Learning and Instruction 105 (2026) 102413

Contents lists available at ScienceDirect

Learning and Instruction

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

Feedback complexities interact with prior knowledge

Fynn Töllner *, Poldi Kuhl, Michael Besser

Leuphana Universität Lüneburg, Universitätsallee 1, 21335, Lüneburg, Germany

ARTICLE INFO

Keywords:
Formative assessment
Feedback
Achievement
Conceptual knowledge
Procedural knowledge

ABSTRACT

Background: Formative assessment is a promising option for addressing heterogeneity in schools. Hereby, feedback is both a core element of formative assessment and beneficial for learning. Regarding its effectiveness, we already know that elaborated feedback is particularly conducive to learning. However, it is unclear what exactly constitutes effective elaborated feedback.
Aims: We investigated the effects of different task-level forms of elaborated feedback, namely feedback based on conceptual or procedural knowledge and knowledge of result feedback on students' achievement.
Sample: Participants were $N = 190$ seventh-grade students from seven secondary schools in Lower Saxony, Germany.
Methods: We conducted a randomized controlled study in a pretest/posttest design. The participants solved math tasks by adding fractions. After incorrect responses, they received feedback. We randomly assigned them to receive elaborated conceptual feedback ($n = 48$), elaborated procedural feedback ($n = 46$), knowledge of result feedback ($n = 50$), or no feedback ($n = 46$). The no-feedback group was the control group.
Results: After controlling for students' prior knowledge, only elaborated feedback was more effective than no feedback, with no significant differences found between the two types of elaborated feedback and the knowledge of result feedback. Additionally, we observed an interaction effect, showing a reduced effect of prior knowledge on student achievement in the conceptual feedback group.
Conclusions: These results imply that feedback can address learner heterogeneity, especially in its elaborated form. Hereby, it is especially important for learners with low prior knowledge to receive adequate feedback, as learners with less prior knowledge especially benefit from conceptual feedback.

1. Introduction

Formative assessment is often referred to as a productive way of promoting student learning (Black & Wiliam, 1998; Hattie, 2012; Kingston & Nash, 2011). This concept consists of two main elements: (1) the diagnosis of learners' performance, and (2) appropriate feedback based on this diagnosis (Andrade, 2010; Heritage, 2007; Mitchell, 2014). Formative assessment therefore stands in contrast to summative assessment, which only reports the status quo (Dolin et al., 2018). In respect to formative assessment, the literature provides several definitions of the term (Filsecker & Kerres, 2012; Yorke, 2003). However, there is insufficient research on the circumstances under which formative assessment is effective (Lui & Andrade, 2022).

Nevertheless, there is a broad consensus that the main purpose of formative assessments is to provide feedback to enhance learning (Timmers et al., 2013) because it informs learners about their current level of achievement in order to improve it (Narciss et al., 2014). This assumed relevance of the feedback component is underpinned by empirical data (Decristan et al., 2015). Regarding feedback, the current state of research suggests that its effectiveness depends on its complexity (Mertens et al., 2022) and learners' prior knowledge (e.g., Fyfe, 2016; Heckler & Mikula, 2016; Lui & Andrade, 2022). As students' heterogeneity in the classrooms increases, posing a challenge for teachers (Borsch, 2018; Nusser & Gehrer, 2020; Parsons et al., 2018), the question arises as to how feedback in heterogeneous learning environments should be provided to adequately address all learners. Therefore, our study examined: (1) the effect of providing feedback, (2) a comparison between non-elaborated and elaborated feedback, (3) the effect of different types of elaborated feedback in heterogeneous learning environments, and (4) potential interactions with learners' prior knowledge.

* Corresponding author.
E-mail addresses: fynn.toellner@leuphana.de (F. Töllner), poldi.kuhl@leuphana.de (P. Kuhl), michael.besser@leuphana.de (M. Besser).

https://doi.org/10.1016/j.learninstruc.2026.102413
Received 6 October 2025; Received in revised form 9 April 2026; Accepted 29 May 2026
Available online 12 June 2026

2. Theoretical background

2.1. Providing useable high-quality feedback

An extensive number of studies investigated the characteristics of effective feedback, with an increase in recent years. Despite all these efforts, there are still calls for more research on how feedback should be delivered (Winstone & Nash, 2023). In this regard, previous research suggests that the content and the quality of the feedback message are most important for students' achievement (Brandmo & Gamlem, 2025). To promote achievement by the provision of high-quality feedback, the feedback message has to be perceived as appropriate (Ruwe & Kuklick, 2026), has to explain why it was given, and should be accompanied by information on how to improve the task at hand (Fong et al., 2024; Van Blankenstein, Dirkx & de Bruycker, 2025). More specifically, feedback should identify incorrect answers and provide a detailed explanation on what is wrong along with help to address the issue (Gao et al., 2025).

A common challenge when providing feedback is that it is often ignored by learners (Jonsson & Panadero, 2018; Maier & Klotz, 2025). In this regard, Tärning et al. (2020) were able to show that learners stop using feedback at various points, with a considerable number of students reading the feedback but not acting upon it. This can be explained in multiple ways: One reason might be that the amount of information needed for a high-quality feedback exceeds the processing resources of the recipients (Ruwe & Kuklick, 2026). The lack of clearness in the feedback message, too late provision, as well as not focussing on enhancing of learning or providing helpful guidance are additional reasons for not using feedback (Paris, 2022).

In order to address the challenge of lacking feedback uptake, recent research has focused on feedback processing and identified elements that have an impact on learning via the provision of feedback (Maier & Klotz, 2025). This led to the development of two feedback processing models (Lipnevich & Smith, 2022; Lui & Andrade, 2022) who differentiate between motivational and cognitive characteristics of learners that are relevant for the effectiveness of feedback. Moreover, the feedback message seems to play a crucial role, since the message is particularly relevant to achieve a positive effect on learning outcomes (Kuklick & Lindner, 2025; Lipnevich & Smith, 2022). A previously published article argues that students are able to differentiate between high- and low-quality feedback (Ruwe & Kuklick, 2026). Accordingly, it is emphasized that the quality of the feedback messages is the best predictor for feedback uptake (Fong et al., 2024; van Blankenstein et al., 2025) and it seems like that feedback which is perceived as more constructive could also be more likely to be actually used (Fong et al., 2021, 2024). Stressing the importance of research on feedback messages which are of high quality and thus perceived as constructive.

2.2. Effectiveness of feedback complexities

Feedback is considered a highly relevant factor in learning and can be used to enhance both higher and lower order learning (Mertens et al., 2022), as it helps to improve learning and to reduce mistakes (e.g. Schiller et al., 2024; Shirah & Sidney, 2023). According to Hattie and Zierer (2019), feedback yields a large effect size of $d = 0.70$. Their meta-analysis also reports a considerable amount of variance regarding the effectiveness of feedback. This can be attributed to the numerous ways in which feedback can be provided. To accomplish a large effect on achievement, one option is to refer the feedback to the task level. In addition, feedback should identify errors and provide information on how to address them (Hattie & Zierer, 2019). Altogether, *providing feedback on tasks is considered to have a positive impact on learning.*

Feedback can be categorized according to complexity, which denotes the knowledge included in the message. A distinction is often made between non-elaborated types–e.g., "knowledge of result" (KR), which only consists of whether an answer is correct or incorrect–and elaborated types, which offer additional knowledge (Narciss, 2006; Shute, 2008). Previous research on knowledge provided in feedback messages has primarily focused on these complexities and their effects. Current research indicates elaborated feedback is particularly conducive to learning and yields larger cognitive effects than non-elaborated feedback (Mertens et al., 2022). For example, in addition to the effects of differing feedback complexities on achievement, students who received elaborated feedback were more likely to correct their errors than students in a KR-feedback group (Kuklick et al., 2023). Moreover, the provision of elaborated feedback has non-cognitive benefits as it helps to reduce the demotivating effect of negative feedback (Fong et al., 2019) and positively affected the emotions triggered by the feedback (Kuklick & Lindner, 2023). Accordingly, *elaborated feedback can be expected to be more effective in terms of learning than non-elaborated feedback*.

However, elaborated feedback is not a uniform construct; rather, it encompasses a wider range of different kinds of knowledge that can be used to create feedback messages. "Elaborated" is an umbrella term about the existence of e.g. an explanation (e.g., Narciss, 2006; Shute, 2008). Furthermore, there is currently no agreement on what the elaborated part should comprise (Martínez et al., 2026). The resulting multitude of meanings leads to difficulties in estimating elaborated feedback's overall effectiveness (Mertens et al., 2022). Additionally, estimating the effectiveness of different types of elaborated feedback is challenging, as previous research has mostly neglected to compare different types of content in elaborated feedback messages. Instead, only one type of elaborated feedback has been typically investigated, with sometimes varying timing (e.g., Attali & van der Kleij, 2017; Fyfe, 2016; Heckler & Mikula, 2016; Kuklick et al., 2023). Notable exceptions are studies that e.g. focussed on different ways how the feedback can be designed (e.g. Kuklick & Lindner, 2025; Steinbach et al., 2025).

