LUMAT
International Journal on Math, Science and Technology Education

1/24
Research article

Effects of guiding feedback on students' performance, calibration, and self-efficacy: Insights from a field study with engineering students

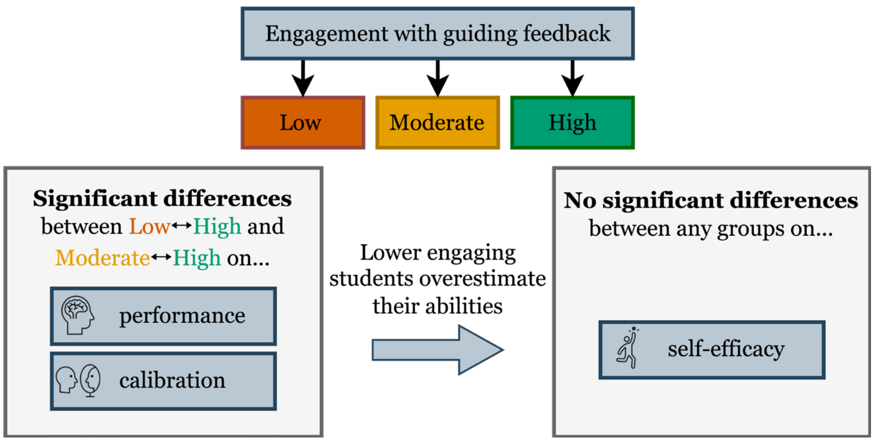
Farhad Razeghpour & Katrin Rolka

Ruhr University Bochum, Germany

Abstract: In technology-based learning environments, informative tutoring feedback (ITF) strategies can be implemented to support students during task completion by providing elaborated feedback such as error-specific hints instead of directly showing correct solutions. However, identifying the underlying causes of errors is challenging, especially for more advanced mathematical tasks in higher education. To counteract this issue, an ITF-strategy called *guiding feedback* has been conceptualized. In the context of guiding feedback, students' answers are examined based on mathematical properties to identify the causes of their errors and provide error-specific feedback. If students' error causes cannot be identified, they can work through the task in a series of sub-steps, allowing the system to gain insights into their procedures and facilitating the provision of specific feedback. A previous field study examined the motivational effects of guiding feedback in a mathematics course with engineering students. Contrary to expectations derived from the literature, students who completed more tasks with guiding feedback did not demonstrate significantly higher self-efficacy than those who completed fewer tasks. To investigate whether students who worked on fewer tasks may have inaccurately judged their abilities, the present research extends the previous findings by further analyzing students' performance and calibration. Based on the number of completed tasks, 196 participants of the field study were grouped into low, moderate, and high engagement categories. We evaluated students' performance by analyzing their assignments and assessed their calibration by comparing their self-efficacy and performance. Results revealed that students in the high engagement group had significantly better performance and calibration than students in the low and moderate engagement groups. Overall, this research provides a more nuanced understanding of the previous findings by showing that even though guiding feedback may not directly boost students' self-efficacy, it ensures that their judgments align with their actual performance.

Keywords: mathematics education, technology-based learning, feedback, performance, self-efficacy, calibration

Correspondence:
farhad.razeghpour@ruhr-uni-bochum.de



Received 28 March 2025 / Accepted 20 March 2026 / Published 29 June 2026
Published by the University of Helsinki, Finland and LUMA Centre Finland.

1 Introduction

The integration of technology plays an increasingly essential role in mathematics education (e.g., Dyrvold & Bergvall, 2023; Erens & Eichler, 2018; Sunzuma, 2023). One key advantage of technology-enhanced learning environments is their potential to automatically generate formative feedback for students, which is provided during their learning processes (Pals et al., 2023). In this context, informative tutoring feedback (ITF) strategies can be implemented to support learners with ongoing constructive guidance by offering elaborated formative feedback, such as error-specific hints, rather than simply presenting correct solutions (Narciss, 2012). However, for more advanced mathematical tasks, it is particularly difficult to determine automatically the underlying causes of students' errors when using assessment systems (Sangwin & Köcher, 2016). This challenge has led to a significant research gap regarding the effects of ITF-strategies in higher mathematics education (Morris et al., 2021; Narciss & Zumbach, 2023).

To address this research gap, we used the Moodle plug-in STACK to conceptualize an ITF-strategy called *guiding feedback* (Razeghpour, 2024a). As part of guiding feedback, students' answers are analyzed to identify the causes of their errors. If typical error causes are identified, students receive error-specific feedback. When they are unclear, guiding feedback offers task loops in which students can work through individual sub-steps. By analyzing students' step-by-step processes, the system gains insights into their procedures, enabling the provision of error-specific feedback.

In a prior field study involving 222 engineering students, the impact of guiding feedback on students' self-efficacy was investigated within a mathematical university course (Razeghpour, 2024b). Contrary to literature-based expectations, students who completed more tasks with guiding feedback did not show significantly higher self-efficacy than their less-engaged peers. However, it remained unclear whether this result might be attributed to misjudgments by the less-engaged students. This study extends the previous research by analyzing additional data from the field study to provide a deeper understanding of the cognitive, metacognitive, and motivational effects of guiding feedback.

1.1 Feedback strategies in instructional contexts

Narciss (2012) defines a feedback strategy as a plan for providing feedback within instructional settings, considering the following aspects:

- situational and individual conditions: e.g., students' prior knowledge
- content of feedback: e.g., error-specific hints
- goals and purposes of feedback: e.g., pointing out mistakes
- timing of feedback: e.g., immediately after entering an answer
- presentation form of feedback: e.g., graphic visualization

When determining the content of a feedback strategy, simple and elaborated components can be combined (Narciss, 2008). Simple feedback includes, for instance, knowledge of result (KR) and knowledge of the correct result (KCR). KR-feedback informs students whether a task was solved correctly, and KCR-feedback provides the correct solution. In contrast, elaborated feedback offers more comprehensive information when answers are incorrect. For example, this can include knowledge about mistakes (KM), which identifies the location of a mistake and its underlying cause, or knowledge on how to proceed (KH), which provides solution-specific hints for task completion.

In the context of ITF-strategies, simple and elaborated forms like KR-, KM-, and KH-feedback are combined to inform learners about their mistakes and help them master the task without presenting the correct solutions (Narciss & Huth, 2004). This concept is grounded in the pedagogical principle that human tutors often support students by assisting them in correcting errors, rather than simply providing correct solutions (Kojo et al., 2018). Prior studies demonstrated beneficial effects of ITF-strategies, but they were primarily focused on basic mathematical tasks for elementary and middle school students (Narciss et al., 2014; Narciss & Huth, 2006). There remains a notable gap in research concerning the application of formative feedback strategies for mathematical tasks in higher education (Morris et al., 2021; Narciss & Zumbach, 2023).

The ITF-strategy guiding feedback addresses this issue by broadening the application of ITF-strategies to more advanced mathematical tasks through the use of the assessment system STACK (Razeghpour, 2024a). When students work on mathematical tasks and answer correctly, they receive positive confirmation in the form of KR-feedback. In contrast, if their answers are incorrect, the system checks whether common error causes led to their responses by employing a computer algebra system (CAS). For instance, in a task involving the determination of a function's extrema, the CAS can assess whether students mistakenly calculated the zeros of the original function instead of its derivative. If this is the case, the feedback can inform students about the underlying cause of their error and point out the relevance of the derivative regarding a function's extrema.

