Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels — Source link ⧉

Maryam Alqassab, Jan-Willem Strijbos, Stefan Ufer

Institutions: Ludwig Maximilian University of Munich, University of Groningen

Related papers:

- Peer Assessment Between Students in Colleges and Universities

- Improving the effectiveness of peer feedback for learning

- The Power of Feedback

- Research on the Effect of Peer Feedback Training in English Writing Teaching—A Case Study of Students in Business English Major

- Peer Feedback in Second Language Writing (2005-2014).

University of Groningen

Training peer-feedback skills on geometric construction tasks

Alqassab, Maryam; Strijbos, Jan-Willem; Ufer, Stefan

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

CrossMark

Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels

Maryam Alqassab[1] · **Jan-Willem Strijbos**[2] ·
Stefan Ufer[3]

Abstract Peer feedback is widely used to train assessment skills and to support collaborative learning of various learning tasks, but research on peer feedback in the domain of mathematics is limited. Although domain knowledge seems to be a prerequisite for peer-feedback provision, it only recently received attention in the peer-feedback literature. In this study, preservice mathematics teachers ($N = 43$) were involved in a peer-feedback training in which they evaluated geometric construction tasks and were (a) trained to provide peer feedback on different levels (i.e. task, process and self-regulation) and (b) scaffolded with worked examples, feedback provision prompts and evaluation rubrics during the training. A quasi-experimental mixed design was implemented with domain knowledge as the between-subject factor and measurement occasion as the within-subject factor. Students' peer-feedback provision skills and their beliefs about peer-feedback provision were measured before and after the training. Students with high and medium domain knowledge provided more peer feedback at the self-regulation level, whereas those low in domain knowledge provided more peer feedback at the task-level after the training. Students' beliefs about peer-feedback provision became less positive after the training, regardless of the level of their domain knowledge.

✉ Maryam Alqassab
maryam.alqassab@psy.lmu.de; http://www.en.mcls.lmu.de/study_programs/reason/people/cohortof2013/phdstudents/alquassab/index.html

Jan-Willem Strijbos
j.w.strijbos@rug.nl; http://www.rug.nl/staff/j.w.strijbos/

Stefan Ufer
ufer@math.lmu.de; http://www.mathematik.unimuenchen.de/~didaktik/index.php?ordner==ufer&data==start

1   Department of Psychology, LMU Munich, Leopoldstr. 44, 80802 Munich, Germany

2   Faculty of Behavioural and Social Sciences, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands

3   Mathematics Institute, LMU Munich, Theresienstr. 39, 80333 Munich, Germany

⌂ Springer

Collaborative learning is considered as an important pedagogical activity that supports students' learning. Yet, successful peer collaboration requires the students to be engaged with and aware of each other's thinking (Kuhn 2015). Involving students in peer feedback is one way to ensure their active engagement during collaborative learning (Phielix et al. 2010). Peer feedback is regarded in this paper as the qualitative variant of peer assessment that is defined as a learning activity where individuals or smallgroup constellations exchange, react to and/or act upon information about their performance on a particular learning task with the purpose to accomplish implicit or explicit shared and individual learning goals.

Providing feedback on the work of a peer is a complex assessment skill that can be improved with training (Sluijsmans et al. 2004). Previous studies showed that peer-feedback skills are influenced by students' individual characteristics such as domain knowledge (Patchan and Schunn 2015) and task characteristics such as complexity (Van Zundert et al. 2012a). However, the role of peer-feedback providers' domain knowledge seems to be ignored in studies conducting peer-feedback trainings. Also, the learning tasks used in most peer-feedback studies are often presentations or written essays. Few studies investigated students' peer-feedback provision skills using complex learning tasks that involve scientific reasoning (e.g. Gan and Hattie 2014; Lavy and Shriki 2014; Van Zundert et al. 2012b).

Particularly in mathematics education, despite the importance of peer-feedback activities for teaching preservice teachers the assessment skills of scientific reasoning tasks like geometric proofs (Lavy and Shriki 2014), no study—to our knowledge—has empirically investigated training preservice teachers in peer-feedback provision skills on proof tasks. Additionally, involving preservice teachers in peer-feedback training does not only influence their assessment skills but can also shape their perspective about peer feedback (e.g. Sluijsmans et al. 2004). Thus, preservice teachers' beliefs about peer-feedback provision should be more systematically investigated when peer-feedback provision trainings are being conducted. In the next sections, we discuss the factors that can influence training peer-feedback provision skills when complex mathematical tasks are used.

Peer feedback and geometric construction tasks

Geometric constructions are useful scientific reasoning tasks because they are discovery tasks in which the results of learned mathematical problems (i.e. theorems or concepts) can be applied to physical objects in the real world (Schoenfeld 1986). Similar to geometry proofs, constructions should be supported by deductions. However, research has shown that students at school and university levels including preservice teachers regard deductive proofs as irrelevant to geometric constructions (Kuzniak and Houdement 2001; Miyakawa 2004; Schoenfeld 1989; Tapan and Arslan 2009).

The failure to use the deductive mathematical knowledge that students already possess is attributed to passive instruction (Schoenfeld 1989). Preservice teachers can be actively engaged with the learning task through involving them in peer-feedback provision. But since they rely on empirical approaches when performing geometric construction tasks, they are more likely to use the same approaches during peer-feedback provision unless they are being trained and supported to use specific types of feedback.

Feedback levels

Hattie and Timperley's (2007) progressive feedback model offers a conceptualization of learner's engagement with the learning task and the learning and self-regulation processes associated with that learning task. According to this model, there are four types (i.e. levels) of feedback that have different effects on learning. The first level is the task-level that refers to the correctness/incorrectness of the solution or the content knowledge used to solve the task (e.g. "your answer is correct"). The second level is feedback about the learning processes and strategies that can be used to solve not only the task at hand but also other similar tasks (process-level, e.g. "your justifications should be based on theorems when you perform constructions"). The third level is self-regulation that directs the learners to monitor and regulate their learning goals (e.g. "what would happen if you change the size of the angle?"). Self-regulation feedback does not provide information to the learner about what should be done; instead, it stimulates learners to act and reflect on their learning.

The fourth level is feedback about the self that is typically used for motivational purposes (self-level). It includes no information about the learning task, but refers to learners' personal characteristics; most of the time in the form of general non-task-related praise (e.g. "you are smart!"). Importantly, this feedback is not the same as what is referred to as self-feedback in the feedback literature (e.g. Butler and Winne 1995). Self-feedback is internal feedback, which is part of self-regulation, whereas feedback about the self is personal and focuses on the self.

Process and self-regulation feedback levels are more beneficial for deeper processing and mastery learning because they stimulate deeper engagement with the learning task (Hattie and Timperley 2007). Feedback at the self-level is regarded as the least useful feedback for learning because it diverts the learner's attention to the self and away from the learning task (Kluger and DeNisi 1996).