However, to the best of our knowledge, there have not yet been comparisons regarding the content of immediate digital task level elaborated feedback. One exception is a study by Martínez et al. (2026) that compared elaborated feedback which either explained why an answer was/would have been correct or why an answer was incorrect/other answers would have been incorrect. This makes it difficult to draw conclusions about which type of knowledge in such a feedback message works best in the context of elaborated feedback. This problem aligns with a call for further research to compare different kinds of elaborated feedback in order to gain insights on how to provide most effective elaborated feedback messages for learning (Swart, Nielen, & Sikkema-de Jong, 2019).

For the creation of a high-quality explanation, Wittwer and Renkl (2008) identified four characteristics in their framework (e.g., Bitzenbauer et al., 2024; Korntreff & Prediger, 2026). According to this framework, explanations should (1) concentrate on concepts and principles, while (2) accounting for learners' prior knowledge, and should be (3) used for further learning, but (4) only provided when they are needed.

In addition, the specificity of the feedback message needs to be considered when providing feedback. Feedback specificity refers to the extent to which the feedback provides clear and detailed information on how to improve the work (Fong et al., 2024). Here, feedback should be specific to the task in question (Koenka et al., 2021) and explain the error to the student (Martínez et al., 2026).

However, prior research has already provided some insights into the characteristics of effective elaborated feedback. Shute (2008), for example, stated that overly long or complex feedback can be useless, as some learners will simply ignore it. This indicates that the knowledge communicated by the feedback must be carefully chosen. Elaboration per se is not what makes the feedback effective; rather, it is how this elaboration is created. This leads to the conclusion that *varying types of elaborated feedback differ also in their effectiveness*.

It is important to note that elaborated feedback is not equally effective for every learner. Research has shown it is particularly important for learners with low prior knowledge (e.g., Fyfe, 2016; Heckler & Mikula, 2016). Consistent with this finding, research has shown that learners with a higher level of prior knowledge benefit more from feedback that contains only the correct answer and basic procedure for solving the task than from elaborated feedback (Smits et al., 2008). Thus, the relevance for feedback providers to know how different types of feedback affect learning is emphasized (Panadero, 2023).

Accordingly, learners' prior knowledge is not only a factor influencing the feedback outcome (e.g. achievement) (Lipnevich & Smith, 2022; Lui & Andrade, 2022), but it is also crucial for the effectiveness of elaborated feedback. This is further underlined by the fundamental relevance of prior knowledge for students' further learning success (Simonsmeier et al., 2022). Moreover, recent research has called for feedback interventions, especially for low performing learners (Maier & Klotz, 2025), emphasizing the importance of investigating the influence of elaborated feedback on learners with low prior knowledge. Overall, this means that *the effect of feedback differs based on learners' prior knowledge.*

2.3. Knowledge types for elaborated feedback

Different types of knowledge can be used in feedback messages to provide elaborated feedback to support learners adequately. A common distinction between knowledge types in learning, teaching, and assessment can be found in the revised version of Bloom's taxonomy of educational objectives (Anderson & Krathwohl, 2001; Bloom, 1956). The taxonomy comprises metacognitive, factual, conceptual, and procedural knowledge (Anderson & Krathwohl, 2001). *Metacognitive knowledge* relates to cognition in general, and involves strategic knowledge, knowledge of cognitive tasks, and self-knowledge (Anderson & Krathwohl, 2001). *Factual knowledge* includes, e.g., terminology or vocabulary without considering the corresponding connections, and can be described as forming the fundamental knowledge of a topic. In mathematics education, for example, this can involve fraction-related terms, such as numerator and denominator.

In contrast, *conceptual knowledge* involves connections among the underlying principles necessary to perform tasks. This type of knowledge focuses on classifications, principles, generalizations, theories, models, and structures (Anderson & Krathwohl, 2001). Furthermore, flexibility and the resulting problem-independence of this knowledge is emphasized (Rittle-Johnson et al., 2001), as well as the necessity for in-depth engagement (Haapasalo & Kadijevich, 2000).

Finally, *procedural knowledge* refers to knowledge of how something is done. This includes subject-specific skills, algorithms, techniques, methods, and knowledge of when to apply such procedures (Anderson & Krathwohl, 2001). This type of knowledge highlights action sequences and their application, as it depends on a certain problem (Rittle-Johnson et al., 2001). Moreover, the application of this knowledge often does not require a high level of attention (Haapasalo & Kadijevich, 2000).

In the academic discourse, there has been a long-standing debate on the relationships between and the developmental order of learners' *conceptual* and *procedural knowledge* (Lenz & Wittmann, 2021; Rittle-Johnson et al., 2015). Overall, the most widely accepted view is that learners' conceptual and procedural knowledge develop together in an iterative process (Rittle-Johnson & Schneider, 2015). In addition, there are debates in various subjects about the role of these types of knowledge. Mathematics is particularly noteworthy in this regard, as these two types of knowledge are considered particularly relevant in this subject (Hiebert, 1986; Rittle-Johnson & Schneider, 2015; Xu et al., 2025). This leads to the conclusion that research involving these types of knowledge seems especially relevant in this context. Furthermore, conceptual and procedural knowledge in mathematics education are considered theoretically and empirically distinct types of knowledge, but it has not yet been clearly established how these types of knowledge influence one another and how they develop (Lenz et al., 2020; Xu et al., 2025).

To the best of our knowledge, little research has been conducted on the differential effects of both types of knowledge when used in feedback messages. Previous research has focused on instructions (e.g., Rittle-Johnson et al., 2016), explanations (e.g., Lachner et al., 2019; Perry, 1991), and the measurement of these knowledge types (e.g., Lenz et al., 2020; Schneider & Stern, 2010). Existing research on feedback has focused on tutoring feedback strategies, where elaborated feedback was offered in combination with mastery learning strategies and tutoring (e.g., Narciss, 2013, 2017). Narciss et al. (2014) compared various sequences of hints and explanations provided with conceptual or procedural knowledge in feedback messages. They found no significant effect of these sequences on the learners' knowledge. Accordingly, the most widespread perspective on the developmental order of conceptual and procedural knowledge (Rittle-Johnson & Schneider, 2015) cannot be translated into guidelines for feedback provision.

In addition, a study which conducted a cluster analyses based on students' individual conceptual or procedural knowledge concluded that it does not seem plausible to base adaptive support strategies on the relative presence of these two knowledge types (Lenz & Wittmann, 2021). Because neither the sequenced administration of these two types of knowledge in feedback messages nor individual adaptation appears to be a promising approach, no conclusions can yet be drawn regarding the superiority of one or the other type of knowledge as a message in elaborated feedback. Correspondingly, we can state that for conceptual and procedural knowledge, the isolated effect in elaborated feedback and how learners can be supported with the provision of these knowledge types through feedback is currently understudied.

3. The present research

In the present study, we investigated which type of knowledge in feedback is effective and whether learners benefit differently depending on their prior knowledge. To this end, we examined the effects of feedback and different feedback messages. The feedback messages differed in their complexity and type of knowledge presented. With this study, we contribute to the literature on the effectiveness of different feedback types and consider individual prior knowledge as a factor that influences the effect of feedback. To this end, we address the following research questions (RQs):

RQ1: Are all types of feedback more conducive to learning than receiving no feedback?
RQ2: Is elaborated feedback more effective in terms of learning than non-elaborated feedback?
RQ3: Does elaborated feedback differ in its effectiveness depending on whether it was created with conceptual or procedural knowledge?
RQ4: Does the effectiveness of the feedback types differ depending on learners' prior knowledge?

4. Method

This study is part of the research project LERN-IF ("Learning process diagnostics and learning supportive performance feedback in inclusive subject teaching"). LERN-IF is funded by the German Federal Ministry of Education and Research (BMBF; funding reference: 01NV2117). Participation in the study was voluntary. Parental consent forms were obtained, and the study was reviewed and approved by the responsible school authority board.

To enhance transparency and ensure reproducibility, and to comply with our funding agency's requirements, we archived the data required for reproducing our analyses at our university's institutional repository (Töllner et al., 2026).