In cases where the causes of errors cannot be identified, guiding feedback informs students that their responses are incorrect and gives them the opportunity to enter a task loop, during which they are guided through the solution process. Considering this scenario for the task of determining extrema, students would progress through individual steps, including identifying the necessary and sufficient conditions for extrema, calculating the first and second derivatives, finding the zeros of the first derivative, and classifying the critical points as maxima or minima. This approach allows the assessment system to gain deeper insights into students' procedures and to provide more detailed feedback on potential error causes. After each sub-step, students can either return to the initial task or continue to the next step. Once all sub-steps are completed, they are encouraged to correct their initial responses.

1.2 Cognitive, metacognitive, and motivational effects of guiding feedback

From a theoretical perspective, the described characteristics of guiding feedback suggest that it can yield positive cognitive, metacognitive, and motivational effects. Focusing first on the cognitive effects, studies have demonstrated that feedback on students' step-by-step processing can significantly enhance their mathematical performance (Corbalan et al., 2010; Kaarakka et al., 2019). By pointing out mistakes while learners engage with tasks, guiding feedback can further facilitate the construction, specification, and structuring of content-related, procedural, and strategic knowledge. This process contributes to a deeper understanding and improved performance (Narciss, 2008).

From a metacognitive perspective, guiding feedback can promote students' ability to accurately judge their performance, which is defined as calibration (Pajares & Graham, 1999). Guiding feedback can foster students' calibration by providing positive reinforcement for correct answers and error-specific hints for incorrect ones (Butler et al., 2008; Osterhage et al., 2019). Accurate calibration is essential, as it plays a central role in the effective management of learning processes (Boekaerts & Rozendaal, 2010). Conversely, poor calibration can have negative effects. Overestimation can result in insufficient learning and reduced motivation (Chen & Zimmerman, 2007), whereas underestimation can hinder the full development of one's potential (Gravill et al., 2006).

Furthermore, guiding feedback can have motivational effects by supporting students in completing their tasks without providing the correct solution outright. This approach can enhance students' self-efficacy, which is defined as their belief in their ability to successfully perform certain tasks (Bandura, 1997). By withholding the correct answers and supporting independent error correction, guiding feedback allows learners to attribute their success to their own abilities, which is essential for increasing their self-efficacy (Stajkovic & Sommer, 2000). Moreover, within the task loops, students receive positive feedback for correctly completed sub-steps, leading to more mastery experiences and further strengthening their self-efficacy (Usher & Pajares, 2006; Zientek et al., 2017). Research has highlighted the significant role of students' self-efficacy in mathematical success, particularly when tackling more advanced tasks (Viholainen et al., 2019), as higher self-efficacy leads to more ambitious goal setting (Honicke & Broadbent, 2016; Saks, 2024) and greater persistence in challenging tasks (Skaalvik et al., 2015).

The potential effects of the presented characteristics of feedback have been demonstrated across studies involving diverse student populations, with research paying particular attention to engineering students (Das, 2023; Lache et al., 2021; Park et al., 2026). Focusing on these students, guiding feedback can be especially effective, as they frequently work on advanced, multi-step tasks that require procedural reasoning (Bergsten et al., 2015; Taraban et al., 2007). As part of guiding feedback, the task loops allow such problems to be broken down into smaller, manageable components, helping students identify the skills needed for successful completion. This process supports students' self-assessment and can enhance both their performance and calibration (Beyer, 2002; Kuklick et al., 2023).

1.3 Previous research on guiding feedback

The theoretically anticipated effects of guiding feedback have been investigated in two studies. In an experimental study, 64 university students of an introductory mathematics course were assigned to one of two groups based on their gender and grade (Razeghpour & Rolka, 2026). One group received guiding feedback, while the other was provided with an alternative ITF-strategy called *summarizing feedback*. Unlike guiding feedback, this ITF-strategy did not allow students to engage in task loops if the causes of students' errors were unclear. Instead, they received detailed explanations for task completion. These explanations included information on how to solve the task, but the final solution was withheld to foster students' active engagement. The results showed that guiding feedback significantly enhanced students' performance, self-efficacy, and calibration in comparison to summarizing feedback.

To investigate the effects of guiding feedback in a regular educational setting, a field study was conducted in a mathematical course with 222 undergraduate engineering students (Razeghpour, 2024b). During the lecture period, students completed several assignments. They were offered five so-called *preparing tasks* that include guiding feedback to support them in practicing for the assignments. Before beginning each assignment, students indicated their self-efficacy on a Likert scale. Contrary to expectations, no significant differences in self-efficacy were observed between students who completed none (0), some (1–4), or all (5) of the preparing tasks with guiding feedback.

The contrasting findings regarding the impact of guiding feedback on students' self-efficacy may stem from methodological differences between the two studies. In the experimental study, all participants completed the same number of tasks and received feedback for each one. In contrast, the students in the field study completed varying numbers of tasks, leading to different amounts of feedback. As a result, students in the field study who only worked on a few preparing tasks may have overestimated their abilities due to a lack of feedback (Labuhn et al., 2010).

Furthermore, while participants in the experimental study (Razeghpour & Rolka, 2026) were grouped based on gender and grade to ensure comparable group compositions, the field study (Razeghpour, 2024b) naturally grouped students according to the number of completed preparing tasks, potentially resulting in pre-existing differences between the groups. Students who overestimated their abilities in the field study may have avoided engaging with the preparing tasks under the assumption that these tasks were unnecessary for enhancing their skills (Chen & Zimmerman, 2007). Overall, a potential overestimation of students who engaged minimally with the preparing tasks might have contributed to the fact that students with higher engagement did not demonstrate significantly higher self-efficacy in the field study.

2 Present research

Building on the theoretical framework of guiding feedback and the mixed findings regarding its impact on students' self-efficacy, this research extends the analysis by additionally examining students' performance and calibration. Specifically, we investigate whether the non-significant effects on self-efficacy observed in the previous research (Razeghpour, 2024b) result from misjudgments by less-engaged students. By clarifying this relationship, the study aims to uncover underlying factors contributing to the non-significant findings and offer a broader perspective beyond self-efficacy alone. Accordingly, the following research questions are formulated:

- RQ1) To what extent does engaging with tasks that include guiding feedback impact students' performance in a mathematics course?
- RQ2) To what extent does engaging with tasks that include guiding feedback impact students' calibration in a mathematics course?

Based on theoretical considerations regarding the effects of guiding feedback (see Chapter 1.2), the following hypotheses are proposed:

- H1) Higher engagement with tasks that include guiding feedback enhances students' performance.
- H2) Higher engagement with tasks that include guiding feedback enhances students' calibration.

Since changes in the evaluation methodology were necessary for a more precise analysis (see Chapter 3.1), students' self-efficacy is reassessed to determine whether the previously observed non-significant findings regarding self-efficacy (Razeghpour, 2024b) are reproduced under the new conditions. Therefore, the additional research question is addressed:

- RQ3) To what extent does engaging with tasks that include guiding feedback impact students' self-efficacy?

It is hypothesized that the previous findings regarding students' self-efficacy will be replicated:

- H3) Higher engagement with tasks that include guiding feedback does not enhance students' self-efficacy.

3 Methods

3.1 Sample and design

This field study was embedded in a mathematics course at a German university. The course focused on probability theory and statistics and was offered annually during the winter semester. It was designed for undergraduate engineering students in their third semester or higher who have successfully completed two prior mathematics courses. Consequently, most participants were in their third or fifth semester and possessed a solid foundational understanding of core mathematical methods.

