Received: 31 March 2023 | Accepted: 24 July 2023

DOI: 10.1002/curj.221

ORIGINAL ARTICLE

The Curriculum Journal BERA

Feedback levels and their interaction with the mathematical reasoning process

Robbert Smit[1] | **Patricia Bachmann**[1] | **Heidi Dober**[2] | **Kurt Hess**[2]

[1]St. Gallen University of Teacher Education, St.Gallen, Switzerland

[2]University of Teacher Education Zug, Zug, Switzerland

Correspondence
Robbert Smit, St. Gallen University of Teacher Education, St. Gallen, Switzerland.
Email: robbert.smit@phsg.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung

Abstract

In our multi-method study, feedback levels derived from the well-known feedback model of Hattie and Timperley were used in conjunction with feedback that was related to subject-specific content; here, mathematical reasoning tasks in primary school. Feedback needs to be aligned with the learning process; in the beginning, more task feedback is valuable. Based on the analyses of videos and questionnaires of 44 teachers of 5th- and 6th-grade primary school classes ($N=804$), we demonstrated that feedback for finding an approach and operationalisation were related to feedback on the task. We further showed that feedback at the task level predicted students' achievement in mathematical reasoning via students' interest in mathematics. It might be concluded that the four levels of feedback should be applied by teachers in such a way that they focus on the current problem that is occurring while the student is solving a task.

KEYWORDS

formative feedback, interest for mathematics, mathematical reasoning, primary class students

INTRODUCTION

In their recent review of feedback models, Panadero and Lipnevich (2022) claim that interactions between different feedback elements, for example messages, implementation or student characteristics, should be studied empirically in the classroom. In this paper, we present a study situated in primary classrooms that explored how feedback messages based on the model of Hattie and Timperley (2007) interacted with feedback that was related to subject-specific content. The aim was to investigate the targeted application of the four levels of their model within the context of working on mathematical reasoning tasks. The question, therefore, refers to micro-processes within the curriculum and how this curriculum is implemented (Thijs & Van Den Akker, 2009). Such empirical findings on developing activities that are educationally effective can be considered research-driven curriculum development (Clements, 2010).

Hattie and Timperley (2007) distinguish four levels of feedback messages that address different aspects of students' learning: task, process, self-regulation and self. According to the authors, the first three levels are most beneficial for student learning. It further seems that feedback needs to be aligned with a student's learning phase, for example in the beginning more task feedback is valuable (Hattie & Clarke, 2019). Brooks et al. (2019) suggested, based on observations in English classrooms, a model for learners at different stages of competence. A novice, for example, requires feedback on the task while learners who are already proficient benefit more from process or self-regulatory feedback. The four levels in the model of Hattie and Timperley (2007) are not specifically described for models of learning particular subjects. Moreover, researchers, mostly look at the impact of these four levels on students' learning, in general, but not on specific phases of learning (Harris et al., 2015; Hattie & Gan, 2011). Despite considerable research related to feedback, much remains under-researched about how domain-specific formative feedback can improve learners' learning processes (Fujita et al., 2018). For example, it remains an open question how the four levels of the model of Hattie and Timperley (2007) are useful for the different steps in problem solving or mathematical reasoning.

Mathematical reasoning can be viewed as a line of thought in a problem-solving process that starts with an assertion and ends with a conclusion (Lithner, 2000). Lithner refers to Schoenfeld (1985) who argues that mathematical problems are defined as tasks requiring students to construct methods that are new to them, in order to solve these problems. Formative feedback supports primary students' learning of mathematical reasoning (Smit et al., 2022), while such feedback seems to foster students' interest in mathematics as well (Harks et al., 2014). In addition, appropriately challenging and contextualized reasoning tasks can trigger students' interest in mathematics (Sullivan et al., 2012). On first contact with such a task, situational interest is aroused, but there is also increased uncertainty and more teacher support may be required. Students might initially feel the need to be simply told what to do; yet when the process is ongoing such feedback is wanted less (Renninger & Su, 2019). Hence, the impact of feedback and the reasoning process might be mediated by students' developing interest in mathematics and the current step in the reasoning process should be taken into account when giving feedback.

With respect to the feedback elements mentioned in the MISCA model of Panadero and Lipnevich (2022), we combined the students' interest (characteristic), the feedback level (message) and the steps of the reasoning process (implementation) to study their interaction in an empirical model. Our research supposes that based on the momentary step in the reasoning process of a student, a specific feedback level is most frequent or adequate to foster the students' interest and in the end, their reasoning competence.

FEEDBACK IN THE CLASSROOM

Since Hattie's (2008) extensive meta-analyses, feedback has been counted among the most influential factors for learning success at school, although its effectiveness depends on a number of differentiated criteria, such as the timing or comprehensiveness of the feedback, the complexity of the task or the personal learning requirements (Kluger & DeNisi, 1996; Shute, 2008). According to Ruiz-Primo and Brookhart (2018), formative feedback is a support that goes beyond a written or oral comment and can take the form of a teaching-learning dialogue between teacher and student with the aim of advancing the learning process. The teacher offers stronger and more binding or more reserved and open hints for the further processing of tasks. Feedback takes place in a feedback dialogue between student and teacher in the sense of socio-constructivist learning (Vygotsky, 1978). This means that learners acquire competencies by acting independently and in gradually self-regulated ways in rich learning opportunities (De Corte et al., 2000).

Formative feedback is a central part of the process of formative assessment (Black & Wiliam, 2009). Formative assessment is seen as part of the lesson-designing process (Dick, 1996). Within this framework, learning goals are clarified, the learning process of the students is observed by the teacher, and difficulties are diagnosed. The teacher determines the state of learning in relation to the learning goals and then provides supportive feedback aligned with these goals (Wiliam & Thompson, 2007). A recent study found that many teachers in Switzerland use formative assessment in mathematics classrooms, but that the time spent using it, except for feedback from the teacher, is relatively short. Other aspects of formative assessment, such as the discussion of learning goals or self-assessment, are rare (Buholzer et al., 2020). If feedback is given in a way that promotes learning, other studies showed that formative assessment not only has a positive effect on subject-specific learning but also on intrinsic motivation (Harlen, 2006) and interest (Harks et al., 2013).

According to the feedback model of Hattie and Timperley (2007), feedback should be given at the task level (feedback on the solution of a task; e.g. you have mentioned a possible solution, but the largest possible sum is required), at the process level (feedback on steps in the process while working on a task) and at the self-regulation level (feedback on the control of the learning process; e.g. on the use of self-regulatory strategies, such as checking the steps of the solution again after completing the task). Feedback on the self-level (e.g. praise) should not be applied as it is inherently ineffective for learning or motivation (Benson-Goldberg & Erickson, 2021; Panadero & Lipnevich, 2022). However, such feedback might help build relations and trust (Hattie & Clarke, 2019; Mandouit & Hattie, 2023).