4.1. Study design

To answer our research questions, we conducted an intervention study between January and May 2024 using a pre/post between-subjects design with two measurement times (Time1/Time2). The two measurements were seven days apart. We collected the data using iPads and iSpring Suite software. All participants attended seventh grade and completed fraction addition tasks. According to the curriculum, this topic had already been covered in the previous school year (Grade 6) in math education.

At Time 1, we used a fractional calculation test to assess the students' prior knowledge. Moreover, we collected the students' demographic data. At the beginning of Time 2, one week later, we randomized on the student level, by assigning students randomly to one of four groups (no feedback, KR feedback, elaborated procedural feedback, and elaborated conceptual feedback). During randomization, we ensured that the four groups were roughly equally represented in all schools.

The students worked in a feedback phase on 26 fraction calculation tasks on iPads in their respective groups. Except for the control group that received no feedback at all, students received immediate task-level feedback according to their group affiliation. Firstly, members of all those groups with feedback (KR feedback, elaborated procedural feedback, and elaborated conceptual feedback) received KR Feedback after each task. Secondly, if they had answered incorrectly, participants in the two elaborated feedback groups, received, in addition, an elaborated feedback according to their group affiliation. Thirdly, in these three groups, students had the chance to correct their answer after an incorrect response. Fourthly, if the students in these groups again answered incorrectly, feedback according to their group membership was provided once again. This time feedback was coupled with the correct solution and a new task was assigned. This gave students the opportunity to use feedback immediately, which made it more useful for them (Jonsson & Panadero, 2018). At the end of Time 2, we administered another fraction calculation test as a posttest.

4.2. Sample

4.2.1. Power analysis

We conducted a priori power analyses using G*Power (Faul et al., 2007) for an analysis of variance (ANOVA) repeated measures, between-factor design with two measurement times (Time1/Time 2) and four groups (no feedback, KR feedback, elaborated procedural feedback, elaborated conceptual feedback). We wanted to detect small effects of about $f = 0.20$ with a high statistical power of 0.95. To do this, the power analysis suggested a sample size of $N = 328$ participants. We chose to aim for small effects as we included multiple types of feedback and wanted to be able to detect the effects of the groups in isolation. We based our assumption on previous research: According to Van der Kleij et al. (2015) KCR feedback yields a small effect size, so it seemed reasonable to look for such an effect as this was a commonality between the groups. Furthermore, Kuklick et al. (2023) compared different types of feedback with regard to the error corrections made and also identified a small effect size for the comparison of elaborated feedback and KR feedback.

4.2.2. Achieved sample

A total of $N = 247$ seventh graders from $n = 7$ non-academic-track schools in Lower Saxony (Germany) participated in this study. Thus, we were unable to achieve the a priori calculated required sample size. We elaborate on this in the discussion. We excluded $n = 18$ students from the analysis because they either had to leave early ($n = 4$), had compliance issues ($n = 9$), or had to rely on a translation application to participate in the study ($n = 5$). At Time 1, $n = 196$ students participated in the study. At Time 2, $n = 190$ students participated. Table 1 provides an overview of the gender, age distribution, and sample sizes of the four feedback groups. The sample was divided approximately equally between the feedback groups in terms of these characteristics.

Table 1
Gender and age distribution across the four feedback groups.

Feedback group	n	Variables: Gender	Variables: Age
No feedback	46	20m; 18f; 1 nb; 7 missing	$M = 12{,}84$ $SD = 0{,}68$
Knowledge of result	50	21m; 18f; 11 missing	$M = 12{,}92$ $SD = 0{,}66$
Procedural feedback	46	17m; 21f; 8 missing	$M = 12{,}62$ $SD = 0{,}55$
Conceptual feedback	48	18m; 21f; 1 nb; 8 missing	$M = 12{,}80$ $SD = 0{,}60$

Note. m = male; f = female; nb = non-binary.

4.3. Instruments and materials

4.3.1. Tasks in the tests and the feedback phase

For the prior-knowledge test, feedback phase, and posttest, students worked on tasks that we developed for the study. In each task, the participants were required to add two fractions. The students answered single-choice questions with five possible answers. Four of these tasks involved common fractions, and one involved mixed fractions. We used four types of tasks that differed in their combinations of numerators and denominators. These were common fractions that: (a) differed in numerator and denominator, (b) had the same numerator but a different denominator, or (c) had a different numerator but the same denominator. For the mixed fractions, we used task (d) with the same numerator but different denominators. The feedback phase comprised 26 tasks.

Both the pre- and posttest consisted of 20 tasks. These were designed so that ten tasks were identical. The remaining ten tasks were of the same type, but the numbers were changed. We developed tasks and distractors based on common error patterns (Hwang, 2016). Cronbach's alpha was very good for both the pretest ($\alpha = .86$; 20 tasks; $n = 157$) and posttest ($\alpha = .91$; 20 tasks; $n = 190$). To assess the students' performance in the prior-knowledge test and the posttest, we computed the sum score of the correct answers, which ranged theoretically between 0 and 20 points.

4.3.2. Feedback used in our study

Depending on the feedback group, the students received no feedback, KR, or one of two types of elaborated feedback. The procedural and conceptual elaborated feedback were structured similarly; there was a short reminder at the top of the presented screen defining numerators and denominators. This served the purpose of aligning factual knowledge by explaining what these terms mean. This knowledge was necessary to understand the subsequent procedural or conceptual knowledge in the provided feedback and should prevent difficulties in understanding. Following this, there was a one-by-one explanation as stepwise feedback seems to be more motivational (Wang et al., 2019).

The *knowledge of result feedback* (Fig. 1, right side, bottom) consisted merely of informing the participants that their attempts were wrong. In the *no feedback* control condition, the students received no indication whether their answers were correct. The students simply continued to work on the tasks, as in the pretest and posttest.

The *procedural feedback* (Fig. 1, right side) explained how the procedure of adding the two fractions must be applied. For this purpose, each required calculation step was presented as a headline. Each step was accompanied by an example and a corresponding explanation. This resulted in instructions students could follow to add the fractions.

The *conceptual feedback* (Fig. 1, left side) explained why the procedure for adding the two fractions worked. The fractions were visualized using an illustration, and each illustration was explained. The first step was to obtain a graphical representation of the fractions and divide them into common divisions. In the last step, the students were told they had to move two rectangles on top of each other to add them.

4.4. Statistical analyses

All statistical analyses were conducted using R (version 4.3.2 [R Core

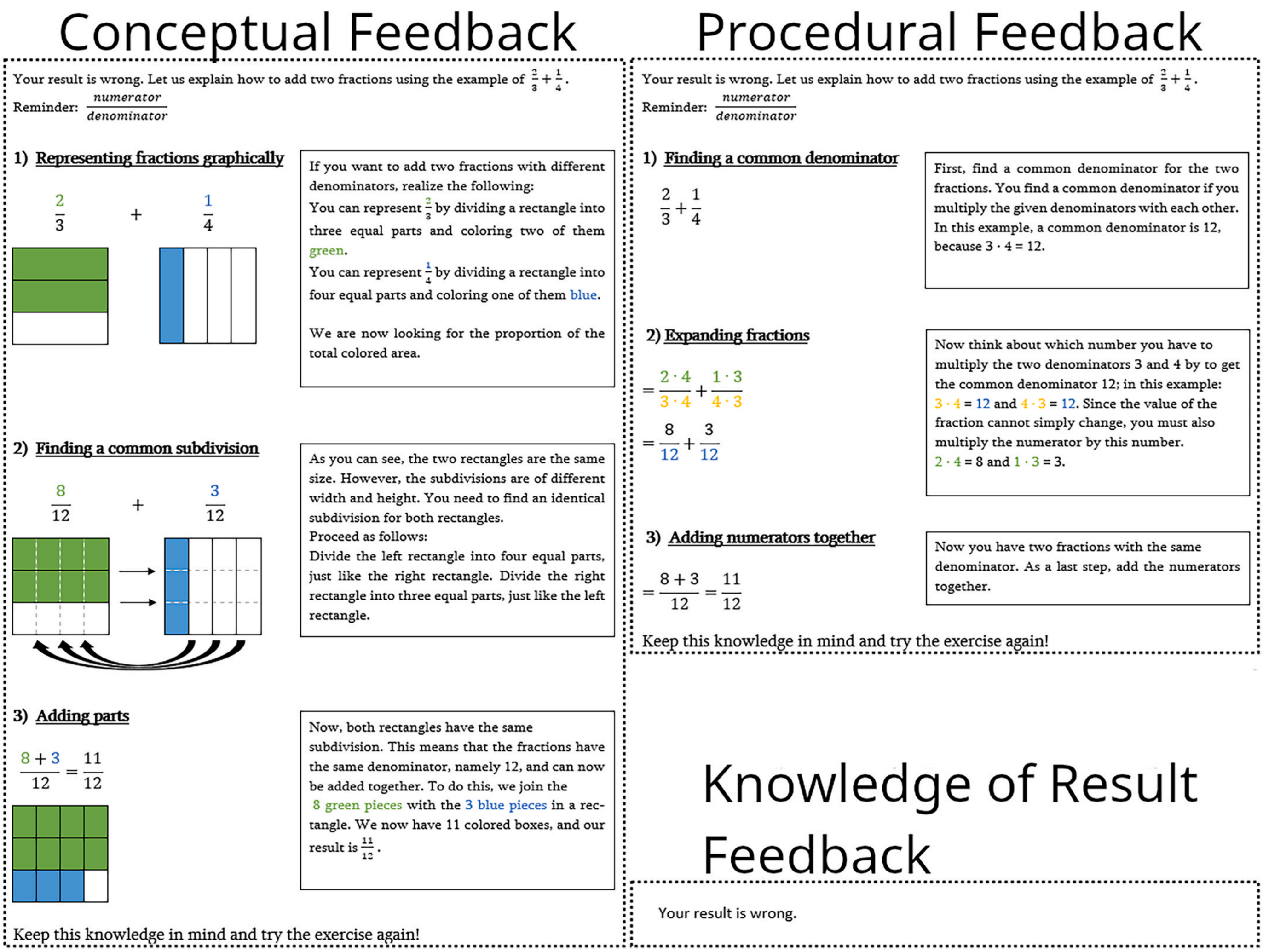

Fig. 1. Examples of conceptual and procedural knowledge in the feedback and the knowledge of result feedback.

Team, 2023]). To handle missing data, we used the MissForest (Stekhoven, 2022) algorithm and predefined the seed to 42 to ensure reproducibility. This iterative imputation method does not formulate distributional assumptions about the data (Stekhoven & Bühlmann, 2012). This algorithm outperforms other established imputation methods (Jaafar et al., 2024) and is a reliable approach for psychological research (Golino & Gomes, 2016). Furthermore, it is suitable for different types of variables and can handle data with up to 30% missing values (Stekhoven & Bühlmann, 2012). Moreover, imputation with MissForest has been proven to preserve the relationships between variables, such as interactions (Stekhoven & Bühlmann, 2012; Tiwaskar et al., 2025).

In all groups there were fewer missing values in the pretest scores (no feedback: 15%; KR: 22%; procedural: 20%; conceptual: 15%) and the posttest (no feedback: 0.02%; KR: 0.02%; procedural: 0%; conceptual: 0%), and the overall missing values were also lower (pretest: 18%, posttest 1%) than the limit would have allowed. Data imputation was based on students who were present at Time 2, when the feedback groups were assigned, imputing feedback groups did not seem plausible. Imputation was performed individually for all feedback groups with regard to data on demographics, cognitive factors, the feedback phase and feedback relevant variables (Lui & Andrade, 2022). Afterwards the distribution of the imputed pre- and posttest scores was visually inspected in comparison to the raw data using histograms.

To answer our research questions, we performed multilevel modelling using the lme4 (Bates et al., 2015) and nlme (Pinheiro, Bates & R Core Team, 2023) packages because of the multilevel data structure in which students are nested in schools. Therefore, we employed a model with students at Level One (L1) and schools at Level Two (L2). Multilevel modelling was deemed appropriate, even when the number of clusters with $n = 7$ was rather small. Even if the benefit of multilevel models is lower with a smaller number of clusters, their use is still appropriate because they are not disadvantageous (Gelman & Hill, 2007). Furthermore, we tested whether a multilevel model was needed (Field et al., 2012), which resulted in a statistically better fit to the data when the intercepts are allowed to vary ($p = .002$). However, adding varying slopes to the model did not improve the fit ($p = .748$). The statistical fit was evaluated using the Bayesian information criterion (BIC). Thus, we continue our analyses using a varying intercept-only model.

We specified the model with the no-feedback control group as the intercept and the other three groups as predictors. Moreover, we added the pretest score as another predictor and allowed for interactions between the pretest score and feedback groups. We defined the posttest achievement score as the dependent variable. Pre- and posttest values were added as raw scores; using the sum score eases the interpretation, as the raw scores translate directly to the number of correctly solved tasks. Consequently, this allows for an immediate interpretation of the models result. Another common reason for centring variables is to establish a meaningful zero point (Enders & Tofighi, 2007). As the raw scores already have such a meaningful zero point, when a student solved all tasks incorrectly, centring was not necessary. The reliability of the pretest and posttest were calculated using the Psych package (Revelle,

Table 2
Feedback groups' pretest and posttest means and standard deviations.

Feedback groups	Pretest		Posttest	
	Mean	SD	Mean	SD
No feedback	6.00	4.02	6.26	4.77
Knowledge of result	6.36	4.90	11.60	5.93
Procedural feedback	4.93	4.07	11.35	5.57
Conceptual feedback	6.94	4.94	12.38	6.10
Total	6.07	4.54	10.44	6.07

Table 3
Correlation between pre- and posttest scores for each feedback group and in total.

Feedback group	Correlation between pre- and posttest scores
No feedback	0.59***
Knowledge of result	0.43**
Procedural feedback	0.40**
Conceptual feedback	0.03
Total	0.31***

Note. *p < .05, **p < .01, ***p < .001.

2024).

The examination of the assumptions for the multilevel models showed that heteroscedasticity was present in the data. However, this is not unusual in intervention studies (Huang, Wiedermann, & Zhang, 2023). We addressed this issue by using sandwich estimators (Field & Wilcox, 2017; Pustejovsky & Tipton, 2018). This made it possible to adjust the standard errors so that they were comparable to those of large samples without violations of assumptions (Schielzeth et al., 2020). To address heteroscedasticity, we used the clubSandwich package (Pustejovsky, 2024). As part of this, we used a Satterthwaite approximation (Satterthwaite, 1946), which is preferable as this type of approximation seems to depend less on the achieved sample size (Luke, 2017). Furthermore, a CR2 estimator was used as a small-sample adjustment to calculate the variance-covariance matrix because recent publications indicate that this estimator is able to account for heteroscedasticity and has been shown to work well for different use cases, even with a small number of clusters (Huang & Li, 2022; Huang, Wiedermann, & Zhang, 2023; Huang, Zhang, & Li, 2023). For further comparison of the groups, we used the linear contrasts in the clubSandwich package (Pustejovsky, 2024) and performed Bonferroni correction. The stats package (R Core Team, 2023) was used for correlation analyses.

5. Results

5.1. Preliminary analyses

The pretest and posttest scores for the test conditions are shown in Table 2. In addition, we performed an ANOVA for the pretest score and found no initial significant differences between the feedback groups (F (3, 186) = 1.63, p = .184). However, the sample seemed to predominantly consist of learners with low prior knowledge, as Table 2 shows. On average, the students solved less than half of the maximum 20 tasks correctly in the pretest. Furthermore, Table 3 provides an overview of the correlations between the two measurement times for the individual feedback groups.

5.2. Results regarding the research questions

To answer our research questions, we set up a multilevel model as described above. Table 4 presents the results of the multi-level model. The control group without feedback represents the intercept, and the estimate can be translated into the number of correctly solved tasks in the posttest. Based on the multilevel model, we conducted a group comparison of the posttest achievement scores, as described in the Methods section. Thus, we controlled for both multilevel structure and prior knowledge (Fig. 2).

5.2.1. RQ1. Are all types of feedback more conducive to learning than receiving no feedback?

After controlling for prior knowledge and the multilevel structure, the data indicates that students who did not receive feedback and solved zero tasks correctly in the pretest answered on average 2.5 tasks correctly in the posttest. In contrast, students with zero correctly solved tasks in the pretest who were assigned to the KR feedback group solved 8.5 tasks correctly, in the procedural feedback group approx. 8 tasks, and in the conceptual feedback group 12 tasks. However, not all of these

Table 4
Effects of feedback group and prior knowledge on posttest performance.

Effect	Estimate	*SE*	95% CI		*p*
			LL	UL	
Intercept	2.509	0.694	0.738	4.280	0.015*
Knowledge of result	6.010	2.172	0.496	11.524	0.038*
Procedural feedback	5.871	0.921	3.442	8.300	0.002**
Conceptual feedback	9.483	0.723	7.501	11.465	<0.001***
Pretest score	0.623	0.136	0.227	1.019	0.013*
Knowledge of result * Pretest score	−0.157	0.375	−1.153	0.839	0.694
Procedural feedback * Pretest score	−0.066	0.189	−0.627	0.496	0.749
Conceptual feedback * Pretest score	−0.609	0.196	−1.195	−0.020	0.046*
Random effects					Variance
School (Intercept)					2.033
Residual					25.426
Number of observations					190
R^2 marginal					0.255
R^2 conditional					0.311
AIC					1174.59
BIC					1207.06
ICC					0.074
RMSE					4.88

Note. *p < .05, **p < .01, ***p < .001; SE = standard error; CI = confidence interval; LL = lower limit; UL = upper limit; AIC = akaike information criterion; BIC = Bayesian information criterion; ICC = intraclass correlation coefficient; RMSE = root-mean-square error of approximation.

