Check for updates

Measuring adaptive teaching in classroom discourse: Effects on student learning in elementary science education

Ilonca Hardy[1*†], Nicola Meschede[2†] and Susanne Mannel[3]

[1]Institute of Early and Primary Education, Educational Sciences, Goethe-University Frankfurt, Frankfurt, Germany, [2]Institute of Elementary Science Education, University of Münster, Münster, Germany, [3]Institute of Early and Primary Education, Educational Sciences, Goethe-University Frankfurt, Frankfurt, Germany

Adaptive teaching is considered fundamental to teaching quality and student learning. It describes teachers' practices of adjusting their instruction to students' diverse needs and levels of understanding. Adaptive teaching on a micro level has also been labeled as contingent support and has been shown to be effective in one-to-one and small-group settings. In the literature, the interplay of teachers' diagnostic strategies and instructional prompts aiming at tailored support are emphasized. Our study adds to this research by presenting a reliable measurement approach to adaptive classroom discourse in elementary science which includes a global index and the single indices of diagnostic strategies, instructional support, and student understanding. Applying this coding scheme, we investigate whether $N=17$ teachers' adaptive classroom discourse predicts $N=341$ elementary school students' conceptual understanding of "floating and sinking" on two posttests. In multilevel regression analyses, adaptive classroom discourse was shown to be effective for long-term student learning in the final posttest, while no significant effects were found for the intermediate posttest. Further, the single index of diagnostic strategies in classroom discourse contributed to long-term conceptual restructuring. Overall, teachers rarely acted adaptively which points to the relevance of teacher professional development.

KEYWORDS

adaptive teaching, contingent support, science education, classroom discourse, coding scheme

Introduction

Adaptive teaching has been repeatedly claimed pivotal to effective classroom instruction and student learning (Corno, 2008; Hermkes et al., 2018; Parsons et al., 2018; Brühwiler and Vogt, 2020; Gallagher et al., 2022). In adaptive teaching, teachers employ prompts, instructional support, and feedback, taking into account individual differences in increasingly heterogeneous classrooms (Parsons et al., 2018). The construct of adaptive

teaching is regarded a broad category involving teachers' planning, implementation, and reflection of instruction (Hardy et al., 2019). On the level of teacher-student-interactions, adaptive teaching is related to the constructs of scaffolding and contingent support (van de Pol et al., 2010, 2011), which aim at detailed descriptions of teachers' tailored support of student learning on the basis of diagnosis and individualized prompts. Effects of adaptive teaching have been found for different instructional environments such as one-to-one tutoring (Pratt and Savoy-Levine, 1998; Mattanah et al., 2005; Pino-Pasternak et al., 2010; Wischgoll et al., 2015), small group work (van de Pol et al., 2015), and classroom instruction (Decristan et al., 2015a), with benefits for students' conceptual understanding in different domains. In classroom instruction, adaptive teaching is challenging as teachers are confronted with the simultaneous support of diverse students in "multiple zones of proximal development" (Hogan and Pressley, 1997, p. 84). While the functions of adaptive teaching in classroom discourse have been described (e.g., Hogan and Pressley, 1997; Smit et al., 2013), there is a lack of respective measurement tools. In this paper, we present a coding scheme for measuring adaptive teaching in classroom discourse. Applying this coding scheme, we investigate effects of adaptive teaching in classroom discourse on elementary school students' conceptual understanding in the science domain of "floating and sinking," disentangling the relative contribution of the facets of teacher diagnosis, realized student understanding, and instructional support to learning outcomes.

Adaptive teaching

In line with socio-constructivist theories of learning, teaching may be viewed as the constant negotiation of a teacher's activity within the social context of the classroom. Adaptive teaching takes students' differing ability levels as "opportunities to learn" rather than "obstacles to overcome" (Corno, 2008, p.171). It is especially the social context of the classroom that allows teachers to orchestrate learning activities based on individual learning prerequisites for the benefit of all students. Parsons et al. (2018) conjecture that teachers who take individual differences into account and adapt their instruction, also "metacognitively reflect on students' needs before, during, and after instruction" (Parsons et al., 2018, p. 209). These teachers are experts on their students' learning prerequisites and projected learning trajectories, as they reflect on successful instructional designs. Importantly, these teachers are also able to flexibly adapt to students' individual differences and situational changes within the complexity of a classroom setting. Thus, adaptive teaching is related to both professional competence of teachers, teacher epistemologies and beliefs on instruction, and their flexibly applied didactical knowledge. Accordingly, in a recent study, Brühwiler and Vogt (2020) found that adaptive teaching competency, conceptualized as planning, diagnosis, didactics, and content knowledge, showed a measurable impact on student learning outcomes. This effect was moderated by the quality of instruction assessed by student ratings on teachers' classroom management and quality of instructional methods. With regard to instructional processes, Corno (2008) differentiates between adaptive teaching on a micro level and a macro level of instruction. On a macro level, curricular adaptations such as differentiated instruction for subgroups of students are implemented on the basis of quantity, quality, method, media, social setting, content, and instructional time. On a micro level, adaptive teaching is concerned with verbally mediated *in situ* teacher-student interactions. Against this background, Hardy et al. (2019) differentiate between intended and implemented adaptive teaching. Intended adaptive teaching refers to the planning component, where teachers acknowledge student differences in designing instructional environments that fit individual needs and learning prerequisites. Implemented adaptive teaching refers to adaptive instructional episodes in which these planned activities are actually taken up by students, resulting in an alignment of intention and *in situ* implementation.

Classroom implementation of adaptive teaching

On a micro level, adaptive teaching is closely related to the constructs of scaffolding and contingent support (van de Pol et al., 2010) as they pursue similar intentions of adaptive teacher moves based on (diagnosed) student understanding. Due to their high situational constraints, teacher actions on a micro level are the ones that are most challenging (Corno, 2008). Hence, it is also the ongoing diagnosis of student understanding during learning activities that is regarded an element of adaptive teaching on a micro level (Brühwiler and Vogt, 2020).

Instructional support

In the literature on teacher support in instructional activity, various conceptualizations are concerned with formats of support. For example, Lazonder and Harmsen (2016) distinguish between teacher prompts, heuristics, scaffolds, and explanations, where explanations and scaffolds are regarded as explicit formats, and prompts and heuristics as implicit formats. These formats showed moderate, yet overall unspecific effects on student outcomes. In an extension, Vorholzer and von Aufschnaiter (2019) propose three dimensions of instructional teacher support, differentiating the degree of autonomy, the degree of conceptual information, and the cognitive level, including their interplay. Whereas these conceptualizations are mostly concerned with typologies of teacher action, the literature on scaffolding is concerned with teacher support in close interplay with individual learners' current levels of task understanding. This kind of support has also been labeled contingent support and is considered a main characteristic of scaffolding, along with the transfer of responsibility (fading) and the use of diagnostic strategies (Puntambekar and Hübscher, 2005; van de Pol et al., 2010; Hermkes et al., 2018). Following this definition, we consider adaptive teaching in instructional discourse as a teacher's contingent support. In line with Vygotsky's

(1978) construct of the zone of proximal development, contingent (adaptive) teacher support operates on the same or slightly higher level of a learner's current level of competence (Wood et al., 1976; van de Pol et al., 2010), aiming at learners' active knowledge construction. Wood et al. (1978) describe the process of adapting support to students' needs by the Contingent Shift Principle. According to this principle, support is contingent if a teacher increases control, or explicit support, when facing a student's failure on a task and decreases control when whitnessing a student's success at a given task.