MATHEMATICAL REASONING AND STUDENTS' INTEREST

What exactly is meant by the concept of mathematical thinking is fuzzy; there is still no definition on which the research community has agreed (Jeannotte & Kieran, 2017). However, there is agreement that the formal processes of proving does not have to take place in mathematics lessons in primary classes, but, according to Jahnke and Ufer (2015), statements should be traced back to reasons. The terms ‘argumentation and reasoning’ are used without distinguishing between these two skills (Whitenack & Yackel, 2002), as they do not have clearly defined boundaries (Hanna, 2014). It can be pointed out that reasoning and argumentation, as individual activities, both refer to the same process in mathematics (Conner et al., 2014). In the present study, analogous to the competence model of the Swiss educational standards for mathematics (Linneweber-Lammerskitten et al., 2010), no distinction is made between arguing and reasoning.

Lithner (2000) describes argumentation as a four-step process consisting of (1) a problematic situation, (2) a strategic decision, (3) its application and (4) a conclusion. A similar process can also be seen in Bezold's (2009, p. 131) ‘Argumentation Model for Researcher Tasks’, in which she—following Toulmin (2003)—identifies building blocks as prerequisites for mathematical argumentation as well as building blocks for actual argumentation. In

Toulmin's terms, when a child follows these steps, he or she is making an argument explaining how all the steps or building blocks are logically connected, which then leads to final reasoning. Jeannotte and Kieran (2017) have proposed a model of mathematical reasoning for use in schools that consists of two main aspects: the structural and the procedural. The structural aspect described by Lithner (2000) and Toulmin (2003) refers to specific steps or building blocks that guide discourse, and the procedural aspect distinguishes between different but interrelated categories of thinking processes, for example generalising, justifying and exemplifying (Stylianides, 2008).

Jahnke and Ufer (2015) think of developing mathematical reasoning skills in primary school as the first step in growing into the culture of argumentation in mathematics. Viholainen (2011) sees argumentation at a primary level as a process of exchanging arguments in order to reach the best conclusion. According to Bezold (2009), argumentation in primary school includes explaining procedures, asserting and testing, as well as predicting and generalising. Hence, mathematical reasoning occurs when exploring patterns and structures or describing relationships or connections. Mathematical reasoning and justification are considered important for mathematics education that focuses on understanding or provides learning opportunities for individual and social approaches to mathematical concepts (Thompson & Schultz-Ferrel, 2008).

By the end of primary grade 6, students in Switzerland should be able to justify or falsify simple statements by examining a concrete example, using existing data or by using obvious arguments (Linneweber-Lammerskitten et al., 2010). This includes, for example the following task:

$$0.8, 0.88, 0.888, 0.8888, \ \ldots$$

In this sequence, is a number always larger than the preceding number? Give reasons for your answer!

In terms of content, our tasks were related to three of the four content areas of the standards for grade 6: number and variables, functional relationships and sizes and measures (Swiss Conference of Cantonal Ministers of Education (EDK), 2011).

Following Bezold (2010), we see describing and justifying as the actual argumentation at the primary level. However, the preliminary steps of the reasoning process, such as understanding the situation, applying strategies and carrying out operations are necessary building blocks in order to be able to draw conclusions and, thus, complete the train of thought of argumentation (Lithner, 2000). Accordingly, the reasoning process, in addition to argumentation, also contains the steps of finding an approach and operationalization (Figure 1). In addition to these three steps, argumentation can also be carried out by non-linguistic means (Bezold, 2010). This includes visual representations (e.g. number lines), free drawings or tables (Boonen et al., 2016).

Students' interest refers to a lasting attraction to certain objects, topics or activities—such as mathematics (Krapp, 2002). Theories on the development of interest distinguish between individual interest and situational interest (Krapp, 2002; Renninger & Hidi, 2011). Individual interest is defined as a relatively permanent inclination to engage with a particular topic over time, whereas situational interest is described as a reaction to certain content (Renninger & Hidi, 2011). Following interest theory, Mayer (1998) concludes that students work more persistently and successfully on mathematical problems that interest them than they do on problems that do not. Based on their developmental stage, from emerging to well-developed interests, students might react differently to feedback (Lipstein & Renninger, 2007). Students who are at an emerging level of interest need concrete and fast-acting feedback, whereas students who have a well-developed (or permanent) interest are more open to critical and peer-feedback. Harks et al. (2014) studied the impact of 'process-oriented' feedback on math achievement in secondary school classes. They showed feedback that helps students

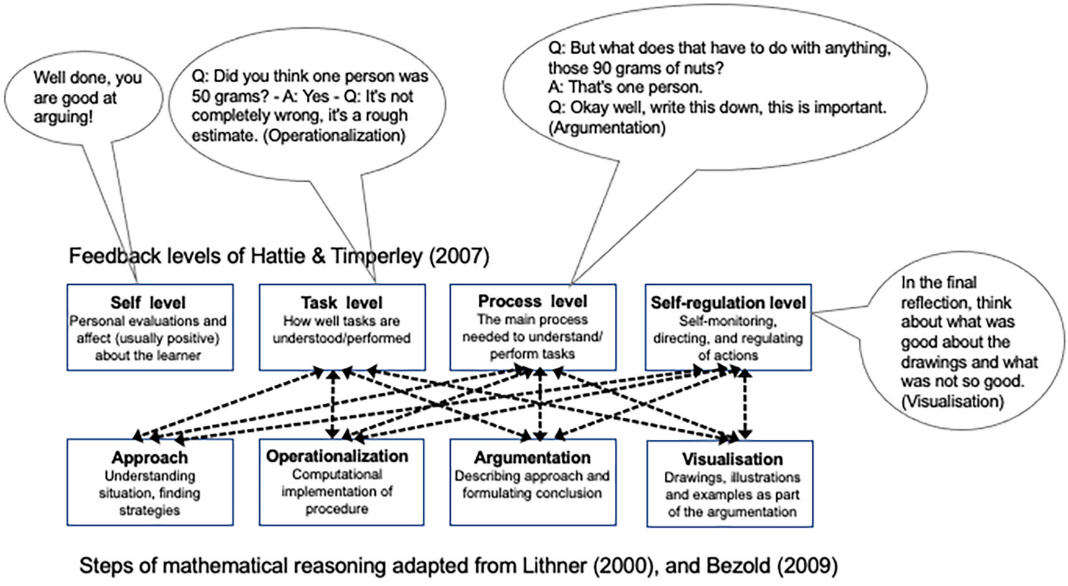

FIGURE 1 Hypothetical relationships between feedback levels, according to Hattie and Timperley (2007), and the steps of mathematical reasoning.

to improve compared to feedback that provides only information about an achieved grade had an effect on students' achievement in mathematics via their interest in the topic. Furthermore, most research has shown that there is a gender effect on students' interest for mathematics in favour of boys (Eriksson, 2020; Frenzel et al., 2010).