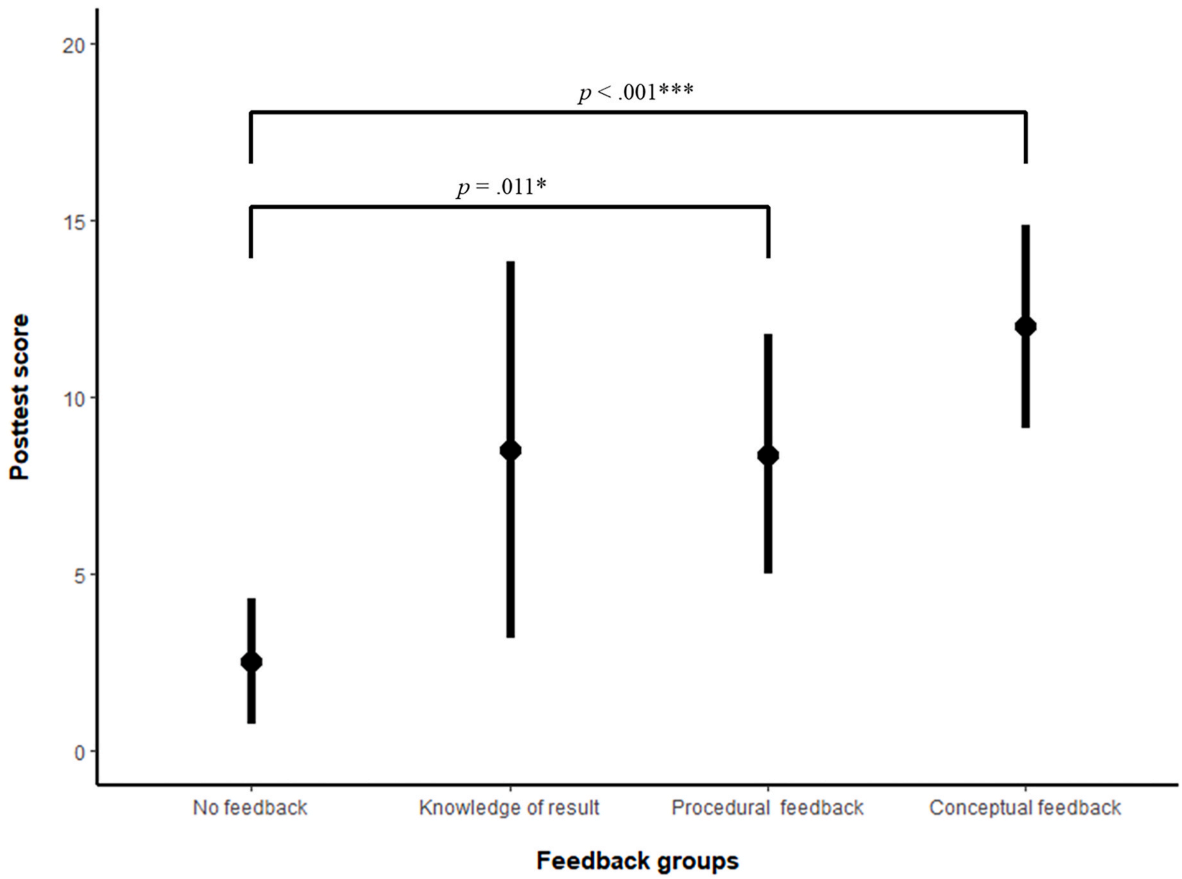

Fig. 2. Estimates and standard errors for the feedback groups' posttest scores after controlling the pretest scores.

differences were statistically significant and thus not all types of feedback were more conducive to learning than receiving no feedback at all. The KR feedback group had a relatively large standard error, resulting in a non-significant difference according to the test for testing linear contrasts $t(5.22) = 2.77$, $p = .226$, 95% CI [−2.95, 14.97] compared to the no feedback group. This relatively large standard error and the corresponding overlap with the no feedback group is shown in Fig. 2 and could mean that this group consisted of students who either benefitted a lot or not at all. Both conceptual $t(4.13) = 13.12$, $p < .001$, 95% CI [6.05, 12.91] and procedural feedback $t(4.60) = 6.38$, $p = .011$, 95% CI [1.80, 9.94], as types of elaborated feedback, were significantly more conducive to learning than providing no feedback group as the test for linear contrasts had revealed.

5.2.2. RQ2. Is elaborated feedback more effective in terms of learning than non-elaborated feedback?

In our study, elaborated feedback did not prove to be superior to KR feedback, as depicted in Fig. 2. It is striking that, as can be seen, the standard error of the procedural feedback completely overlaps with the standard error of the KR feedback, and the standard errors of the conceptual feedback and KR feedback overlaps for the most part. Accordingly, as the test for linear contrasts demonstrated, neither conceptual feedback $t(4.08) = -1.78$, $p = .888$, 95% CI [−12.79, 5.85] nor procedural feedback $t(4.64) = -0.06$, $p = 1.00$, 95% CI [−9.78, 9.50] differed significantly from KR feedback.

5.2.3. RQ3. Does elaborated feedback differ in its effectiveness depending on whether it was created with conceptual or procedural knowledge?

Regarding the provision of different types of elaborated feedback, the test for linear contrasts revealed no significant differences between procedural and conceptual feedback $t(3.46) = -2.75$, $p = .361$, 95% CI [−10.75, 3.53]. As shown in Fig. 2, both elaborated feedback groups share a noticeable amount of overlap in their standard errors.

5.2.4. RQ4. Does the effectiveness of the feedback types differ depending on learners' prior knowledge?

An analysis of the interactions of the feedback groups with the pretest score revealed a significant interaction only for the conceptual feedback group. Additionally, Table 4 shows that the estimate of this negative interaction is approximately as large as the estimate of the pretest. This means, that for the students who were assigned to the no feedback group, each subsequently correctly solved task in the pretest translated into 0.623 additionally correct solved tasks in the posttest. However, while this relationship did not differ between the no feedback group, the KR group, and the procedural feedback group, it did differ for the conceptual feedback group. In this group, an additional correctly solved task in the pretest only translates to 0.014 additionally correctly solved tasks in the posttest, due to the estimate of −0.609 of the interaction term. This indicates that especially those students who performed less well on the pretest benefited from this type of feedback. Moreover, the effect of the interaction approximately cancelled out the effect of the pretest. That is, regardless of their prior knowledge, students who received conceptual feedback achieved a relatively high posttest score compared to the other feedback groups. Consequently, students with relatively low prior knowledge can achieve a similar posttest score as their more knowledgeable peers. This finding made this type of feedback relatively more effective for these students than other feedback types.

6. Discussion

6.1. Summary and contribution of our study

Our study compared various feedback complexities in terms of their effectiveness on achievement outcomes depending on learners' prior knowledge. Our research has shown that elaborated feedback can contribute to tackling the challenge of increasing student heterogeneity in classrooms. Thus, the consideration of prior knowledge is particularly relevant for providing fitted, elaborated feedback. We were able to show that learners with low prior knowledge benefited from feedback enriched with conceptual knowledge.

Regarding RQ 1, on the benefit of receiving feedback, our study concluded that feedback is not automatically effective. Earlier research has also reported a relatively large confidence interval and the conclusion that the provision of feedback alone is not automatically sufficient to achieve learning (Hattie & Zierer, 2019). Our findings confirmed this finding. In our study, however, elaborated feedback was significantly more effective than no feedback. This result is consistent with other studies, which also emphasize the relevance of elaborated feedback (Wagner et al., 2024). In contrast, non-elaborated feedback that reported only the correctness of an answer did not differ significantly from no feedback in our study. A possible explanation for this is that feedback, which merely reports the correctness of an answer, does not provide all students with enough information to improve. Another reason for the lack of a significant difference between the KR and no feedback group is that elaborated feedback seems to be advisable when it comes to familiar knowledge in repetitive tasks (Mertens et al., 2022). Another explanation could be, that only the elaborated feedback provided enough information to be perceived as appropriate, a prerequisite of feedback in order to be effective (Ruwe & Kuklick, 2026). Furthermore, elaborated feedback is especially important for learners with low levels of prior knowledge (Fyfe, 2016; Heckler & Mikula, 2016). Since the students in our sample answered, on average, less than half of the tasks in the pretest correctly, our sample possibly consisted mostly of students with low prior knowledge, for which KR feedback was insufficient.

Addressing the differences between elaborated and non-elaborated feedback in RQ 2, our results showed no significant difference between the elaborated and KR feedback groups. This result might be because the elaborated feedback was quite long, and students may have ignored it (Shute, 2008). Another reason may be this study's specific setting. Previous research has indicated that, in some middle school math contexts, giving the correct result is a valid alternative to providing elaborated feedback (Mertens et al., 2022). As all feedback groups would let the learners know which answer was correct after the second attempt, it is possible that this information was already sufficient for growth in achievement and that the benefit of the additional elaboration was not large enough to make a statistically significant impact. However, this result could also be due to our insufficient sample size according to G*Power.

Concerning RQ 3, regarding the effect of different types of elaborated feedback, we did not find significant differences between conceptual and procedural feedback. Accordingly, both types of feedback provide students with sufficient in-depth help to address the issue and to constitute effective feedback for them (Gao et al., 2025). Thus, our results are only partially in line with previous research showing the benefits of conceptual explanations (e.g., Lachner et al., 2019). Although conceptual feedback resulted, on average, in the most correctly solved tasks in our study, it was not statistically significantly better than procedural feedback. A possible explanation for this is the sample size, which according to G*Power was too small. Thus, we can draw only limited conclusions regarding the isolated effects of conceptual or procedural knowledge in feedback.

With RQ 4 addressing the relevance of prior knowledge, we contribute to the current state of research in two ways: First, our findings are in line with existing models of the effects of feedback (Lipnevich & Smith, 2022; Lui & Andrade, 2022) and recent meta-analyses emphasizing the importance of prior knowledge for achievement after learning (Simonsmeier et al., 2022). Our work highlights the relevance of students' prior knowledge for the effect of feedback on their achievement. This finding is consistent with other publications, which also emphasize the role of prior knowledge for the effectiveness and use of feedback (Brandmo & Gamlem, 2025; Koenka et al., 2026). Second, our findings contribute to the body of research on the interplay between prior knowledge and the complexity of feedback (e.g., Fyfe, 2016; Heckler & Mikula, 2016) and, thus, how elaborated feedback should be conceptualized. As effective feedback shows learners how they can improve (Fong et al., 2024), it seems like conceptual knowledge in feedback accomplishes this, particularly for learners with less prior knowledge. Our results highlight the importance of not only providing complex but also conceptual knowledge feedback, as demonstrated by the significant interaction between prior knowledge and the conceptual feedback group. The finding that conceptual feedback nearly nullifies the effect of prior knowledge makes it particularly suitable for learners with low prior knowledge. This leads to the conclusion that offering the (right) feedback can be an effective method of assisting learners with less prior knowledge.

6.2. Practical implications

Our findings on the benefits of different types of feedback have the following practical implications for secondary mathematics education. However, the extent to which these can be transferred to other contexts is unknown, as is discussed in the limitations section. Feedback is an essential aspect of formative assessment (Heritage, 2007); therefore, we conclude two main findings regarding its effective use. First, our study seems to show once again that (elaborated) feedback matters: students benefited from the provision of elaborated feedback. Thus, we advise teachers to provide elaborated feedback, even though they perceive finding and investing the necessary time as a challenge (Töllner et al., 2025). Secondly, the type of knowledge used to provide feedback is important. In our study, elaborated feedback was significantly more conducive to learning than providing no feedback. This finding is further supported by the role of prior knowledge on the effect of the feedback. Based on our results, it appears that especially those learners who perform poorly in class should receive teacher feedback based on conceptual knowledge. This could also address the challenge identified by Maier and Klotz (2025) that feedback must be designed primarily to support learners with less prior knowledge in order to reduce existing differences in achievement.

Another implication concerns the creation of AI-generated feedback. AI-generated feedback is less beneficial for learning than feedback provided by teachers (Weidlich et al., 2025). This could indicate that the provided AI feedback was lacking quality. To improve the quality of the AI feedback, our findings could be considered when creating the prompts to generate the feedback. In doing so, the AI-generated feedback could be prompted to include conceptual knowledge within the feedback, especially for learners with less prior knowledge to enhance the quality of the AI-generated feedback.

6.3. Limitations and future directions

Like any other study, this study has some limitations. One limitation is that we only examined a specific population in a specific subject; therefore, generalizability is questionable. In addition, neither the tests were high-stakes assessments nor was the specific topic relevant to students' mathematics lessons at that time. Hence, one might question the students' motivation to perform well. Future studies should further investigate the role of subject matter by systematically varying the topic. Additionally, future studies should explicitly address the questionable ecological validity of our study by conducting experiments in real teaching situations. This would also make it possible to address the limitation that our findings cannot reveal whether the students' learning gains were permanent or if they must be considered a short-term effect.

Another limitation relates to the lack of measurements that allow to draw conclusions on the time spent using the feedback. In future studies, the actual time spend on viewing the feedback should be recorded, using log data or eye tracking technology. This data could be used to investigate possible effects of the different texts used in the elaborated feedback types and draw conclusions on their processing. Moreover, we focused exclusively on learners' prior knowledge as a relevant learner characteristic. As there are more known aspects influencing feedback effectiveness, future studies should incorporate additional elements of prominent feedback processing models (Lipnevich & Smith, 2022; Lui & Andrade, 2022). For example, it could be investigated to what extend the effectiveness of (elaborated) feedback depends on the motivational aspects of feedback processing. In addition, future research should examine the varying effectiveness of feedback in relation to further individual characteristics such as gender or feedback literacy.

Furthermore, our study was underpowered, as we did not reach the a priori calculated required sample size. The main reason for this was the difficulty of recruiting more schools. Although we addressed this issue using statistical methods, such as selecting approximation methods for smaller samples and performing an adjustment for small-samples, this must still be considered when interpreting the results, as significant differences may not have been found because of the insufficient sample size. However, the fact that we nevertheless found these results makes them appear to be even more meaningful.

Based on these limitations, our study reveals several starting points for future research. In addition to further studies examining whether our results can be generalized to other samples and subjects, future studies should investigate what type of non-elaborated feedback is most beneficial for learners with more prior knowledge through the systematic variation of non-elaborated feedback. Thus, it may be worthwhile to gain more insight regarding the large variance in the group receiving KR feedback. Several factors may have played a role here and appear to be promising starting points for investigating this circumstance.

Firstly, the tasks themselves could have been a factor, as it could be argued that our tasks were repetitive, which may have influenced the result. Furthermore, it could have been relevant whether the feedback triggered memories of how to solve these tasks, meaning that the KR and possibly the subsequent KCR were sufficient for some learners. Determining which factors influence this seems to be a promising direction for further studies. Future research should also investigate whether procedural or conceptual knowledge is more effective for learning in certain contexts. Another research direction that deserves further attention is the investigation of the extent to which feedback effectiveness depends on how well aligned the types of knowledge conveyed in the previous instruction and feedback are. Hereby, it is possible that the effect of feedback is influenced by how familiar the students are with the type of knowledge presented in the feedback, since if the knowledge presented in the feedback is not known from the previous instruction, it may pose a new learning object that must be mastered before it can be used to solve the tasks.

Second, the way the feedback was created may have influenced our results. That said, this was an intentional design choice to adequately present the different types of knowledge. Accordingly, the texts of the conceptual feedback were longer and appeared thus more complex. However, our aim was not to compare feedback texts at different length, but rather to gain insights into how different types of knowledge in feedback influence learning. Consequently, these differences in text length resulted solely from the differing requirements to illustrate those knowledge types. This is also emphasized by other studies on different knowledge types, where explanations containing conceptual knowledge also rely on longer texts (see Lachner et al., 2019).

Lastly, future research could use our results in learning analytics-based feedback, as the combination of learning analytics and feedback enables an individualized provision of feedback. Additionally, a recent study identified the need to investigate how feedback can be used most effectively within such a system (Tepgec et al., 2025). Accordingly, our results regarding the relevance of prior knowledge and effect of varying knowledge types could be considered here.

7. Conclusion

In summary, our study investigated the effects of different complexities in feedback messages on students' achievement. The results imply that feedback can be an effective means of addressing learners with different levels of prior knowledge. This is particularly underlined by the fact that no feedback did not lead to any change in achievement. Moreover, only the provision of elaborated feedback led to a significant difference in achievement compared to no feedback. Furthermore, in terms of the effectiveness of different types of elaborated feedback, we conclude that only conceptual feedback reduces the relevance of prior knowledge, thus making it especially effective for learners with low prior knowledge.

CRediT authorship contribution statement

Fynn Töllner: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Poldi Kuhl:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Michael Besser:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This research is part of the LERN-IF project, which is funded by the Federal Ministry of Education, Family Affairs, Senior Citizens, Woman and Youth (BMBFSFJ) within the funding line "Förderbezogene Diagnostik in der inklusiven Bildung (InkBi 2)" as part of the framework programme "Empirische Bildungsforschung" under the funding reference 01NV2117. The authors are responsible for the content of this publication.

Acknowledgements

We would like to thank Friederike von Kienitz for her support in preparing and running the assessments, Kyra Renftel for her contribution in early stages of study planning, and Tino Paulsen for his technical assistance in the data preparation.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (complete ed., (reprint))*. New York: Longman.

Andrade, H. L. (2010). Summing up and moving forward: Key challenges and future directions for research and development in formative assessment. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 344–352). New York, Abingdon, Oxon, England: Routledge.

Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education, 110*, 154–169. https://doi.org/10.1016/j.compedu.2017.03.012

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bitzenbauer, P., Höfler, S., Veith, J. M., Winkler, B., Zenger, T., & Kulgemeyer, C. (2024). Exploring the relationship between surface features and explaining quality of YouTube explanatory videos. *International Journal of Science and Mathematics Education, 22*(1), 25–48. https://doi.org/10.1007/s10763-022-10351-w

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain*. New York: David McKay.

Borsch, F. (2018). *Alle lernen gemeinsam! Pädagogisch-psychologisches Wissen für den inklusiven Unterricht*. Göttingen: Vandenhoeck & Ruprecht. https://doi.org/10.13109/9783666711466

Brandmo, C., & Gamlem, S. M. (2025). Students' perceptions and outcome of teacher feedback: A systematic review. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1572950

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., … Hardy, I. (2015). Embedded Formative Assessment and Classroom Process Quality: How do they interact in promoting science understanding? *American Educational Research Journal, 52*(6), 1133–1159. https://doi.org/10.3102/0002831215596412

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In J. Dolin, & R. Evans (Eds.), *Transforming assessment. Through an interplay between practice, research and Policy* (pp. 53–80). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3_3.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles, London, New Delhi, Singapore, Washington, DC: Sage Publications.

Field, A., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.013

Filsecker, M., & Kerres, M. (2012). Repositioning formative assessment from an educational assessment perspective: A response to Dunn & Mulvenon (2009). *Practical Assessment, Research and Evaluation, 17*(16). https://doi.org/10.7275/xrkr-b675

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review, 31*(1), 121–162. https://doi.org/10.1007/s10648-018-9446-6

Fong, C. J., Schallert, D. L., Lin, S., & Altan, S. (2024). How constructive is that feedback? Associations with undergraduates' future time perspectives moderated by student characteristics'. *Studies In Educational Evaluation, 81*, Article 101351. https://doi.org/10.1016/j.stueduc.2024.101351

Fong, C. J., Schallert, D. L., Williams, K. M., Williamson, Z. H., Lin, S., Kim, Y. W., & Chen, L.-H. (2021). Making feedback constructive: The interplay of undergraduates' motivation with perceptions of feedback specificity and friendliness. *Educational Psychology, 41*(10), 1241–1259. https://doi.org/10.1080/01443410.2021.1951671

Fyfe, E. R. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior, 63*, 568–574. https://doi.org/10.1016/j.chb.2016.05.082

Gao, X., Schunn, C. D., Noroozi, O., Gulikers, J., Banihashem, S. K., & Biemans, H. (2025). Consistency in providing peer feedback: The role of individual and situational factors. *Assessment & Evaluation in Higher Education, 50*(8), 1173–1187. https://doi.org/10.1080/02602938.2025.2517617

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: Its impact on item fit and difficulty of the Rasch model. *International Journal of Research and Method in Education, 39*(4), 401–421. https://doi.org/10.1080/1743727X.2016.1168798

Haapasalo, L., & Kadijevich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik, 21*(2), 139–157. https://doi.org/10.1007/BF03338914

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. London: Routledge.

Hattie, J., & Zierer, K. (2019). *Visible learning insights*. London, New York: Routledge Taylor & Francis Group. https://doi.org/10.4324/9781351002226

Heckler, A. F., & Mikula, B. D. (2016). Factors affecting learning of vector math from computer-based practice: Feedback complexity and prior knowledge. *Physical Review Physics Education Research, 12*(1). https://doi.org/10.1103/PhysRevPhysEducRes.12.010134

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140–145. https://doi.org/10.1177/003172170708900210

Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics*. New York: Routledge. https://doi.org/10.4324/9780203063538.

Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods, 54*(3), 1181–1199. https://doi.org/10.3758/s13428-021-01627-0

Huang, F. L., Wiedermann, W., & Zhang, B. (2023). Accounting for heteroskedasticity resulting from between-group differences in multilevel models. *Multivariate Behavioral Research, 58*(3), 637–657. https://doi.org/10.1080/00273171.2022.2077290

Huang, F. L., Zhang, B., & Li, X. (2023). Using robust standard errors for the analysis of binary outcomes with a small number of clusters. *Journal of Research on Educational Effectiveness, 16*(2), 213–245. https://doi.org/10.1080/19345747.2022.2100301

Hwang, J. (2016). *An examination of error patterns associated with middle school students' solution pathways when solving fraction computation involving addition*. Dissertation, Special Education https://etda.libraries.psu.edu/files/final_submissions/12513.

Jaafar, N. N., Rosdi, M. N. A., Jamaludin, K. R., Ramlie, F., & Talib, H. A. (2024). Imputation analysis of time-series data using a random forest algorithm. In W. H. M. Isa, I. M. Khairuddin, M. A. Mohd Razman, S.'A. Saruchi, S.-H. Teh, & P. Liu (Eds.), *Intelligent manufacturing and mechatronics* (pp. 51–60). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-8819-8_4.

Jonsson, A., & Panadero, E. (2018). Facilitating students' active engagement with feedback. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 531–553). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316832134.026.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology, 41*(7), 922–947. https://doi.org/10.1080/01443410.2019.1659939

Koenka, A. C., McManus, D. N., & Nicolai, K. D. (2026). Leveraging the power and potential of feedback: Conceptual contributions and areas for expansion. *Contemporary Educational Psychology, 84*, Article 102427. https://doi.org/10.1016/j.cedpsych.2025.102427

Korntreff, S., & Prediger, S. (2026). How can students' engagement with instructional videos on generalizing with algebraic expressions be scaffolded? Design research for specifying the content-specific focus of scaffolds. *The Journal of Mathematical Behavior, 82*, Article 101317. https://doi.org/10.1016/j.jmathb.2026.101317

Kuklick, L., Greiff, S., & Lindner, M. A. (2023). Computer-based performance feedback: Effects of error message complexity on cognitive, metacognitive, and motivational outcomes. *Computers & Education, 200*, Article 104785. https://doi.org/10.1016/j.compedu.2023.104785

Kuklick, L., & Lindner, M. A. (2023). Affective-motivational effects of performance feedback in computer-based assessment: Does error message complexity matter? *Contemporary Educational Psychology, 73*, Article 102146. https://doi.org/10.1016/j.cedpsych.2022.102146

Kuklick, L., & Lindner, M. A. (2025). How to enhance elaborated feedback in computer-based assessment: The role of multimedia and emotional design factors. *Contemporary Educational Psychology, 82*, Article 102396. https://doi.org/10.1016/j.cedpsych.2025.102396

Lachner, A., Weinhuber, M., & Nückles, M. (2019). To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of principle-oriented explanations. *Contemporary Educational Psychology, 58*, 175–185. https://doi.org/10.1016/j.cedpsych.2019.03.008

Lenz, K., Dreher, A., Holzäpfel, L., & Wittmann, G. (2020). Are conceptual knowledge and procedural knowledge empirically separable? The case of fractions. *British Journal of Educational Psychology, 90*(3), 809–829. https://doi.org/10.1111/bjep.12333

Lenz, K., & Wittmann, G. (2021). Individual differences in conceptual and procedural fraction knowledge: What makes the difference and what does it look like? *International Electronic Journal of Mathematics Education, 16*(1). https://doi.org/10.29333/iejme/9282. em0615.