With regard to successful scaffolding episodes, Pea (2004) and Reiser (2004) point to relevant cognitive functions of teacher support. Problematizing aims at a central element of teaching quality, i.e., students' cognitive activation to promote their higher order thinking processes (e.g., by provoking cognitive conflicts or justification of ideas; Reiser, 2004; Praetorius et al., 2018). Structuring aims at reducing complexity of the learning situation by means of focusing, highlighting, or summarizing relevant information. According to Reiser (2004), structuring and problematizing are complementary mechanisms that may be in tension and thus have to be carefully balanced. Whereas too much structure may prevent students from engaging actively in a task, problems that are too complex might lead to frustration. Similarly, Pea (2004) refers to modeling and focusing as higher-order functions of scaffolding. While focusing is used to channel learners' attention to relevant aspects, modeling is used to familiarize learners with advanced reasoning and solution procedures. Overall, teacher actions of modeling, problematizing, focusing and structuring involve a high degree of support intended to support active task involvement by students (Wood et al., 1978). Scaffolding has been shown to be effective for student learning in tutoring situations (Pratt and Savoy-Levine, 1998; Mattanah et al., 2005; Pino-Pasternak et al., 2010) as well as in small group work (e.g., van de Pol et al., 2010, 2014). However, scaffolding seems to be scarce in regular classrooms (van de Pol et al., 2010, 2011).

The role of diagnosis

Given the complexity of adaptive *in situ* interactions, the use of diagnostic strategies is considered a main characteristic of contingent support (e.g., Puntambekar and Hübscher, 2005). The use of diagnostic strategies is assumed to enable teachers to implement support based on a student's current level of understanding. In the model of contingent teaching, van de Pol et al. (2011) describe phases and steps of contingent teaching in teacher-student-interactions which illustrate the cyclic and tailored interplay between diagnostics and instructional strategies. First, a teacher applies diagnostic strategies in order to gather information of students' current levels of understanding, and the student responds to these teacher prompts. Second, the teacher checks the diagnosis if necessary. Third, the teacher offers support, i.e., intervention strategies, on the basis of the student's understanding. The fundamental role of diagnosis for adaptive teaching is also emphasized in the literature on formative

assessment. The construct of formative assessment points to the practices in which teachers gather information on students' understanding and use this information to improve student learning (Black and Wiliam, 1998, 2009; Bell and Cowie, 2001). Research shows that the implementation of formative assessment has a positive impact on student learning and motivation (e.g., Black and Wiliam, 1998; Kingston and Nash, 2011; Decristan et al., 2015a; Hondrich et al., 2018; Lee et al., 2020). Typically, authors distinguish between formal and informal formative assessment practices with the latter taking place in teacher-student-interactions. Shavelson et al. (2008) use the term "on-the-fly" to describe the flexible, immediate and sometimes improvisational nature of informal formative assessment. It is especially within classroom discourse that opportunities for informal assessment arise. Ruiz-Primo and Furtak (2007) describe a teacher's activities within these conversations in their ESRU model of informal formative assessment which shows considerable overlap with the model of contingent teaching (van de Pol et al., 2011): The teacher Elicits a response (e.g., by asking questions or calling for explanations); the Student responds; the teacher Recognizes the student's response (e.g., by acknowledging students' contribution and interpreting its degree of correctness); and then immediately Uses the information to offer support. As the ESRU model was developed in the context of scientific inquiry (e.g., Lazonder and Harmsen, 2016), the authors' conceptualization of "use" implies actively engaging students in the learning process, for example by challenging their thinking or by contrasting ideas. Thereby, similar to the scaffolding function of problematizing, it stresses a student's active knowledge construction. According to these authors, more than one iteration of the cycle of eliciting, recognizing, and using may be needed in order to reach an intended level of student understanding.

Measurement approaches

Whereas the models described above provide theoretical considerations for contingent support, approaches to its measurement aim at disentangling the complex interplay of discourse functions in teacher-student interactions empirically. Hence, most approaches use fine-grained interaction-based codings in one-to-one-settings or group work with small samples. Based on the model of contingent teaching, van de Pol et al. (2011) developed a coding scheme for the analysis of three teachers' interactions with their students during group work. They coded teacher utterances with respect to the three steps of contingent teaching separately; in addition, they rated protocol fragments with regard to their degree of contingency "when the teacher was judged to use information gathered about the student's or students' understanding in his provision of support to the student(s)" (van de Pol et al., 2011, p. 49f). Results show that contingency appeared to be scarce. In particular, the investigated teachers barely used diagnostic strategies (see also Elbers et al., 2008; Lockhorst et al., 2010). Moreover, a variation of intervention strategies and a low

level of instructive task approaches were related to contingency. Further measurement approaches applied the Contingent Shift Principle (Wood et al., 1978) to teacher-student-interactions. van de Pol et al. (2012) not only coded a teacher's behavior in terms of the degree of control, but also assessed the learners' understanding. In order to determine the contingency of support they defined coding rules which consider the adequacy of teacher control relative to student understanding. In a study based on this coding scheme, van de Pol et al. (2015) showed that contingent support affects students' learning outcomes in small group work. In addition, the effectiveness of scaffolding was determined by the time that group members worked with each other. Hermkes et al. (2018) refined the measurement approach by van de Pol et al. (2012). For example, van de Pol et al. (2012) defined three-turn sequences (teacher-student-teacher) as units of analysis for contingency. As this may ignore the complexity of small group interactions, Hermkes et al. (2018) divided the interactions into episodes according to the "Student Level of Attainment." That is, every time a shift in students' level of understanding is observed (in terms of higher or lower understanding of the task), a new episode begins. Moreover, whereas van de Pol et al. (2012) coded student understanding "according to the apparent judgement of the teacher" (van de Pol et al., 2012, p. 94), Hermkes et al. (2018) judged the students' understanding based on its objective subject-related correctness. Their analyses indicate a valid measurement approach and revealed contingent interactional patterns in five preservice teachers' activities in small group work. Still, the authors did not measure the effect of contingent support on learning outcomes.

In what way may the reported approaches to measuring contingent support be applied to classroom settings? In this regard, Hogan and Pressley (1997) refer to the consideration of multiple zones of development within larger groups of students, their diverse communication styles, curriculum and time constraints, the need for student assessment, the ownership of ideas, and the uncertainty of endpoints in classroom discussions. Smit et al. (2013) add that the layered, distributed, and cumulative nature of scaffolded discourse in the classroom is most distinctive compared to one-to-one settings. Ruiz-Primo and Furtak (2007) applied their ESRU-model to the classroom discourse of three science teachers and mainly found incomplete ESRU cycles. Though teachers elicited and recognized student understanding, they often did not use this information for their instructional support. More recently, Furtak et al. (2017) added sequential analysis to their ESRU model approach by examining four classrooms. They found differential patterns of teacher reactions to students' utterances, indicating differences in contingency.