Feedback levels in peer-feedback research Due to its comprehensibility, several researchers used the Hattie and Timperley (2007) model to train or analyse peer feedback of high school students in reading, mathematics and chemistry (e.g. Gan 2011; Gan and Hattie 2014; Harris et al. 2015). These studies reported that peer feedback was dominated by the task- and self-levels (Gan and Hattie 2014; Harris et al. 2015). Nonetheless, when prompted and/or trained to provide peer feedback at the different levels on chemistry lab reports, students could provide more peer feedback at the process and the self-regulation levels (Gan 2011). However, it is unclear if this also applies to complex mathematics tasks such as geometric constructions.

Engaging preservice mathematics teachers in peer-feedback provision in which they are instructed to provide peer feedback on the process and self-regulation levels can be beneficial for their assessment skills of geometric constructions, as the provision of these types of peer feedback requires deeper processing of the task. The progressive nature of the feedback levels in the Hattie and Timperley (2007) model can aid the student to move beyond the surface features of the construction (i.e. how it looks) and consider other processes through which the construction was created, as well as rely on deductive reasoning to be able to provide peer feedback at the process and self-regulation levels. Nonetheless, since most of students' peer feedback is at the task- and self-levels (Gan and Hattie 2014), they need to be trained to provide peer feedback at the higher levels.

Training peer-feedback skills

Peer feedback is usually a new practice to most students including preservice teachers. Students often experience uncertainty about their ability to assess the work of a peer and wish to have more support (Cheng and Warren 1997). There is evidence that training preservice teachers to assess the work of peers resulted in better peer assessment skills (e.g. Sluijsmans et al. 2004). Similarly, training students with peer-feedback prompts was found to improve feedback provided at the process and self-regulation levels on chemistry lab reports (Gan 2011). Thus, providing peer feedback at the higher levels can be trained, but students need some instructional support especially when complex learning tasks such as geometric constructions are being used.

Instructional support for peer feedback training Van Zundert et al. (2012a) recommended training domain knowledge before training assessment skills when a complex learning task is used. However, in a typical face-to-face classroom, instruction time is usually limited and sequential training of domain knowledge and peer-feedback skills might not be feasible. One way to account for this challenge is to support students using domain knowledge scaffolds and peer-feedback scaffolds (i.e. tools used to help students succeed in performing challenging tasks; Quintana et al. 2004).

An efficient domain knowledge scaffold that can support students during peer-feedback provision is that of a worked example, which typically includes the problem with its solution (Renkl 2014). Several studies in geometry and algebra education showed that teaching students using worked examples works better than the conventional teaching, especially for students with low domain knowledge (e.g. Carroll 1994; Paas and Van Merriënboer 1994; Reiss et al. 2008). Using a worked example as a domain knowledge scaffold can be expected to reduce the demand introduced by the complexity of the task and, consequently, help students to focus on the peer-feedback provision process.

Feedback scaffolds that can be used to support the peer-feedback provision activity are prompts and an evaluation rubric. Structuring the peer-feedback activity through feedback provision prompts can lead to more peer-feedback provision at the process and the self-regulation levels (Gan and Hattie 2014; Gielen et al. 2010). Yet, students can still focus on one aspect of the peer solution. Providing students with an evaluation rubric (i.e. task-specific criteria) against which they judge the peer solution can help them to focus on the essential parts of the solution. The use of evaluation rubrics was found to increase the accuracy of peer assessment (Panadero et al. 2013) and can be used in combination with the feedback provision prompts to provide peer feedback at different levels.

In sum, structured peer-feedback training seems to have a potential to improve preservice mathematics teachers' feedback skills on peer solutions to geometric construction tasks. Yet, this assumption requires empirical examination. Importantly, the peer-feedback providers' level of domain knowledge is likely to influence the degree to which they can benefit from the peer-feedback training as well as their perspective about peer-feedback provision. In the next sections, we discuss the role of domain knowledge during peer-feedback provision and students' beliefs about peer-feedback provision.

Domain knowledge and peer-feedback provision

Providing feedback on peer solutions to a learning task requires an understanding of the task. Van Zundert et al. (2012b) showed that domain knowledge is a prerequisite for assessing the

work of a peer, especially when the learning task is complex. Thus, the type of provided peer feedback is likely to be influenced by the provider's domain knowledge. Patchan and Schunn (2015) reported that, in academic writing, peer feedback by students with low domain knowledge was dominated by praise, whereas peer feedback provided by high domain knowledge students involved more criticism. However, there is still limited evidence about whether the type of peer feedback provided on complex mathematical tasks, like geometric constructions, is also influenced by domain knowledge.

Students' beliefs about peer feedback

According to the Interactional Framework of Feedback (Strijbos and Müller 2014), providers' and recipients' individual characteristics including their beliefs are equally important for feedback. While recipients' beliefs influence the processing of the feedback message, the providers' beliefs are expected to shape the composition of the feedback message (Strijbos and Müller 2014). Students' beliefs about the helpfulness of peer feedback were found to be positively associated with (self-reported) self-regulation and negatively associated with GPA (Brown et al. 2016). Similarly, students' beliefs about the usefulness of peer feedback were found to be positively associated with perceived peer-feedback accuracy and trust in one-self as a provider and in the peer as a recipient (Rotsaert et al. 2017).

In studies implementing peer-feedback activities, changes in students' beliefs about peer feedback are often investigated given that they might be associated with students' insecurities regarding peer-feedback usefulness or their ability to provide peer feedback. Some studies that measured students' beliefs about peer-feedback usefulness or their confidence regarding peer-feedback provision before and after the peer-feedback activities reported positive changes (e.g. Cheng and Warren 1997; Sluijsmans et al. 2004). In contrast, in a more recent EFL writing study, Wang (2014) reported a decrease in students' beliefs about peer-feedback usefulness that was attributed to several factors among them domain knowledge.

So far, most studies have focused only on changes in beliefs about peer-feedback usefulness or confidence regarding peer-feedback provision after involving students in peer-feedback provision (e.g. Sluijsmans et al. 2004). Peer-feedback related epistemological beliefs (e.g. critical evaluation, self-reflection; see Nicol et al. 2014) are also important when investigating peer-feedback provision, and changes in these beliefs are yet to be explored. While domain knowledge is suggested to play a role in how students' beliefs about peer-feedback provision change (Wang 2014), there seems to be no study to our knowledge that empirically tested this.

Current research and hypotheses

The aim of this study was to investigate the impact of a structured peer-feedback training on preservice mathematics teachers' feedback provision skills on peer solutions to geometric construction tasks and on their peer-feedback provision beliefs taking into account their domain knowledge. We formulated the following hypotheses:

- H1a: Structured peer-feedback training leads to improved peer feedback at the higher levels (i.e. process and self-regulation).

- H1b: Higher levels of domain knowledge result in providing more peer feedback at the process- and self-regulation levels and less peer feedback at the task- and self-levels after a structured peer-feedback training.
- H2a: Structured training results in changes in beliefs about peer-feedback provision.
- H2b: Students' level of domain knowledge influences the direction of changes in beliefs.

Method

Participants

The participants were 58 middle school preservice mathematics teachers from a large university in southern Germany. Participation was a course requirement, and the students received no additional compensation. The study ran throughout the semester. Data collection and peer-feedback training took place over several sessions. Only 43 out of the 58 students were present at all measurement occasions (9 males, 34 females, mean age 22.51, SD = 2.36) and were included in the analyses.