During the lecture period, mathematical assignments were conducted on a weekly basis, with a single attempt permitted within designated time frames. Students earned bonus points for correctly solved assignments that contributed to an improvement in their final course grade. To support students in practicing for the assignments, students were offered voluntary preparing tasks that include guiding feedback. As it was not feasible to create preparing tasks for all assignments due to time constraints, five assignments were selected in collaboration with professors of mathematics and mathematics education, which had a medium level of difficulty, covered different areas of probability theory and statistics, and were spread over the entire lecture period. These selected five assignments focused on the binomial distribution, the central limit theorem, the normal distribution, confidence intervals and hypothesis tests. For these five assignments, students were also asked to assess their self-efficacy before completing them.

In the previous study by Razeghpour (2024b), students were considered participants if they reported their self-efficacy for at least one assignment and completed at least one of them. For the present study, only students who reported their self-efficacy for more than half of the selected assignments and completed more than half of them were included in the analysis. This criterion was set to strengthen the reliability of the findings, resulting in a final sample of 196 students. These students were naturally grouped according to the number of completed preparing tasks.

To enhance the validity of group comparisons, the classification criteria from the previous research were adjusted to ensure more meaningful distinctions and that each group contained enough participants for more robust comparisons. The revised classification resulted in the following distribution:

- low engagement (0–1 preparing tasks completed): 31 participants
- moderate engagement (2–3 preparing tasks completed): 33 participants
- high engagement (4–5 preparing tasks completed): 132 participants

3.2 Preparing tasks

For each of the five selected assignments, a corresponding preparing task was developed. Both the assignments and the preparing tasks covered the same mathematical competencies and maintained a comparable level of difficulty, but they differed in their numerical values. The preparing tasks could be completed multiple times, with the numerical values changing after each successful completion, allowing students to decide how many similar tasks they wished to complete before progressing to the related assignment.

To design guiding feedback for the preparing tasks, we identified potential causes of errors and developed error-specific feedback for them. To determine these error causes, we analyzed students' responses from previous lecture periods, focusing on their most common mistakes. For all other error causes, for which no specific feedback was developed, we created task loops with individual sub-steps. When defining the sub-steps, we specified the competencies required for the preparing tasks to ensure comprehensive coverage of all necessary skills. Additionally, feedback texts for both correct and incorrect answers were designed for all sub-steps. The development process was supervised by professors of mathematics and mathematics education from three German universities.

Furthermore, previous versions of the preparing tasks were evaluated one year earlier in the same university course through semi-structured interviews with five students (Lache et al., 2021). The interviews addressed both content-related and design-related aspects of the tasks. They were recorded, transcribed, and analyzed to support the further development of the tasks. The findings indicated that the tasks were generally well designed, with students particularly appreciating the opportunity to work through sub-steps and receive immediate feedback. At the same time, they identified areas for improvement, as some tasks were perceived as insufficiently precise and the length of feedback messages was occasionally considered excessive. These results were subsequently used to refine and enhance the preparing tasks.

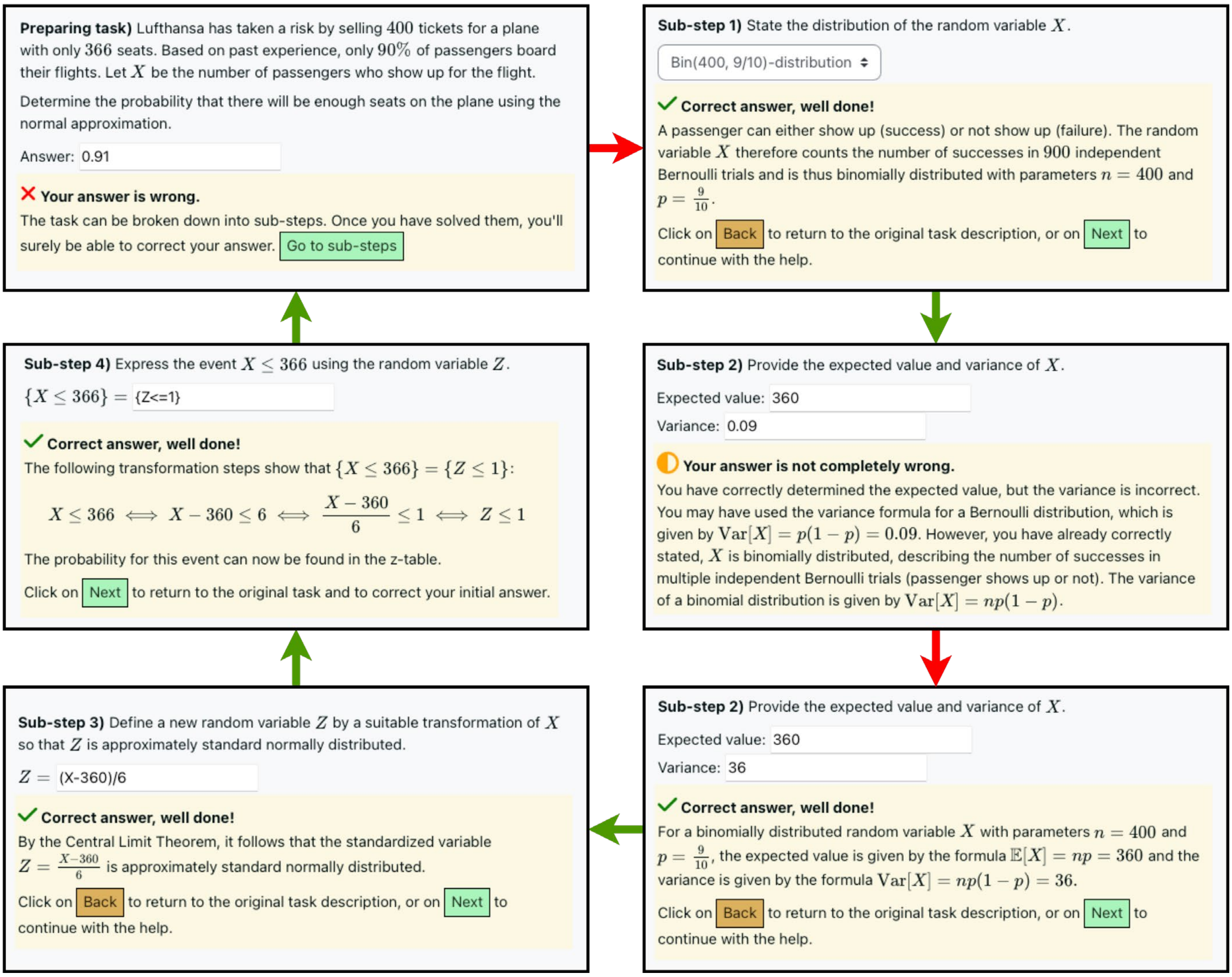
To illustrate the material, Figure 1 visualizes the completion of a preparing task. In this task, students were required to determine the probability that, under given conditions, only a certain proportion of people who have purchased tickets will take their flight. In the exemplary scenario shown in Figure 1, it was assumed that the task had not been solved correctly and that the cause of the error could not be identified. Consequently, the system indicated that the response was incorrect and offered the opportunity to complete the task through structured sub-steps (see Figure 1, top left). The sub-steps involved: (1) determining the distribution of the random variable, (2) calculating the expected value and variance, (3) defining a new normally distributed random variable through an appropriate transformation, and (4) expressing the event using this new random variable.

During the completion of the sub-steps, an error occurred in the second sub-step, where the variance was calculated incorrectly (see Figure 1, center right). The system determined that the error was likely caused by applying the formula for calculating the variance of a Bernoulli distribution instead of a binomial distribution. The feedback explained that the given case did not correspond to a Bernoulli distribution and supplied the correct formula for the variance of a binomial distribution. By engaging with the feedback, the second sub-step could be solved correctly in a subsequent attempt (see

Figure 1, bottom right). The remaining steps were completed correctly, and after the fourth sub-step, the initial task was revisited and could be corrected (see Figure 1, center left).