The present study

Previous research suggests that the implementation of adaptive teaching on a micro level of instruction affects student learning. However, these findings in the traditions of scaffolding and contingent support are mainly based on one-to-one tutorial settings or small group work and they are limited to small samples. In addition, the distinct contribution of facets of adaptive teaching in classroom discourse to student learning outcomes have not been investigated empirically. The present study adds to prior research by investigating the effects of adaptive teaching in classroom discourse on elementary school students' science learning. We aim at developing a reliable measurement approach to analyze adaptive classroom discourse and at disentangling the relative contribution of diagnosis, instructional support, and displayed student understanding, to student learning outcomes in the domain of "floating and sinking." We address the following research questions:

1. May a reliable instrument to adaptive classroom discourse be devised on the basis of existing approaches?
2. Does teachers' adaptive classroom discourse predict elementary school students' conceptual understanding of "floating and sinking" on two posttests?
3. What is the contribution of the indices of diagnostic strategies, instructional support, and student understanding in classroom discourse to students' conceptual understanding in the posttest measures?

Based on the theoretical background outlined above, we hypothesize that teachers' adaptive classroom discourse predicts student learning. We expect a separate contribution of the three indices of adaptive discourse to student conceptual understanding as they have been empirically related to student learning in prior research.

Materials and methods

Sample

The data base consists of $N = 17$ transcribed science lessons from 17 teachers with their respective students. In each class, the third lesson of unit 1 of a curriculum on "floating and sinking" was videotaped using a standardized procedure. The average class size was 20.1 students ($SD = 3.44$). The mean age of the participating teachers was 39.5 years ($SD = 9.3$). Of the teachers, 16 were female. The mean age of the 341 third-grade students was 8.3 years ($SD = 0.6$); 49% of the students were female. All classes were mixed gender classes. All students came from public primary schools in Germany, located in urban and rural areas.

The sample is taken from a larger sample of $N = 54$ teachers who took part in an extensive professionalization study on the effectiveness of three teaching approaches (scaffolding, formative assessment, peer tutoring) and an intervened control group (parental counseling) for teacher and student development in elementary science (see Decristan et al., 2015a,b for a detailed description of the intervention and results on student outcomes). In the context of the professionalization, the teachers implemented

a curriculum with two units on the science topic of "floating and sinking" within 4 months, including a pretest, an intermediate posttest after unit 1 and a final posttest after unit 2. For the purpose of this analysis, we employed all teachers with consent to be videotaped of the control group and the professionalization group of scaffolding (total $N = 17$). In both groups, the professionalization started with a workshop on the subject matter of "floating and sinking" addressing the scientific concepts of density and buoyancy (4.5 h). Then, the group of scaffolding participated in three consecutive workshops with a focus on instructional strategies such as eliciting and supporting scientific argumentation, student cognitive activation, and task differentiation in the context of the curriculum on "floating and sinking" (3×4.5 h). The control group instead participated in three workshops on parental counseling, a topic that is not related to the science curriculum. For the purpose of this study, we combined the two groups in order to maximize variance with regard to teachers' discursive patterns of science talk within the given curricular unit.

Curriculum

The participating teachers of both the control group and the scaffolding group implemented a curriculum with two extensive instructional units on the topic of "floating and sinking" of empirically established effectiveness for student learning, see Hardy et al., 2006; Decristan et al., 2015a,b. Both units consisted of 9 lessons of 45 min each, combinable as double lessons of 90 min. According to the standard schedule, each unit was expected to span two and a half weeks. The units focused on the concept of density (unit 1) and the concepts of buoyancy force and displacement (unit 2). To standardize the implementation of the curriculum, teachers were provided with the following materials: (a) a manual with detailed lesson plans and lesson goals, (b) each lesson's worksheets for individual and group student work, and (c) boxes of complementary material for student experiments (e.g., objects of the same class of material, differing in weight and/or size; see Decristan et al., 2015a, for a detailed description of the curriculum of unit 1). In the curriculum, students are confronted with authentic and challenging tasks such as "Why would an iron ship float in water, but an iron cube sink in water?" Students are engaged in sequenced experiments based on inquiry-based learning targeting principles such as water displacement, material classes, and buoyancy (Hardy et al., 2006). The focus of classroom discourse is on the joint construction of knowledge, allowing students to share initial hypotheses, insights from science experiments, and their conclusions based on observed outcomes.

The focus and lesson goal of the videotaped lesson 3 is on the concept of density to describe different solid objects' floating and sinking in water. In detail, students first saw a demonstration experiment by the teacher in which two objects of differing densities floated or sunk, respectively, in water. Then, an extended period of classroom talk in which the students used their previously stated hypotheses and arguments to explain the objects' behavior in water took place. The students also worked on tasks determining the weight of standard cubes of different material, comparing them, and finally positioning them from light to heavy standard cubes. In the second part of the lesson, the teacher asked the students to come up with ideas to visualize these objects' densities. In this context, classroom talk on the students' ideas on visualizations and their usefulness to represent different densities took place. Finally, the students worked on tasks applying the concept of density with different objects to predict their floating and sinking in water.

Measures

Student conceptual understanding

Students' conceptual understanding in the domain of "floating and sinking" was assessed by a pretest, an intermediate posttest (PT1), and a final posttest (PT2), see Decristan et al. (2015a,b). PT1 was administered following the instructional unit 1; PT2 was administered following the instructional unit 2, ~4 months after the administration of the pretest. The pretest consisted of 16 items (expected *a posteriori*/plausible value [EAP/PV] reliability = 0.52), the intermediate posttest and the final posttest each included 13 items (expected *a posteriori*/plausible value [EAP/PV] reliability = 0.70 in the intermediate posttest and 0.76 in the final posttest). There were seven items common to the pretest and the two posttests. All tests included multiple-choice items and two free-response items. For example, students were asked to give an explanation to the question of why a large, heavy ship of iron does not sink in water. The items were adapted from tests on "floating and sinking" by Hardy et al. (2006) and Kleickmann et al. (2010). As in previous work, students' answers were scored according to three levels of conceptual understanding as naïve conceptions (0), conceptions of everyday life (1), or scientific conceptions (2). The free-response items were double-coded ($\kappa = 0.87$). Pre- and posttests were separately scaled using a Partial Credit Model. Weighted likelihood estimates were used to estimate student ability parameters. As the pretest's reliability was not sufficient, it was not considered in the subsequent statistical analyses; instead, the student control measures were employed as predictors on the student level. The intra-class correlations of PT1 and PT2 ($\text{ICC}_{\text{PT1}} = 0.008$; $\text{ICC}_{\text{PT2}} = 0.013$) indicated that a substantial amount of variance was located at the classroom level.