Presented research question

In the present study, we investigated whether and how different formative feedback levels and steps of mathematical reasoning interacted (see Figure 1) to explain students' competence for mathematical reasoning as defined in the Swiss standards. According to Hattie and Gan (2011), feedback on the self-level is not effective for learning. Therefore, there are no interaction relationships between the self-level with the steps of the reasoning process depicted in Figure 1.

As this is an exploratory study, we did not have a specific hypothesis about which steps of mathematical reasoning were related to which feedback levels, but we assumed, based on Mandouit and Hattie (2023), that teachers' feedback for the approach and operationalisation steps might be related to the task level, as this is important information for the student to know 'How am I doing'? On the other hand, when a student has reached the end of the line of thought (Lithner, 2000), feedback at the process level is needed to find out whether argumentation is complete. Competent problem solvers use visualization to understand a problem (Krawec, 2014).

We further hypothesized that the frequency of feedback on a specific level would indirectly predict students' achievement in mathematical reasoning via students' interest in mathematics. This supposed mediating effect of students' interest is based on the work of the research group on the project 'Conditions and Consequences of Classroom Assessment' (Co2CA) (Harks et al., 2014; Rakoczy et al., 2008; Rakoczy et al., 2019).

The following research questions (RQ) emerged from the studies described in the previous paragraphs:

RQ1: Is feedback at the task level related to the approach and operationalisation steps, and does mediation by students' interest in mathematics predict achievement in mathematical reasoning?
RQ2: Is feedback at the process level related to the approach and argumentation steps, and does mediation by students' interest in mathematics predict achievement in mathematical reasoning?

METHOD

The present observational study was part of a project titled 'Feedback for Mathematical Reasoning', which was conducted by two teacher–education universities in Switzerland (St. Gallen and Zug). The project began in 2018 and ended in 2022. All participating in-service teachers followed two workshops on mathematical reasoning and supporting feedback. We used a survey with a longitudinal pre-/post-test design and video data retrieved from a lesson in which students worked on reasoning tasks. Longitudinal quantitative methods were useful for testing hypotheses. Teachers' actions were observed easily by means of videos and the inclusion of multiple classrooms permitted generalisation of the findings. Students' interests, which are subjective dispositions, were measured as both individual and class-related variables.

We measured the students' competence in mathematical reasoning at the beginning of the implementation phase (T1). In the same week, the students had to answer a questionnaire about their attitudes and related aspects of learning in the classroom. This included, for example, the methods of formative assessment. Towards the end of the implementation phase in each class, a lesson sequence of students working individually or in small groups on mathematical reasoning tasks was recorded (around 30 min). The teacher was requested to provide individual support for students based on their level of learning. Although we scheduled the videotaped lessons using the same series of reasoning tasks, some recordings contained different reasoning tasks than those that were planned because some teachers were behind with respect to the project's schedule.

Satisfaction, attitudes and perceptions of learning in the classroom were surveyed a second time at the end of the implementation phase (T2). In this study, we limited ourselves to students' interest in mathematics. Finally, the students' competence in mathematical reasoning was measured a second time as well.

Participants

To find the participating in-service teachers with their classes, we used personal enquiries and an advertisement in a regional teachers' magazine. The sample for our study included 44 full-time teachers. Their classes were from two regions in central and eastern Switzerland. Of the teachers, 17 were men and 27 were women, their mean age was 39 years, and their mean length of service was 13 years. Eighteen teachers managed a 5th grade class, 13 were responsible for a 6th grade class and the remaining 13 teachers taught multi-grade classes. Five of the 13 multi-grade teachers even involved their 4th-grade students in the project. Thus, we obtained 44 class datasets at two points in time, consisting of 42 4th graders, 454 5th graders and 308 6th graders (804 students in all). Approximately 51% of the students were boys and 49% were girls. The official school language in the region of the study was German; however, German was not the main language spoken at home for 26% of the students. All teachers were informed about the study's procedure, and parental consent was obtained for the video recordings. All data were anonymised, and data on the performance of the class were communicated to the teachers. The survey was conducted in accordance with the internal guidelines of the ethics committee of the implementing universities.

Procedure

First, the same team members conducted separate workshops at each of the two teacher-education universities. During these two workshops, theoretical knowledge related to the content and teaching of mathematical reasoning was presented. These workshops also served as the initial stage of the curriculum development process by laying common ground. Teachers' experiences so far in using argumentation tasks in class allowed for a stimulating discussion about their impact on student learning. The implementation phase following the workshop included 9 weeks of training for the students. In concrete terms, there was a lesson every week in which the competence of mathematical argumentation was worked on. We asked the teachers to follow a given script which consisted of a detailed lesson plan. Although for research reasons a congruent approach is desirable, we have given teachers the freedom to adapt the script according to the needs of the class and the local situation. These lesson plans had a socio-constructivist orientation, meaning that the learners constructed new knowledge through engagement in activities and mathematical reasoning within a community (Ball & Bass, 2000). A collaborative group or peer work activity (e.g. placemat activities) to enhance students' dialogue, was part of almost every lesson (Knudsen et al., 2014; Mercer & Sams, 2006).

The first part of the script was about the teachers discussing the quality of the different argumentation examples with the class and thus also clarifying the objectives. During the lesson's implementation, the students engaged in peer- and self-assessments of their work. Generally, classroom instruction and group work were most frequent; whereas, towards the end, the students worked more individually on tasks. In the last weeks of the training phase, teachers were asked to give feedback to each student on how good his/her competence in mathematical reasoning already is. All teachers were given a set of mathematical reasoning tasks, along with possible solutions. These exercises covered quantities and units, basic operations, numerical sequences, the decimal system, estimations and proportions. Some of these tasks were related to reality and contextualized, whereas others were purely mathematical.