Design

We implemented a quasi-experimental mixed design to investigate the impact of peer-feedback providers' domain knowledge (between-subjects, addressing H1b and H2b) on changes in the levels of their written peer feedback and their beliefs about peer-feedback provision after a structured peer-feedback provision training on geometric construction tasks (within-subject, addressing H1a and H2a) (see Fig. 1).

Materials

Geometric construction tasks The domain tasks consisted of a set of geometric objects (e.g. a line, a circle and an angle) and asked for the construction of a specified object (e.g. a
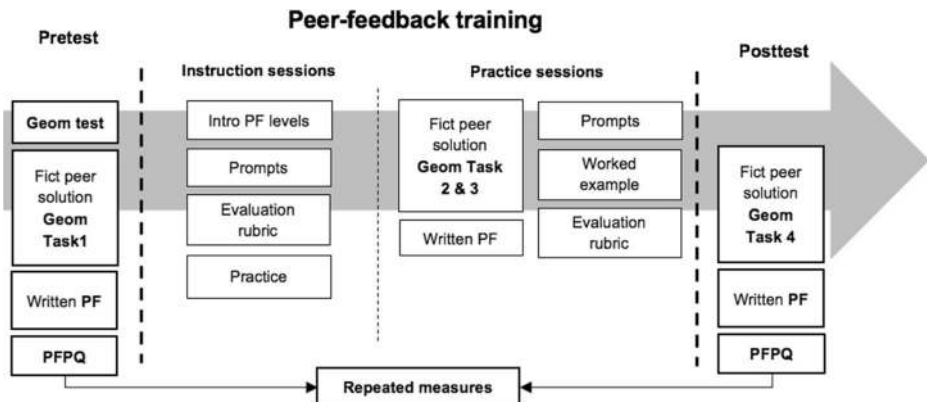
Fig. 1 A detailed illustration of the study design showing different activities that took place during the peer feedback training in sequence. *Geom test* = domain knowledge test, *Fict peer solution* = fictional peer solution to geometric construction task(s), *Geom Task* = geometric construction tasks, *PFPQ* = peer feedback provision beliefs questionnaire, *Intro PF levels* = introduction to peer-feedback levels

tangent to the circle that intersects the line in a congruent angle). Solving the task requires (a) performing the construction, (b) describing the construction step by step and (c) providing reasoning to show why the construction yields the specified object. The participants were required to provide written peer feedback on fictional peer solutions to these domain tasks.

Fictional peer solutions One fictional peer solution was used for the pretest and one for the posttest. The geometric construction tasks on which the peer solutions were created were similar (constructing a line tangent to a circle and parallel/ perpendicular to another line for the pretest and the posttest, respectively). Both fictional peer solutions contained (a) some correct steps, but partly followed an incorrect strategy, (b) correct descriptions of some but not all steps, (c) correct and incorrect reasoning steps and (d) vague language in some parts of the solution. The graphical construction in the peer solutions matched the description but was not completely accurate.

Feedback provision prompts A visual organiser (developed by Gan 2011) with progressive prompts reflecting different levels of feedback according to Hattie and Timperley's (2007) model was used (see Hattie and Gan 2011 for the visual organiser). We extended Gan's visual organiser with additional prompts: mostly knowledge integration/self-reflection prompts (adopted from Chen et al. 2009; King 2002; Nückles et al. 2009; see Table 1). Most of the added prompts were at the self-regulation level because it is assumed that students benefit more from providing knowledge integration/self-reflective questions, as they need to think deeply about the learning task (King 2002).

Evaluation rubric The evaluation rubric consisted of a set of criteria that could be used in combination with the feedback prompts to judge the peer solution and produce written peer feedback. More specifically, the students had to judge (a) the construction of the geometric object, (b) the description of the construction and (c) the reasoning provided to prove the construction true (see Fig. 2). The feedback provision prompts could be applied to each section of the evaluation rubric that directed the students to all parts of the peer solution.

Worked example All students received a standard worked example of the geometric construction task they had to provide written peer feedback on during the peer-feedback training. Since the purpose of this study was to improve students' peer-feedback skills, and

Table 1 Prompts added to Hattie and Gan (2011) feedback level visual organiser and their sources

Prompts	Peer-feedback level	Source
Which parts were not sufficiently clarified?	Task	Nückles et al. (2009)
How can you best explain…?	Self-regulation	
How would you use…. to…..?	Self-regulation	King (2002)
What would happen if you….?		
How does….tie to what we learned before?		
What conclusions can you draw about…?		
Explain why…..		
Explain how….		
How are …and … similar?		
How are… and ….different?		
What can you infer from….?	Self-regulation	Chen et al. (2009)
What can you think of from….?		

Geometry Construction Tasks – Evaluation Rubric

| *Construction procedure* |
| Success criteria: |
| ☐ The construction does only use known operations. |
| ☐ The construction can be performed with any set of starting objects. |
| ☐ The construction does indeed yield an object with the specified properties. |

| *Reasoning for construction* |
| Success criteria: |
| ☐ Where not obvious, reasons are given why the construction does indeed work in the described way (e.g., why intersection points exist,...). |
| ☐ Reasons are given, why the construction yields an object with the specified properties. |
| ☐ There is some reasoning as to how many different solutions to the task exist. |

| *Description of the construction* |
| Success criteria: |
| ☐ The description is clearly understandable and not ambiguous. |
| ☐ The description makes correct use of mathematical symbols and terminology. |

Fig. 2 Evaluation rubric given to students in the training sessions to judge the construction procedure of the geometric figure, the description of construction and the reasoning provided for the construction

not their domain knowledge, we provided them with the worked example to ensure that those with low domain knowledge can still practice providing written peer feedback at the higher levels (i.e. process and self-regulation).

Measures

Domain knowledge test To measure students' domain knowledge, a geometry basic knowledge test (Ufer, Heinze and Reiss 2008) consisting of 49 true/false items measuring different topics (e.g. properties of an equilateral triangle, properties of a parallelogram, transversals) was used at the pretest ($M = 37.93$, $SD = 5.15$, Cronbach's $\alpha = 0.77$, maximum score of 49). Based on students' scores, they were grouped into the lowest ($M = 32.05$, $SD = 3.17$), middle ($M = 38.33$, $SD = 1.32$) and highest ($M = 43.67$, $SD = 1.65$) one third of the sample.

Peer-feedback provision questionnaire Students' beliefs about (a) learning from peer-feedback provision (LPF) (e.g. "I learn from providing peer feedback"), (b) confidence regarding peer-feedback provision (CPF) (e.g. "I feel confident when providing positive feedback to my peers") and (c) engaging in reasoning during peer-feedback provision (RPF) (e.g. "Providing peer feedback helps me to be critical about my own arguments") were measured before and after the peer-feedback training using the peer-feedback provision questionnaire. This questionnaire was developed for the current study, and it consists of 40 items—five of which were adapted from Linderbaum and Levy's (2010) Feedback Orientation Scale. The items were scored on a 6-point Likert scale (strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree). Means were calculated for each subscale for further analyses.