Student control measures

We assessed students' degree of scientific competence, their cognitive ability, and their language proficiency as control measures prior to instruction. These measures were administered at the beginning of the school year, ~4 months before the implementation of the instructional unit 1 (see Decristan et al., 2015a). For science competence, we used an adapted version of the TIMSS-test (Martin et al., 2008) in elementary school. The test consisted of 13 items in total [expected *a posteriori*/plausible value (EAP/PV)

reliability = 0.70], covering the domains of physics, chemistry, and geography. For cognitive ability, we employed the CFT 20-R (Weiß, 2006), which is a standardized German version of the Culture Fair Intelligence Test, consisting of 56 items (Cronbach's α = 0.72). For language proficiency, we employed a self-constructed test targeting German word and sentence comprehension, adapted from German diagnostic tests of language comprehension (Elben and Lohaus, 2001; Petermann et al., 2010; Glück, 2011), with a total of 20 items (Cronbach's α = 0.72).

Coding scheme of adaptive classroom discourse

Of the N = 17 teachers, each two phases of classroom discourse in the videotaped lesson 3 (total of 90 min) of the curricular unit 1 were transcribed. Phase 1 occurred directly at the beginning of the lesson. The teacher presented an experiment with two cubes of different materials and size but the same weight. The teacher then explored ideas on whether and why these cubes would float or sink in water in a class discussion. Phase 2 occurred in the middle of this lesson after the students had already worked on tasks with cubes of different materials to understand the concept of density and before a phase of individual work on application tasks took place. In phase 2, the teacher asked the students to come up with ideas for the visual representations of the concept of density. The two phases were selected as representative for science classroom discourse involving opportunities for active student participation, conceptual restructuring, and scientific reasoning (Osborne, 2014).

In step 1 of the coding procedure, we determined units of analysis in the transcribed discourse of phases 1 and 2 of the science lessons. The units of analysis were based on the processes of scientific reasoning of posing questions, hypothesizing, describing and reporting, and interpreting results. With the switch to a different process of scientific reasoning, a new unit of analysis is marked. Altogether, seven units were segmented for phases 1 and 2 in each of the transcripts. If a scientific reasoning process did not occur in the transcribed discourse, it was indicated as missing. Thereby, our units of analysis are broader compared to assessment approaches of contingent support in one-to-one-settings or small group work (e.g., van de Pol et al., 2012; Hermkes et al., 2018) in order to account for the cumulative nature of classroom discourse. The procedure was based on Furtak et al. (2010) who devised a coding scheme for scientific reasoning in science classroom discourse in which teacher-student interaction is first segmented into reasoning units and then analyzed based on defined criteria for argumentation levels. In our sample, the coding of units of analysis was done by one rater and was then communicatively validated with a second rater. For the purpose of the analyses presented here, the two raters' final judgment of units of analysis was used.

In step 2 of the coding procedure, we applied a coding scheme on adaptive classroom discourse to the identified units of analysis. The coding scheme is based on existing measurement approaches, extending them to classroom discourse in an extensive cycle of theory-based deduction of categories and their empirical validation The coding scheme includes three indices which describe central facets of adaptive classroom discourse based on the ESRU-model (Ruiz-Primo and Furtak, 2007) and the model of contingent teaching (van de Pol et al., 2011): (a) diagnostic strategies, (b) instructional support, and (c) student understanding. The indices draw on observable verbalizations of teachers and students and were double-coded for each unit of analysis by independent raters with a high interrater agreement (diagnostic strategies: κ = 0.81, instructional support: κ = 0.74; student understanding: κ = 0.86). In addition, in line with van de Pol et al. (2012) and Hermkes et al. (2018), we defined coding rules for the combination of these three indices in order to determine a global index of adaptive classroom discourse (see also Hardy et al., 2020). Table 1 gives an overview of the coding scheme.

Diagnostic strategies

We coded the frequency of diagnostic strategies within a unit of analysis ranging from "0 = no occurrence" to "3 = full occurrence." Diagnostic strategies were teacher prompts and questions targeting explication of student understanding, procedural explanations or rephrasing student answers. Following van de Pol et al. (2011) this code also includes a teacher's checking of his or her diagnosis, i.e., the teacher verifies if she understood the student correctly or elicits more information on student understanding.

Instructional support

According to van de Pol et al. (2011), a greater variation in instructional strategies enables a teacher to adapt support to students' individual learning needs. This may hold especially true for classroom discourse where multiple zones of proximal development have to be considered (Hogan and Pressley, 1997). Moreover, following the step "using information" of the ESRU-model (Ruiz-Primo and Furtak, 2007), the teacher should recognize the student's previous response when offering support. Hence, we differentiated whether the teacher's support is related to students' responses (level 1 = no relation to student response; level 2–5 = relation to student response) and the degree to which teachers used a variation of instructional strategies. This variation ranges from simple strategies of support with mainly high degree of support (i.e., structuring, e.g., by providing answers, modeling) and multiple extended strategies which combine strategies offering higher support and strategies promoting students' active knowledge construction (i.e., problematizing, e.g., by prompting, contrasting ideas).

Student understanding

First, a content-specific correct answer to a teacher task was defined (see Hermkes et al., 2018). Student answers were then related to this intended answer. The frequency of correct student responses with respect to the defined answer was coded for each unit of analysis, ranging from "0 = no occurrence" to "2 = full occurrence." If content-based student participation did not occur, the respective unit was not coded.

Adaptive classroom discourse (global index)

In order to determine the degree of adaptive teaching in classroom discourse, we used a combination of the three indices,

TABLE 1 Coding scheme of adaptive classroom discourse.

Index	Levels	Example and annotation
(a) Diagnostic strategies	0 = no occurrence. 1 = occasional occurrence: One diagnostic strategy. 2 = frequent occurrence: Two diagnostic strategies but missed opportunities. 3 = full occurrence: Three or more diagnostic strategies without missed opportunities.	Level 3: T: Do you have any ideas on how to visualize the cubes' different weight on the blackboard? How may you change the pictured cubes in a way rendering the cubes' different weights perceptible to us? … So you can see that some of them are light, some of them are heavy. Maya? S: Using different colors. T: Marking them by different colors! **How would you do that?** *[diagnostic strategy]* S: We could paint the heavy ones in red and the light ones in blue. T: **Do I get it right? You think red is a more appropriate color for cubes that sink in water?** *[checking the diagnosis]*
(b) Instructional support	0 = no occurrence. 1 = simple strategy (without student recognition): Providing the solution of the central task(s). 2 = simple strategy (with student recognition): Providing the solution of the central task(s). 3 = simple strategies with low variation (with student recognition): Using simple strategies which mainly offer high support (structuring, modeling). 4 = extended strategies with intermediate variation (with student recognition): Occasionally promoting students' thinking (e.g., cognitive conflict, contrasting students' answers) in addition to simple strategies. 5 = multiple extended strategies (with student recognition): Frequently promoting students' thinking (e.g., cognitive conflict, contrasting students' answers) in addition to simple strategies.	Level 1: T: **Let us make the cubes' weight visible with the help of different points. The Styrofoam cube only weighs one gram. Let us draw only a few dots in here, for that reason. (Drawing two dots in the Styrofoam cube on the blackboard).** *[solution of the task <u>without any student recognition</u>]* Level 5: S1: I would draw three or four dots in all the cubes that sink in water. I would give two to that one and only one [dot] to the one on top. T: **What do you think of Alexander's idea? Elisa?** *[Promoting thinking]* […] T: **Alexander, can you relate your own idea to the one of Elisa. What did you do?** *[promoting thinking]* S1: I changed the number of dots. T: That's an interesting idea. **We could change the number of dots.** *[structuring]* T: **Kevin, which one would get only a few dots?** *[promoting thinking]*
(c) Student understanding	0 = no occurrence: The majority of students' responses does not correspond to the intended answer (e.g., student answers were mostly incorrect and/or a deterioration of understanding within the unit of analysis). 1 = occasional occurrence: Both intended and unintended student responses occur. 2 = full occurrence: The majority of students' responses corresponds to the intended answer (e.g., answers were mostly correct and/or an improvement from unintended to intended student responses in the unit of analysis).	Level 2: T: How can you change the pictured cubes in a way rendering the cubes' different weights perceptible to us? So you can see that some of them are light, some of them are heavy. Maya? S4: Using different colors. T: Marking them by different colors! How would you do that? S4: **We could paint the heavy ones in red and the light ones in blue.** *[intended student answer]* S5: **We could paint all the cubes sinking in water in blue.** *[unintended student answer]* T: Now we have two different options. How shall we proceed and why?