In addition to the exemplary completion of the task, Figure 1 also illustrates that beyond single-choice options (see Figure 1, sub-step 1), the use of STACK allowed for the integration of free-input fields (see Figure 1, sub-steps 2–4), enabling the independent formulation of mathematical results.

Figure 1. Exemplary processing of a preparing task



Note. Illustration of a preparing task (top left) and its four sub-steps. Green arrows denote correct inputs, while red arrows indicate incorrect inputs.

To facilitate students' entries of mathematical expressions, an overview comprising four categories (Numbers, operations, and sets; Trigonometric functions; Mathematical constants; Vectors and matrices) was integrated into the HTML block of the Moodle course. After students selected a category, a table was displayed, providing guidance on how to input the corresponding mathematical expressions (see Figure 2). Embedding this support within the HTML block ensured its accessibility for students while working on preparing tasks or assignments.

Figure 2. Support for entering mathematical expressions

▼ Numbers, operations, and sets

Math Expression	Input
$1{,}234$	`1.234`
$a+b$	`a+b`
$a-b$	`a-b`
$a \cdot b$	`a*b`
$\frac{a}{b}$	`a/b`
a^b	`a^b`
$\sqrt{a}$	`sqrt(a)`
$\{a,b,c\}$	`{a,b,c}`

▶ Trigonometric functions

▶ Mathematical constants

▶ Vectors and matrices

Note. Overview of the four categories for which support was provided in entering mathematical expressions. The table for the first category (Numbers, operations, and sets) is displayed.

3.3 Measures

To assess students' performance, we collected their scores on five digital assignments. The assignments were created within the Moodle learning platform using STACK, enabling an automated evaluation of students' responses. To prevent students from copying answers from their peers, the numerical values in the assignments were randomized using the CAS (Härterich, 2019), ensuring that students received different versions of the assignments while keeping the task requirements consistent. For each assignment, students could receive a total score ranging from 0 to 1 point. In cases where assignments consisted of multiple parts, the total possible score was evenly distributed among the parts. Thus, in an assignment with four parts, students could earn a maximum of 0.25 points per part. To ensure an accurate assessment of students' performance, the system acknowledged minor discrepancies between the expected solutions and students' answers caused by incorrect rounding. These responses were not considered entirely incorrect, resulting in only a half-point deduction. In addition, for multi-part assignments, consequential errors were considered, preventing multiple deductions when intermediate results were relevant for subsequent parts.

To assess students' self-efficacy, an eight-point Likert scale was employed, enabling students to rate their confidence in solving each assignment. The scale ranged from 1 ("I do not trust myself at all") to 8 ("I trust myself completely") and was adapted from Pajares and Graham (1999), whose study demonstrated its high reliability (Cronbach's $\alpha = 0.93$). For each assignment, students were asked to indicate their confidence in solving it on the Likert scale. This task-specific evaluation of self-efficacy follows Bandura's guidelines (2006).

To investigate students' calibration, both accuracy and bias scores were calculated. While accuracy reflects the degree of precision in one's judgments, ranging from complete inaccuracy to complete accuracy, bias indicates the direction of these judgments, ranging from under- to overestimation (Pajares & Graham, 1999). To calculate accuracy and bias scores, the performance scale was transformed to align with the self-efficacy scale. The bias score for each of the five assignments was computed as the difference between students' self-efficacy ratings and their transformed performance scores. This resulted in a scale ranging from -7 to 7, where -7 indicated a maximum underestimation, 0 an exact estimation, and 7 a maximum overestimation of one's abilities. The accuracy score was derived by subtracting the absolute value of the bias score from 7, yielding a scale ranging from 0 to 7. In this case, 0 indicated a highly inaccurate estimation, while 7 represented a highly accurate self-assessment.

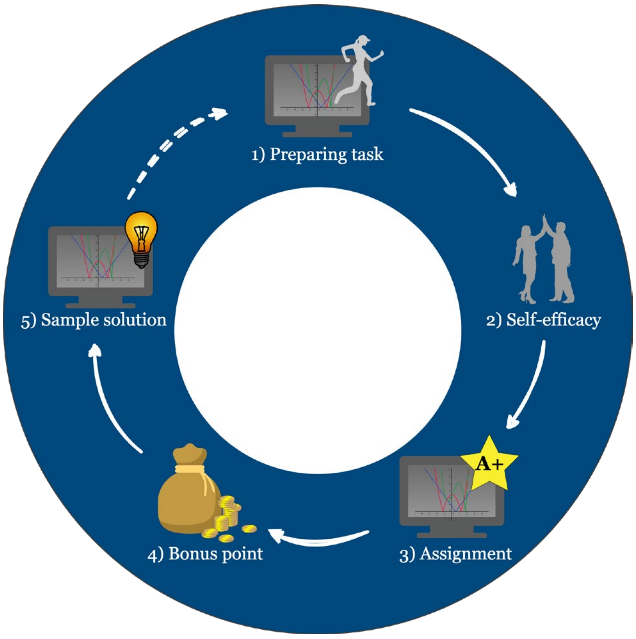
3.4 Procedure

In the mathematical course, weekly assignments focusing on probability theory and statistics were conducted. The current study focuses on five selected weeks for which students were offered preparing tasks. The course schedule of these weeks is illustrated by the cycle in Figure 3.

In each of these five weeks, students had the free choice to work on the preparing task (see Figure 3, No. 1) to practice for the current assignment. After either completing or opting to skip the preparing task, students were asked to assess their self-efficacy for the corresponding assignment on the Likert scale (see Figure 3, No. 2). Once students rated their self-efficacy, they could begin working on the assignment itself (see Figure 3, No. 3). An 11-day period was designated for the completion of each assignment. Students earned a bonus point for each correctly completed assignment, which contributed to an improvement in their final course grade (see Figure 3, No. 4). After the submission deadline of each assignment, students received KR-feedback indicating whether their solutions were correct, and KCR-feedback explaining how to solve the assignment correctly (see Figure 3, No. 5).

The course schedule shown in Figure 3 is merely an illustrative case. Since completing the preparing tasks and rating self-efficacy on the Likert scale were voluntary, students could also begin the course schedule by reporting their self-efficacy or working on the assignments. The last arrow in Figure 3, from the sample solution to the preparing task, symbolizes the transition to the next selected week and is therefore dashed, distinguishing it from the other arrows. At the end of the lecture period, students were classified into three groups (low, moderate, and high engagement) based on the number of completed preparing tasks.

Figure 3. Course schedule



Note. The course schedule is presented as a cycle, starting at the top with working on a preparing task, then proceeding clockwise to reporting self-efficacy, completing an assignment, earning a bonus point, and receiving a sample solution.

4 Results

To address the research questions regarding the effects of guiding feedback on students' performance (RQ1), calibration (RQ2), and self-efficacy (RQ3), we analyzed their responses to the assignments and their self-efficacy ratings on the Likert scale. Based on these data, students' calibration was assessed by computing accuracy and bias scores. As stated in the hypotheses, higher engagement with preparing tasks was expected to significantly enhance students' performance (H1) and calibration (H2) but not their self-efficacy (H3).