The instrument (with initially 44 items) was piloted with an independent sample of students (N = 83). Parallel analysis following (O'Connor's 2000) procedure revealed a three-factor solution that corresponded to the theoretical structure of the questionnaire (Hinkin 1998). The items were subjected to multiple principle component analyses (PCAs) because the ratio of the sample size to the number of items was too small to include all of the 44 items in one analysis. Three rounds of PCA were conducted with two theoretically distinct scales as the components in each round (LPF vs. CPF, CPF vs. RPF and LPF vs. RPF). Items were retained only if they loaded meaningfully on the intended theoretical component with factor loadings > 0.40 and had a value of at least 0.50 for the diagonal elements of the anti-image correlation matrix (Field 2009).

The procedure resulted in the exclusion of four items from the RPF scale. The three scales supported by the PCAs consisted of (a) ten items for LPF (Cronbach's α = 0.87), (b) 17 items measuring CPF (Cronbach's α = 0.91) and (c) 13 items for RPF (Cronbach's α = 0.88). PCA was not conducted for the current study as the sample size was not sufficient, but the scales' reliabilities were equally high for the present sample: LPF (Cronbach's α_{pre} = 0.83, Cronbach's α_{post} = 0.90), CPF (Cronbach's α_{pre} = 0.90, Cronbach's α_{post} = 0.96) and RPF (Cronbach's α_{pre} = 0.71, Cronbach's α_{post} = 0.87).

Peer-feedback levels The peer-feedback levels at the pretest and posttest were coded using a coding scheme based on Hattie and Timperley's (2007) model. Previous studies showed the applicability of this feedback model to peer feedback and developed coding schemes that were successfully used to analyse peer feedback (e.g. Gan and Hattie 2014; Harris et al. 2015). The coding scheme (Table 2) was adapted from Gan and Hattie (2014). Although peer feedback at the self-level was not prompted for, it was included in the coding because peer feedback often includes statements of this nature (Harris et al. 2015).

Prior to coding students' written peer feedback, it was segmented following the procedure by Strijbos et al. (2006), with the smallest meaningful segment as the unit of analysis. Two coders (first author and a student-assistant) independently segmented 10% of the data reaching an acceptable percentage agreement level (81.80% lower bound, 82.30% upper bound), after which the first author segmented the remainder. The same two coders independently coded 10% of the segments (Krippendorff's α = 0.76). The remaining segments were then coded by the first author. Proportions of peer feedback at each level (task, process, self-regulation, self) were calculated for further analyses.

Procedure

In the first session, all participants answered the domain knowledge test and provided written peer feedback on a fictional peer solution to a geometric construction task to measure their baseline peer-feedback skills. Then, they completed the inventory on their beliefs about peer-feedback provision. In sessions two to five, all participants received peer-feedback provision training that consisted of two stages. In the first stage, two instructional sessions were held each lasting 45 min. In the first session, the notion of peer feedback was discussed with the students. The students shared their thoughts about peer feedback, its benefits, how it should look like and their insecurities regarding peer feedback. Then, the feedback levels (task, process, self-regulation and self) were

Table 2 Scheme used to code the written peer feedback based on the feedback levels adapted from Gan and Hattie (2014)

Peerfeedback level	Definition	Example
Task	Regarding correctness/incorrectness of the solution. Can be regarding mathematical knowledge of the task or telling the peer how to solve the task. It can inform the learner about what is missing/ unclear in the solution. Stylistic comments about the structure of the solution or spelling, grammar mistakes also belong to this category.	"Shifting the line this way is not correct". "You did not describe how you have constructed the tangent".
Process	Refers to learning processes and strategies that can be used/or used to solve the task. Involves questions that prompt the receiver to use a learning strategy to better improve the answer or to provide further explanations and justifications for their answers. Can be surface (related to parts of task) or deep (more general).	Process surface: "How do you know that the line is tangent to the circle?" Process deep: "If you give reasons, please also explain why you did something and why it should work like that".
Self-regulation	Reflective statements that require thinking about conceptual understating, and hence lead to further information seeking, involvement in self-evaluation or monitor learning progress to reach learning goal. Feedback at this level is mostly but not always probing questions that require further actions from the learner. This type of feedback leads the learners closer to the self-feedback and self-error detection.	"You very well understood and applied the rules". "Is your explanation sufficient?" "What would happen if you change the size of the angle?"
Self-level (praise)	Refers to the self and does not involve the learning task. This can refer to some characteristics of the learner and most of the time is in the form of praise, with the purpose of encouragement.	"Well done!" "Keep it up!"

introduced and discussed with the students. At that point, the students also received the feedback provision prompts accompanied with the evaluation rubric. In the remaining part of the first session and in the second session, the students were involved in several individual and group activities to better understand the different levels of feedback. They had to (a) identify each feedback level in written peer feedback comments, (b) transform one feedback level to a higher level and (c) work in groups to provide written peer feedback on a solution, as well as share and discuss their peer feedback with the rest of the class. In the second stage of the peer-feedback training—which also involved two sessions—we provided the students with worked examples in addition to the feedback prompts and the evaluation rubric that they received at an earlier stage. The students received a fictional peer solution and practised providing written peer feedback on that solution with the help of the instructional scaffolds (i.e. feedback prompts, evaluation rubric and worked example). At the end of the semester, each participant provided written peer feedback on a fictional peer solution in the absence of all instructional scaffolds and answered the peer-feedback provision questionnaire again.

Analyses

We used repeated measure multivariate analysis of variance (MANOVA) followed by repeated measure ANOVAs to analyse peer feedback provided by preservice mathematics teachers. The peer-feedback data violated the assumption of normality, and thus, we applied a semi-parametricrepeated measure MANOVA test using the MANOVA.RM package in R (Friedrich et al. 2017). The test implements the rank-based ANOVAtype statistic (ATS) tests for factorial designs (Brunner et al. 1999). For the post hoc tests, we also used the ATS using the R package nparLD (Noguchi et al. 2012).

In the ATS tests for within-subject factors and interactions involving within-subject factors, the denominator degrees of freedom used for the approximation of the distribution is assumed to be infinity (Brunner et al. 1999) because the degrees of freedom used in conventional ANOVA produce conservative measures (Bathke et al. 2009). A measure for the effect size for the ATS is the relative effect, that can be interpreted as the probability that a randomly chosen observation from the sample would result in a smaller value for a specific peer-feedback level (e.g. task or process) than a randomly chosen observation from a domain knowledge group (e.g. low, medium or high) for a specific measurement occasion (i.e. pretest or posttest).

Changes in students' beliefs about peer-feedback provision were analysed using parametric MANOVAs followed by repeated measure ANOVAs. Following recommendations by Lakens (2013), partial eta-squared (η_p^2) is used as a measure of effect size. Values of 0.01, 0.06 and 0.14 indicate small, medium and large effects, respectively (Cohen 1988). Hypotheses 1a and 1b were tested as directional, whereas hypotheses 2a and 2b were tested as non-directional.