defining specific coding rules (van de Pol et al., 2012; Hermkes et al., 2018) from "1 = low adaptive" to "3 = highly adaptive." Units of analysis were assigned a code of low adaptive (level 1) when

1. diagnostic strategies occurred at least occasionally (level 1) as these form the basis for tailoring support,
2. instructional support was related to students' responses (levels 2–5) as this is an indicator of the actual use of diagnostic information, and
3. student understanding occurred at least occasionally (levels 1–2) as support should only be regarded as

contingent if it at least partly relates to correct understanding.

If one of these conditions was not fulfilled, the level of "0 = non adaptive" was assigned. The more diagnostic strategies a teacher used, the higher was the rating of adaptive classroom discourse. That is, given that instructional support was related to students' responses, student understanding occurred at least occasionally, and diagnostic strategies occurred frequently, adaptive classroom discourse was coded as intermediate (level 2). With full occurrence of diagnostic strategies, highly adaptive classroom discourse was assigned (level 3).

Results

We conducted several multilevel regression analyses in Mplus 7 (Muthén and Muthén, 1998–2012), using robust likelihood estimation in order to investigate the predictive power of our measures of contingent support on students' performance in PT1 and PT2 on "floating and sinking." We were interested in individual-level (i.e., students' science competence, cognitive ability, and language proficiency) and classroom-level (codings of adaptive classroom discourse) variables on student performance in the posttest measures. Individual-level variables and classroom-level variables were standardized ($M = 0$, $SD = 1$). Student level variables were introduced as grand-mean centered level 1 predictors representing variance within and between classes (Lüdtke et al., 2009). Regression analyses were estimated as doubly-manifest models according to the framework proposed by Marsh et al. (2009), with single manifest indicators for the scales at the individual level. The codings of adaptive classroom discourse (global index, single indices) were aggregated within each classroom to examine differences in codings among classes. There were no missing data at level 2. The amount of missing data at level 1 was only small and non-systematical (see also prior publications of this study of Decristan et al., 2015a,b).

With regard to research question 1, Table 2 displays the descriptives for the different measures of adaptive classroom discourse. It shows that for the total of $N = 119$ units in the sample of $N = 17$ teachers, there was a high degree of variance both for the global index and the three indices of diagnostic strategies, instructional support, and student understanding. As indicated by the mean and the corresponding standard deviation, in the majority of units the level of adaptive classroom discourse (global index) was rather low. The same holds for the index of diagnostic strategies and the index of instructional support. At the level of individual units, the relative frequencies reveal that the full range of levels was used in each of the three single codes of diagnostic strategies (level 0: 66%; level 1: 16%; level 2: 5%; level 3: 0.8%; missing: 11.8%), instructional support (level 0: 22.7%; level 1: 14.3%; level 2: 19.3%; level 3: 24.4%; level 4: 5.0%; level 5: 2.5%;

missing: 11.8%), and student understanding (level 0: 5%; level 1: 65.5%; level 2: 5.9%; missing: 23.5%).

To investigate our research question 2, we computed two multilevel regression models to predict student performance on PT1 (Models 1a and 2a) and on PT2 (Models 1b and 2b). Table 3 displays the correlations of dependent variables. Table 4 displays the outcomes of multilevel regressions. In Model 1a, we introduced student measures (science competence, cognitive abilities, language proficiency) at level 1. In Model 2a, we additionally introduced the measure of adaptive classroom discourse (global index) at level 2 to test the predictive power of coded adaptive classroom discourse beyond student control variables. Likewise, in model 1b, we introduced student measures at level 1 to predict

TABLE 2 Descriptives of indices of adaptive classroom discourse.

Index	Range	M	SD	Min	Max
Global index	0–3	0.29	0.60	0	3
Diagnostic strategies	0–3	0.32	0.62	0	3
Instructional support	0–5	1.80	1.39	0	5
Student understanding	0–2	1.01	0.38	0	2

Total number of units = 119; missings = 14.

TABLE 3 Intercorrelations of student variables.

Variable	1	2	3	4
1 Science competence	–			
2 Language proficiency	0.561	–		
3 Cognitive ability	0.369	0.392	–	
4 Intermediate posttest	0.402	0.374	0.297	–
5 Final posttest	0.396	0.402	0.313	0.485

TABLE 4 Multilevel regression analyses predicting student performance on the posttests from individual-level variables and global index of adaptive classroom discourse.

Predictor	Model 1a (PT1)		Model 2a (PT1)		Model 1b (PT2)		Model 2b (PT2)	
	B	SE	B	SE	B	SE	B	SE
Individual level								
Science competency	0.25***	0.05	0.25***	0.05	0.23***	0.04	0.22***	0.04
Language proficiency	0.18***	0.05	0.18***	0.05	0.22***	0.06	0.22**	0.07
Cognitive ability	0.14**	0.05	0.13**	0.05	0.14*	0.06	0.14*	0.06
Classroom level								
Global index			0.50	0.94	–		0.44*	0.22
R^2 (within)	0.20***	0.05	0.21***	0.05	0.23***	0.05	0.22***	0.05
R^2 (between)			0.25	0.95			0.19	0.19

***$p \leq 0.001$; **$p < 0.01$; *$p < 0.05$; one-tailed testing. PT1 = intermediate posttest; PT2 = final posttest.

performance on PT2; in model 2b, we additionally introduced the measure of adaptive classroom discourse (global index).

All measures at the individual level were significantly related to student performance on the two posttests (Models 1a and 1b). The global index of adaptive classroom discourse was shown to be a statistically significant predictor on PT2 (Model 2b), after controlling for individual pretest performance on level 1. In PT1 (Model 2a), the global index does not significantly contribute to the prediction of student performance, while the variables on level 1 remain significant.

In research question 3, the predictive power of the single indices of diagnostic strategies, instructional support, and student understanding were investigated. In Models 3a, 4a, and 5a, we entered our single indicators for contingent support (3a: diagnostic strategies; 4a: instructional support, 5a: student understanding) at level 2 into multilevel regression models, after having entered student measures (science competence, cognitive abilities, language proficiency) at level 1 to predict student performance on PT1. In Models 3b, 4b, and 5b we used the same procedure to predict student performance on PT2, see Table 5.

