Teaching and Teacher Education 127 (2023) 104099



Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate



Research paper

Digital video-based peer feedback training: The effect of expert feedback on pre-service teachers' peer feedback beliefs and peer feedback quality



Christopher Neil Prilop [a, *], Kira Elena Weber [b]

[a] Aarhus University, Danish School of Education – Educational Theory and Curriculum Studies, Jens Chr. Skous Vej 4, 8000, Aarhus, Denmark
[b] IPN – Leibniz Institute for Science and Mathematics Education, Olshausenstr. 62, 24118, Kiel, Germany

ARTICLE INFO

Article history:
Received 8 November 2022
Received in revised form
5 March 2023
Accepted 7 March 2023
Available online 23 March 2023

Keywords:
Pre-service teacher education
Peer feedback
Beliefs
Intervention study
Content analysis
Professional vision

ABSTRACT

We investigate the effects of peer feedback training (enhanced by expert feedback) on pre-service teachers' feedback quality and beliefs. Intervention group 1 (IG1) analysed classroom videos before the feedback training, while intervention group 2 (IG2) additionally received expert feedback on their video analyses. IG1 and IG2 were compared to a control group (CG). Results show pre-service teachers in IG1 and IG2 improved their peer feedback quality and beliefs in comparison to the CG. Expert feedback (IG2) lead to greater increases of peer feedback quality than without (IG1). This study highlights the role of expert feedback when training peer feedback.

© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The ability to provide and seek peer feedback has been emphasised as a crucial part of teachers' professional role that enables them to continuously and collaboratively develop their expertise (Hammerness et al., 2005). Peer feedback concerning classroom practice may be outlined in three steps: first, a peer's teaching performance is observed; second, the peer is provided with information about their classroom practice; third, the peer reflects on the information provided, with the goal of improving future performance (Prilop, Weber, & Kleinknecht, 2020; Tripp & Rich, 2012). However, studies in other domains have shown that low-quality feedback decreased participants' performance (Kluger & DeNisi, 1996). Hence, to ensure the development of pre-service teachers' expertise and enable pre-service teachers to provide effective feedback, they need to acquire feedback expertise during teacher education.

Feedback expertise can be described as the ability to provide high-quality feedback, and having a strong belief in the power of receiving and providing feedback (Alqassab, Strijbos, & Ufer, 2018; Prilop, Weber, & Kleinknecht, 2019). Studies (e.g., Gielen & De Wever, 2015; Prins, Sluijsmans, & Kirschner, 2006) show that high-quality feedback takes into account a set of criteria. Feedback needs to pertain to assessment criteria, be specific, include suggestions, pose reflective questions, include positive and negative comments, and be formulated in the first person. Also, feedback-related beliefs significantly impact feedback efficacy, because individuals' beliefs affect their decision-making and actions (Bandura, 1986). The belief in the usefulness of feedback, one's ability to provide feedback or the reasoning potential of feedback, determines how peer feedback messages are processed or composed (Alqassab et al., 2018).

Prior studies in other domains have shown that feedback quality and feedback beliefs can be fostered by providing multiple practice occassions and using feedback prompts (e.g., Cheng & Warren, 1997; Gielen & De Wever, 2015; Rotsaert, Panadero, Schellens, & Raes, 2018; Sluijsmans, Brand-Gruwel, Van Merriënboer, & Martens, 2004). However, the complexity of the feedback focus seems to influence feedback beliefs. During peer feedback training that focused on a complex task, pre-service teachers' feedback beliefs

* Corresponding author.
E-mail addresses: cnp@edu.au.dk (C.N. Prilop), k.weber@leibniz-ipn.de (K.E. Weber).

https://doi.org/10.1016/j.tate.2023.104099
0742-051X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

decreased (Alqassab et al., 2018), whereas interventions that focused on less complex tasks strengthened students' feedback beliefs (Cheng & Warren, 1997; Sluijsmans et al., 2004).

With regard to peer feedback quality, expert feedback may play an important modelling role. Van Blanckenstein et al. (2019) showed that higher education students that were presented with an expert model of feedback improved the quality of their feedback. A study by Prilop, Weber, & Kleinknecht, (2020) indicates that expert feedback may have a twofold effect on pre-service teachers' feedback expertise. On the one hand, observing an expert providing feedback may lead to pre-service teachers' learning in situ (Bandura, 1986); on the other hand, expert feedback may also foster pre-service teachers' professional vision. A recent study (Prilop et al., in press) indicates that teachers' professional vision, defined as the 'ability to notice and interpret significant interactions in the classroom' (Sherin, Russ, Sherin, & Colestock, 2008, p. 28), is a determinant of the quality of pre-service teachers' peer feedback. Hence, apart from multiple practice occasions or feedback prompts, expert feedback may influence the development of pre-service teachers' feedback quality by simultaneously providing a modelling example and fostering professional vision.

So far, peer feedback training in pre-service teacher education has not concentrated on enhancing feedback expertise (quality and beliefs) that concerns a complex feedback focus (classroom practice). Furthermore, as expert and peer feedback were combined in Prilop, Weber, & Kleinknecht, (2020) study, the significance of the twofold effect of expert feedback on the quality of pre-service teachers' peer feedback is unclear. Moreover, in this previous study (Prilop, Weber, & Kleinknecht, 2020), pre-service teachers did not receive specific feedback training before giving feedback. Therefore, we investigated whether pre-service teachers' peer feedback expertise (quality and beliefs) increased more through participation in (a) peer feedback training or (b) peer feedback training with added expert feedback. Peer feedback training for both groups included multiple opportunities to practice and feedback prompts. Pre-service teachers in the same cohort that were taking a different course acted as a control group. This study uses this approach to present characteristics of effective peer feedback training in pre-service teacher education, and provides a basis for future research in fostering pre-service teachers' peer feedback expertise on classroom practice.

2. Peer feedback

Peer feedback on teaching situations consists of the ability to convey critical assessments of a (pre-service) teacher's classroom practices, to stimulate reflection (Hammerness et al., 2005). While observing a fellow teacher in their classroom (feedback recipient), the feedback provider needs to make a criteria-based evaluation of the teaching performance. Subsequently, the feedback provider needs to constructively communicate apparent strengths and weaknesses to the feedback recipient (Sluijsmans, Brand-Gruwel, Van Merriënboer, & Bastiaens, 2003).

According to the prominent feedback model by Narciss (2013), characteristics of the source of external feedback, that is, of the peer providing a colleague with feedback on their teaching, determine the quality of the feedback. Therefore, a peers' feedback beliefs and domain knowledge, or the 'will and skill in tutoring' (Narciss, 2013, p. 16), can impact the effectiveness of the feedback. A peers' beliefs or domain knowledge affect how the feedback message is composed, which then determines how it is perceived by the feedback recipient. The feedback can consist of evaluative (i.e., a summative judgement of the performance) and/or tutoring content (i.e., information about strategies or suggestions for how to proceed).

1.1. Peer feedback expertise

Expert–novice comparisons that relate to feedback illustrate how feedback expertise develops (Prilop et al., 2019, Prilop, Weber, & Kleinknecht, 2021; Prins et al., 2006). In medical education, Prins et al. (2006) found that experts made more use of assessment criteria, provided more specific comments, and feedback was more positive and presented in the first person. Furthermore, novices often did not ask any reflective questions (59%) or provide suggestions for how to proceed (44%). In teacher education, Prilop et al. (2019), Prilop, Weber, Prins, & Kleinknecht, (2021) were able to replicate some of Prins et al.'s findings. Experts' feedback was more specific, included more questions, and used the first person more frequently than novices' feedback. Novices' feedback very rarely comprised specific descriptions of teaching situations, or questions. However, in contrast to Prins et al.'s findings, Prilop et al. found that experts and novices mainly provided negative feedback. Prilop et al. (2019), Prilop, Weber, Prins, & Kleinknecht, (2021) concluded that the difference between novices and experts in terms of specificity and questions may be based on the level of the participants' professional vision, that is, the ability to, first, perceive relevant situations, second, rapidly interpret the situation and third, decide on alternate actions (Blömeke, Gustafsson, & Shavelson, 2015; Sherin & van Es, 2009). In a recent cross-sectional study, Prilop et al. (in press) were able to support this assumption. Prilop et al. found that pre-service teachers' professional vision positively predicted feedback quality concerning questions and specificity. This is supported by results from the field of professional vision studies, which show that experts perceive, interpret and decide on alternate strategies more efficiently than novices (Gold, Hellermann, & Holodynski, 2016; Wolff, van den Bogert, Jarodzka, & Boshuizen, 2015).

For (pre-service) teachers' peer feedback expertise, the foregoing means that (pre-service) teachers need to be able to experience authentic peer feedback situations in teacher education to develop their expertise (Berliner, 1988; Chi, 2011). The importance of experiencing authentic situations is also highlighted in Bandura's Social Cognitive Theory (1986). People learn specific skills in situ by observing models. In teacher education, expert teachers often assume the role of 'modelers of practice' (Clarke, Triggs, & Nielsen, 2014, p. 177) and pre-service teachers try to imitate expert practices. However, Grossman and McDonald (2008) emphasise that in practice-based teacher education, pre-service teachers require multiple practice occasions to develop fluidity. Furthermore, after observing modelling instances of specific practices, pre-service teachers should develop their skills by 'approximations of practice' (Grossman & McDonald, 2008, p. 190). Approximations of practice can be described as authentic training situations that take place in a sheltered environment, i.e., without the pressure of action and with the opportunity for reflection.