Table 1 presents the descriptive statistics of each group for all dependent variables. Across all pairwise comparisons, groups with higher engagement exhibited greater performance, accuracy, and self-efficacy scores. For bias, groups with higher engagement had lower scores in each comparison, indicating that students in the low and moderate engagement groups overestimated their abilities, while those in the high engagement group slightly underestimated their abilities. Furthermore, for all variables, score differences between the low and moderate engagement groups were smaller than those between the moderate and high engagement groups. The descriptive statistics provided initial support for the proposed hypotheses, showing substantial differences between the groups in performance and both calibration measures, but only marginal differences in self-efficacy.

Table 1. Descriptive statistics for performance, calibration, and self-efficacy

Dependent variable	**Group**	***M***	***SD***
Performance	Low engagement	0.28	0.26
	Moderate engagement	0.34	0.16
	High engagement	0.61	0.29
Accuracy	Low engagement	3.48	1.47
	Moderate engagement	3.97	1.38
	High engagement	5.30	0.89
Bias	Low engagement	1.92	3.06
	Moderate engagement	1.44	2.01
	High engagement	-0.21	1.17
Self-efficacy	Low engagement	4.67	1.96
	Moderate engagement	4.84	1.54
	High engagement	5.21	1.82

Note. Overview of the means (*M*) and standard deviations (*SD*) for each dependent variable and each group. Performance ranges from 0 to 1; Accuracy scale ranges from 0 to 7; Bias ranges from -7 to 7, Self-efficacy ranges from 1 to 8.

To evaluate whether the differences are statistically significant, proper tests are necessary. Since the available data were not normally distributed according to Shapiro-Wilk tests, the use of parametric methods such as ANOVA was not appropriate to examine whether the groups differ significantly. Instead, a Kruskal–Wallis test was conducted for each variable. This is a nonparametric method for independent samples that does not require the data to be normally distributed. The results revealed significant differences between the groups for performance ($\chi^2(2) = 43.70$, $p < .001$, $\eta^2 = .21$), accuracy ($\chi^2(2) = 54.30$, $p < .001$, $\eta^2 = .27$), and bias ($\chi^2(2) = 33.42$, $p < .001$, $\eta^2 = .16$). In contrast, differences in self-efficacy were not statistically significant ($\chi^2(2) = 3.31$, $p = .191$, $\eta^2 = .01$). All p-values were adjusted using the Holm-Bonferroni method to control the family-wise error rate.

The findings provide clear support for hypotheses H1 and H2, demonstrating that statistically significant differences exist between at least two groups regarding performance ($p < .001$) as well as both calibration measures ($p < .001$). Furthermore, the result for self-efficacy ($p = .191$) supports Hypothesis H3, indicating that self-efficacy does not differ significantly across the groups.

In a subsequent step, performance and both calibration measures were analyzed in greater detail to determine which group differences were statistically significant. Because the data were not normally distributed, Student's *t*-test was not used. Instead, pairwise

Mann–Whitney *U* tests with continuity correction were conducted for each dependent variable (see Table 2). The p-values were adjusted again using the Holm-Bonferroni method to control the family-wise error rate. Significant differences were found for all three variables (performance, accuracy, and bias) between the moderate and high engagement groups and between the low and high engagement groups. In contrast, no significant differences were observed between the low and moderate engagement groups. The effect sizes were medium for all significant comparisons (Cohen, 1988, p. 82).

Table 2. Mann–Whitney *U* tests with continuity correction

Dependent variable	**Group comparison**	***z***	***p***	***r***
Performance	Low and moderate engagement	2.00	.098	.21
	Moderate and high engagement	4.56	< .001	.38
	Low and high engagement	5.01	< .001	.41
Accuracy	Low and moderate engagement	1.34	.200	.16
	Moderate and high engagement	5.07	< .001	.41
	Low and high engagement	5.93	< .001	.48
Bias	Low and moderate engagement	1.34	.220	.16
	Moderate and high engagement	3.47	< .001	.36
	Low and high engagement	4.18	< .001	.34

Note. Overview of the test statistics (z), the p-values (p), and the effect sizes (r) of the Mann–Whitney *U* tests for each group comparison.

5 Discussion

This study investigated three research questions regarding the effects of guiding feedback on students' performance (RQ1), calibration (RQ2), and self-efficacy (RQ3) in a mathematics course. The results revealed that higher engagement with preparing tasks that include guiding feedback was associated with significantly higher performance and calibration but not self-efficacy. In this chapter, these results are discussed by focusing first on the significant effects related to performance, accuracy, and bias (see Chapter 5.1.1), followed by the non-significant findings for self-efficacy (see Chapter 5.1.2). Subsequently, limitations of the study (see Chapter 5.2) and implications for teaching and research are addressed (see Chapter 5.3).

5.1 Effects on dependent variables

5.1.1 Performance, accuracy, and bias

The hypotheses for research questions RQ1 and RQ2, that higher engagement with tasks that include guiding feedback enhances students' performance and calibration, were supported by the study. The following section explains possible factors that may have contributed to the confirmation of these hypotheses. The discussion begins by addressing the first research question regarding students' performance, followed by the second research question on calibration, considering both students' accuracy and bias. Finally, this chapter addresses why the effects of guiding feedback are particularly pronounced at a high level of engagement, with significant differences observed only when the high engagement group was compared with the other groups.

Regarding performance (RQ1), the process of deconstructing mathematical tasks into discrete sub-steps may have facilitated students' acquisition of skills and construction of knowledge, thereby enhancing their performance (Corbalan et al., 2010; Kaarakka et al., 2019). This aligns with cognitive load theory, which focuses on optimizing working memory during learning (Paas et al., 2003; Sweller, 2011). According to the theory, *extraneous load* is caused by inefficiently structured materials and should be minimized, while *germane load* involves cognitive resources for processing and understanding the material and should be maximized. The provision of guiding feedback could have contributed to a reduction in extraneous load and an increase in germane load by allowing students to focus on specific sub-steps to correct their answers. By focusing on relevant sub-steps, students might have avoided engaging with content that was unnecessary for error correction, which allowed them to allocate their cognitive resources toward addressing their mistakes. Additionally, they received error-specific feedback at each step, which supported them in refining their solutions and deepening their understanding (Narciss, 2008).

Focusing on students' accuracy (RQ2), the findings indicate that high engagement with preparing tasks was associated with significantly higher self-assessment precision. This finding could stem from the fact that students who engaged highly with the preparing tasks became more familiar with the skills required for the tasks before evaluating their abilities on the Likert scale. While working on the preparing tasks, students received positive feedback for correct answers and error-specific feedback for incorrect ones, which likely facilitated a more accurate self-assessment (Butler et al., 2008; Osterhage et al., 2019).

In addition, a reciprocal relationship between performance and accuracy (Ewers & Wood, 1993; Nietfeld & Schraw, 2002) could have amplified the observed effects. According to the reciprocal relationship, performance increases resulting from high engagement with the preparing tasks contributed to the enhancement of accuracy, and vice versa. On the one hand, the greater performance in the high engagement group might have supported the understanding and interpretation of error-specific feedback, enabling students to make more precise self-assessments. On the other hand, their clearer understanding of their areas for improvement may have encouraged them to engage more intensively with the preparing tasks, resulting in greater performance compared to students in the low and moderate engagement groups.

Furthermore, significant effects were observed for students' bias (RQ2). Specifically, while students in the high engagement group slightly underestimated their abilities, those in the low and moderate engagement groups overestimated their abilities. These tendencies may be linked to the performance levels of the groups. The high engagement group achieved approximately twice the performance level of the low and moderate engagement groups. As a result, high-performing students had fewer opportunities to overestimate their abilities, since overestimation could only occur near the upper end of the self-efficacy scale. In contrast, low-performing students had a broader range for potential overestimation, as they could misjudge their abilities to a greater extent. The inverse relationship between performance and bias has been empirically confirmed in the field of mathematics (Champion, 2010; Chen, 2003) as well as in other academic domains (Bol et al., 2005; Hacker et al., 2000; Nietfeld et al., 2005).