As may be seen in Table 5, the three indices of adaptive classroom discourse differentially predict students' posttest performance. While in PT1, the three indices do not contribute significantly to an explanation of student outcomes, in PT2, the regression weights of the respective regression models are statistically significant. There is a significant effect of the index of diagnostic strategies on student performance even when controlling for individual measures, contributing to an explanation of between-classroom variation. Also, the indices of instructional support and student understanding significantly contribute to the prediction of student performance in PT2.

Discussion

The present study extended existing approaches on adaptive teaching to the analysis of classroom discourse on a micro level of instruction. Adaptive teaching is regarded a pivotal element of successful learning environments in which teachers base their instructional support on individual student learning prerequisites and needs (Corno, 2008; Parsons, 2012; Gallagher et al., 2022). In prior research based on the model of contingent teaching (van de Pol et al., 2011) and the ESRU model (Ruiz-Primo and Furtak, 2007), teacher moves of contingent support are described as an interplay between their use of (formative) diagnostics and appropriate instructional strategies. In line with these models, we constructed an instrument for adaptive classroom discourse and applied it to a sample of elementary school teachers' classroom teaching. Specifically, we investigated the effects of adaptive classroom discourse on students' conceptual understanding in the domain of "floating and sinking" in two posttests, employing a global index and single indices for the dimensions of diagnostic strategies, instructional support, and student understanding in respective units of analysis.

The measurement approach for adaptive classroom discourse developed in this study proved to be reliable. This is indicated by a high interrater agreement for each of the indices as well as variance in the global index and the single indices, thus capturing differences in adaptive classroom discourse among the participating teachers of this study. Yet, the overall level of adaptive classroom discourse proved to be rather low. With regard to the effects of adaptive classroom discourse on students' learning outcomes we confirmed our hypothesis that the global index is a statistically significant predictor of the final posttest administered ~4 months after the

TABLE 5 Multilevel regression analyses predicting student performance on the posttests from individual-level variables and three indices of adaptive classroom discourse.

Predictor	Model 3a (PT1)		Model 4a (PT1)		Model 5a (PT1)		Model 3b (PT2)		Model 4b (PT2)		Model 5b (PT2)	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Individual level												
Science competency	0.25***	0.05	0.25***	0.05	0.25***	0.05	0.22***	0.04	0.22***	0.04	0.22***	0.04
Language proficiency	0.18***	0.05	0.18***	0.05	0.17***	0.04	0.22***	0.07	0.22***	0.06	0.21**	0.06
Cognitive ability	0.14**	0.05	0.14**	0.05	0.14***	0.05	0.15**	0.06	0.14**	0.06	0.15**	0.06
Classroom level												
Diagnostic strategies	0.89	1.25					0.70***	0.13				
Instructional support			0.08	0.65					0.40*	0.24		
Student understanding					0.54	0.75					0.42*	0.23
R^2 (within)	0.21***	0.05	0.21***	0.05	0.20***	0.04	0.22***	0.05	0.23***	0.05	0.22***	0.05
R^2 (between)	0.79	2.22	0.01	0.11	0.29	0.81	49***	0.18	0.16	0.19	0.17	0.19

***$p \leq 0.001$; **$p < 0.01$; *$p < 0.05$; one-tailed testing. PT1 = intermediate posttest; PT2 = final posttest.

implementation of the curriculum. This provides evidence that contingent support is not only relevant in one-to-one situations or small group work (e.g., Pratt and Savoy-Levine, 1998; Mattanah et al., 2005; Pino-Pasternak et al., 2010) but also in classroom discourse with the goal of advancing conceptual understanding in science. However, the positive effect of adaptive classroom discourse was not confirmed for the intermediate posttest, suggesting that repeated learning opportunities were necessary for adaptive discourse to unfold effects on student conceptual understanding. Following Smit et al. (2013), this may be explained by the long-term cognitive nature of whole-class scaffolding. They argue that whole-class scaffolding is often distributed over several teaching episodes and that its effect on student learning is cumulative in the use of many diagnostic and responsive actions over time, which results in long-term learning processes. In addition, the rather delayed effect of adaptive classroom discourse might be due to the complex nature of conceptual change in the domain of science. For example, Li et al. (2021) found a delayed, but no immediate effect of collaborative argumentation versus individual argumentation on postgraduate students' conceptual change in science lessons. The authors explain this result with the robustness of students' preconceptions. Also, Hardy et al. (2006) found long-term effects on third-graders' conceptual understanding of "floating and sinking" in a quasi-experimental intervention study only for the intervention group using scaffolding within a sequenced and structured curriculum involving multiple opportunities for teacher-guided conceptual restructuring. Research in science education has long established that students tend to enter the classroom with preconceptions that are not in line with scientific concepts (Duit and Treagust, 2003; Schneider et al., 2012; Amin et al., 2014). In order to develop scientific understanding, students need to reorganize or even revise their existing knowledge structures which is considered a complex and long-term process. The results by Li et al. (2021) also underline the distinct role of collaborative argumentation and discourse for conceptual change. Our study adds to these findings by pointing to the relevance of teachers implementing adaptive science classroom discourse. In order to develop scientific understanding, children's preconceptions need to form the basis of dialog in which they can examine competing explanations, advance claims, justify their conceptions, and where they are challenged by new ideas (Osborne, 2010; Osborne et al., 2019). Hence, the nature of teachers' classroom discourse, adapting instructional moves to individual students' conceptions presented in class, is pivotal.

Our global index of adaptive classroom discourse considered the degree to which teachers used diagnostic strategies, the degree to which they employed instructional support of varying sophistication, and the degree to which students displayed conceptual understanding in the respective units of analysis. These elements are considered central facets of contingent support within the literature on scaffolding and formative assessment (Ruiz-Primo and Furtak, 2007; van de Pol et al., 2011). Results of multiple regression analyses with the three single indices, controlling for individual student variables of science competence, language

proficiency, and cognitive abilities, indicate that each contributes to predicting student conceptual understanding in the posttest. As with the global index, there were no significant effects of the three indices on the intermediate posttest. Among the three indices, diagnostic strategies turned out to have the greatest predictive power on student understanding in the posttest (PT2), as indicated by a large regression weight and a substantial contribution to between-group variance. In order to support students' conceptual change toward scientifically sound concepts, teachers need to diagnose their preconceptions and interpret their adequacy in discourse (Ruiz-Primo and Furtak, 2007). The particular role of diagnosis is also stressed in the literature on formative assessment. For example, Decristan et al. (2015b) showed that curriculum-embedded formative assessment promotes student learning but also interacts with classroom process quality. In addition, our finding is in line with research on interactions in small-group and one-to-one-settings. For example, Chiu (2004) found that teachers' interventions had the largest positive effect on small groups' subsequent problem-solving when the teacher evaluated the students' work before offering support. In a related finding by Wischgoll et al. (2019), unsuccessful tutoring situations were those in which tutors failed to offer support responsive to students' level of understanding which the authors attribute to a lack of tutors' use of diagnostics.