Lipnevich, A. A., & Smith, J. K. (2022). Student—Feedback interaction model: Revised. *Studies In Educational Evaluation, 75*, Article 101208. https://doi.org/10.1016/j.stueduc.2022.101208

Lui, A. M., & Andrade, H. L. (2022). The next Black box of formative assessment: A model of the internal mechanisms of feedback processing. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.751548

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Maier, U., & Klotz, C. (2025). Students ignore their mistakes: Elaborated error feedback processing in a digital learning system. *Contemporary Educational Psychology, 82*, Article 102395. https://doi.org/10.1016/j.cedpsych.2025.102395

Martínez, T., García, A., Cerdán, R., & Vidal-Abarca, E. (2026). Repairing errors with elaborative feedback in computerised learning environments. *Journal of Computer Assisted Learning, 42*(2). https://doi.org/10.1002/jcal.70199

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology, 114*(8), 1743–1772. https://doi.org/10.1037/edu0000764

Mitchell, D. (2014). *What really works in special and inclusive education: Using evidence-based teaching strategies* (2nd ed.). London, New York: Routledge Taylor & Francis Group.

Narciss, S. (2006). I. In *nformatives tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. Pädagogische Psychologie und Entwicklungspsychologie* (Vol. 56). Münster, New York, München, Berlin: Waxmann.

Narciss, S. (2013). Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review, 23*, 7–26.

Narciss, S. (2017). Conditions and effects of feedback viewed through the lens of the interactive tutoring feedback model. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 173–189). Singapore: Springer. https://doi.org/10.1007/978-981-10-3045-1_12.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56–76. https://doi.org/10.1016/j.compedu.2013.09.011

Nusser, L., & Gehrer, K. (2020). Addressing heterogeneity in secondary education: Who benefits from differentiated instruction in German classes? *International Journal of Inclusive Education*, 1–18. https://doi.org/10.1080/13603116.2020.1862407

Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist, 58*(3), 193–204. https://doi.org/10.1080/00461520.2023.2223642

Paris, B. (2022). Instructors' perspectives of challenges and barriers to providing effective feedback. *Teaching and Learning Inquiry, 10*. https://doi.org/10.20343/teachlearninqu.10.3
Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., … Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research, 88*(2), 205–242. https://doi.org/10.3102/0034654317743198
Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*(4), 449–468. https://doi.org/10.1016/0885-2014(91)90049-J
Pinheiro, J., Bates, D., & R Core Team. (2023). *_nlme: Linear and Nonlinear (Version R package version 3.1-163) (computer software)*.
Pustejovsky, J. E. (2024). *clubSandwich: Cluster-robust (Sandwich) Variance estimators with Small-Sample (Version R package version 0.6.2) (computer software)*.
Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics, 36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004
R Core Team. (2023). *R: A Language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.r-project.org/.
Revelle, W. (2024). *Psych: Procedures for Psychological. Psychometric, and personality research (version R package version 2.4.1)* (Computer software).
Rittle-Johnson, B., Fyfe, E. R., & Loehr, A. M. (2016). Improving conceptual and procedural knowledge: The impact of instructional content within a mathematics lesson. *British Journal of Educational Psychology, 86*(4), 576–591. https://doi.org/10.1111/bjep.12124
Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014.
Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x
Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362. https://doi.org/10.1037//0022-0663.93.2.346
Ruwe, T., & Kuklick, L. (2026). Quality counts? Examining the role of feedback provider and feedback quality on students' feedback perceptions. *British Journal of Educational Technology, 57*(1), 272–298. https://doi.org/10.1111/bjet.70011
Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. https://doi.org/10.2307/3002019
Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., … Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434
Schiller, R., Fleckenstein, J., Mertens, U., Horbach, A., & Meyer, J. (2024). Understanding the effectiveness of automated feedback: Using process data to uncover the role of behavioral engagement. *Computers & Education, 223*, Article 105163. https://doi.org/10.1016/j.compedu.2024.105163
Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology, 46*(1), 178–192. https://doi.org/10.1037/a0016701
Shirah, J. F., & Sidney, P. G. (2023). Computer-based feedback matters when relevant prior knowledge is not activated. *Learning and Instruction, 87*, Article 101796. https://doi.org/10.1016/j.learninstruc.2023.101796
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795
Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, 57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700
Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*(2), 183–193. https://doi.org/10.1080/10494820701365952
Steinbach, M., Fleckenstein, J., Kuklick, L., & Meyer, J. (2025). (De)motivating zero-performing students with negative feedback: Does the salience of performance information matter? *Journal of Computer Assisted Learning, 41*(4). https://doi.org/10.1111/jcal.70070
Stekhoven, D. J. (2022). *missForest: Nonparametric Missing Value Imputation using Random Forest (Version R package version 1.5.)* (Computer software).
Stekhoven, D. J., & Bühlmann, P. (2012). Missforest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597
Swart, E. K., Nielen, T. M., & Sikkema-de Jong, M. T. (2019). Supporting learning from text: A meta-analysis on the timing and content of effective feedback. *Educational Research Review, 28*, Article 100296. https://doi.org/10.1016/j.edurev.2019.100296
Tärning, B., Lee, Y. J., Andersson, R., Månsson, K., Gulz, A., & Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *The Journal of the Learning Sciences, 29*(4–5), 511–549. https://doi.org/10.1080/10508406.2020.1770092
Tepgec, M., Heil, J., & Ifenthaler, D. (2025). Feedback literacy matters: Unlocking the potential of learning analytics-based feedback. *Assessment & Evaluation in Higher Education, 50*(1), 50–66. https://doi.org/10.1080/02602938.2024.2367587
Timmers, C. F., Braber-van den Broek, J., & van den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education, 60*(1), 25–31. https://doi.org/10.1016/j.compedu.2012.07.007
Tiwaskar, S., Rashid, M., & Gokhale, P. (2025). Impact of machine learning-based imputation techniques on medical datasets- a comparative analysis. *Multimedia Tools and Applications, 84*(9), 5905–5925. https://doi.org/10.1007/s11042-024-19103-0
Töllner, F., Kuhl, P., & Besser, M. (2026). *Materialien der Interviewstudie und der Interventionsstudie im Projekt "Lernprozessbegleitende Diagnostik und lernförderliche Leistungsrückmeldung im inklusiven Fachunterricht (LERN-IF)"*. https://doi.org/10.48548/pubdata-2919 [Data set].
Töllner, F., Renftel, K., Kuhl, P., & Besser, M. (2025). Teachers' use of formative assessment in inclusive mathematics education in German secondary schools. In T. Ehmke, & J. C.-K. Lee (Eds.), *Innovative teaching and classroom processes* (pp. 256–269). London: Routledge. https://doi.org/10.4324/9781003436065-22.
Van Blankenstein, F. M., Dirkx, K. J. H., & Bruycker, N. M. F. de (2025). Ask your peer! how requests for peer feedback affect peer feedback responses. *Educational Research and Evaluation, 30*(1–2), 36–57. https://doi.org/10.1080/13803611.2024.2376832
Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475–511. https://doi.org/10.3102/0034654314564881
Wagner, S., Sibley, L., Weiler, D., Burde, J.-P., Scheiter, K., & Lachner, A. (2024). The more, the better? Learning with feedback and instruction. *Learning and Instruction, 89*, Article 101844. https://doi.org/10.1016/j.learninstruc.2023.101844
Wang, Z., Gong, S.-Y., Xu, S., & Hu, X.-E. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Computers & Education, 136*, 130–140. https://doi.org/10.1016/j.compedu.2019.04.003
Weidlich, J., Gotsch, F., Schudel, K., Marusic-Würscher, C., Mazzarella, J., Bolten, H., … Merki, K. M. (2025). Teacher, peer, or AI? Comparing effects of feedback sources in higher education. *Computers and Education Open, 9*, Article 100300. https://doi.org/10.1016/j.caeo.2025.100300
Winstone, N. E., & Nash, R. A. (2023). Toward a cohesive psychological science of effective feedback. *Educational Psychologist, 58*(3), 111–129. https://doi.org/10.1080/00461520.2023.2224444
Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49–64. https://doi.org/10.1080/00461520701756420
Xu, C., Di Lonardo Burr, S., Rodgers, L. J., Wylie, J., Si, J., Huang, B., … Li, H. (2025). A longitudinal investigation of the codevelopment and bidirectional relations among whole number arithmetic and conceptual and procedural fraction knowledge. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000942. Advance online publication.
Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477–501. https://doi.org/10.1023/A:1023967026413

Fynn Töllner is a research associate at Leuphana University Lüneburg and puts a special emphasis on feedback.

Poldi Kuhl is a professor of educational psychology at Leuphana University Lüneburg. She focuses on inclusive education, digital teaching and learning, and teacher professionalisation.

Michael Besser is a professor of educational research in Mathematics Didactics at Leuphana University Lüneburg. He focuses on teaching and learning in mathematics education, professional development of (prospective) teachers, and teachers' and learners' competencies.