Instruments

Instrument for rating the video episodes

Our manual for coding videos consisted of two main categories: 1. Four feedback levels according to Hattie and Timperley (2007) and 2. Four steps of mathematical reasoning adapted from Lithner (2000) and Bezold (2009). Coding was applied to situations where teachers and students interacted and where instructional moves (Ruiz-Primo & Brookhart, 2018) were visible. Indicators such as ‘the teacher starts talking to a child’ set the start or, for example ‘the teacher turns away from the child, her or his eyes wander elsewhere’ set the end of an interaction. Instructional moves refer to the notion that feedback is more than just providing information to the student. Before teachers provide feedback, they need to collect information on the students' learning level, and after the feedback, teachers should check on the understanding or use of it by the student.

Questionnaire

Data on the teachers' and students' attitudes and their perceptions of the class were collected using a questionnaire. The items on the teacher's side related to her or his understanding of learning and various questions on the implementation of instruction such as diagnosing, giving feedback and the use of mathematical word problems. On the student side, questions were asked about interest and self-efficacy in mathematics. The questionnaire also asked about the use of self-regulation and the ability to explain the mathematical procedure. Furthermore, the teacher's behaviour in class was to be assessed, that is diagnosing, giving feedback and making learning goals transparent. The four items on students' interest for mathematics were selected from Rakoczy et al. (2005). Our construct refers to individual interest or the permanent attraction to topics or subjects (Krapp, 2002; Renninger & Hidi, 2011). A sample item was: 'I like mathematics'. Answers were reported using a Likert scale ranging from 1 (do not agree at all) to 6 (fully agree). The scale's Cronbach's α was 0.80 at T1 and 0.94 at T2, indicating good internal consistency.

Mathematical reasoning test

All of the items measuring mathematical reasoning were adapted from other standards-based tests or developed by the project's team (see examples below). The items were aligned with the basic competences of Switzerland (Swiss Conference of Cantonal Ministers of Education (EDK), 2011). As the items were open-ended, we expected a student to work approximately 35–40 min on the 10 items during each session. All of the responses to the test items were rated on a scale with four levels of competence, based on a rubric and a detailed manual with a description for each level. The project's procedure and reliability were examined in a pilot project (Smit & Birri, 2014). Satisfactory inter-rater reliability within each rating team was reached: Kappa >0.70. The complete number of test items comprised 14 items, which were presented in two testlets of 10 items each at the two time points. We applied item response theory (IRT) as an alternative to classical test theory. In IRT, item difficulty is assumed to be the characteristic that influences the person's responses to items, and the person's ability is the characteristic that influences the estimates of item difficulty (Bond et al., 2021). To serve as an anchor for IRT calibration, six items were used repeatedly. We applied the WLSMV estimator for ordinal data. Upon completing the IRT analyses (graded response model), final person measures, based on Bayesian plausible values (Von Davier et al., 2009), were computed for the reasoning tests. For students with missing data on one of the two data points plausible values were estimated. The factor score for mathematical reasoning at T1 was standardized, meaning it was fixed at 0 (Var = 1), and the plausible values for the person parameters were calculated from a standard normal distribution. Hence, at T1 the mean was 0, and the SD was 0.81 at both measurement points. At T2, the mean was higher: M = 0.60.

Sample item	
$4 \cdot 6 = 24$	This is a multiplication with two numbers and the result 24.
240: 10 = 24	This is a division with two numbers and the result 24.
Laura claims that there are more divisions than multiplications with the result 24. Is she right? Give reasons.	

Data analysis

Analysis of the video data

In the first step, we identified relevant lesson sequences (teacher–student interactions), and so-called episodes of feedback (Ruiz-Primo & Min, 2013), to examine our ratings. Two coders viewed the recorded lessons of the 44 teachers and identified feedback sequences in which the teachers' feedback helped students move from their current state of understanding to the next step towards mastery of the goal' (Ruiz-Primo & Brookhart, 2018, p. 50). Therefore, teachers were required to ask about and learn what students were thinking. While doing this, the teachers had the learning goals in mind and provided feedback accordingly. Furthermore, 'the student needed to understand and use it [the feedback]' (Ruiz-Primo & Brookhart, 2018, p. 52). The two raters achieved an inter-coder reliability of 73%–83% or a Kappa of 0.64–0.71 for this basic coding, which can be interpreted as good rater agreement (Kersting et al., 2009).

In the second step, the basic codings were applied to the two main categories of 'feedback levels' and 'steps of the reasoning process'. This means the two raters tried to assign each feedback sequence to one of the four levels and one of the four steps. Simple coding by means of weighting was used by the same two coders. The larger proportion of time determined the category. For the four feedback levels, the inter-coding reliability was 79%–89% and the Kappa value was 0.68–0.85. Similar to the four steps of mathematical reasoning, the inter-coding reliability was 72%–92% and the Kappa value was 0.66–0.90.

Questionnaire

Multi-level analysis was used in this study because the units of analysis included teachers and students nested within classrooms. In this approach, it is assumed that the teachers have an influence on the students and that the individual students in turn describe the characteristics of the class. Thus, a variable can be defined simultaneously at the student level and the class/teacher level, and the effects at the class level and individual level may differ.

We applied structural equation modelling in combination with multi-level analysis. The models were estimated with aggregated manifest variables for students' interest (means calculated from parcels of items) to reduce the number of parameters calculated in the complex models, owing to the sample size for the class level (Boivard & Koziol, 2012).

The questionnaire completed by the students on their perceptions of interest contained missing values. Given that these missing values were not due to the study design, we assumed they occurred randomly. Thus, we initially applied the full-information-maximum-likelihood procedure as a model-based treatment of missing data (Enders, 2010).

RESULTS

We start with some descriptive information about the video analyses. Quite similar to other studies (Brooks et al., 2019; Dirkx et al., 2021; Monteiro et al., 2019), feedback on the task level was most frequent (Table 1), while feedback on the process level also occurred quite frequently in the observed lessons. Almost absent was feedback on self-regulation and self-level. Regarding the steps of mathematical reasoning, it appeared that feedback for the first step—finding an approach—was given most often by the teacher. Feedback for the second step (Operationalisation) and third step (Argumentation) were only a third as frequent as feedback for the first step, while feedback for visualisation by the teacher

TABLE 1 Means, standard deviations and intercorrelations of frequencies for feedback levels and steps of mathematical reasoning.

	M	SD	Self-level	Task level	Process level	Self-regulation level	Approach	Operationalisation	Argumentation
Self-level	0.53	1.19							
Task level	82.02	26.48	0.13						
Process level	29.84	11.52	−0.09	0.23					
Self-regulation level	0.34	1.20	−0.13	−0.21	−0.04				
Approach	60.68	26.10	−0.02	0.69**	0.38*	0.00			
Operationalisation	21.66	18.15	0.20	0.58**	0.09	−0.22	−0.04		
Argumentation	18.89	13.74	−0.15	−0.02	0.35*	−0.12	−0.22	−0.11	
Visualisation	8.16	8.23	−0.02	0.04	0.18	−0.09	0.04	−0.05	−0.03

Note: $N = 44$; Mean frequencies per 30 min of lesson time.
Abbreviation: SD, standard deviation.
$^{**}p < 0.01$; $^{*}p < 0.05$.

was almost absent. Hence, the teachers were mainly busy giving feedback for initiating the task-solving process. Once the students had started, the teachers seemed to let them work.

Analysing the correlations, we detected some possible relationships between the frequencies of our two rating categories (levels and reasoning steps). It showed that feedback on the task level and feedback for the approach and the operationalisation correlated the highest (Table 1). A further relationship existed between the process level and the steps approach and argumentation. However, these correlation coefficients were only of medium size. According to Cohen (1992), correlations ranging from $r=0.30$ to 0.49 can be interpreted as medium values, and those above 0.49 are considered large values.

The aim of the following analysis was to clarify whether a certain feedback level occurred during the multiple steps of the reasoning process. Based on these first results and in line with theoretical considerations, we tested two multiple linear regression models for the frequency of feedback levels explained by feedback for the steps of mathematical reasoning. Preliminary analyses were conducted to ensure no violation of normality or homoscedasticity. In the first model (Table 2), the frequency of feedback for the reasoning steps approach and operationalisation explained 83% of the variance for the frequency of feedback on the task level F (2, 41) = 105.21, $p < 0.001$. Both predictors were significant and showed high Beta values. In the second model (Table 3), the frequency of feedback for the steps approach and argumentation explained 35% of the variance for the frequency of feedback on the argumentation level, F (2, 41) = 10.79, $p < 0.001$. The variance explained in the second model was much lower than that explained in the first model, and the Beta effects ranged from small to medium (Cohen, 1988).

The two regression models in Tables 2 and 3 each served as a starting point for the next procedures with the aim of answering RQ1 and RQ2. We analysed the data using two structural equation models (SEM), with each of the three feedback aspects and their effects on mathematical reasoning mediated by the students' interest in mathematics.

First, we present the changes over time of the variables (mathematical reasoning and students' interest). The results were more or less the same for both models. On the class level, mathematical reasoning grew over time from $M_{T1}=0.00$ (SD = 0.81) to $M_{T2}=0.60$ (SD = 0.81), while students' interest in mathematics remained about the same, $M_{T1}=4.21$ (SD = 1.25) and $M_{T2}=4.17$ (SD = 1.43). Mathematical reasoning showed highly stable effects over time ($\beta=0.87$), and students' interest was comparatively less stable ($\beta=0.44$) over time.

We continue with RQ1. Our first SEM tested the relationship between the three feedback aspects (task, approach and operationalisation) and their effects on mathematical reasoning mediated by students' interest in mathematics (Table 4 and Figure 2). This model (SEM1) had very good fit values: χ^2/df = 0.50 (16.40/17), CFI/TLI = 1.00, RMSEA = 0.00, SRMRw = 0.04 and SRMRb = 0.05. The fit values indicate that the model is consistent with the data.

Next, we discuss the relationships between the variables at the class level. Feedback for approach ($\beta=0.71$) and operationalisation ($\beta=0.61$) predicted feedback on the task level, as presented in the regression model before (Table 2). Furthermore, feedback on the task level

TABLE 2 Results of the regression analyses for feedback on the task level.

	B	SE *B*	*Β*	*T*	*p*
Approach	0.72	0.06	0.71	11.20	<0.001
Operationalisation	0.89	0.09	0.61	9.61	<0.001

Note: $N=44$.

TABLE 3 Results of the regression analyses for feedback on the process level.

	B	SE *B*	*β*	*t*	*p*
Approach	0.21	0.06	0.48	3.72	<0.001
Argumentation	0.38	0.11	0.46	3.51	0.001

Note: $N=44$.

TABLE 4 Results of the two multi-level models' predictions of students' mathematical reasoning outcomes.

	SEM 1		SEM 2	
	Task level		Process level	
	Beta	SD	Beta	SD
Student level				
Female → Interest math T1	−0.13**	0.03	−0.13**	0.03
Female → Interest math T2	−0.10	0.05	−0.10	0.05
Interest math T1 → Interest math T2	0.59**	0.05	0.59**	0.05
Interest math T2 → Math reasoning T2	0.04*	0.02	0.04*	0.02
Math reasoning T1 → Math reasoning T2	0.09*	0.04	0.09*	0.04
Interest math T1 ↔ Math reasoning T2	0.19**	0.03	0.19**	0.03
Teacher/Class level				
FB approach → FB task level	0.71**	0.09	—	—
FB operation → FB task level	0.61**	0.07	—	—
FB approach ↔ FB operation	−0.04	0.14	—	—
FB approach → FB process level	—	—	0.46**	0.12
FB argumentation → FB process level	—	—	0.48**	0.15
FB approach ↔ FB argumentation	—	—	−0.22	0.13
FB task level → Interest math T2	0.33**	0.12	—	—
FB process level → Interest math T2	—	—	−0.07	0.25
Interest math T1 → Interest math T2	0.44*	0.22	0.40	0.26
Interest math T2 → Math reasoning T2	0.23*	0.11	0.24*	0.11
Math reasoning T1 → Math reasoning T2	0.87**	0.04	0.87**	0.04
Interest math T1 ↔ Math reasoning T2	−0.08	0.19	−0.08	0.19

Note: $N_{within}=804$; $N_{between}=44$.
$^{**}p<0.01$; $^{*}p<0.05$.

predicted students' interest in mathematics ($\beta=0.33$). Finally, mathematical reasoning was explained by students' interest ($\beta=0.23$), controlling for previous interest in mathematics at T1. This means that classes where teachers were providing more feedback, compared to

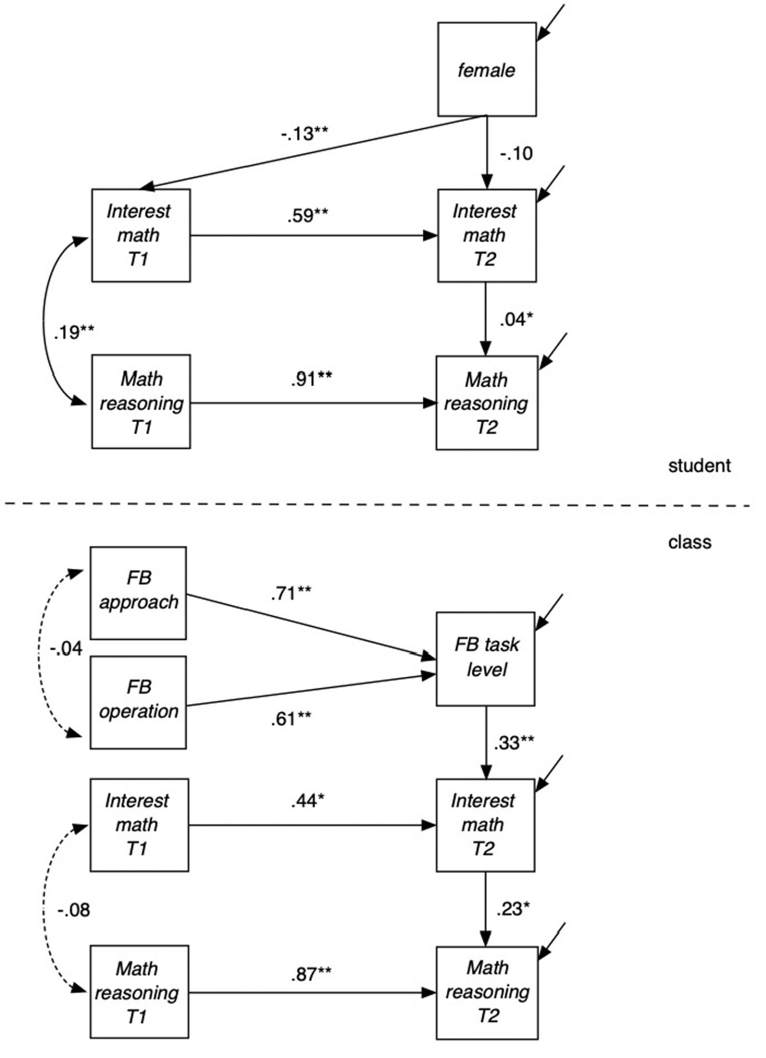

FIGURE 2 Longitudinal multi-level model for the effects of feedback on the task level on mathematical reasoning (SEM 1). $N_{between} = 44$, $N_{within} = 804$; $^{**}p < 0.01$, $^{*}p < 0.05$.

other teachers, whereas students were busy finding an approach or an operationalisation showed higher interest and higher mathematical reasoning competence at the end of the implementation phase. Thus, the relationships at the class level were thus as expected in RQ1.

On the individual level, students' interest predicted mathematical reasoning as well, but the effect size was very low ($\beta = 0.04$). Within the class, a students' interest did change a little over the implementation phase while a students' reasoning competence was extremely stable ($\beta = 0.91$). Hence, almost all the variance in mathematical reasoning competence was explained by previous competence and other variables did not contribute much on the individual level. Regarding control variables, girls showed lower interest in mathematics than boys did, which is in line with other results (Eriksson, 2020; Frenzel et al., 2010). The effect size was small, however, and significant only at T1.

We constructed SEM 2 similar to the way we constructed SEM 1 to answer RQ2. The relationships between the three feedback aspects (process, approach and argumentation) and their effects on mathematical reasoning, which was mediated by students' interest in mathematics, were tested using SEM 2. From our previous linear regression analysis (Table 3), we knew, however, that the effects would be smaller than they were in SEM 1. SEM 2 had good fit values, as well: χ^2/df = 0.10 (24.80/17), CFI/TLI = 0.99, RMSEA = 0.02, SRMRw = 0.04 and SRMRb = 0.12. For SEM 2 on the individual level, the structure was the same as SEM 1; changes only occurred at the class level. Contrary to SEM 1, there was no effect ($\beta = -0.07$) of the frequency of feedback for the process level on students' interest in mathematics (Table 4). In other words, if teachers were providing more feedback on the process level, in comparison to other teachers, then this did not lead to class differences in mathematical reasoning competence at the end of the implementation phase. Thus, the postulated relationships of RQ2 were not supported.

DISCUSSION

The results of our study can help teachers design their lessons in a more optimal way, by tailoring the type of feedback more specifically to the student's learning process and mathematical reasoning (Clements, 2010). In our exploratory study, we intended to examine feedback given by teachers during mathematical reasoning lessons from two different viewpoints. We demonstrated, based on video data from the classroom, that two of the feedback levels of Hattie and Timperley (2007) can be used specifically in subject specific-aspects of learning. Overall, feedback on the process level occurred three times less often in the video data than feedback on the task, whereas the other two levels were marginal. This showed that during the process of solving mathematical reasoning tasks, feedback on the task is given frequently by the teacher or demanded by the student to find an approach or an operationalisation in the reasoning process. Similarly, feedback on the process level was mostly given for discussing the approach or the argumentation in the reasoning process. We further connected the video data to students' mathematical reasoning tests and demonstrated that task-level feedback had an indirect effect on reasoning competence that is mediated by students' interest in mathematics. However, for feedback at the process level, we did not observe such an effect on their achievement in mathematical reasoning.

These results empirically confirmed—at least partially—the hypothesized matrix of feedback for learning by Brooks et al. (2019) by applying the concept to a concrete topic of a specific subject. However, it seems to be more difficult to demonstrate the final effect of this adaptive use of different feedback levels on learning success.

Why could we find an effect for feedback on the task level but not on the process level? What can be learned from the research by Mandouit and Hattie (2023) is that students look first for feedback about whether they have done well, have made errors and where they need to improve—all questions related to feedback on the task. Furthermore, such 'easy to correct' feedback also stimulates positive emotions and persistence. According to Mandouit and Hattie (2023), feedback on the process level—although also perceived as useful—can trigger negative emotions for some students, as they perceive such feedback as discouraging and demotivating. Negative emotions might inhibit students' engagement and the development of interest in mathematical problems (Schukajlow, 2015). Mayer (1998) points out that even if students possess domain-specific knowledge, they might have a problem finding an approach to solve a mathematical problem and need help from the teacher. It might be concluded that once students have found an approach and an operationalization, they might be satisfied with their interest and not curious enough for deeper thinking with respect to further improvement or feedback on the process (Mandouit & Hattie, 2023). This possibility might explain the different results for RQ1 and 2.

Limitations

Each research method has its flaws and limitations. Questionnaires, in general, are limited by their inadequate measures of abstract constructs (Johnson & Christensen, 2012). To ensure adequate measurement of complex constructs, we used a summated rating scale and multiple methods. We believe that the study's video analysis, the test and the questionnaire complemented each other and facilitated a more accurate interpretation of what was happening in the classroom. Nevertheless, we only captured one lesson from each teacher. Thus, it remains unclear to some extent whether the lesson that was observed was typical of the teacher's teaching or feedback behaviour. Feedback at the level of self-regulation was very rarely given. This was also the case in other studies (Brooks et al., 2019; van den Bergh et al., 2013). The question arises as to whether teachers can succeed at all in engaging in a retrospective discussion with learners about their work in a normal lesson. Presumably, the teacher sees it as the main task to get all 20 or so pupils into an active learning state and help them find an approach to the task.

Following the research of Fyfe and Brown (2018), it would be worthwhile to distinguish teachers' feedback in relation to the students' competence levels, as competent students need little or more feedback for self-regulation and less competent students need more feedback on the task and also modelling by the teacher (Brooks et al., 2019). It could be that our videos included mostly interactions with less-competent students. Determining this would require the teacher of each class to identify students according to their competence level in each coded video sequence, so it is important to consider how we could address this methodological challenge in future research.

Another limitation is that we did not study the emotional aspects of learning mathematical reasoning in our study. In the MISCA model of Panadero and Lipnevich (2022), feedback can have different functions, that is it can support students cognitively, motivationally and emotionally. Boekaerts and Corno (2005) found that favourable appraisals of tasks and opportunities for learning (e.g. relevance or interest) motivate students and promote perseverance, whereas unfavourable appraisals (e.g. difficult, irrelevant or stressful) lead to a focus on well-being. Consequently, students anticipating negative feelings try to avoid failure by dawdling or waiting for the teacher's explanation of the task, instead of trying to find their own solutions. The teachers who participated in our study told us that towards the end of the implementation phase, some of their students lost interest and did not persevere. Comforting feedback from the teacher in combination with praise can help students persist in their learning tasks (Eriksson et al., 2017). One might adopt a more differentiated view of Hattie's statement that feedback on the self-level is not beneficial for learning (Mandouit & Hattie, 2023). Perhaps, the problem is that teachers provide too much feedback at the self-level without adding feedback on the task or the process at the same time (Hattie & Gan, 2011; Tunstall & Gipps, 1996).

Finally, we did not consider the use of feedback by the students which is an important point when thinking about feedback being used effectively to enhance achievement and understanding (Mandouit & Hattie, 2023).

Conclusions

As suggested in the MISCA model of Panadero and Lipnevich (2022), we studied interactions between different elements of feedback that practitioners and researchers need to keep in mind. We looked at the students' characteristics (interest), the feedback message (the four levels) and implementation during the four steps of the reasoning process, and studied their interaction in an empirical model. Related to these analyses was the question of whether the intended (micro-) curriculum actually appeared in the theoretically implemented and attained form (Thijs & Van Den Akker, 2009). It became apparent that the ideal form of the components of the curriculum had to be adapted to the conditions of the subject, content and conditions of the classroom.

So far, Hattie and Timperley's feedback levels have mostly been studied independently of subject-specific learning processes and communicated to practitioners. Teachers might think that all levels can be equally useful in any learning phase. Our study showed that the feedback levels for guiding students can be assigned to specific phases of the reasoning process for optimal results. The level of feedback should focus on a current problem as it occurs during the process of solving a task. However, researchers in other subject areas should examine this assumption to ensure its validity. Teachers in other subjects may have a need to know which level of feedback is optimal for a specific phase of the learning process; for example, the process of scientific inquiry or the writing process during language instruction might come to mind.

ACKNOWLEDGEMENTS

Open access funding provided by Padagogische Hochschule St Gallen.

FUNDING INFORMATION

The present project was funded by the Swiss National Science Foundation (Project no. 100019_179230).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in SwissUBase at https://www.swissubase.ch/de/catalogue/studies/13873/18001/overview, reference number 13873.

ETHICS STATEMENT

Data collection followed ethical standards of the swissethics committee and was in accordance with the Declaration of Helsinki. Informed consent was obtained from the participants in the study.

ORCID

Robbert Smit https://orcid.org/0000-0003-2922-916X

REFERENCES

Ball, D. L., & Bass, H. (2000). Making believe: The collective construction of public mathematical knowledge in the elementary classroom. In D. C. Philipps (Ed.), *Constructivism in education: Opinions and second opinions on controversial issues. Yearbook of the National Society for the study of education* (pp. 193–224). University of Chicago Press.

Benson-Goldberg, S., & Erickson, K. A. (2021). Praise in education. In *Oxford research encyclopedia of education*. Retrieved August 14, 2023, from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-1645

Bezold, A. (2009). Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule. Kovac.

Bezold, A. (2010). *Mathematisches Argumentieren in der Grundschule fördern. Was Lehrkräfte dazu beitragen können*. Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Mathe_Bezold.pdf

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, *54*(2), 199–231.

Boivard, J. A., & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 495–511). The Guildford Press.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the rasch model* (4th ed.). Routledge.

Boonen, A., Reed, H., Schoonenboom, J., & Jolles, J. (2016). It's not a math lesson-we're learning to draw! Teachers' use of visual representations in instructing word problem solving in sixth grade of elementary school. *Frontline Learning Research*, *4*(5), 34–61.

Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education (Online)*, *44*(4), 14–32.

Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A., & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. Unterrichtswissenschaft. https://doi.org/10.1007/s42010-020-00083-7

Clements, D. H. (2010). Linking research and curriculum development. In *Handbook of international research in mathematics education* (pp. 603–639). Routledge.

Cohen, J. (1988). The effect size. In *Statistical power analysis for the behavioral sciences* (pp. 77–83). Routledge.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Identifying kinds of reasoning in collective argumentation. *Mathematical Thinking and Learning*, *16*(3), 181–200.

De Corte, E., Verschaffel, L., & Op't Eynde, P. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687–726). Academic Press. https://doi.org/10.1016/B978-012109890-2/50050-0

Dick, W. (1996). The Dick and Carey model: Will it survive the decade? *Educational Technology Research and Development*, *44*(3), 55–63. https://doi.org/10.2307/30221035

Dirkx, K., Joosten-ten Brinke, D., Arts, J., & van Diggelen, M. (2021). In-text and rubric-referenced feedback: Differences in focus, level, and function. *Active Learning in Higher Education*, *22*(3), 189–201. https://doi.org/10.1177/1469787419855208

Enders, C. K. (2010). *Applied missing data analysis*. Guildford.

Eriksson, E., Björklund Boistrup, L., & Thornberg, R. (2017). A categorisation of teacher feedback in the classroom: A field study on feedback based on routine classroom assessment in primary school. *Research Papers in Education*, *32*(3), 316–332. https://doi.org/10.1080/02671522.2016.1225787

Eriksson, K. (2020). Gender differences in the interest in mathematics schoolwork across 50 countries. *Frontiers in Psychology*, *11*, 578092. https://doi.org/10.3389/fpsyg.2020.578092

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507–537. https://doi.org/10.1111/j.1532-7795.2010.00645.x

Fujita, T., Jones, K., & Miyazaki, M. (2018). Learners' use of domain-specific computer-based feedback to overcome logical circularity in deductive proving in geometry. *ZDM*, *50*(4), 699–713. https://doi.org/10.1007/s11858-018-0950-4

Fyfe, E. R., & Brown, S. A. (2018). Feedback influences children's reasoning about math equivalence: A meta-analytic review. *Thinking & Reasoning*, *24*(2), 157-178. https://doi.org/10.1080/13546783.2017.1359208

Hanna, G. (2014). Mathematical proof, argumentation, and reasoning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 404–408). Springer. https://doi.org/10.1007/978-94-007-4978-8_102

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2013). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, *1-22*, 269–290. https://doi.org/10.1080/01443410.2013.785384

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, *34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384

Harlen, W. (2006). The role of assessment in developing motivation for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 61–80). Sage.

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

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge Table of contents only. http://www.loc.gov/catdir/toc/ecip0819/2008021702.html

Hattie, J., & Clarke, S. (2019). *Visible learning feedback*. Routledge.

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction* (pp. 263–285). Routledge.

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

Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 331–355). Springer. https://doi.org/10.1007/978-3-642-35119-8_12

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2009). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*(1–2), 172–181. https://doi.org/10.1177/0022487109347875

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Knudsen, J., Lara-Meloy, T., Stevens Stallworth, H., & Wise Rutstein, D. (2014). Advice for mathematical argumentation. *Mathematics Teaching in the Middle School*, *19*(8), 494–500. https://doi.org/10.5951/mathteacmiddscho.19.8.0494

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383–409. https://doi.org/10.1016/S0959-4752(01)00011-1

Krawec, J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. *Journal of Learning Disabilities*, *47*(2), 103–115. https://doi.org/10.1177/0022219412436976

Linneweber-Lammerskitten, H., Wälti, B., Smit, R., Marc, V., Pochon, L.-O., Frapolli, A., Cadorin, L., Moser Opitz, E., Hirt, U., & Keller, R. (2010). Basisstandards für die Mathematik. Unterlagen für den Anhörungsprozess. 25.10.2010.

Lipstein, R. L., & Renninger, K. A. (2007). "Putting things into words": The development of 12-15-year-old students' interest for writing. *Writing and Motivation*, *19*, 113–140.

Lithner, J. (2000). Mathematical reasoning in task solving. *Educational Studies in Mathematics*, *41*(2), 165–190. http://www.jstor.org/stable/3483188

Mandouit, L., & Hattie, J. (2023). Revisiting "The Power of Feedback" from the perspective of the learner. *Learning and Instruction*, *84*, 101718. https://doi.org/10.1016/j.learninstruc.2022.101718

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*(1), 49–63. https://doi.org/10.1023/A:1003088013286

Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education*, *20*(6), 507–528. https://doi.org/10.2167/le678.0

Monteiro, V., Mata, L., Dos Santos, N. D. N., Sanches, C., & Gomes, M. (2019). Classroom talk: The ubiquity of feedback. *Frontiers in Education*, *4*, 140.

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Rakoczy, K., Buff, A., Lipowsky, F., Hugener, I., Pauli, C., & Reusser, K. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"*. Gesellschaft zur Förderung Pädagogischer Forschung.

Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction: Grading and feedback in mathematics classrooms. *Zeitschrift für Psychologie/Journal of Psychology*, *216*(2), 111–124.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, *60*, 154–165. https://doi.org/10.1016/j.learninstruc.2018.01.004

Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist*, *46*(3), 168–184.

Renninger, K. A., & Su, S. (2019). Interest and its development, revisited. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 205–225). Oxford University Press.

Ruiz-Primo, M. A., & Brookhart, S. M. (2018). *Using feedback to improve learning*. Routledge.

Ruiz-Primo, M. A., & Min, L. (2013). Examining formative feedback in the classroom context: New research perspectives. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 215–232). SAGE Publications, Inc. https://doi.org/10.4135/9781452218649

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Schukajlow, S. (2015). Effects of enjoyment and boredom on students' interest in mathematics and vice versa. Proceedings of the 39th Psychology of Mathematics Education conference, Hobart, Australia.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation*, *43*(December), 5–13. https://doi.org/10.1016/j.stueduc.2014.02.002

Smit, R., Dober, H., Hess, K., Bachmann, P., & Birri, T. (2022). Supporting primary students' mathematical reasoning practice: The effects of formative feedback and the mediating role of self-efficacy. *Research in Mathematics Education*, *1-24*, 1–24. https://doi.org/10.1080/14794802.2022.2062780

Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, *28*(1), 9–16.

Sullivan, P., Clarke, D., & Clarke, B. (2012). *Teaching with tasks for effective mathematics learning*. Springer.

Swiss Conference of Cantonal Ministers of Education (EDK). (2011). *Basic competences in mathematics. National standards*. EDK http://edudoc.ch/record/96784/files/grundkomp_math_d.pdf

Thijs, A., & Van Den Akker, J. (2009). *Curriculum in development*. Netherlands Institute for Curriculum Development (SLO).

Thompson, D. R., & Schultz-Ferrel, K. (2008). *Introduction to reasoning and proof. Grades 6-8*. Heinemann.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, *22*(4), 389–404.

van den Bergh, L., Ros, A., & Beijaard, D. (2013). Feedback during active learning: Elementary school teachers' beliefs and perceived problems. *Educational Studies*, *39*(4), 418–430. https://doi.org/10.1080/03055698.2013.767188

Viholainen, A. (2011). The view of mathematics and argumentation. In *Proceedings of the 7th congress of the European Society for Research in mathematics education, 9th-13th February*. University of Rzeszów.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. In *IERI monograph series* (Vol. *2*, pp. 9–36). IEA-ETS Research Institute (IERI).

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Whitenack, J., & Yackel, E. (2002). Making mathematical arguments in the primary grades: The importance of explaining and justifying ideas. *Teaching Children Mathematics*, *8*(9), 524–528.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.

How to cite this article: Smit, R., Bachmann, P., Dober, H., & Hess, K. (2023). Feedback levels and their interaction with the mathematical reasoning process. *The Curriculum Journal*, *00*, 1–19. https://doi.org/10.1002/curj.221