Even though our study points to the relevance of adaptive classroom discourse, the descriptive results revealed that teachers' instructional support and their use of diagnostic strategies were rather low. This is in line with existing research (Ruiz-Primo and Furtak, 2007; Elbers et al., 2008; Lockhorst et al., 2010; van de Pol et al., 2011) which also indicates that teachers tend to have difficulties in acting contingently and applying diagnostic strategies. van de Pol (2012) suggests several reasons with respect to the rare use of diagnostic strategies, for example time constraints, a lack of knowledge on diagnostic strategies, and cognitive overload due to managing several instructional processes at the same time. According to Doyle (1986), teachers in general are confronted with a variety of often unpredictable, complex, and uncertain situations. This may particularly apply to classroom discourse. Especially in classroom discourse, contingent support is assumed to be challenging as teachers are faced with a variety of conceptions, thus in need of constant assessment and re-assessment of actualized student understanding for tailored responses (e.g., Hogan and Pressley, 1997). For example, Brühwiler and Blatchford (2011) found a significant correlation between class size and teachers' accuracy of diagnosing students' achievement indicating that teachers diagnose student achievement less accurately in large groups. Apart from a low frequency of employed diagnostic strategies, the level of teachers' instructional support in our data was rather low. According to Corno (2008), teachers have to continually adapt reactions based on assessments, as they teach with "thought and action intertwined" (p. 163). While in our sample, teachers did offer support, they rarely

recognized students' responses when doing so and they generally showed little variation in instructional strategies, mainly using strategies of high support. In a video study on physics instruction, Seidel and Prenzel (2006) also found that teachers barely took up students' statements in classroom discourse and predominantly asked closed questions so that students acted as suppliers or keyword givers to the teacher. Smith et al. (2004) report similar findings for numeracy and literacy instruction with teachers mostly using directive discourse moves, explanations or closed questions. Thus, teachers' tendencies to provide rather high support in a quick reaction to student utterances in classroom discourse may explain our findings of low levels of sophisticated instructional support.

Limitations

Although the present study contributes to understanding the construct of adaptive teaching in classroom discourse, there are some limitations to be discussed. First, the sample was generated from a larger sample of an extensive professionalization study. Some of the teachers who took part in this research were in the control group of the professionalization study and others were trained with respect to scaffolding with a focus on instructional support (see Decristan et al., 2015a). Though this was necessary to maximize variance in teachers' discursive patterns in order to validate the new assessment approach, we did not use data from a naturally observed setting. Second, in the context of the professionalization study the teachers were provided with a curriculum including material and learning tasks which were prepared by the research group. Whereas this offered the opportunity to investigate the micro level of adaptive teaching while to a certain degree standardizing instruction at the macro level, the teachers may have used different instructional strategies in a natural setting. Further research is needed in order to explore the interplay between the macro level and the micro level of adaptive teaching. For example, one may assume that decisions such as task differentiation, group work, or individualized instruction will influence the type of *in situ* instructional support provided by teachers. Third, while we found that the developed instrument was reliable and showed predictive power in terms of student learning, it needs to be further validated. In particular, validity with regard to different instructional settings, content domains, and samples of teachers needs to be established. This is particularly important as the global index of adaptive classroom discourse and the three single indices differed in their predictive power with regard to student learning in the science content investigated in this study. Also, empirical relations to constructs of classroom discourse such as scientific argumentation would help to build discriminant validity of the instrument. Fourth, for analyzing adaptive classroom

discourse, we used a procedure that coded a sequence of utterances within predetermined units of analysis. Though we found evidence for a reliable measurement approach which can be used in further research, it does not allow for determining teachers' responsiveness to and interaction with individual students in detail. An even more fine-grained analysis of classroom discourse, e.g., by applying the Contingent Shift Principle, is a conceivable extension of our coding scheme and may be used to analyze individual students' interactions with their teacher. Fifth, while we coded $N = 119$ discourse units in total, we only considered two phases of classroom discourse per teacher in instructional unit 1. We do not have evidence on teachers' adaptive classroom discourse in instructional unit 2. Praetorius et al. (2014) found that whereas teachers' quality of classroom management and personal support were stable across lessons, their cognitive activation showed high variability. Hence, future research should clarify the stability of teachers' adaptive classroom discourse across lessons. Sixth, we found a conceivably high influence of adaptive teaching on student outcomes on an individual level beyond the relevant individual student prerequisites of cognitive ability, language proficiency, and science competency. However, there might have been additional variables affecting student learning in the long run. Therefore, future research should include variables on the classroom level such as the use of constructive teacher feedback and on the individual level such as student motivational states in order to get a comprehensive picture of instructional influences on student learning. Finally, while our sample size is rather large in relation to other studies analyzing micro-level scaffolding, the sample size, including the skewed distribution of our indices, only meet the requirements for multilevel regressions. In a larger sample, cross-level interactions of indices of adaptive classroom discourse with individual student prerequisites may be analyzed to shed light on adaptive classroom discourse with regard to variables on a student level.

Conclusion

The present study adds to existing research by providing evidence that adaptive classroom discourse affects students' learning. Our results point to the potential of teachers' use of diagnostic strategies and sophisticated instructional support for long-term conceptual restructuring, but they also show that these discourse moves were rarely evidenced in teacher talk. In terms of teachers' professional development, this means that teachers not only need to gain knowledge on instructional strategies, as was the case for part of our sample. They also need to learn how to diagnose students' conceptions in classroom discourse and how to adapt their instructional support based on this information. What is more, teachers need to be prepared to apply this knowledge to multiple students' needs and prerequisites in the complexity of classroom discourse. Recent research shows that

teachers can change their discourse practices through professional development (e.g., van de Pol et al., 2012; Böheim et al., 2021; Borko et al., 2021) and that the use of videos as a basis for teacher reflection is a promising route.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Goethe University Frankfurt, Department of Psychology. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

IH contributed to the design and implementation of the original study. NM, SM, and IH developed the coding scheme. NM and SM coded the data. SM and IH performed statistical data analyses. IH and NM wrote a first draft of the manuscript. Data at the student level was collected within the project IGEL (Individuelle Förderung und adaptive Lern-Gelegenheiten in der Grundschule) of the IDeA Center, as referenced in prior publications by Decristan et al.

(2015a,b). All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amin, T. G., Smith, C., and Wiser, M. (2014). "Student conceptions and conceptual change: three overlapping phases of research" in *Handbook of Research on Science Education*. eds. N. G. Lederman and S. K. Abell (New York, NY: Routledge), 57–81.