1.2. Peer feedback beliefs

1.2.1. Defining peer feedback beliefs

In general, beliefs can be characterised as an individuals' 'convictions, philosophy, tenats, or opinions' (Haney, Lumpe, & Czerniak, 2003, p. 367). Consequently, (pre-service) teachers' beliefs can have a significant impact on their decision-making and actions (Bandura, 1986; Pajares, 1992). In the domain of peer feedback, this means that feedback beliefs influence how feedback is processed or composed (Strijbos & Müller, 2014). Although a feedback provider's beliefs can influence the quality of peer feedback, a feedback recipient's beliefs can determine the uptake of the information provided (Alqassab et al., 2018; Narciss, 2013). To date, studies have focused on three factors of peer feedback beliefs: usefulness of feedback (e.g., Cheng & Warren, 1997), confidence in

providing feedback (e.g., Rotsaert et al., 2018), and reasoning potential of feedback (Alqassab et al., 2018).

1.2.2. Changing peer feedback beliefs

Studies in more well researched domains, such as self-efficacy beliefs, indicate which sources can impact beliefs (e.g., Bandura, 1994; Gold, Hellermann, & Holodynski, 2017; Tschannen-Moran & McMaster, 2009). In general, beliefs form because of mastery experience, vicarious experiences, verbal persuasion and physiological cues (Bandura, 1994). Research indicates that mastery experience can have the strongest influence on participants' beliefs, especially when combined with verbal persuasion (e.g., feedback). However, vicarious experiences have also been shown to change beliefs to some extent (Gold et al., 2017).

In peer feedback research, most studies investigated changes in beliefs about the usefulness of feedback or in confidence in providing peer feedback during feedback practice (e.g., Cheng & Warren, 1997; Rotsaert et al., 2018; Sluijsmans et al., 2004; Wang, 2014). To our knowledge, only Alqassab et al. (2018) also analysed students' beliefs concerning the reasoning potential of peer feedback. In a variety of designs, researchers showed that mastery experience (i.e., practising peer feedback) enhanced participants' beliefs about peer feedback usefulness and increased participants' confidence in providing peer feedback (Cheng & Warren, 1997; Rotsaert et al., 2018; Sluijsmans et al., 2004). However, Wang's (2014) and Alqassab et al. (2018) studies contradict these findings. In a longitudinal study, Wang found that students' beliefs in the usefulness of peer feedback decreased, whereas pre-service teachers' peer feedback beliefs concerning usefulness, confidence and reasoning became less positive in Alqassab et al.'s study. Both Wang and Alqassab et al. attribute their findings to the need for comprehensive domain knowledge to be able to provide peers with high-quality feedback. Consequently, when students or pre-service teachers start training peer feedback by providing feedback to each other, this may be detrimental to peer feedback beliefs, as they do not yet possess comprehensive domain knowledge. Pre-service teachers may not be able to learn from each other vicariously and may not make successful mastery experiences, owing to a lack of knowledge (Bandura, 1994).

1.3. Peer feedback quality

1.3.1. Defining peer feedback quality

Extant studies that have measured feedback quality (e.g., Prilop, Weber, Prins, & Kleinknecht, 2021; Prins et al., 2006; S. Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; M. Gielen & De Wever, 2015) have used a set of criteria to assess peer feedback quality. First, feedback needs to take into account pre-defined criteria that are appropriate to the specific context. Feedback to a fellow teacher must be based on the actual performance and agreed-upon observation goals or assessment criteria. Second, judgements of a teachers' performance need to elaborate on specific, relevant classroom practices (S. Gielen et al., 2010). Third, besides summative evaluations of a teachers' performance, the feedback provider needs to offer constructive suggestions, such as knowledge about task constraints, concepts, mistakes, strategies or knowledge on how to proceed (Narciss, 2013), to enable the teacher to know which steps to take next. Fourth, peer feedback should include reflective questions that actively engage feedback recipients (S. Gielen et al., 2010; Nicol & Macfarlane-Dick, 2006). Fifth, peer feedback should consist of positive and negative comments (M. Gielen & De Wever, 2015). Finally, high-quality peer feedback should be formulated in the first person (Prins et al., 2006).

1.3.2. Fostering high-quality peer feedback

Specific training conditions aside, studies in higher education (e.g., Rotsaert et al., 2018; Sluijsmans et al., 2004) or pre-service teacher education (Prilop, Weber, & Kleinknecht, 2020) have shown that participants improve the quality of their peer feedback by practising it. In Sluijsmans et al.'s (2004) study, peer feedback training led to a significantly higher use of assessment criteria. Rotsaert et al. (2018) showed that peer feedback training resulted in students making greater use of negative verifications (negative evaluation), suggestive elaborations (suggestions) and informative elaborations (information specifying evaluation). In both Sluijsmans et al.'s and Rotsaert et al.'s studies, participants received detailed evaluation criteria and general guidelines for giving feedback, rather than specific prompts.

Greater improvements in peer feedback quality could be achieved through peer feedback prompts (e.g., for school students, Gan & Hattie, 2014; for higher education students, Gielen & De Wever, 2015). Feedback prompts can be defined as guiding questions or sentence beginnings that highlight or suggest content, or remind the peer feedback provider of crucial peer feedback components (Chen, Wei, Wu, & Uden, 2009). Various studies(Alqassab et al., 2018; Gan & Hattie, 2015; Gielen & De Wever, 2015) have shown that elaborate feedback prompts lead to higher-quality peer feedback. In a study on the effects of peer feedback prompts with varying structuring degree, Gielen and De Wever (2015) found that highly structured prompts were most effective for improving participants' peer feedback quality. Gan and Hattie (2015) showed that under prompted conditions, school students suggested more improvements to their peers and pinpointed mistakes more often than without prompts. In Alqassab et al.'s (2018) study, students with medium or high levels of domain knowledge improved their peer feedback during prompt-supported training at the self-regulation level, for example, by incorporating reflective questions.

Studies (Prilop, Weber, & Kleinknecht, 2020; van Blankenstein, Truțescu, van der Rijst, & Saab, 2019) have recently indicated that in addition to feedback prompts, modelling (i.e., an expert providing exemplary feedback) can improve peer feedback quality. In a study by van Blankenstein et al. (2019) higher education students practised giving peer feedback on oral presentations in a condition with modelling examples or a condition without. In the modelling examples condition an expert demonstrated feedback appropriate for a good or bad presentation. Van Blanckenstein et al. (2019) showed that after one week, the students in the modelling example condition used assessment criteria more often and made more judgements than students in the other condition. However, these students did not improve in categories such as making suggestions for improvement. Van Blanckenstein et al. attributed this finding to assessment criteria and making judgements being easier to copy and imitate following an expert model. Students perceive that experts use assessment criteria and that expert feedback includes judgement of the performance. Students can easily copy these components of expert feedback and imitate it, whereas suggestions for improvement cannot be copied, but require creativity and more in-depth knowledge. Although the experts in van Blanckenstein et al.'s study explicitly modelled high-quality feedback, experts in a study by Prilop, Weber, & Kleinknecht, (2020) provided pre-service teachers with feedback on their classroom practice without explicitly modelling high-quality feedback. Nevertheless, pre-service teachers seemed to have learnt from the expert feedback in situ and with regard to the quality of their suggestions, improved their peer feedback more than pre-service teachers in a condition that only offered peer feedback. However, it is unclear whether pre-service teachers learnt from expert feedback models in situ or whether the improvement in peer feedback quality was due to the combination of peer and expert

feedback, or the increase of pre-service teachers' professional vision.

3. Research questions

High-quality peer feedback can effectively promote (pre-service) teachers' expertise (Hammerness et al., 2005). To be able to provide high-quality feedback, teachers' need to develop their peer feedback expertise early in their careers. Although feedback training (e.g., Rotsaert et al., 2018; Sluijsmans et al., 2004) has been shown to foster pre-service teachers' peer feedback expertise, research (Prilop, Weber, & Kleinknecht, 2020; van Blankenstein et al., 2019) indicates that expert feedback may also enhance pre-service teachers' feedback expertise during feedback training. Therefore, we investigated whether pre-service teachers improve their feedback expertise more by participating in peer feedback training without (IG1 – Practice) or with expert feedback (IG2 – Expert Practice). Both courses were also compared to a control group (CG) that did not take part in peer feedback training.

The following research questions were investigated.

1) How does feedback training impact pre-service teachers' beliefs in peer feedback provision (learning, reasoning, confidence)?

We assumed that the peer feedback beliefs of the pre-service teachers in the intervention groups (Practice, Expert Practice) would generally show positive development during the training, as they are provided with multiple practice occasions and peer feedback prompts, that is, making successful mastery experiences (Bandura, 1994; Cheng & Warren, 1997; Gielen & De Wever, 2015; Rotsaert et al., 2018; Sluijsmans et al., 2004). Although Alqassab et al. (2018) showed that feedback training could also have detrimental effects, they attributed this finding to a lack of domain knowledge. As pre-service teachers in our design trained their domain knowledge and professional vision before practising peer feedback, and practised in fictitious situations, we cautiously assume a positive effect of feedback training on pre-service teachers' peer feedback beliefs.

2) How does additional expert feedback during feedback training influence pre-service teachers' beliefs about providing peer feedback (learning, reasoning, confidence), compared to conditions without expert feedback?

To our knowledge, there is no research on the effect of expert feedback on peer feedback beliefs. Though expert modelling as a vicarious experience may impact pre-service teachers' peer feedback beliefs(Gold et al., 2017), research in other domains indicates that mastery experience has a stronger impact on beliefs than vicarious experiences (Tschannen-Moran & McMaster, 2009). As there is no research from the domain of peer feedback, we did not formulate a specific hypothesis for research question 2.

3) How does feedback training impact pre-service teachers' feedback quality?

We expected intervention group participants (Practice, Expert Practice) to improve the quality of their peer feedback, as they practiced providing feedback several times, and were supported by peer feedback prompts (Alqassab et al., 2018; Gan & Hattie, 2014; Gielen & De Wever, 2015; Prilop, Weber, & Kleinknecht, 2020; Rotsaert et al., 2018).

4) How does additional expert feedback during feedback training influence pre-service teachers' feedback quality, compared to conditions without expert feedback?

We presumed that participants in IG2 (Expert Feedback) would profit from expert feedback (Prilop, Weber, & Kleinknecht, 2020; van Blankenstein et al., 2019) and, consequently, improve their peer feedback quality more than participants in IG1 (Practice).

2. Methods

2.1. Design of the quasi-experimental study

The study was conducted with a quasi-experimental pre-post-design. The intervention groups (IG1, IG2) participated in the 'Video-based development of teacher knowledge' course, whereas the control group (CG) participated in the 'Understanding communication – Transactional analysis for teachers' course. Fig. 1 shows the timetable of the quasi-experimental study.

2.2. Participants

Participants in the study were pre-service teachers in their tenth semester at a German university. Eighty-eight pre-service teachers were invited to participate: 30 pre-service teachers in each intervention group (IG1, IG2) and 28 in the control group (CG). Data from 20 pre-service teachers were excluded from the study, because they did not fully complete the tests, or did not participate in the pre-test and/or post-test (IG1: $n = 3$; IG2: $n = 6$; CG: $n = 11$). Pre-service teachers were provided with their test results after the post-test. In the 'Video-based development of teacher knowledge' courses, this was part of a reflective process, and can be seen as an additional incentive for participation. At the same time, the large number of exclusions from the CG can be attributed to the test not being linked to the course content, i.e., leading to less motivation to participate. Consequently, the final sample of the study consisted of 68 pre-service teachers: 27 pre-service teachers (63% female; $M_{Age} = 25.89$ years, $SD_{Age} = 1.89$) in IG1, 24 (74.1% female; $M_{Age} = 25.75$ years, $SD_{Age} = 2.03$) in IG2 and 17 (51.9% female; $M_{Age} = 28.12$ years, $SD_{Age} = 5.89$) in the CG.[1] Although pre-service teachers' mean age differed between the control group and the intervention groups, additional variables, such as teaching experience, $F(2,65) = 2.868$, $p = .06$ ($M_{CG} = 2.06$, $SD_{CG} = 1.30$; $M_{IG1} = 1.74$, $SD_{IG1} = 0.98$; $M_{IG2} = 1.33$, $SD_{IG2} = 0.64$), self-assessed feedback expertise, $F(2,65) = 0.028$, $p = .93$ ($M_{CG} = 2.94$, $SD_{CG} = 0.56$; $M_{IG1} = 2.96$, $SD_{IG1} = 0.76$; $M_{IG2} = 2.92$, $SD_{IG2} = 0.72$) and the number of times pre-service teachers received feedback prior to the intervention, $F(2,65) = 1.632$, $p = .20$ ($M_{CG} = 1.71$, $SD_{CG} = 0.85$; $M_{IG1} = 2.41$, $SD_{IG1} = 1.56$; $M_{IG2} = 2.04$, $SD_{IG2} = 1.12$), did not differ significantly. Furthermore, the groups' comparability was checked for feedback quality and feedback beliefs (please see 5.1).

2.3. Design of the 'video-based development of teacher knowledge' course

Research (e.g., Alqassab et al., 2018) indicates that low levels of domain knowledge can have a substantial impact on peer feedback quality and beliefs. Therefore, a sequential training of domain knowledge and peer feedback expertise was applied in the course (van Zundert, Sluijsmans, Könings, & van Merriënboer, 2012). In the first half of the course, pre-service teachers analysed classroom

[1] A sub-sample of the sample used in this study was also subject to analyses concerning other dependent variables (Prilop, Weber, & Kleinknecht, 2021).

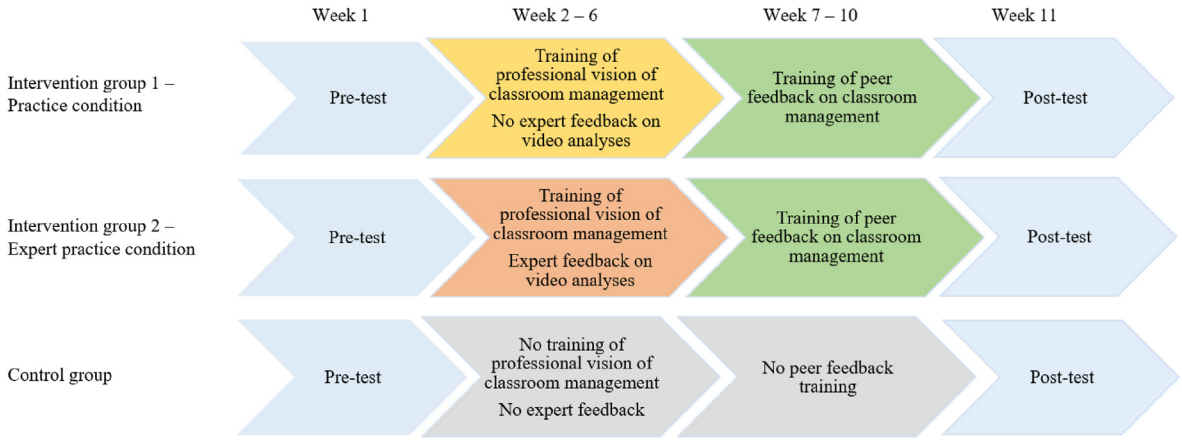


Fig. 1. Timetable of the quasi-experimental study for CG, IG1 and IG2.

videos that presented facets of classroom management, to acquire comprehensive domain knowledge of classroom management, which they could then apply to peer feedback practice. In the second half of the course, pre-service teachers composed peer feedback on fictitious video- and text-based situations that concerned classroom management situations.

We used classroom management as participants' focus, because it has been identified as a crucial dimension of teaching quality and is subject-independent (König, 2015). It can be defined as a teachers' ability to create and maintain learning environments that elicit a high level of student engagement (Brophy, 2010; Gold & Holodynski, 2017). A large body of studies indicates that classroom management skills have a significant impact on school students' outcomes (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016) and on teachers' mental health (Friedman, 2006). Although the impact of classroom management skills is widely acknowledged, research indicates that pre-service teachers struggle to effectively process classroom management events (e.g.,Wolff et al., 2017). To foster pre-service teachers' classroom management skills, various studies have successfully used video-based approaches (e.g., Gold, Pfirrmann, & Holodynski, 2021; Hellermann, Gold, & Holodynski, 2015) similar this study's design. These studies showed that video-based approaches could promote pre-service teachers' professional vision of classroom management, as well as their knowledge of classroom management.

Apart from an introductory session, each course session followed the same structure. Groups of four to five pre-service teachers briefly (5 min) presented their analysis of a classroom video (first half of the course, weeks 2–6) or their peer feedback on a video- and text-based fictitious classroom management situation (second half of the course, weeks 7–10) and summarised the theoretical and practical input of the previous session. After the presentation, all pre-service teachers had 15 min to revise the previous week's analyses or peer feedback. Then, the course instructor provided approximately 20 min of theoretical and practical input (e.g., visual peer feedback organiser; see Appendix). This was followed by a 10-min joint analysis of a video sequence, or preparation of peer feedback. Finally, pre-service teachers analysed a video sequence or composed peer feedback based on a video- and text-based peer feedback situation in a digital learning environment. This phase lasted for approximately 35 min. Participants could continue working on their analyses or peer feedback in the digital learning environment after the individual course sessions had ended. Participants had to provide their analyses and peer feedback as an appendix to the course exam, which consisted of a theory- and research-based essay on how their skills developed during the course. This guaranteed that participants completed the analyses and peer feedback.

The intervention was conducted in two 'Video-based development of teacher knowledge' courses. To control for course-instructor effects, pre-service teachers in each course were divided into members of intervention group 1 and intervention group 2. Members of intervention group 2 received expert feedback in addition to the course content outlined above. Expert feedback was provided on three occasions (weeks 3, 5 and 7). Expert feedback focused on pre-service teachers' analyses of video-based classroom management situations. Written feedback was provided in the digital learning environment (for a detailed description of the expert feedback provided, please see 4.3.3). Pre-service teachers had time to take the feedback into account during the 15-min revision phase at the beginning of the individual course session. In week one and week eleven, pre-service teachers from the intervention groups and the control group took part in video- and text-based tests. Pre-service teachers were presented with the tests through the online platform Unipark (Questback, 2017). Test items were presented in random order, to control for sequence effects. Test time was approximately 45–60 min.

2.3.1. Digital learning environment

Digital learning environments that incorporate classroom video need to be carefully designed to prevent attentional biases and to limit cognitive overload (Derry, Sherin, & Sherin, 2014, pp. 785–812; Syring et al., 2016). Therefore, video material with high-quality sound and images was specifically developed for the intervention. Classroom videos were filmed by the research team at a local school, to offer authentic classroom management representations (Sherin & van Es, 2009). Teaching situations were chosen based on a systematic coding procedure (for details, see Prilop, Weber, & Kleinknecht, 2021).

The digital learning environment was Moodle-based and divided into sections for each course session. During the course sessions that focused on classroom management, pre-service teachers were presented with a classroom management video sequence for analysis, with a specific focus to concentrate on. After analysing the video sequence, pre-service teachers in IG2 received expert feedback on their analysis, and subsequently revised and extended their analysis. In IG1, pre-service teachers did not receive expert feedback, but still revised and extended their analysis, based on the group presentations at the beginning of each course session.

During the course sessions that focused on peer feedback, pre-

service teachers were presented with a classroom video accompanied by a peer feedback situation, including a reflection by the teacher shown in the video. Pre-service teachers were supposed to imagine that they had to provide their colleague with feedback concerning the teaching sequence and the colleague's reflection. Pre-service teachers had to upload their written peer feedback to the digital learning environment.

2.3.2. Content-related and methodological support

Pre-service teachers were provided with content-related and methodological support for the peer feedback situations. Content-related and methodological support were supposed to decrease task complexity, i.e., limit pre-service teachers' cognitive load (Syring et al., 2016), and support pre-service teachers' use of peer feedback quality criteria (Gielen & De Wever, 2015).

Content-related support was based on material provided during the course sessions, which focused on classroom management. As the peer feedback situations revolved around classroom management teaching situations, this specific lens guided pre-service teachers' peer feedback (Santagata & Angelici, 2010). The reflections uttered by the fictitious colleague concentrated on certain facets of classroom management, indicated by key words (e.g., *I think a lot about the transitions. I want the lesson to go as smoothly as possible. But somehow, I have the feeling that I could structure such a transition phase even better*). Pre-service teachers were able to clearly link the key word *transitions* to the classroom management facet of managing momentum (Gold & Holodynski, 2017), and consequently, could focus on this facet in their peer feedback. The individual facets were comprehensively addressed in weeks 2–6 of the course.

From week 7, pre-service teachers were provided with a visual organiser with progressive prompts in the form of guiding questions (see Appendix) to support the peer feedback practice. The guiding questions focused on individual categories of peer feedback quality. For example, the guiding questions, *Where should he/she have acted differently? How should he/she have acted differently? Why should he/she have acted differently?* were supposed to guide pre-service teachers to provide specific peer feedback that included elaborate suggestions concerning the fictitious situations.

2.3.3. Expert feedback

As outlined above, pre-service teachers in IG2 received expert feedback three times (weeks 3, 5 and 7) during the intervention period. Expert feedback was composed according to quality criteria defined by Prins et al. (2006) and Prilop et al. (2019): assessment criteria, specificity, suggestions, questions, use of the first person and balance of positive/negative remarks. Consequently, the same criteria were used to assess pre-service teachers' pre- and post-test peer feedback.

All feedback (see Fig. 2) was provided by the same expert to prevent differences in feedback composition. Feedback presented the pre-service teachers with a summative comment regarding the assessment criteria, highlighted specific positive aspects, but also aspects to focus on during revision, and added a reflective question. Feedback was provided in the first person, and comprised a balanced number of positive and negative remarks. The expert took approximately 15 min per student to compose the feedback.

To guarantee that the information provided by the expert would not be part of the group presentation at the beginning of the courses, expert feedback was uploaded to the digital learning environment shortly before the course session started (i.e., for analyses composed in week 6, participants received expert feedback in week 7). This also ensured equal time-on-task, because all the pre-service teachers had only 15 min for revision, following the group presentation.

2.4. Ethical standards

Quasi-experimental studies raise certain ethical questions. To meet the highest ethical standards, the following measures were taken. In general, in accordance with the Office for Human Research Protections' guidelines (2009), research conducted in a conventional educational setting with the goal of comparing instructional techniques does not require to be overseen by the ethical review board at our institution. Nonetheless, ethical questions were discussed with members of the department and a member of the ethical review board. To reduce pressure to participate, several steps were taken. First, pre-service teachers were informed that participation was entirely voluntary and anonymous. Anonymity was guaranteed by anonymous links to the test. The pre- and post-test were matched using an eight-letter-digit code that pre-service teachers had to generate themselves, so that the research team did not know individual participants' results. Second, pre-service teachers were free to take the test during a course session or during the week following the first or eleventh week, as it was an online test. Our university does not have compulsory attendance, i.e., there is no pressure to attend a course, and, thus, to take a test. Third, providing only IG2 members with expert feedback could have presented unequal learning opportunities. In Germany, the length of a semester is 14 weeks and the intervention period lasted the first eleven weeks of the semester. Pre-service teachers in IG1 were provided with expert feedback in weeks 12–14. Fourth, the test did not add to the pre-service teachers' course load, as it was incorporated into course sessions.

2.5. Instruments

2.5.1. Peer feedback beliefs

We used Alqassab et al.'s peer feedback provision questionnaire (2018) to assess pre-service teachers' feedback beliefs. The questionnaire assesses pre-service teachers' beliefs about *learning from peer feedback provision* (e.g., 'I learn from providing peer feedback': Cronbach's $\alpha_{pretest} = .62$, Cronbach's $\alpha_{post\text{-}test} = 0.79$), *confidence regarding peer feedback provision* (e.g., 'I feel confident when providing positive feedback to my peers'; Cronbach's $\alpha_{pretest} = .71$, Cronbach's $\alpha_{post\text{-}test} = 0.69$) and *engaging in reasoning during peer feedback provision* (e.g., 'Providing peer feedback helps me to be critical about my own arguments'; Cronbach's $\alpha_{pretest} = .28$, Cronbach's $\alpha_{post\text{-}test} = 0.77$). Cronbach's α of the pretest scale for *engaging in reasoning during peer feedback provision* indicated limited reliability, therefore the results for this scale need to be interpreted with caution. However, as only Alqassab et al. have focused on the reasoning potential of peer feedback has only been focused on by Alqassab et al., this may be a belief that reliably develops only after focusing on peer feedback extensively. All scales consisted of five items, which were rated on a 6-point Likert scale (1 = strongly disagree, to 6 = strongly agree).

2.5.2. Peer feedback quality

A video- and text-based instrument (Prilop et al., 2019) was used to assess the quality of pre-service teachers' peer feedback (for a detailed description of the instrument, see Prilop et al., 2019, Prilop, Weber, & Kleinknecht, 2020). The instrument includes a peer feedback situation that focuses on classroom management. Based on a 1-min video recording of a second-year primary-school class, information on the teacher depicted, a text-based self-reflection and utterance of the teacher, pre-service teachers had to compose peer feedback for the teacher shown. Peer feedback was limited to 200 words.

Written peer feedback that pre-service teachers composed for the pre- and post-tests was coded using a coding scheme (Prilop

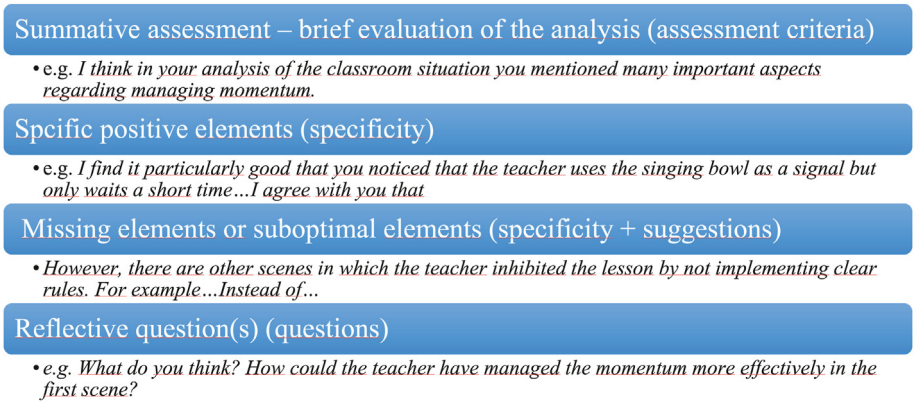


Fig. 2. Composition of expert feedback.

et al., 2019) based on Prins et al.'s work (2006). Adapted versions of Prins et al.'s coding scheme have been applied in a variety of domains (e.g., M. Gielen & De Wever, 2015; S. Gielen et al., 2010). The coding scheme qualitatively evaluates whether a set of criteria is present in the peer feedback, instead of quantifying the number of criteria, hence, each peer feedback is treated as a unit of analysis. This approach follows guidelines for quantifying qualitative data from Chi (1997) by capturing the 'semantics of the inference' (p. 10).

Our coding scheme consists of six categories (see Table 1). It distinguishes between good, average, and sub-optimal peer feedback. Good feedback was coded as '2', average feedback as '1' and sub-optimal feedback as '0'. An example of code 2 in the *suggestions* category is 'Perhaps it would be an idea to stand at the front of the class during such station learning, so you have everything in view'. In the *specificity* category, 'I had the impression that you did not notice Max and Anna playing with the water when you were explaining the task to Charlotte' would have been coded as '2'.

A member of the research team trained two student assistants as coders. Ten randomly-selected peer feedback were chosen for practice. Differences between coders' evaluations were discussed before 20% of the sample was double coded (Zottmann et al., 2013). Subsequently, Cohen's kappa (κ) was calculated to establish inter-coder reliability (Fleiss & Cohen, 1973). As kappa values indicated substantial inter-coder reliability (see Table 1), the remaining sample was coded by one student assistant.

2.6. Methods of analysis

The data were analysed by applying one-way multivariate analyses of variance (MANOVA) and planned contrasts. All analyses were performed with SPSS27 software. Furthermore, the alpha value was set at $p < .05$ for all statistical analyses. Normal distribution of data was slightly violated in the CG. As parametric analysis tools such as MANOVA are robust with regard to this kind of violation (Finch, 2005), we decided to apply it. Sphericity was provided in the MANOVA. Sufficient statistical power was established for all analyses using the GPower program (Faul, Erdfelder, Buchner, & Lang, 2009).

3. Results

3.1. Assessment of pre-service teachers' peer feedback expertise prior to intervention

We analysed possible pre-test differences for the analyses between CG vs. IG1 vs. IG2 and the planned contrast, CG vs. IG1 & IG2 and IG1 vs. IG2, by performing ANOVAs and *t*-tests. Pretest differences could have an impact on estimated effects.

The one-way ANOVA for CG vs. IG1 vs. IG2 showed no pretest differences between conditions for pre-service teachers' peer feedback beliefs concerning *learning*, $F\,(2,65) = 0.315$, $p = .73$, *reasoning*, $F\,(2,65) = 0.054$, $p = .95$, and *confidence*, $F\,(2,65) = 0.402$, $p = .67$. The planned comparisons (CG vs. IG1 & IG2; IG1 vs. IG2) also revealed no pretest differences.

The one-way ANOVA for CG vs. IG1 vs. IG2 showed that most peer feedback quality categories were comparable: *assessment criteria*, $F\,(2,65) = 0.48$, $p = .62$, *specificity*, $F\,(2,65) = 1.894$, $p = .16$, *suggestions*, $F\,(2,65) = 0.87$, $p = .42$, *questions*, $F\,(2,65) = 0.70$, $p = .50$ and *first person*, $F\,(2,65) = 0.70$, $p = .50$. However, the peer feedback category *positive/negative* revealed significant pretest differences, $F\,(2,65) = 7.367$, $p < .001$, which was confirmed by the one-way ANOVA for the planned comparison of IG1 & IG2 ($F\,(1,49) = 10.569$, $p = .002$).

Consequently, results of the analyses concerning the peer feedback category *positive/negative* should be disregarded, owing to significant pretest differences.

Table 1
Content analysis of pre-service teachers' peer feedback: category, feedback quality, codes and inter-coder reliability (Cohen's kappa).

Category	Good feedback	Code	Average feedback	Code	Sub-optimal feedback	Code	κ
Assessment criteria	Reference to classroom management, including terminology	2	Reference classroom management without terminology	1	No reference to classroom management	0	1.00
Specificity	Specific situation is elaborated on	2	Specific teaching phase is elaborated on	1	No specific situation or phase elaborated on	0	1.00
Suggestions	Alternatives presented with explanation	2	Alternatives presented without explanation	1	No alternatives presented	0	.68
Questions	Activating question posed	2	Clarifying question posed	1	No questions posed	0	.88
First person	Written in the first person throughout feedback	2	Occasionally written in the first person	1	Not written in the first person	0	.84
Positive/ negative	Balanced positive and negative feedback	2	Mainly positive feedback	1	Mainly negative feedback	0	.88

3.2. Effects on pre-service teachers' peer feedback beliefs and quality

3.2.1. Effects on pre-service teachers' peer feedback beliefs

For research question 1, we performed one-way repeated measures multivariate analyses of variance (MANOVA) to assess the effect of time on all peer feedback belief factors (dependent variables: *learning*, *reasoning*, *confidence*) simultaneously. The conditions (CG, IG1, IG2) were analysed individually (see Table 2). Although the MANOVA for the CG revealed no significant effects because Wilks' Lambda failed to reach significance level, large, significant effects for IG1 and large, significant effects for IG2 were found for *confidence*. The results show that practising peer feedback with or without expert feedback (IG1, IG2) lead to an increase of pre-service teachers' *confidence* beliefs in providing peer feedback.

For research question 2 (differences of peer feedback beliefs between groups), we conducted one-way repeated measures MANOVA to assess interaction effects on all peer feedback belief factors (dependent variables: *learning*, *reasoning*, *confidence*; independent variables: CG, IG1, IG2) simultaneously (see Table 3). No significant differences were found for the comparison of CG, IG1 and IG2, as Wilks' Lambda did not reach the significance level. The planned comparison of CG vs. IG1 and IG2 revealed a medium, significant effect for *confidence*, whereas Wilks' Lambda failed to reach significance level for the planned comparison of IG1 vs. IG2. Hence, peer feedback practice with or without expert feedback led to an increase of peer feedback beliefs concerning *confidence*, in contrast to a course that did not focus on peer feedback (CG).

3.2.2. Effects on the quality of pre-service teachers' peer feedback

For research question 3, one-way repeated measures multivariate analyses of variance (MANOVA) were performed, to assess the effect of time on all peer feedback-quality categories (dependent variables: *assessment criteria*, *specificity*, *suggestions*, *questions*, *first person*, *positive/negative*) simultaneously. Analyses were conducted for each condition (CG, IG1, IG2) individually (see Table 4). The MANOVA for CG and IG1 showed no significant effects, as Wilks' Lambda failed to reach significance level. Significant, large effects were found for IG2 with regard to *assessment criteria*, *specificity* and *questions*. Although pre-service teachers in IG2 profited from peer feedback practice and expert feedback in three categories (*assessment criteria*, *questions*, *positive/negative*), a significant decrease was found for *specificity*. The significant increase of pre-service teachers' peer feedback quality with regard to *positive/negative*, should be disregarded, owing to the significant pretest differences between conditions.

Table 2
Means (M), standard deviations (SD) and results of one-way repeated measures MANOVA for time effects on feedback beliefs.

	Pre-test		Post-test					
	M	*SD*	*M*	*SD*	*F*	*df*	*p*	η_p^2
Learning								
CG	4.49	0.73	4.82	0.80	1.153	1,16	.30	.07
Practice (IG1)	4.75	0.67	4.84	0.68	0.368	1,26	.55	.01
Expert Practice (IG2)	4.72	0.62	4.76	0.50	0.090	1,23	.77	.004
Reasoning								
CG	4.67	0.28	4.68	0.67	0.006	1,16	.94	.00
Practice (IG1)	4.65	0.49	4.76	0.70	1.102	1,26	.30	.04
Expert Practice (IG2)	4.63	0.49	4.70	0.44	0.505	1,23	.48	.02
Confidence								
CG	3.88	0.76	3.87	0.86	0.006	1,16	.94	.00
Practice (IG1)	3.89	0.73	4.34	0.72	7.509	1,26	.01	.22
Expert Practice (IG2)	3.72	0.75	4.43	0.50	16.042	1,23	<.001	.41

Note: CG = control group, Practice = Feedback practice group (IG1), Expert practice = Feedback practice with expert feedback group (IG2), $n_{CG} = 17$, $n_{Practice} = 27$ and $n_{Expert\ Practice} = 24$; feedback beliefs Min = 1, Max = 6; Wilks' Lambda: CG: $F(3,14) = 0.466$, $p = .71$, $\eta_p^2 = 0.09$; Practice: $F(3,24) = 3.088$, $p = .046$, $\eta_p^2 = 0.28$; Expert Practice: $F(3,21) = 5.172$, $p = .008$, $\eta_p^2 = 0.43$.

Table 3
Results of one-way repeated measures MANOVA for interaction effects on all feedback belief categories for all groups and planned comparisons.

	CG vs. Practice (IG1) vs. Expert Practice (IG2)			CG vs. Practice (IG1) & Expert Practice (IG1)			Practice (IG1) vs. Expert Practice (IG2)		
	F(2,65)	*p*	η_p^2	*F(1,66)*	*p*	η_p^2	*F(1,49)*	*p*	η_p^2
Learning	0.301	.74	.009	0.551	.46	.008	0.066	.80	.001
Reasoning	0.167	.85	.005	0.284	.60	.004	0.058	.81	.001
Confidence	3.953	.024	.11	6.545	.013	.09	1.188	.28	.024

Note: Wilks' Lambda: CG vs. Practice vs. Expert-Practice, $F(6,60) = 2.043$, $p = .07$, $\eta_p^2 = 0.09$; CG vs. Practice & Expert-Practice, $F(6,61) = 3.625$, $p = .018$, $\eta_p^2 = 0.15$; Practice vs. Expert-Practice, $F(6,44) = 0.554$, $p = .65$, $\eta_p^2 = 0.03$.

Table 4
Means (M), standard deviations (SD) and results of one-way repeated measures MANOVA for time effects on peer feedback quality.

	Pre-test		Post-test		*F*	*df*	*p*	η_p^2
	M	*SD*	*M*	*SD*				
Assessment criteria								
CG	0.94	0.43	0.71	0.47	2.133	1,16	.16	.12
Practice (IG1)	1.04	0.52	1.07	0.47	0.088	1,26	.77	.003
Expert Practice (IG2)	0.92	0.41	1.38	0.58	11.644	1,23	.002	.34
Specificity								
CG	0.53	0.87	0.29	0.69	1.662	1,16	.22	.09
Practice (IG1)	0.44	0.85	0.44	0.80	0.000	1,26	1.00	.00
Expert Practice (IG2)	0.92	0.97	0.29	0.69	4.276	1,23	.04	.17
Suggestions								
CG	0.53	0.80	0.71	0.92	0.459	1,16	.51	.03
Practice (IG1)	0.89	0.93	1.37	0.88	4.201	1,26	.05	.14
Expert Practice (IG2)	0.83	0.96	1.21	0.93	2.309	1,23	.14	.09
Questions								
CG	0.24	0.56	0.18	0.53	0.105	1,16	.75	.01
Practice (IG1)	0.11	0.32	0.78	0.80	17.333	1,26	<.001	.40
Expert Practice (IG2)	0.08	0.41	1.25	0.79	38.862	1,23	<.001	.63
First person								
CG	1.47	0.87	1.00	1.00	5.885	1,16	.027	.27
Practice (IG1)	1.26	0.90	1.44	0.80	0.572	1,26	.46	.02
Expert Practice (IG2)	1.42	0.88	1.58	0.78	1.353	1,23	.26	.06
*Positive/Negative**								
CG	0.18	0.53	0.18	0.53	0.000	1,16	1.00	.00
Practice (IG1)	0.78	0.85	0.81	0.74	0.024	1,26	.88	.001
Expert Practice (IG2)	0.17	0.38	1.00	0.51	41.071	1,23	<.001	.64

Note: CG = control group, Practice = Feedback practice group, Expert practice = Feedback practice with expert feedback modelling group, $n_{CG} = 17$, $n_{Practice} = 27$ and $n_{Expert\ Practice} = 24$, peer feedback quality categories Min = 0 and Max = 2; Wilks' Lambda: CG: $F(6,11) = 1.506$, $p = .26$, $\eta_p^2 = 0.45$; Practice: $F(6,21) = 2.532$, $p = .053$, $\eta_p^2 = 0.42$; Expert Practice: $F(6,18) = 13.462$, $p < .001$, $\eta_p^2 = 0.82$; *significant time effects disregarded, due to significant pretest differences between conditions.

For research question 4 (differences in peer feedback quality between groups), we performed one-way repeated measures MANOVA to assess interaction effects on all peer feedback quality categories (dependent variables: *assessment criteria*, *specificity*, *suggestions*, *questions*, *first person*, *positive/negative*; independent variables: CG, IG1, IG2) simultaneously (see Table 5). Significant, large interaction effects were found for the comparison of CG vs. IG1 vs. IG2 (*assessment criteria*, *questions*). The medium interaction effects of the *positive/negative* category were disregarded, owing to pretest differences. The planned comparison of CG vs. IG1 & IG2 revealed a significant, large interaction effect for *questions* and significant, medium effects for *assessment criteria* and *first person*. The planned comparison of IG1 vs. IG2 revealed significant, medium interaction effects for *assessment criteria* and *questions*. The significant, large interaction effect for the *positive/negative* peer feedback quality category was disregarded, owing to significant

Table 5
Results of one-way repeated measures MANOVA for interaction effects on all peer feedback quality categories for all groups and planned comparisons.

	CG vs. Practice (IG1) vs. Expert Practice (IG2)			CG vs. Practice (IG1) & Expert Practice (IG2)			Practice (IG1) vs. Expert Practice (IG2)		
	F(2,65)	*p*	η_p^2	*F(1,66)*	*p*	η_p^2	*F(1,49)*	*p*	η_p^2
Assessment criteria	5.899	.004	.15	6.163	.02	.09	5.282	.03	.10
Specificity	2.017	.14	.06	0.034	.85	.00	3.395	.07	.07
Suggestions	0.348	.71	.01	0.601	.44	.01	0.098	.76	.00
Questions	10.497	<.001	.24	15.708	<.001	.19	4.169	.047	.08
First person	2.728	.07	.08	5.536	.02	.08	0.004	.95	.00
Positive/Negative*	6.360	.01	.09	2.217	.14	.03	7.849	.007	.14

Note: Wilks' Lambda: CG vs. Practice vs. Expert-Practice, $F(6,60) = 7.147$, $p < .001$, $\eta_p^2 = 0.42$; CG vs. Practice & Expert-Practice, $F(6,61) = 3.988$, $p = .002$, $\eta_p^2 = 0.28$; Practice vs. Expert-Practice, $F(6,44) = 3.269$, $p = .01$, $\eta_p^2 = 0.31$; *significant interaction effects disregarded owing to significant pretest differences between conditions.

pre-test differences between conditions.

Briefly, the quality of pre-service teachers' peer feedback (*assessment criteria*, *questions*, *first person*) in the courses that focused on peer feedback practice improved significantly, when compared to that of the CG. Including expert feedback in peer feedback training (IG2) significantly improved the quality of pre-service teachers' peer feedback (*assessment criteria*, *questions*), more than without expert feedback (IG1).

The quantitative results can be illustrated with the following example, which shows the change in one student's feedback from pre-to post-test. This illustration should not be regarded as an in-depth qualitative analysis, but instead illustrates how peer feedback quality developed in the categories of *assessment criteria*, *specificity*, *suggestions* and *questions*, for students in intervention group 2.

Following the utterance by the fictitious peer ('Of course, some children are a bit unfocused, but overall, I think I did a good job keeping the students working. How do you feel about that?'), the pretest peer feedback of participant ANFR176 shows that this participant perceived the teachers' classroom management negatively. Although ANFR176 is very specific about the teacher' shortcomings, ANFR176 only alludes to possible alternatives without any explanation. ANFR176 references the assessment criteria by mentioning 'withitness', a central skill of effective classroom management, but fails to link it to the classroom situations in question. As could be seen in most pretest peer feedback, ANFR176 did not use activating questions in the feedback.

> I have to say that I perceived your withitness differently in this lesson. You kept conversations with the students very short, and quickly went on to another child, so that no real conversation could take place. You also didn't go into detail about the individual students, for example, you wanted to see a student's research book, but you didn't look at it; instead, you went on to another child. Even when a child said it was busy pushing, you didn't respond to the situation. I'm saying that in this lesson you didn't get close enough to the children, you just jumped back and forth for a few moments. (student ANFR176, pretest)

ANFR176's post-test feedback shows that this participant focuses on positive and negative aspects of the classroom situation. The focus of the feedback is much more clearly connected to withitness through an explanation of where it is lacking ('lose sight of the rest of the students') and suggestions for how it may be better achieved ('try to look at all students to give them a feeling of my presence'). This suggestion is then linked to an activating question, to engage the feedback recipient. However, the pretest feedback was more specific, as it clearly outlined several relevant situations, whereas the post-test feedback was not as precise.

> Dear colleague, I see that you act with a calmness in class. I noticed that you address the learners, and thus give them your attention, which is positive. On the other hand, I've seen you lose sight of the rest of the students during this time of the conversations, resulting in decreased withitness. Many students still work attentively at their tasks, which is supported by your calm nature. I often also stand by the learners who need help with the work, but I always try to look at all students, to give them a feeling of my presence. Would that be an idea for implementation in your classroom? (student ANFR176, post-test)

4. Discussion

Our study investigated the impact of two peer feedback trainings on pre-service teachers' peer feedback beliefs and peer feedback quality, in comparison to a control group. Participants in the control group took part in a course that did not focus on peer feedback. Intervention group 2 (Expert Practice) received additional expert feedback on video analyses before repeatedly practising peer feedback provision, whereas intervention group 1 (Practice) did not receive expert feedback. We hypothesised that pre-service teachers' beliefs about, and the quality of, peer feedback would increase in IG1 and IG2, but with stronger effects in IG2, because of the expert feedback. Improvements were not expected in the control group.

4.1. Development of pre-service teachers' peer feedback beliefs

With regard to peer feedback beliefs, in IG1 and IG2 pre-service teachers' beliefs increased significantly only for *confidence*, whereas no significant changes were detected for *learning* and *reasoning*. Pre-service teachers' peer feedback belief concerning *confidence* increased significantly in IG1 and IG2, in comparison to the CG. No differences were found between IG1 and IG2.

Our hypotheses were only partially confirmed. Our findings confirm studies (Cheng & Warren, 1997; Sluijsmans et al., 2004) that showed that participants require multiple practice occasions to develop confidence in providing peer feedback. In Alqassab et al.'s (2018) study, pre-service teachers' peer feedback beliefs decreased. Alqassab et al. attributed this finding to a lack of domain knowledge when practising providing peer feedback in a complex domain. In this study, we took two precautions to prevent negative effects on pre-service teachers' peer feedback beliefs. First, before they engaged in peer feedback practice, pre-service teachers' domain knowledge of the focus of the feedback was comprehensively trained by analysing videos of classroom management until session 6. Second, pre-service teachers did not provide peer feedback to

peers, but practised providing peer feedback in fictitious peer feedback situations. Hence, our study provides the insight that peer feedback beliefs can also be fostered in complex domains by taking domain knowledge into account and offering a more sheltered practice environment (Grossman & McDonald, 2008). Expert feedback did not have an effect on pre-service teachers' peer feedback beliefs. As research in other domains shows, mastery experience, i.e., the successful application of skills, has the most significant impact on people's beliefs (Bandura, 1994). Hence, the vicarious experience of observing expert feedback does not influence pre-service teachers' peer feedback beliefs as much as mastery experience. Peer feedback beliefs concerning *learning* and *reasoning* remained at a high level, indicating that the value of receiving and providing peer feedback seems to have been established previously.

4.2. Development of the quality of pre-service teachers' peer feedback

With regard to peer feedback quality, the quality of pre-service teachers' peer feedback improved significantly only in the *questions* category for IG1, and the *assessment criteria* and *questions* categories for IG2. No significant improvements were detected in the CG. The quality of peer feedback for the *specificity* category actually decreased for members of IG2. The quality of pre-service teachers' peer feedback related to the *assessment criteria* and *questions* categories improved significantly in IG1 and IG2, in comparison to the CG. A significant difference in development was found for the *assessment criteria* and *questions* categories, when comparing pre-service teachers' peer feedback quality in IG1 and IG2. Participants in the condition with expert feedback (IG2) significantly improved the quality of their peer feedback in these categories. The *positive/negative* category was disregarded, owing to significant differences between conditions.

Our results partially support our hypotheses that peer feedback training can enhance the quality of pre-service teachers' peer feedback and that expert feedback can enhance the effect of peer feedback training. The opportunity to practise peer feedback multiple times with the aid of peer feedback prompts can significantly improve pre-service teachers' peer feedback quality. Our results align with Alqassab et al.'s (2018) finding that pre-service teachers improved their use of *questions* when supported by prompts aimed at self-regulatory feedback. In contrast to Sluijsmans et al. findings (2004), the quality of pre-service teachers' peer feedback regarding *assessment criteria* did not improve in IG1. Furthermore, pre-service teachers did not improve the quality of peer feedback with regard to *suggestions*, in contrast to participants in Gan and Hattie's (2015) and Rotsaert et al.'s (2018) studies. The complex task of providing high-quality *suggestions* may require an abundance of authentic teaching experiences that require more teaching experience (Berliner, 1988; Chi, 2011; van Blankenstein et al., 2019).

Expert feedback seems to be necessary to foster peer feedback quality. The findings indicate a twofold effect of expert feedback as modelling examples and expert feedback that enhances professional vision. On the one hand, the effect on the quality of pre-service teachers' peer feedback related to *assessment criteria* and *questions* seems to derive from pre-service teachers imitating the use of *assessment criteria* and *questions* from the expert model (van Blankenstein et al., 2019). In Prins et al.'s (2006) study, experts showed higher-quality use of *assessment criteria*, whereas experts in Prilop et al.'s (2019) study applied significantly more high-quality *questions* than novices. This suggests that pre-service teachers learn from expert feedback in situ, as novices mimic feedback practices (Bandura, 1986; Clarke et al., 2014). In a previous study, Prilop, Weber, & Kleinknecht, (2020) showed that analysing classroom video recordings, and receiving peer and expert feedback improved the quality of pre-service teachers' peer feedback with regard to *specificity*. Furthermore, Prilop et al. (in press) showed that professional vision determines the quality of the *specificity* and *questions* categories. Our study partially contradicts these finding, as the quality of peer feedback related to *specificity* decreased for pre-service teachers in IG2. This could be an effect of the intervention design. In Prilop, Weber, & Kleinknecht, (2020) study, pre-service teachers did not take part in peer feedback training before providing their peers with feedback. Receiving detailed input on individual peer feedback quality categories could lead pre-service teachers to focus their attention on surface features of peer feedback, such as *assessment criteria* and *questions*. However, the impact on the quality of pre-service teachers' peer feedback with regard to *questions* may also be attributed to an additional increase of pre-service teachers' professional vision, due to expert feedback (Prilop, Weber, & Kleinknecht, 2020; in press). This is supported by the expert–novice-comparison of (pre-service) teachers' peer feedback quality (Prilop et al., 2019) and the peer feedback determinants study (Prilop et al., in press), which indicate that professional vision predicts the quality of pre-service teachers' peer feedback with respect to *questions*.

4.3. Limitations, future directions and implications for teacher education

Our findings highlight two crucial factors that need to be taken into account, to enhance peer feedback beliefs. First, feedback training should be preceded by a phase of building domain knowledge and various types of knowledge, such as conceptual, procedural, and conditional knowledge. Pre-service teachers need to be able to make mastery experiences, which is unlikely when domain knowledge is lacking (Bandura, 1994). Second, pre-service teachers should be provided with the opportunity to practise peer feedback in a sheltered environment (Grossman & McDonald, 2008). As practising peer feedback in authentic situations can harm peer feedback beliefs (Alqassab et al., 2018), peer feedback training should start with solitary practice applying fictitious situations.

With regard to peer feedback quality, the different approaches taken by IG1 and IG2 each indicate a need for more in-depth research. Future peer feedback training should try to more effectively foster the quality of pre-service teachers' *suggestions* or *specificity*. Though the descriptive quality of *suggestions* increased, no significant improvement was found. Regarding *specificity*, our study showed a decrease or stagnation in quality. The visual peer feedback organiser used in our study may not sufficiently scaffold pre-service teachers' *suggestions* and *specificity*, which is a potential limitation of this training. Although some studies (e.g., Gan, 2011) showed that elaborate scaffolding elements successfully improved peer feedback quality, other studies (Alqassab et al., 2018; Gielen et al., 2010) indicate that too much scaffolding may also harm the quality of peer feedback. Consequently, future studies should investigate the effect of various degrees of scaffolding in peer feedback training. Regarding peer feedback training with expert feedback, studies should investigate when and how expert feedback should be provided, and whether expert feedback can be substituted with standardised expert feedback. Expert feedback is resource intensive, i.e., to scale peer feedback training with expert feedback, standardised versions from which pre-service teachers can learn may be a viable option.

Quasi-experimental designs have some limitations that should be considered in future research. First, a waiting control group could be included, to ensure that all participants wish to participate in the same intervention. This could also enhance the CG's motivation to participate in the pre- and the post-test, and limit the possible bias of convenience sampling. Second, the design of this

study could be shortened, to allow for equal intervention time for IG1 and IG2 participants. IG2 participants would benefit from expert feedback in the first round of the intervention, and IG1 participants could benefit from expert feedback in a second subsequent round. This switching replication design would also allow for more in-depth analysis by using a third measurement point. Third, in quasi-experimental designs, time-on-task may vary from participant to participant. We tried to control for this limitation by incorporating the expert feedback into the 15-min time slot during the course session, and by making completed analyses and feedback preconditions for taking the exam at the end of the course. However, some students may have worked more with the feedback or course materials after the sessions. In future studies, this limitation should be checked by asking students about the amount of time they invested in the course. Though an experimental design could remove this constraint, an experimental study would offer less insight into what is possible in actual teaching. Fourth, our results could benefit from an in-depth qualitative analysis, to provide a more nuanced understanding of how pre-service teachers' feedback changed. This is outside the scope of the present article, however it should be considered for more qualitative study designs. Combined with interviews following the intervention, a qualitative approach could shed some light on which components of the intervention (e.g., multiple practice occasions, expert feedback, visual organiser) helped pre-service teachers improve their feedback.

4.4. Conclusions

Our study sheds light on the effects of peer feedback training in pre-service teacher education. Our results show that peer feedback training should combine domain knowledge acquisition with sheltered peer feedback practice. Furthermore, our findings indicate that expert feedback can further enhance peer feedback training by improving specific aspects of the quality of pre-service teachers' feedback (*assessment criteria* and *questions*). Expert feedback seems to have a twofold effect on the quality of pre-service teachers' peer feedback. Our study indicates that expert feedback influences the development of the quality of pre-service teachers' peer feedback by simultaneously providing a modelling example and fostering professional vision. Hence, our study presents the characteristics of effective peer feedback training that enables pre-service teachers to use the full potential of peer feedback.

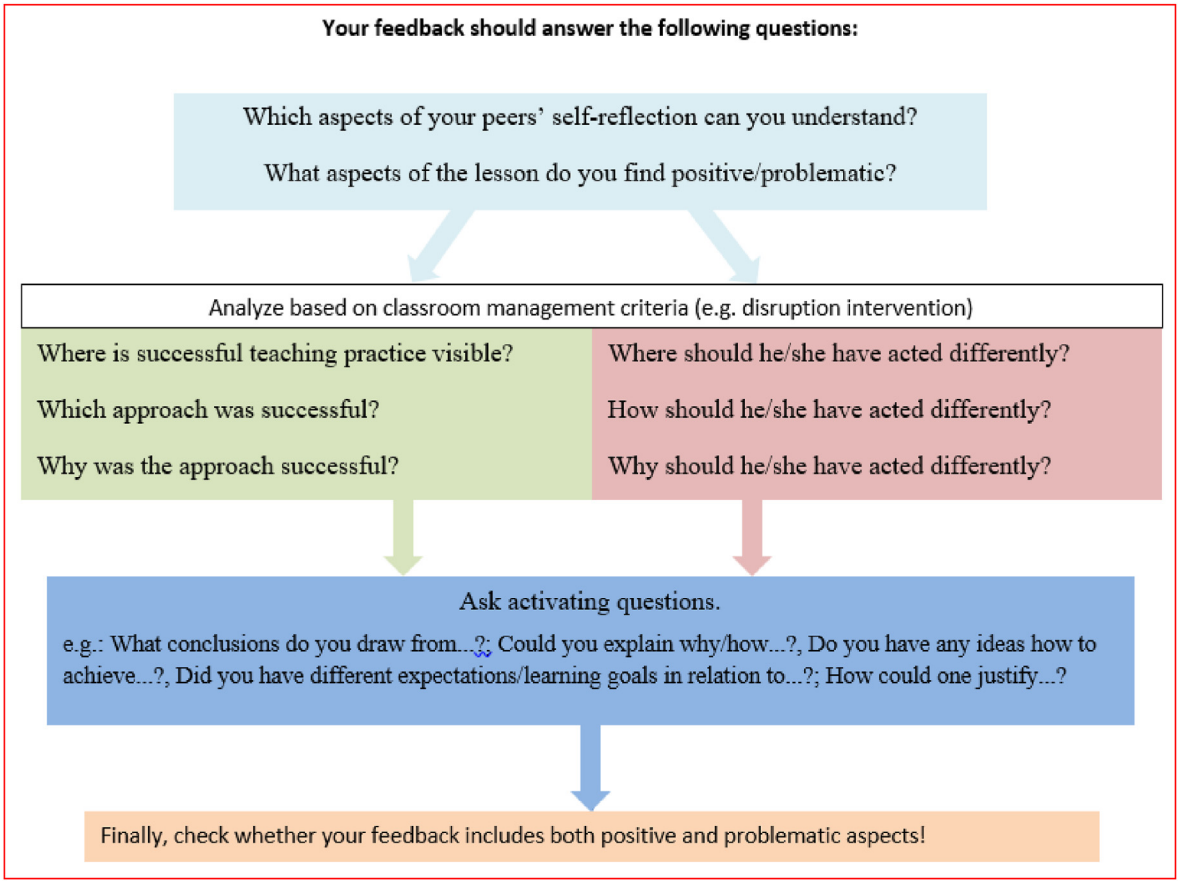
Data availability

The authors do not have permission to share data.

Acknowledgements

We would like to thank the pre-service teachers who participated in this study. Furthermore, we are grateful to our student assistants for administering the blended-learning environment and coding written peer feedback.

Appendix



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

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). New York: Academic Press.

Berliner, D. C. (1988). HuntMemorial lecture presented at the annual meeting of the American association of colleges for teacher education. In *The development of expertise in pedagogy*. New Orleans: Charles W.

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

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Brophy, J. (2010). Classroom management as socializing students into clearly articulated roles. *Journal of Classroom Interaction, 45*(1), 41–45.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*(2), 233–239. https://doi.org/10.1080/03075079712331381064

Chen, N.-S., Wei, C.-W., Wu, K. T., & Uden, L. (2009). Effects of high level prompts and peer assessment on online learners' reflection levels. *Computers & Education, 52*(2), 283–291. https://doi.org/10.1016/j.compedu.2008.08.007

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Chi, M. T. H. (2011). Theoretical perspectives, methodological approaches, and trends in the study of expertise. In Y. Li, & G. Kaiser (Eds.), *Expertise in mathematics instruction* (pp. 17–39). Boston, MA: Springer.

Clarke, A., Triggs, V., & Nielsen, W. (2014). Cooperating teacher participation in teacher education: A review of the literature. *Review of Educational Research, 84*(2), 163–202. https://doi.org/10.3102/0034654313499618

Derry, S., Sherin, M., & Sherin, B. (2014). The Cambridge handbook of multimedia learning. In R. Mayer (Ed.), *Multimedia learning with video*. Cambridge, UK: Cambridge University Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA. Test Statistics when Assumptions Are Violated. *Methodology, 1*(1), 27–38. https://doi.org/10.1027/1614-1881.1.1.27.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*(3), 613–619. https://doi.org/10.1177/001316447303300309

Friedman, I. A. (2006). Classroom management and teacher stress. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice, and contemporary issues* (pp. 925–945). Mahwah, NJ: Erlbaum.

Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science, 42*(6), 861–878. https://doi.org/10.1007/s11251-014-9319-4

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

Gold, B., Hellermann, C., & Holodynski, M. (2016). Professionelle Wahrnehmung von Klassenführung - vergleich von zwei videobasierten Erfassungsmethoden. In K. Schwippert, & D. Prinz (Eds.), *Der forschung-der lehre-der bildung: Aktuelle entwicklungen der Empirischen bildungsforschung* (pp. 103–118). Münster: Waxmann.

Gold, B., Hellermann, C., & Holodynski, M. (2017). Effekte videobasierter Trainings zur Förderung der Selbstwirksamkeitsüberzeugungen über Klassenführung im Grundschulunterricht [Effects of video-based trainings for promoting self-efficacy in elementary classroom management]. *Zeitschrift für Erziehungswissenschaft, 20*(19), 115–136.

Gold, B., Pfirrmann, C., & Holodynski, M. (2021). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education, 72*(4), 431–447. https://doi.org/10.1177/0022487120963681

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184–205. https://doi.org/10.3102/0002831207312906

Hammerness, K. M., Darling-Hammond, L., Bransford, J., Berliner, D. C., Cochran-Smith, M., McDonald, M., et al. (2005). How teachers learn and develop. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness, & H. Duffy (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 358–389). San Francisco, CA: Jossey-Bass.

Haney, J. J., Lumpe, A. T., & Czerniak, C. M. (2003). Constructivist beliefs about the science classroom learning environment: Perspectives from teachers, administrators, parents, community members, and students. *School Science & Mathematics, 103*(8), 366–377. https://doi.org/10.1111/j.1949-8594.2003.tb18122.x

Hellermann, C., Gold, B., & Holodynski, M. (2015). Die Wirkung der Analyse eigener und fremder Unterrichtsvideos auf das strategische Wissen und die professionelle Wahrnehmung [Classroom management skills in teacher education. The effect of analysis of own and other classroom videos on strategic knowledge and professional vision]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47*(2), 97–109. https://doi.org/10.1026/0049-8637/a000129

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

König, J. (2015). Measuring classroom management expertise (CME) of teachers: A video-based assessment approach and statistical results. *Cogent Education, 2*(1), Article 991178. https://doi.org/10.1080/2331186X.2014.991178

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive feedback model. *Digital Education Review, 23*, 7–26. https://doi.org/10.1344/der.2013.23.7-26

Office for Human Research Protections. (2009). Part 46, Protection of human subjects, Code of federal regulations. *hhs.gov. p. US 45 CFR 46, 101*. Retrieved 12 April 2019.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Prilop, C. N., Weber, K., & Kleinknecht, M. (2019). Entwicklung eines video- und textbasierten Instruments zur Messung kollegialer Feedbackkompetenz von Lehrkräften [Development of a video- and text-based instrument for the assessment of teachers' peer feedback competence]. In T. Ehmke, P. Kuhl, & M. Pietsch (Eds.), *Lehrer. Bildung. Gestalten: Beiträge zur empirischen Forschung in der Lehrerbildung* (pp. 153–163). Weinheim Basel: Beltz Juventa Verlag.

Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2020). Effects of digital video-based feedback environments on pre-service teachers' feedback competence. *Computers in Human Behavior, 102*, 120–131. https://doi.org/10.1016/j.chb.2019.08.011.

Prilop, C. N., Weber, & Kleinknecht, M. (2021). The role of expert feedback in the development of pre-service teachers' professional vision of classroom management in an online blended-learning environment. *Teaching and Teacher Education, 99*, 103276. https://doi.org/10.1016/j.tate.2020.103276.

Prilop, C. N., Weber, K. E., Prins, F., & Kleinknecht, M. (2021). Connecting feedback to self-efficacy: Receiving and providing peer feedback in teacher education. *Studies In Educational Evaluation, 70*(2), 101062. https://doi.org/10.1016/j.stueduc.2021.101062.

Prilop, C. N., Weber, K. E., Rotsaert, T., & Keppens, K. (2023). (in press). Knowledge, self-efficacy and professional vision: How do they predict pre-service teachers' peer feedback quality? In R. Stahnke, & A. Gegenfurtner (Eds.), *Teacher professional vision: Empirical perspectives*. Abingdon, Oxon, New York, NY: Routledge.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289–303. https://doi.org/10.1007/s10459-005-3250-z

Questback. (2017). *Unipark - EFS survey, version summer 2017*. Köln: Questback GmbH.

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). 'Now you know what you're doing right or wrong!' Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education, 33*(2), 255–275. https://doi.org/10.1007/s10212-017-0329-x

Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education, 61*(4), 339–349. https://doi.org/10.1177/0022487110369555

Sherin, M. G., Russ, R. S., Sherin, B. L., & Colestock, A. (2008). Professional vision in action: An exploratory study. *Issues in Teacher Education, 17*(2), 27–46.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*, 20–37. https://doi.org/10.1177/0022487108328155

Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merriënboer, J. J. G., & Bastiaens, T. J. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies In Educational Evaluation, 29*(1), 23–42. https://doi.org/10.1016/S0191-491X(03)90003-4

Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education & Teaching International, 41*(1), 59–78. https://doi.org/10.1080/1470329032000172720

Strijbos, J.-W., & Müller, A. (2014). Personale Faktoren im Feedbackprozess [personal factors in the feedback process]. In H. Ditton, & A. Müller (Eds.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder [Feedback. Theoretical foundation, empirical findings, practical fields of application]* (pp. 83–134). Münster: Waxmann.

Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M., & Schneider, J. (2016).

Fallarbeit als Angebot - fallbasiertes Lernen als Nutzung. Empirische Ergebnisse zur kognitiven Belastung, Motivation und Emotionen bei der Arbeit mit Unterrichtsfällen [Case work as proposition – case-based learning as usage. Empirical results on cognitive load, motivation and emotions when working with classroom cases]. *Zeitschrift für Padagogik, 62*, 86–108. https://doi.org/10.3262/ZP1601086

Tripp, T. R., & Rich, P. J. (2012). The influence of video analysis on the process of teacher change. *Teaching and Teacher Education, 28*(5), 728–739. https://doi.org/10.1016/j.tate.2012.01.011

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal, 110*(2), 228–245. https://doi.org/10.1086/605771

Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. (2015). Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education, 66*(1), 68–85. https://doi.org/10.1177/0022487114549810

Zottmann, J. M., Stegmann, K., Strijbos, J.-W., Vogel, F., Wecker, C., & Fischer, F. (2013). Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching experience on knowledge convergence. *Computers in Human Behavior, 29*(5), 2100–2108. https://doi.org/10.1016/j.chb.2013.04.014

van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & van Merriënboer, J. J. D. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology, 32*(1), 127–145. https://doi.org/10.1080/01443410.2011.626122