Besides the general significance for performance, accuracy, and bias, it is remarkable that the pairwise Mann–Whitney *U* tests revealed significant differences only when comparing the high engagement group to the other groups. The descriptive statistics also supported this finding, indicating that differences between the low and moderate engagement groups were relatively small in terms of performance, accuracy, and bias (see Table 1). The effects of the preparing tasks appear to be more pronounced at a high level of engagement. A possible explanation is that students' high engagement with the preparing tasks may also be positively related to the amount of time and effort invested in completing them. Students in the high engagement group may have been more motivated to complete the preparing tasks, leading to deeper involvement with the provided guiding feedback. The degree of interaction with feedback is referred to as feedback processing, which encompasses not only the comprehension of feedback but also the implementation of appropriate corrective actions based on it (Narciss, 2012). Feedback processing is a critical factor in learning (Jonsson, 2013; Winstone et al., 2017). Students in the high engagement group potentially benefited more effectively from guiding feedback by engaging intensively with it, as this allowed them to gain valuable insights that supported the improvement of their performance (Narciss, 2008; Timms et al., 2016). Moreover, intensive feedback processing could have fostered students' reflection on their knowledge and skills, ultimately enhancing their calibration (Butler & Winne, 1995).

Overall, the results concerning performance and calibration are consistent with prior findings on the generally positive impact of feedback in engineering education (Das, 2023; Lache et al., 2021; Park et al., 2026) and align with previous evidence on the effectiveness of ITF-strategies (Narciss et al., 2014; Narciss & Huth, 2006; Razeghpour & Rolka, 2026). It is plausible that the multi-step structure of the assignments, commonly employed in engineering settings, enhanced the effectiveness of guiding feedback. This structure made it possible for the corresponding preparing tasks to be particularly well divided into distinct sub-steps. These sub-steps enabled students to more effectively evaluate their competencies relative to task demands and identify their strengths and weaknesses, thereby further enhancing both performance and calibration (Beyer, 2002; Kuklick et al., 2023).

5.1.2 Self-efficacy

The hypothesis for research question RQ3—positing that higher engagement with tasks that include guiding feedback does not enhance students' self-efficacy—was supported, as no significant differences were observed between the groups. This section discusses potential factors that may have contributed to this lack of statistical significance. Furthermore, these results are contextualized against the seemingly contrasting findings of the experimental study by Razeghpour and Rolka (2026), which identified significant differences in self-efficacy.

As noted in the introduction, there are essential methodological differences between the present field study and the previous experimental study that help explain the observed findings. In the experimental study, a direct comparison was made between two different feedback strategies, such that groups differed only with respect to the type of feedback received. In contrast, in this field study, students in the low and moderate engagement groups worked on fewer preparing tasks and consequently received less feedback than students in the high engagement group. This difference in feedback frequency may be a critical factor, as particularly in mathematical contexts, students tend to overestimate their abilities—a phenomenon observed across various age groups (García et al., 2016; Pajares & Miller, 1994; Sheldrake et al., 2022). Students who engaged more frequently with the preparing tasks received more critical feedback, which presumably contributed to a reduction in self-efficacy among those who initially overestimated their abilities. In contrast, students who engaged less with the preparing tasks received fewer critical insights and were therefore less aware of their weaknesses (Labuhn et al., 2010). In this regard, a reduction of overly high self-efficacy in the high engagement group, alongside the persistence of inflated self-efficacy in the lower engagement groups, may explain why no significant differences in self-efficacy were observed between groups.

Additionally, the present research also differs from the experimental study because there was no controlled group allocation, and therefore pre-existing differences between the groups likely contributed to the results regarding self-efficacy as well. It is possible that students with an initially high bias were more inclined to avoid the preparing tasks. Research suggests that students' overestimation of their actual abilities is often linked to insufficient effort and the adoption of ineffective learning strategies (Chen & Zimmerman, 2007). Consequently, students with such a bias may have neglected the preparing tasks, mistakenly perceiving them as unnecessary for enhancing their knowledge and skills. This would also explain why the high engagement group did not have significantly higher self-efficacy than the low or moderate engagement groups.

Furthermore, pre-existing performance differences between the groups could have also contributed to the results regarding self-efficacy. Lower-performing students often demonstrate an overly high level of self-efficacy, a phenomenon known as the Dunning-Kruger effect (Kruger & Dunning, 1999). This effect is attributed to insufficient cognitive and metacognitive abilities to recognize one's own weaknesses (McIntosh et al., 2022; Schlösser et al., 2013). The Dunning-Kruger effect has been empirically validated across various domains, particularly within mathematics education (Erickson & Heit, 2015; Yang Hansen et al., 2024). It is therefore plausible that an initially low performance in the low and moderate engagement groups contributed to their relatively high levels of self-efficacy, leading to the observation that the high engagement group did not exhibit a significantly higher self-efficacy than the low or moderate engagement groups.

Nevertheless, it should be emphasized that it cannot be completely ruled out that the preparing tasks may have nevertheless elevated students' self-efficacy, as only the post-differences between the groups were assessed. Consequently, it is possible that the preparing tasks and the provision of guiding feedback enhanced students' self-efficacy in the high engagement group. However, it is likely that students' overestimation in the low and moderate engagement groups contributed to the absence of a significantly higher self-efficacy in the high engagement group. This assumption is supported by the fact that guiding feedback encouraged students to independently correct their errors, which allowed them to attribute their success to their own abilities and thereby could have promoted their self-efficacy (Stajkovic & Sommer, 2000). Moreover, the positive feedback for successfully completed sub-steps as well as for the overall task could have contributed to boosting students' self-efficacy in the high engagement group (Usher & Pajares, 2006; Zientek et al., 2017). However, this potential increase in self-efficacy was likely limited by remaining within a realistic self-assessment, so that the self-efficacy of students with high engagement was not significantly higher than that of students in the low and moderate engagement groups.

5.2 Limitations

As part of this field study, students were naturally grouped based on their voluntary participation in the preparing tasks. This caused the high engagement group to be larger than the others, as most students intensively used the preparing tasks to practice for the assignments. Consequently, the disparity in group sizes restricted the comparability between the groups. The natural grouping of students may also have led to the observed effects not being solely attributable to engagement with the preparing tasks, as pre-existing differences between the groups likely influenced the outcomes as well. Specifically, students with initially low performance may have engaged less with the preparing tasks, which could have contributed to amplifying the evaluated differences between the groups. Studies have indicated that students with lower academic performance often have negative attitudes towards mathematics (Gjicali & Lipnevich, 2021; Kunwar, 2021), which may result in a lack of interest in participating in supplementary educational activities.

Furthermore, limitations existed in the assessment of performance, as the submitted solutions in the assignments were evaluated by the STACK system based solely on the final answers, rather than students' detailed solution processes. This limited the capacity to assign appropriate scores, especially in cases where incorrect answers still contained partially correct approaches.

Finally, it remained unclear to what extent the students actually engaged with the tasks. Students were classified based on the number of preparing tasks they completed, which represents a purely quantitative measure. It would be particularly interesting to investigate to what extent this quantitative measure correlates with qualitative aspects, such as students' feedback processing. However, such an analysis would require more extensive data, which were not collected in this field study.