Bell, B., and Cowie, B. (2001). The characteristics of formative assessment in science education. *Sci. Educ.* 85, 536–553. doi: 10.1002/sce.1022

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assess. Educ.* 5, 7–74. doi: 10.1080/0969595980050102

Black, P., and Wiliam, D. (2009). Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* 21, 5–31. doi: 10.1007/s11092-008-9068-5

Böheim, R., Schnitzler, K., Gröschner, A., Weil, M., Knogler, M., Schindler, A.-K., et al. (2021). How changes in teachers' dialogic discourse practice relate to changes in students' activation, motivation and cognitive engagement. *Learn. Cult. Soc. Interact.* 28:100450. doi: 10.1016/j.lcsi.2020.100450

Borko, H., Zaccarelli, F. G., Reigh, E., and Osborne, J. (2021). Teacher facilitation of elementary science discourse after a professional development initiative. *Elem. Sch. J.* 121, 561–585. doi: 10.1086/714082

Brühwiler, C., and Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learn. Instr.* 21, 95–108. doi: 10.1016/j.learninstruc.2009.11.004

Brühwiler, C., and Vogt, F. (2020). Adaptive teaching competency. Effects on quality of instruction and learning outcomes. *J. Educ. Res. Online* 12, 119–142. doi: 10.25656/01:19121

Chiu, M. M. (2004). Adapting teacher interventions to student needs during cooperative learning: how to improve student problem solving and time on-task. *Am. Educ. Res. J.* 41, 365–399. doi: 10.3102/00028312041002365

Corno, L. (2008). On teaching adaptively. *Educ. Psychol.* 43, 161–173. doi: 10.1080/00461520802178466

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., et al. (2015a). Impact of additional guidance in science education on primary students' conceptual understanding. *J. Educ. Res.* 108, 358–370. doi: 10.1080/00220671.2014.899957

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., et al. (2015b). Embedded formative assessment and classroom process quality: how do they interact in promoting students' science understanding? *Am. Educ. Res. J.* 52, 1133–1159. doi: 10.3102/0002831215596412

Doyle, W. (1986). "Classroom organization and management" in *Handbook of Research on Teaching*. ed. M. C. Witttrock. *3rd* ed (New York: Macmillan), 392–431.

Duit, R., and Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *Int. J. Sci. Educ.* 25, 671–688. doi: 10.1080/09500690305016

Elben, C. E., and Lohaus, A. (2001). *Marburger Sprachverständnistest für kinder ab 5 Jahren [Marburger Test of Language Comprehension for Children 5 Years of Age and Older]*. Göttingen, Germany: Hogrefe.

Elbers, E., Hajer, M., Jonkers, M., Koole, T., and Prenger, J. (2008). "Instructional dialogues: participation in dyadic interactions in multicultural classrooms" in *Interaction in Two Multicultural Mathematics Classrooms: Mechanisms of Inclusion and Exclusion*. eds. J. Deen, M. Hajer and T. Koole (Amsterdam: Aksant), 141–172.

Furtak, E., Hardy, I., Beinbrech, C., Shemwell, J., and Shavelson, R. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educ. Assess.* 15, 175–197. doi: 10.1080/10627197.2010.530553

Furtak, E. M., Ruiz-Primo, M. A., and Bakeman, R. (2017). Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educ. Meas. Issues Pract.* 36, 28–38. doi: 10.1111/emip.12143

Gallagher, M. A., Parsons, S. A., and Vaughn, M. (2022). Adaptive teaching in mathematics: a review of the literature. *Educ. Rev.* 74, 298–320. doi: 10.1080/00131911.2020.1722065

Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6–10 [Test of Vocabulary of 6- to 10-year-old Children]*. München, Germany: Elsevier, Urban & Fischer.

Hardy, I., Decristan, J., and Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *J. Educ. Res.* 11, 169–191. doi: 10.25656/01:18004

Hardy, I., Jonen, A., Möller, K., and Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *J. Educ. Psychol.* 98, 307–326. doi: 10.1037/0022-0663.98.2.307

Hardy, I., Mannel, S., and Meschede, N. (2020). "Adaptive Lernumgebungen [Adaptive learning environments]," in *Heterogenität in Schule und Unterricht [Heterogeneity in Schools and Instruction.* eds. M. Kampshoff and C. Wiepcke (Stuttgart: Kohlhammer), 17–26.

Hermkes, R., Mach, H., and Minnameier, G. (2018). Interaction-based coding of scaffolding processes. *Learn. Instr.* 54, 147–155. doi: 10.1016/j.learninstruc.2017.09.003

Hogan, K., and Pressley, M. (1997). "Scaffolding scientific competencies within classroom Communi-ties of inquiry" in *Scaffolding Student Learning: Instructional Approaches and Issues.* eds. K. Hogan and M. Pressley (Louiseville, Quebic: Brookline Books), 74–108.

Hondrich, A. L., Decristan, J., Hertel, S., and Klieme, E. (2018). Formative assessment and intrinsic motivation: the mediating role of perceived competence. *Z. Erzieh.* 21, 717–734. doi: 10.1007/s11618-018-0833-z

Kingston, N., and Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educ. Meas. Issues Pract.* 30, 28–37. doi: 10.1111/j.1745-3992.2011.00220.x

Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., and Beinbrech, C. (2010). Die Modellierung naturwissenschaftliche Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion [Modeling scientific competence of primary school children: theoretical background and test construction]. *Zeitschrift für Didaktik der Naturwissenschaften* 16, 263–281. doi: 10.25656/01:3385

Lazonder, A. W., and Harmsen, R. (2016). Meta-analysis of inquiry-based learning: effects of guidance. *Rev. Educ. Res.* 86, 681–718. doi: 10.3102/0034654315627366

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., and Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: a systematic review. *Appl. Meas. Educ.* 33, 124–140. doi: 10.1080/08957347.2020.1732383

Li, X., Li, Y., and Wang, W. (2021). Long-lasting conceptual change in science education: the role of U-shaped pattern of argumentative dialogue in collaborative argumentation. *Sci. Educ.*, 1–46. doi: 10.1007/s11191-021-00288-x

Lockhorst, D., Wubbels, T., and van Oers, B. (2010). Educational dialogues and the fostering of pupils' independence: the practices of two teachers. *J. Curric. Stud.* 42, 99–121. doi: 10.1080/00220270903079237

Lüdtke, O., Robitzsch, A., Trautwein, U., and Kunter, M. (2009). Assessing the impact of learning environments: how to use student ratings of classroom or school characteristics in multilevel modeling. *Contemp. Educ. Psychol.* 34, 120–131. doi: 10.1016/j.cedpsych.2008.12.001

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthen, B., et al. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivar. Behav. Res.* 44, 764–802. doi: 10.1080/00273170903333665

Martin, M. O., Mullis, I. V. S., Foy, P., Olson, J. F., Erberber, E., Preuschoff, C., et al. (2008). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mattanah, J. F., Pratt, M. W., Cowan, P. A., and Cowan, C. P. (2005). Authoritative parenting, parental scaffolding of long-division mathematics, and children's academic competence in fourth grade. *J. Appl. Dev. Psychol.* 26, 85–106. doi: 10.1016/j.appdev.2004.10.007

Muthén, L. K., and Muthén, B. O. (1998–2012). *Mplus (version 7) [computer software]*. Los Angeles, CA: Muthén & Muthén.

Osborne, J. (2010). Arguing to learn in science: the role of collaborative, critical discourse. *Science* 328, 463–466. doi: 10.1126/science.1183944

Osborne, J. (2014). "Scientific practices and inquiry in the science classroom" in *Handbook of Research on Science Education*. eds. N. G. Lederman and S. K. