedback, because the provider and the receiver have no personal connection. Therefore, anonymity should be considered carefully when designing peer feedback interventions (Panadero 2016).

The lack of an effect of the modelling example on elaborations is a pity, because research shows that elaborating on judgments is a relevant learning activity for the provider of feedback (Van Popta et al. 2017). Elaborating on judgments can be considered a type of generative learning because it challenges students to explain their thoughts in a coherent manner to others (Fiorella and Mayer 2016). This can be accomplished, for instance, by promoting interaction between the provider and the receiver of the feedback. Narciss (2017) argues that the process of feedback should always be interactive rather than unidirectional from the provider to the receiver. She also suggests several strategies to promote interaction in feedback. Promoting interaction between the provider and receiver of

feedback can promote the understanding of the received feedback and enable receivers to generate personal goals for their future learning (Nicol and Macfarlane-Dick 2006).

The lack of an effect of the modeling example on the amount of feed-forward raises the question to what extent the feedback was useful for those who received the feedback at T2. Topping (1998) argued that feedback is effective only when it can be acted upon by the receiver. Zhang et al. (2017) showed that peer comments are associated to a large degree with revisions that students made in draft versions of their papers, a finding that underlines the importance of using peer feedback as feed-forward to make improvements. However, receivers of feedback may also be able to derive feed-forward from positive or negative judgments. Van der Pol et al. (2008) found in one study that the more judgments and feed-forward students gave in their peer feedback, the more the receivers of the peer feedback changed their own work. This suggests that not only feed-forward is used to make revisions, but judgments as well. In other words, judgments may be perceived as implicit feed-forward. In the present study, some judgments may have been interpreted as implicit feed-forward. For instance, a negative judgment that there was “little eye contact,” may have been perceived as “I should make more eye contact.” Phrasing the feed-forward explicitly (e.g., “make more eye contact”) may even have been redundant in this case.

Despite the efforts made to conduct this study in a controlled and authentic setting, there are some limitations. First, there was no baseline measure for peer feedback, so the quality of the peer feedback after the intervention could not be compared to the quality of feedback before the intervention. There were, however, no indications that students’ ability levels differed between the two conditions, as self-reported GPA of secondary school final examinations did not differ significantly between the two experimental conditions. Although self-reported GPA may be a less valid measure of actual GPA for students with low GPAs (Kuncel et al. 2005), we do believe that self-reported GPA was fairly accurate, because it can be calculated fairly easily as the average grade on the final examination, which is a standardized exam in the Netherlands.

Second, the contexts of T1 and T2 differed. As explained previously, this may have affected the results of the study. Future studies should create more similar contexts over the course of the study. This can be established by using a setting in which students only provide each other with peer feedback, instead of feedback in a simulated setting followed by peer feedback in an authentic setting. Such a method would require a more thorough redesign of a course, something that was not possible in this study.

In sum, the results of this study indicate that modeling examples can be useful to increase the use of assessment criteria in feedback, although this effect may not persist when the use of assessment criteria is not reinforced. Modeling examples may also be useful to promote critical thinking when students provide feedback in a simulated setting. However, this effect may disappear when students provide peer feedback in an authentic setting, perhaps because of friendship bias. The effectiveness of modeling examples may be increased by stimulating elaboration on judgments even further. This can be established by promoting interaction between the provider and receiver of the feedback. Future research should therefore focus more on peer interaction, for instance by adding a discussion between the provider and the receiver of the feedback after both have observed a modeling example and given each other peer feedback.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Training form

Goals (feed-up)	Feed-back and feed-forward	
A. Presentation	Feed-back	Feed-forward
Introduction		
Presents the context		
Presents an overview of the content		
Checks the audience's prior knowledge on the subject		
Attracts the audience's attention		
Triggers the audience's curiosity		
Structure		
Breaks up the story into different parts		
Indicates transitions		
Summarizes intermediately		
Gives clear examples		
Makes comparisons		
Repeats difficult and/or important points		
Conclusion		
Summarizes the story		
Repeats or presents key messages		
PowerPoint		
Not too much text on the slides		
Text is clearly readable		
Not too many bullet points		
Illustrations match the story		
Use of a clearly readable text font		
Goals (feed-up)	Feed-back and feed-forward	
B. Speaker	Feed-back	Feed-forward
Interaction with audience		
Engages with the audience		
Asks questions of the audience		
Good eye contact with the audience		
Looks across the whole room		
Enthusiasm		
Tells the story with enthusiasm		
Makes the audience enthusiastic about the subject		
Use of voice		
Speaks clearly, articulates well		
Speaks dynamically: alternates between high and low, soft and loud, slow and rapid		
Stresses key aspects		
Raises curiosity by using short pauses		
Body language		
Takes a stable position		
Renders an open attitude		
Uses gestures that support the story		
Shows vivid mimicry		

References

- Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education*, 33(1), 11–30. <https://doi.org/10.1007/s10212-017-0342-0>.
- Carvalho, A. (2013). Students' perceptions of fairness in peer assessment: Evidence from a problem-based learning course. *Teaching in Higher Education*, 18(5), 491–505. <https://doi.org/10.1080/13562517.2012.753051>.
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–84. <https://doi.org/10.1037/a0021950>.
- De Grez, L., & Valcke, M. (2013). Student response system and how to make engineering students learn oral presentation skills. *International Journal of Engineering Education*, 29(4), 940–947.
- De Grez, L., Valcke, M., & Roozen, I. (2014). The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. *Higher Education Research & Development*, 33(2), 256–271. <https://doi.org/10.1080/07294360.2013.832155>.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350. <https://doi.org/10.1080/03075079912331379935>.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120. <https://doi.org/10.3102/0034654312474350>.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>.
- Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315–325. <https://doi.org/10.1016/j.chb.2015.06.019>.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304–315. <https://doi.org/10.1016/j.learninstruc.2009.08.007>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certainty. *Educational Psychology Review*, 1(4), 279–308. <https://doi.org/10.1007/bf01320096>.
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75(1), 63–82.
- Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525–536. <https://doi.org/10.1111/j.1467-8535.2009.00968.x>
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275. <https://doi.org/10.1007/s11251-011-9177-2>.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30–43. <https://doi.org/10.1016/j.jslw.2008.06.002>.
- Mayer, R. E. (1985). Structural analysis of science prose: Can we increase problem-solving performance? In B. K. Britton & J. B. Black (Eds.), *Understanding expository text: A theoretical and practical handbook for analyzing explanatory text* (pp. 65–87). Hillsdale, NJ: Erlbaum.
- McLaren, B. M., van Gog, T., Gao, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior*, 55(Part A), 87–99. <https://doi.org/10.1016/j.chb.2015.08.038>.
- 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. 125–144). New Jersey: Lawrence Erlbaum Associates.
- Narciss, S. (2017). Conditions and effects of feedback viewed through the lens of the Interactive Tutoring Feedback Model. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education*. Singapore: Springer.

- 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*(2), 199–218. <https://doi.org/10.1080/03075070600572090>.
- Nievelstein, F., van Gog, T., van Dijk, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*(2), 118–125. <https://doi.org/10.1016/j.cedpsych.2012.12.004>.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1.
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Brown (Eds.), *Handbook of human and social conditions in assessment* (1st ed., pp. 247–266). New York: Routledge.
- Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195–203. <https://doi.org/10.1016/j.stueduc.2013.10.005>.
- Peters, O., Körndle, H., & Narciss, S. (2018). Effects of a formative assessment script on how vocational students generate formative feedback to a peer's or their own performance. *European Journal of Psychology of Education, 33*(1), 117–143. <https://doi.org/10.1007/s10212-017-0344-y>.
- Pond, K., Ul-Haq, R., & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education & Training International, 32*(4), 314–323. <https://doi.org/10.1080/1355800950320403>.
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1.
- Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*(3), 313–322. <https://doi.org/10.1037/0022-0663.77.3.313>.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*(1), 54–61. <https://doi.org/10.1037/0022-0663.79.1.54>.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. <https://doi.org/10.3102/0034654307313795>.
- Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*(5), 443–454. <https://doi.org/10.1080/0260293022000009311>.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. <https://doi.org/10.3102/00346543068003249>.
- Van Blankenstein, F. M., Dolmans, D. H. J. M., Van der Vleuten, C. P. M., & Schmidt, H. G. (2013). Relevant prior knowledge moderates the effect of elaboration during small group discussion on academic achievement. *Instructional Science, 41*(4), 729–744. <https://doi.org/10.1007/s11251-012-9252-3>.
- Van den Berg, I., Admiraal, W., & Pilot, A. (2006a). Peer assessment in university teaching: Evaluating seven course designs. *Assessment & Evaluation in Higher Education, 31*(1), 19–36. <https://doi.org/10.1080/02602930500262346>.
- Van den Berg, I., Admiraal, W., & Pilot, A. (2006b). Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education, 11*(2), 135–147. <https://doi.org/10.1080/103562510500527685>.
- Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804–1817. <https://doi.org/10.1016/j.compedu.2008.06.001>.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>.
- Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24–34. <https://doi.org/10.1016/j.edurev.2016.10.003>.
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*(4), 316–327. <https://doi.org/10.1016/j.learninstruc.2009.08.009>.

- Van Zundert, M. J., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270–279. <https://doi.org/10.1016/j.learninstruc.2009.08.004>.
- Voerman, L., Meijer, P. C., Korthagen, F., & Simons, R. J. (2015). Promoting effective teacher-feedback: From theory to practice through a multiple component trajectory for professional development. *Teachers and Teaching, 21*(8), 990–1009. <https://doi.org/10.1080/13540602.2015.1005868>.
- Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education, 33*(1), 165–184. <https://doi.org/10.1007/s10212-017-0348-7>.
- Witrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2.
- Zhang, F., Schunn, C. D., & Baikadi, A. (2017). Charting the routes to revision: An interplay of writing goals, peer comments, and self-reflections from peer reviews. *Instructional Science, 45*(5), 679–707. <https://doi.org/10.1007/s11251-017-9420-6>.
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147.
- Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660–668.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.