Results

Data inspection

The standardised skewness and kurtosis were within the acceptable range of ± 3 (Tabachnick and Fidell 2013) for all the peer-feedback provision questionnaire subscales, and there were no extreme outliers. However, the standardised skewness and kurtosis values were outside the acceptable range for the proportions of task-level peer feedback (skewness = − 4.18, kurtosis = 5.20), process-level peer feedback (skewness = 4.92, kurtosis = 6.68), self-regulation level peer feedback (skewness = 7.82, kurtosis = 13.63) and self-level peer feedback (skewness = 11.19, kurtosis = 29.33) in the pretest and for self-regulation level peer feedback (skewness = 3.49, kurtosis = 3.28) and self-level peer feedback (skewness = 5.77, kurtosis = 6.17) in the posttest. Seven univariate outliers were identified: one outlier for task-level peer feedback, one for process-level peer feedback, one for self-regulation level peer feedback and two outliers for self-level peer feedback in the pretest. One univariate outlier was identified for self-regulation level peer feedback and one for self-level peer feedback in the posttest. The outliers were checked to ensure that they were actual values. All outliers were retained as non-parametric tests were used to analyse the peer-feedback level variables.

Peer-feedback levels after the training

We performed a 2 × 3repeated measure MANOVA with domain knowledge and measurement occasion (pretest vs. posttest) as the independent variables, and peer-feedback levels in terms

of task, process, self-regulation and self as the dependent variables. According to the ATS, there was a significant three-way interaction between domain knowledge, measurement occasion and peer-feedback levels, $F(4.26, \infty) = 4.16$, $p = .001$. There were significant interactions between domain knowledge and peer-feedback levels, $F(3.29, 49.85) = 3.50$, $p = .019$; and between measurement occasion and peer-feedback levels, $F(2.49, \infty) = 9.62$, $p = .000$. There was no significant interaction between domain knowledge and measurement occasion, $F(1.84, \infty) = 0.04$, $p = .955$. There was a significant main effect for peer-feedback type, $F(2.07, 49.85) = 283.38$, $p = .000$; and no significant main effects for measurement occasion, $F(1, \infty) = 0.05$, $p = .825$; and for domain knowledge, $F(1.63, 49.85) = 0.01$, $p = .972$.

Follow-up 2×3 repeated measure ANOVAs with domain knowledge and measurement occasions as the independent variables and peer-feedback type as the depended variable were computed for each peer-feedback type with Bonferroni corrections.

Peer feedback at task-level There was a significant interaction between domain knowledge and measurement occasion, $F(1.86, \infty) = 4.99$, $p = .032$ (Table 3, Fig. 3a). There was no significant main effect for measurement occasion, $F(1, \infty) = 4.07$, $p = .128$ and no significant main effect for domain knowledge, $F(1.88, 34.87) = 3.95$, $p = .124$, on peer feedback at the task-level. Post hoc multiple comparisons with Bonferroni correction revealed a significant difference between low domain knowledge students ($M_{Rank} = 57.40$, $SD = 0.33$) and medium domain knowledge students ($M_{Rank} = 23.93$, $SD = 0.31$), $F(1, \infty) = 11.78$, $p = .002$. No significant differences were found between low domain knowledge students and high domain knowledge students ($M_{Rank} = 36.00$, $SD = 0.46$), $F(1, \infty) = 2.87$, $p = .273$ or between medium and high domain knowledge students, $F(1, \infty) = 1.21$, $p = .815$.

Peer feedback at process-level There was no significant interaction between domain knowledge and measurement occasion, $F(1.99, \infty) = 1.65$, $p = .772$. Similarly, there were no significant main effects of domain knowledge, $F(1.95, 37.35) = 1.82$, $p = .708$ or measurement occasion, $F(1, \infty) = 0.45$, $p = 1$, on peer feedback at the process level (Table 3, Fig. 3b).

Table 3 Means and standard deviations of peer feedback levels (task, process, self-regulation and self) before (pretest) and after (posttest) the training for each domain knowledge group

Peer feedback level	Group	n	Pretest		Posttest	
			M	SD	M	SD
Task	Low	15	0.87	0.17	0.94	0.12
	Medium	14	0.83	0.22	0.66	0.18
	High	14	0.86	0.18	0.76	0.23
Process	Low	15	0.08	0.15	0.04	0.10
	Medium	14	0.14	0.23	0.16	0.15
	High	14	0.08	0.14	0.10	0.13
Self-regulation	Low	15	0.05	0.11	0.02	0.04
	Medium	14	0.03	0.08	0.17	0.19
	High	14	0.01	0.04	0.12	0.13
Self	Low	15	0	0	0.01	0.03
	Medium	14	0	0	0.01	0.04
	High	14	0.04	0.07	0.03	0.05
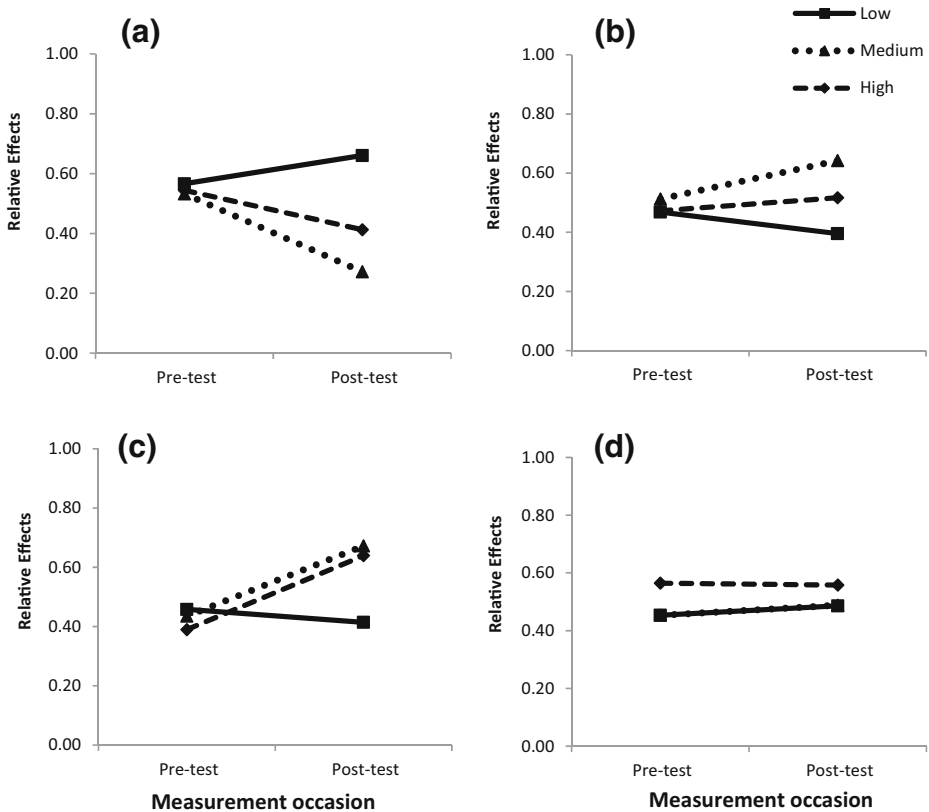
Fig. 3 Relative effects of peer-feedback levels: **a** task, **b** process, **c** self-regulation and **d** self, for each domain knowledge group (low, medium and high) before (pretest) and after (posttest) the peerfeedback training

Peer feedback at self-regulation level There was a significant interaction between domain knowledge and measurement occasion, $F(1.91, \infty) = 4.72$, $p = .040$. There was a significant main effect of domain knowledge, $F(1.90, 35.38) = 2.29$, $p = .472$ and of measurement occasion, $F(1, \infty) = 11.16$, $p = .004$, on peer feedback at the self-regulation level (Table 3, Fig. 3c). Post hoc multiple comparisons with Bonferroni correction showed a significant difference between low domain knowledge students ($M_{Rank} = 36.10$, $SD = 0.25$) and high domain knowledge students ($M_{Rank} = 55.50$, $SD = 0.40$), $F(1, \infty) = 9.02$, $p = .008$; and between low domain knowledge and medium domain knowledge students ($M_{Rank} = 58.32$, $SD = 0.44$), $F(1, \infty) = 5.90$, $p = .045$. No significant difference was found between medium and high domain knowledge students, $F(1, \infty) = 0.01$, $p = 1$.

Peer feedback at self-level There was no significant interaction between domain knowledge and measurement occasion, $F(1.82, \infty) = 0.32$, $p = 1$. Similarly, there was no significant main effect for domain knowledge, $F(1.30, 18.81) = 2.60$, $p = .468$ or for measurement occasion, $F(1, \infty) = 0.74$, $p = 1$, on peer feedback at the self-level (Table 3, Fig. 3d; note that lines for low and medium domain knowledge groups are superimposed in the figure).

Changes in beliefs about peer-feedback provision

We performed a 2×3 MANOVA with domain knowledge and measurement occasion (pretest vs. posttest) as the independent variables and peer-feedback provision beliefs (LPF, CPF and RPF) as the dependent variables. The results revealed significant multivariate main effects for measurement occasion, Pillai's Trace $= 0.337$, $F(1, 40) = 20.31$, $p = .000$, $\eta_p^2 = 0.34$ and for peer-feedback provision beliefs, Pillai's Trace $= 0.39$, $F(2, 39) = 12.64$, $p = .000$, $\eta_p^2 = 0.34$. There were no significant multivariate interactions between measurement occasion and domain knowledge, Pillai's Trace $= 0.01$, $F(2, 40) = 0.17$, $p = .844$, $\eta_p^2 = 0.01$; peer-feedback provision beliefs and measurement occasion, Pillai's Trace $= 0.02$, $F(2, 39) = 0.47$, $p = .627$, $\eta_p^2 = 0.02$; peer-feedback provision beliefs and domain knowledge, Pillai's Trace $= 0.01$, $F(4, 80) = 0.11$, $p = .978$, $\eta_p^2 = 0.01$; or between peer-feedback provision beliefs, domain knowledge and measurement occasion, Pillai's Trace $= 0.06$, $F(4, 80) = 0.66$, $p = .620$, $\eta_p^2 = 0.03$.

Follow-up 2×3 repeated measure ANOVAs with Bonferroni correction were performed for each peer-feedback provision belief as the dependent variable and measurement occasion (pretest, posttest) and level of domain knowledge (low, medium and high) as the independent variables. There were significant main effects of measurement occasion on LPF, $F(2, 40) = 10.93$, $p = .006$, $\eta_p^2 = 0.21$; CPF, $F(2, 40) = 9.81$, $p = .009$, $\eta_p^2 = 0.20$; and on RPF, $F(2, 40) = 26.19$, $p = .000$, $\eta_p^2 = 0.40$ (Table 4). There were no significant interactions between domain knowledge and measurement occasions on LPF, $F(2, 40) = 0.80$, $p = 1$, $\eta_p^2 = 0.04$; for CPF, $F(2, 40) = 0.16$, $p = .851$, $\eta_p^2 = 0.008$; and for RPF, $F(2, 40) = 0.33$, $p = 1$, $\eta_p^2 = 0.02$. There were no significant main effects of domain knowledge on LPF, $F(2, 40) = 1.23$, $p = .690$, $\eta_p^2 = 0.07$; CPF, $F(2, 40) = 0.94$, $p = 1$, $\eta_p^2 = 0.04$; and RPF, $F(2, 40) = 1.70$, $p = .60$, $\eta_p^2 = 0.07$. Students' beliefs about peer-feedback provision significantly decreased after the training regardless of the levels of their domain knowledge (see Table 4).

Discussion

This study investigated (a) whether preservice mathematics teachers, with different levels of domain knowledge, benefited differentially from a structured peer-feedback training and (b)

Table 4 Means and standard deviations of beliefs about peer feedback provision for each domain knowledge group before (pretest) and after (posttest) the training

	Low		Medium		High	
	n	M (SD)	n	M (SD)	n	M (SD)
LPF						
Pretest	15	4.13 (0.75)	14	4.33 (0.48)	14	4.55 (0.57)
Posttest		3.90 (0.82)		3.78 (0.99)		4.23 (0.64)
CPF	15	3.74 (0.59)	14	3.84 (0.60)	14	4.03 (0.64)
Pretest						
Posttest		3.28 (0.82)		3.54 (0.93)		3.68 (1.00)
RPF	15	4.15 (0.38)	14	4.29 (0.61)	14	4.29 (0.49)
Pretest						
Posttest		3.72 (0.48)		3.77 (0.97)		4.11 (0.52)

LPF learning from peer feedback provision, *PCF* confidence about peer feedback provision, *RPF* engaging in reasoning during peer feedback provision

how students' beliefs about peer-feedback provision changed in response to the training. The students received training on providing peer feedback at different levels (i.e. task, process and self-regulation) and practised providing written peer feedback on two fictional peer solutions with the help of three instructional scaffolds (worked example, feedback provision prompts, evaluation rubric).

Improvement of peer-feedback levels

The results indicated an increase in peer feedback provided at the highest level (i.e. self-regulation), but only for medium and high domain knowledge students (hypothesis 1a was partially supported; hypothesis 1b was not supported). This finding suggests that engaging preservice teachers in peer-feedback provision activities structured with instructional scaffolds can help them process solutions to geometric construction tasks at deeper levels. However, domain knowledge appears to be a prerequisite as suggested by Van Zundert et al. (2012b). This is also supported by the finding that low domain knowledge students ended up providing more peer feedback at the task-level. Even when trained with a set of instructional scaffolds including a worked example, providing higher levels of peer feedback seems to be challenging for low domain knowledge students.

In contrast to the study by Gan and Hattie (2014), peer feedback at the process-level did not significantly increase after the training in our study. One explanation might be the type of task. In Gan and Hattie's (2014) study, the object of peer feedback was a chemistry lab report of an experiment which might elicit procedural comments to a larger extent. Conversely, in the case of geometric constructions in our study, many processes and strategies are rather implicit and therefore might be harder to provide peer feedback on. A second explanation might be the research sample. Gan and Hattie's (2014) conducted their study with high school students, whereas our study was conducted with preservice teachers who can be regarded as having more domain knowledge than high school students and might not have realised the importance of providing process-related peer feedback to their peers. This difference requires empirical investigation by comparing the peer feedback preservice mathematics teachers provide to that provided by high school students on geometric construction tasks.

Although no significant differences were found between medium and high domain knowledge students in the process and self-regulation peerfeedback levels after the training, medium domain knowledge students tended to descriptively provide slightly more peer feedback at those higher levels. Furthermore, only the medium domain knowledge group provided significantly less task-level peer feedback than the low domain knowledge group after the training. These findings may suggest that medium domain knowledge students benefited most from the structured training in providing more feedback at the higher levels. However, the relatively lower amount of peer feedback at the process and self-regulation levels provided by the high domain knowledge students does not necessarily indicate that they did not process the peer solution deeply. It might be the case that self-regulation level peer feedback has already been internalised by students with high domain knowledge, and they might not consider it relevant to provide peer feedback at that level (i.e. akin to experts' automatization of cognitive processes; see Nathan et al. 2001). Alternatively, they might not be able to verbalise self-regulation peer feedback as their mastery of performing the construction task might preclude them from verbalising all individual steps of the procedure (Ericsson and Charness 1994; Ericsson and Crutcher 1991). Conversely, through the process of providing feedback to a fictional peer, the medium domain knowledge students—who still have room for

improvement—might have realised the importance of the self-regulation peer feedback for their own improvement and consequently used it more than their high domain knowledge counterparts. Further research with larger sample sizes is required to explore if such differences between high and medium domain knowledge with respect to higher peer-feedback levels exist and, if so, find explanations for them.

Decrease in beliefs about peer-feedback provision

Students' beliefs about peer-feedback provision (LPF, CPF and RPF) all decreased after the training with a medium effect size (hypothesis 2a was supported). This finding is consistent with the study by Wang (2014) in which students' beliefs about the usefulness of peer feedback decreased over repeated peer-feedback activities. Moreover, Wang (2014) suggested two related factors that contributed to this decrease in perceived usefulness: lack of domain knowledge and domain skills. In the present study, this decrease was observed for all domain knowledge levels, so we can conclude that hypothesis 2b was not supported. It might be the case that the geometric construction tasks were difficult for all participants and their beliefs became less positive regardless of their domain knowledge. Another potential explanation is provided by the self-assessment literature in which it is often reported that students who are low-performing on the target skill tend to over-estimate their performance (Panadero et al. 2016). Students in this study might have over- or under-estimated their ability to provide peer feedback due to the lack of or limited previous experience with peer feedback. When they were then introduced to the peer-feedback levels and repeatedly experienced producing peer feedback at the higher levels on complex geometric construction tasks, they might have realised that providing peer feedback was more complex than they expected. The decrease in beliefs observed in this study is inconsistent with previous studies in which students' beliefs became more positive after the peer-assessment/feedback activities (e.g. Cheng and Warren 1997; Sluijsmans et al. 2004). Two factors might contribute to these contradictory findings: (a) the complexity of the peer-feedback object (i.e. essay vs. geometric constructions) (Van Zundert et al. 2012b) and (b) the type of the peer-feedback product (i.e. grade, or feedback; low or high peer-feedback levels). Furthermore, while many peer-feedback studies—including our study—investigated changes in students' beliefs about peer-feedback provision, very few studies attempted to examine the impact of students' peer-feedbackrelated beliefs on their performance or on the type of peer feedback they provide. Future research could examine (a) the impact of the complexities of the task and the peer-feedback product on students' beliefs about peer-feedback provision and (b) the impact of different peer-feedback related beliefs on the peer-feedback type and learning outcomes.

Methodological limitations

Although the training improved peer-feedback provision at higher levels, separate effects of the instructional scaffolds could not be determined. A study with a larger sample size is required to systematically vary different combinations of instructional scaffolds in peer-feedback training and then compare peer feedback between different conditions. Furthermore, students' domain knowledge was determined by testing their factual knowledge that is a good predictor of their performance in geometric tasks (Ufer et al. 2009). Yet, their meta-cognitive reasoning that is expected to be closely related to self-reflection comments (i.e. self-regulation feedback) was not tested. A combination of

domain knowledge and metacognitive measures would be an informative measure of students' knowledge level for future studies. Additionally, this study focused on peer-feedback provision skills with no direct measures of training effects on students' task-specific performance (i.e. geometric construction task). While peer-feedback skills are essential for learners, it is important to investigate if training these skills can improve task-specific learning outcomes as well. Future studies could investigate whether peer-feedback provision training fosters performance on construction tasks.

Practical implications

The findings of the current study show that domain knowledge is an important factor when it comes to the type of peer feedback (i.e. progressively higher levels). Therefore, it is important to take into account students' basic domain knowledge when designing peer-feedback activities in classrooms, as it might be challenging for low domain knowledge students to provide higher levels of peer feedback even with the help of instructional scaffolds. Teachers should not assume that the instructional scaffolds automatically solve the problem of the lack of knowledge required to provide peer feedback, especially for peer feedback at the higher levels (i.e. process and self-regulation). A progressive training in which only domain knowledge is trained followed by peer-feedback skill training seems to be more beneficial for students with low domain knowledge (Van Zundert et al. 2012a).

References

Bathke, A. C., Schabenberger, O., Tobias, R. D., & Madden, L. V. (2009). Greenhouse-Geisser adjustment and the ANOVA-type statistic: cousins or twins? *The American Statistician, 63*(3), 239–246.

Brown, G. T. L., Peterson, E. R., & Yao, E. S. (2016). Student conceptions of feedback: impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology.* doi:10.1111/bjep.12126.

Brunner, E., Munzel, U., & Puri, M. L. (1999). Rank score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis, 70*(2), 286–317. doi:10.1006/jmva.1999.1821.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*(3), 360–367.

Chen, N., Wei, C., Wua, K., & Uden, L. (2009). Effects of high level prompts and peer assessment on online learners' reflection levels. *Computers & Education, 52*(2), 283–291. doi:10.1016/j.compedu.2008.08.007.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New Jersey: Lawrence Erlbaum Associates.

Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychologist, 49*(8), 725–747. doi:10.1037/0003-066X.49.8.725.

Ericsson, K. A., & Crutcher, R. J. (1991). Introspection and verbal reports on cognitive processes—two approaches to the study of thinking: a response to Howe*. *New Ideas in Psychology, 9*(1), 57–91. doi:10.1016/0732-118X(91)90041-J.

Field, A. P. (2009). Discovering statistics using SPSS (3rd ed). Los Angeles: SAGE.

Friedrich, S., Konietschke, F., & Pauly, M. (2017). GFD: an R-package for the analysis of general factorial designs. *Journal of Statistical Software, 79*(1), 1–18. doi:10.18637/jss.v079.c01.

Gan, M. (2011). The effects of prompts and explicit coaching on peer feedback quality (Unpublished doctoral dissertation). The University of Auckland, New Zealand.

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. doi:10.1007/s11251-014-9319-4.

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.

Harris, L. R., Brown, G. T. L., & Harnett, J. (2015). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: applying Hattie & Timperley's feedback model. *Assessment in Education: Principles, Policy and Practice, 22*(2), 265–281. doi:10.1080/0969594X.2014.976541.

Hattie, J. A., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi:10.3102/003465430298487.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121. doi:10.1177/109442819800100106.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory Into Practice, 41*(1), 33–39. doi:10.1207/s15430421tip4101_6.

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.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*(1), 46–53. doi:10.3102/0013189 X15569530.

Kuzniak, A., & Houdement, C. (2001). Pretty (good) didactical provocation as a tool for teacher's training in geometry. In J. Novotná (Ed.), *Proceedings of European research in mathematics education II* (pp. 292–303). Mariánské Lázně Czech Republic: Charles University, Faculty of Education.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1–10. doi:10.3389/fpsyg.2013.00863.

Lavy, I., & Shriki, A. (2014). Engaging prospective teachers in peer assessment as both assessors and assesses: the case of geometrical proofs. *International Journal for Mathematics Teaching and Learning*. Retrieved from http://www.cimt.plymouth.ac.uk/jour-nal/lavey2pdf.

Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the feedback orientation scale (FOS). *Journal of Management, 36*(6), 1372–1405. doi:10.1177/0149206310373145.

Miyakawa, T. (2004). Reflective symmetry in construction and proving. In M. J. Høines & A. B. Berit (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics education* (pp. 337–344). Bergen: Bergen University College.

Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001). Expert blind spot: when content knowledge eclipses pedagogical content knowledge. In L. Chen (Ed.), *Proceedings of the Third International Conference on Cognitive Science* (pp. 644–648). Beijin: University of Science and Technology on China Press.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. doi:10.1080 /02602938.2013.795518.

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: an R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software, 50*(12), 1–23.

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259–271. doi:10.1016/j.learninstruc.2008.05.002.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers, 32*(3), 396–402.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133.

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of 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. doi:10.1016/j.stueduc.2013.10.005.

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The future of student self-assessment: a review of known unknown and potential directions. *Educational Psychology Review, 28*(4), 803–830. doi:10.1007/s10648-015-9350-2.

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591–614. doi:10.1007/s11251-015-9353-x.

Phielix, C., Prins, F. J., & Krischner, P. A. (2010). Awareness of group performance in a CSCL-environment: effects of peer feedback and reflection. *Computers in Human Behavior, 26*, 151–161. doi:10.1016/j.chb.2009.10.011.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337–386. doi:10.1207/s15327809jls1303_4.

Reiss, K. M., Heinze, A., Renkl, A., & Groß, C. (2008). Reasoning and proof in geometry: effects of learning environment based on heuristic worked-out examples. *ZDM Mathematics Education, 40*(3), 455–467. doi:10.1007/s11858-008-0105-0.

Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. doi:10.1111/cogs.12086.

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation, 53*, 29–40. doi:10.1016/j.stueduc.2017.02.003.

Schoenfeld, A. H. (1986). On having and using geometric knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics* (pp. 225–264). Hillsdale,NJ: Erlbaum.

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education, 20*(4), 338–355. doi:10.2307/749440.

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 and Teaching International, 41*(1), 59–57. doi:10.1080/1470329032000172720.

Strijbos, J. W., & Müller, A. (2014). Personale faktoren im feedbackprozess. In H. Ditton & A. Müller (Eds.), *Feedback und rückmeldungen: theoretische grundlagen, empirische befunde, praktische anwendungsfelder [Feedback and evaluation: theoretical foundations, empirical findings, practical implementation]* (pp. 87–134). Münster: Waxmann.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: what are they talking about? *Computers & Education, 46*(1), 29–48. doi:10.1016/j.compedu.2005.04.002.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Essex, England: Pearson.

Tapan, M. S., & Arslan, C. (2009). Preservice teachers' use of spatio-visual elements and their level of justification dealing with a geometrical construction problem. *US-China Education Review, 6*(3), 54–60.

Ufer, S., Heinze, A., & Reiss, K. (2008). Individual predictors of geometrical proof competence. In Figueras, O., Cortina, J.L., Alatorre, S., Rojano, T., Sepulveda, A. (Eds.), *Proceedings of the Joint Meeting of PME 32 and PME-NA XXX, Vol. 4* (pp. 361–368). Mexico: Cinvestav-UMSNH.

Ufer, S., Heinze, A., & Reiss, K. (2009). Mental models and the development of geometric proof competency. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education, Vol. 5* (pp. 257–264). Thessaloniki: PME.

Van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2012a). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies, 38*(5), 541–557. doi:10.1080/03055698.2012.654920.

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

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: a longitudinal inquiry. *Assessing Writing, 19*, 80–96. doi:10.1016/j.asw.2013.11.008.

Maryam Alqassab. Department of Psychology, LMU Munich, Germany Leopoldstr. 44, 80802 Munich, Germany. E-mail:maryam.alqassab@psy.lmu.de, website: http://www.en.mcls.lmu.de/study_programs/reason/people/phdstudents/alquassab/index.html

Current Research themes:

Peer assessment and peer feedback (PF). Role of Individual factors in PF activities. Mathematical argumentation and proofs

Most relevant publications:

Alqassab, M., Strijbos, J., & Ufer, S. (2015, August). *The impact of peer feedback on mathematical reasoning: role of domain-specific ability and emotions.* Paper presented at the 16th Biennial EARLI Conference for Research on Learning and Instruction, Limassol, Cyprus.

Alqassab, M., & Strijbos, J. W. (2016, August). *The diversity of peer assessment revisited.* Poster presented at the 8th Biennial Conference of EARLI SIG 1: Assessment & Evaluation, Munich, Germany.

Alqassab, M., Strijbos, J. W., & Ufer, S. (2016). Exploring the composition process of peer feedback. In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P.(Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS), Volume 2* (pp.862-865). Singapore: International Society of the Learning Sciences.

Prof. Dr. Jan-Willem Strijbos. Faculty of Behavioural and Social Sciences, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands. E-mail: j.w.strijbos@rug.nl, website: http://www.rug.nl/staff/j.w.strijbos/

Current Research themes:

Assessment and Feedback. Technology-Enhanced Learning (TEL). (Computer-Supported) Collaborative Learning (CS)CL. Peer assessment. Peer feedback. Feedback dialogues. Communities of learners/communities of practice. Instructional design. Analysis methods for (CS)CL. Content analysis of discourse

Most relevant publications:

Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, 31(5), 422-434. doi: 10.1111/jcal.12091

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. doi: 10.1016/j.stueduc.2013.10.005

Strijbos, J-W., Narciss, S., & Duennebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291-303. doi: 10.1016/j.learninstruc.2009.08.008

Prof. Dr. Stefan Ufer Mathematics Institute, LMU Munich, Theresienstr. 39, 80333 Munich, Germany. E-mail: ufer@math.lmu.de, website: http://www.mathematik.unimuenchen.de/~didaktik/index.php?ordner=ufer&data=start

Current research themes:

Mathematics education. Mathematical argumentation and proofs. Secondary-tertiary transition

Most relevant publications:

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research 2* (2), 28-45.

Sommerhoff, D., Ufer, S., Kollar, I. (2016). Proof Validation Aspects and Cognitive Student Prerequisites in Undergraduate Mathematics. In Csikos, C., Rausch, A., & Szitanyi, J. (Eds.): *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education,* Vol. 4, pp. 219-226. Szeged, Hungary: PME.

Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: the role of transactivity. *Instructional Science 44*(5), 477-500.