5.3 Implications

The findings of this field study carry significant implications for both teaching and research. From a pedagogical standpoint, it is noteworthy that most engineering students in this study (132 out of 196) demonstrated a high level of engagement with the preparing tasks. As highlighted in the introduction, tasks that include guiding feedback are particularly valuable for this group of students, as their university education frequently involves advanced, multi-step tasks that require substantial procedural reasoning (Bergsten et al., 2015; Taraban et al., 2007).

The positive effects associated with high engagement in preparing tasks underscore the need to integrate such materials more systematically into engineering curricula. The fact that approximately 67% of participants engaged intensively with the voluntary preparing tasks reflects a clear demand for these learning resources among engineering students. At the same time, it is crucial for instructors to recognize that there is a tendency for students who might benefit most from such support to avoid these opportunities, as they overestimate their abilities. Although this group is relatively small, it is important to address this in teaching practice by emphasizing the relevance and benefits of preparing tasks more explicitly during lectures.

From a research perspective, further studies are required to explore in greater depth how students engage with guiding feedback. As mentioned in the limitations (see Chapter 5.2), qualitative insights would be particularly valuable, as benefits of feedback can only be fully realized when students engage with it meaningfully. For this purpose, students need to be observed while engaging with guiding feedback, asked to think aloud, or interviewed afterward. Furthermore, given the recent developments, it is also desirable to explore how generative artificial intelligence (GenAI) can be effectively integrated into guiding feedback. AI-supported tools have the potential to adapt to students' individual learning processes (Khazanchi et al., 2025; Opesemowo & Adewuyi, 2024), which could optimize ITF-strategies and enhance their effectiveness.

6 Conclusion

To examine the effects of guiding feedback, a field study was conducted in a mathematics course at a German university. Students were offered five voluntary preparing tasks with guiding feedback across several weeks, aimed at practicing topics from probability theory and statistics. Students were asked to assess their self-efficacy on a Likert scale for five assignments and to complete the assignments afterward to demonstrate their performance. The assignments covered the same competencies as the preparing tasks but differed in their numerical values. After students indicated their self-efficacy and completed the assignments, calibration was determined based on the differences between perceived and actual performance, using accuracy and bias scores.

The present research extended previous findings on the effects of guiding feedback (Razeghpour, 2024b) by additionally analyzing students' performance and calibration, with refinements to the analysis methodology to obtain more precise results. These refinements included adjustments to the sample and the grouping of students. Data from 196 students who reported their self-efficacy for at least half of the assignments and completed at least half of them were analyzed, with students grouped based on the number of completed preparing tasks into low (0–1 tasks), moderate (2–3 tasks), and high engagement (4–5 tasks).

The results of this study confirmed the previous findings regarding self-efficacy, as no significant differences in self-efficacy were observed between the groups, even after refinements in the analysis. However, significant differences were found in terms of performance, accuracy, and bias between the low and high engagement groups, as well as between the moderate and high engagement groups. Students in the high engagement group performed significantly better and showed more accurate calibration compared to those in the low and moderate engagement groups. While the low and moderate engagement groups strongly overestimated their abilities, the high engagement group slightly underestimated their abilities.

Overall, the current research contributes to a deeper understanding of the effects of guiding feedback by showing that, although no significant differences in self-efficacy were observed between the groups, students in the high engagement group assessed their abilities more accurately. In contrast, participants in the low and moderate engagement groups tended to overestimate their performance, suggesting that comparing self-efficacy levels alone is insufficient. These findings indicate that engaging with preparing tasks and receiving guiding feedback does not necessarily increase self-efficacy, but rather promotes a more realistic alignment between self-assessment and actual performance.

Research ethics

Author contributions

F.R.: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing - Original Draft, Writing - Review & Editing.

K.R.: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing - Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Informed consent statement

Informed consent was obtained from all research participants.

Data availability statement

Data will be made available on request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Information Age Publishing.

Bergsten, C., Engelbrecht, J., & Kågesten, O. (2015). Conceptual or procedural mathematics for engineering students: Views of two qualified engineers from two countries. *International Journal of Mathematical Education in Science and Technology*, *46*(7), 979–990. https://doi.org/10.1080/0020739X.2015.1075615

Beyer, S. (2002). The effects of gender, dysphoria, and performance feedback on the accuracy of self-evaluations. *Sex Roles*, *47*, 453–464. https://doi.org/10.1023/A:1021600510857

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*(5), 372–382. https://doi.org/10.1016/j.learninstruc.2009.03.002

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Champion, J. K. (2010). *The mathematics self-efficacy and calibration of students in a secondary mathematics teacher preparation program*. [Doctoral dissertation, University of Northern Colorado].

Chen, P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, *14*(1), 77–90. https://doi.org/10.1016/j.lindif.2003.08.003

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education*, *75*(3), 221–244. https://doi.org/10.3200/JEXE.75.3.221-244

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Corbalan, G., Paas, F., & Cuypers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*, *55*(2), 692–703. https://doi.org/10.1016/j.compedu.2010.03.002

Das, D. K. (2023). Exploring the impact of feedback on student performance in undergraduate civil engineering. *European Journal of Engineering Education*, *48*(6), 1148–1164. https://doi.org/10.1080/03043797.2023.2238188

Dyrvold, A., & Bergvall, I. (2023). Static, dynamic and interactive elements in digital teaching materials in mathematics: How do they foster interaction, exploration and persistence? *LUMAT: International Journal on Math, Science and Technology Education*, *11*(3), 103–131. https://doi.org/10.31129/LUMAT.11.3.1941

Erens, R., & Eichler, A. (2018). Role of technology in calculus teaching: Beliefs of novice secondary teachers. In B. Rott, G. Törner, J. Peters-Dasdemir, A. Möller, & Safrudiannur (Eds.), *Views and beliefs in mathematics education* (pp. 221–231). Springer International Publishing. https://doi.org/10.1007/978-3-030-01273-1_20

Erickson, S., & Heit, E. (2015). Metacognition and confidence: Comparing math to other academic subjects. *Frontiers in Psychology*, *6*, 742. https://doi.org/10.3389/fpsyg.2015.00742

Ewers, C. A., & Wood, N. L. (1993). Sex and ability differences in children's math self-efficacy and prediction accuracy. *Learning and Individual Differences*, *5*(3), 259–267. https://doi.org/10.1016/1041-6080(93)90006-E

García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A., & Torrance, M. (2016). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition and Learning*, *11*(2), 139–170. https://doi.org/10.1007/s11409-015-9139-1

Gjicali, K., & Lipnevich, A. A. (2021). Got math attitude? (In)direct effects of student mathematics attitudes on intentions, behavioral engagement, and mathematics performance in the U.S. PISA. *Contemporary Educational Psychology*, *67*, 102019. https://doi.org/10.1016/j.cedpsych.2021.102019

Gravill, J. I., Compeau, D. R., & Marcolin, B. L. (2006). Experience effects on the accuracy of self-assessed user competence. *Information & Management*, *43*(3), 378–394. https://doi.org/10.1016/j.im.2005.10.001

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. https://doi.org/10.1037/0022-0663.92.1.160

Härterich, J. (2019). Using randomized quizzes in undergraduate linear algebra and multivariable calculus. In *Contributions to the 1st International STACK conference 2018*. Friedrich-Alexander-Universität Erlangen-Nürnberg. https://doi.org/10.5281/zenodo.2582873

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, *17*, 63–84. https://doi.org/10.1016/j.edurev.2015.11.002

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, *14*(1), 63–76. https://doi.org/10.1177/1469787412467125

Kaarakka, T., Helkala, K., Valmari, A., & Joutsenlahti, M. (2019). Pedagogical experiments with MathCheck in university teaching. *Lumat: International Journal of Math, Science and Technology Education*, *7*(3), 84–112. https://doi.org/10.31129/LUMAT.7.3.428

Khazanchi, R., Di Mitri, D., & Drachsler, H. (2025). The effect of AI-based systems on mathematics achievement in rural context: A quantitative study. *Journal of Computer Assisted Learning*, *41*(1), e13098. https://doi.org/10.1111/jcal.13098

Kojo, A., Laine, A., & Näveri, L. (2018). How did you solve it? – Teachers' approaches to guiding mathematics problem solving. *Lumat: International Journal of Math, Science and Technology Education*, *6*(1), 22–40. https://doi.org/10.31129/LUMAT.6.1.294

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

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*, 104785. https://doi.org/10.1016/j.compedu.2023.104785

Kunwar, R. (2021). A study on low-performing students' perception towards mathematics: A case of secondary-level community school students of Nepal. *Researcher: A Research Journal of Culture and Society*, *5*(1), 125–137. https://doi.org/10.3126/researcher.v5i1.41384

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*, 173–194. https://doi.org/10.1007/s11409-010-9056-2

Lache, J., Rolka, K., Kallweit, M., Dehling, H., & Meißner, D. (2021). Open educational resources for engineering statistics. In H.-U. Heiß, H.-M. Järvinen, A. Mayer, & A. Schulz (Eds.), *Blended learning in engineering education: Challenging, enlightening – and lasting?* (pp. 995–1004). SEIFI.

McIntosh, R. D., Moore, A. B., Liu, Y., & Della Sala, S. (2022). Skill and self-knowledge: Empirical refutation of the dual-burden account of the Dunning–Kruger effect. *Royal Society Open Science*, *9*(12), 191727. https://doi.org/10.1098/rsos.191727

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), e3292. https://doi.org/10.1002/rev3.3292

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). Lawrence Erlbaum.

Narciss, S. (2012). Feedback strategies. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1289–1293). Springer. https://doi.org/10.1007/978-1-4419-1428-6_283

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brünken (Eds.), *Instructional design for multimedia learning* (pp. 181–195). Waxmann.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, *16*(4), 310–322. https://doi.org/10.1016/j.learninstruc.2006.07.003

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

Narciss, S., & Zumbach, J. (2023). Formative assessment and feedback strategies. In J. Zumbach, D. A. Bernstein, S. Narciss, & G. Marsico (Eds.), *International handbook of psychology learning and teaching* (pp. 1359–1386). Springer International Publishing. https://doi.org/10.1007/978-3-030-28745-0_63

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education*, *74*(1), 7–28.

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research*, *95*(3), 131–142. https://doi.org/10.1080/00220670209596583

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(7), em2478. https://doi.org/10.29333/ejmste/14762

Osterhage, J. L., Usher, E. L., Douin, T. A., & Bailey, W. M. (2019). Opportunities for self-evaluation increase student calibration in an introductory biology course. *CBE—Life Sciences Education*, *18*(2), 16. https://doi.org/10.1187/cbe.18-10-0202

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*(2), 124–139. https://doi.org/10.1006/ceps.1998.0991

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Pals, F. F. B., Tolboom, J. L. J., & Suhre, C. J. M. (2023). Development of a formative assessment instrument to determine students' need for corrective actions in physics: Identifying students' functional level of understanding. *Thinking Skills and Creativity*, *50*, 101387. https://doi.org/10.1016/j.tsc.2023.101387

Park, L. E., Ward, D. E., Moore-Russo, D., Rickard, B., Vessels, V., & Hundley, J. (2026). Positive feedback as a lever to boost students' STEM outcomes. *Personality and Social Psychology Bulletin*, *52*(1), 176–197. https://doi.org/10.1177/01461672241265954

Razeghpour, F. (2024a). Conception of two informative tutoring feedback strategies for mathematical tasks with STACK. In P. Iannone, F. Moons, C. Drüke-Noe, E. Geraniou, F. Morselli, K. Klingbeil, M. Veldhuis, & S. Olsher (Eds.), *Proceedings of FAME 1 – Feedback & Assessment in Mathematics Education* (pp. 254–261). Utrecht University & ERME. https://doi.org/10.5281/zenodo.14231455

Razeghpour, F. (2024b). Exploring the influence of guiding feedback on mathematical self-efficacy: Insights from a field study with engineering students. *LUMAT-B: International Journal on Math, Science and Technology Education*, *9*(2). https://urn.fi/urn:nbn:fi:hulib:editori:lumatb.v9i21

Razeghpour, F., & Rolka, K. (2026). *Informative tutoring feedback strategies in higher education: Effects on students' performance, self-efficacy, and calibration* [Unpublished manuscript].

Saks, K. (2024). The effect of self-efficacy and self-set grade goals on academic outcomes. *Frontiers in Psychology*, *15*, 1324007. https://doi.org/10.3389/fpsyg.2024.1324007

Sangwin, C., & Köcher, N. (2016). Automation of mathematics examinations. *Computers & Education*, *94*, 215–227. https://doi.org/10.1016/j.compedu.2015.11.014

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, *39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2022). Implications of under-confidence and over-confidence in mathematics at secondary school. *International Journal of Educational Research*, *116*, 102085. https://doi.org/10.1016/j.ijer.2022.102085

Skaalvik, E. M., Federici, R. A., & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research*, *72*, 129–136. https://doi.org/10.1016/j.ijer.2015.06.008

Stajkovic, A. D., & Sommer, S. M. (2000). Self-efficacy and causal attributions: Direct and reciprocal links. *Journal of Applied Social Psychology*, *30*(4), 707–737. https://doi.org/10.1111/j.1559-1816.2000.tb02820.x

Sunzuma, G. (2023). Technology integration in geometry teaching and learning: A systematic review (2010–2022). *LUMAT: International Journal on Math, Science and Technology Education*, *11*(3), 1–18. https://doi.org/10.31129/LUMAT.11.3.1938

Sweller, J. (2011). Cognitive load theory. In Jose P. Mestre & Brian H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Taraban, R., Anderson, E. E., DeFinis, A., Brown, A. G., Weigold, A., & Sharma, M. P. (2007). First steps in understanding engineering students' growth of conceptual and procedural knowledge in an interactive learning context. *Journal of Engineering Education*, *96*(1), 57–68. https://doi.org/10.1002/j.2168-9830.2007.tb00915.x

Timms, M., DeVelle, S., & Lay, D. (2016). Towards a model of how learners process feedback: A deeper look at learning. *Australian Journal of Education*, *60*(2), 128–145. https://doi.org/10.1177/0004944116652912

Usher, E. L., & Pajares, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology*, *31*(2), 125–141. https://doi.org/10.1016/j.cedpsych.2005.03.002

Viholainen, A., Tossavainen, T., Viitala, H., & Johansson, M. (2019). University mathematics students' self-efficacy beliefs about proof and proving. *Lumat: International Journal of Math, Science and Technology Education*, *7*(1), 148–164. https://doi.org/10.31129/LUMAT.7.1.406

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Yang Hansen, K., Thorsen, C., Radišić, J., Peixoto, F., Laine, A., & Liu, X. (2024). When competence and confidence are at odds: A cross-country examination of the Dunning–Kruger effect. *European Journal of Psychology of Education*, *39*(2), 1537–1559. https://doi.org/10.1007/s10212-024-00804-x

Zientek, L. R., Fong, C. J., & Phelps, J. M. (2017). Sources of self-efficacy of community college students enrolled in developmental mathematics. *Journal of Further and Higher Education*, *43*(2), 183–200. https://doi.org/10.1080/0309877X.2017.1357071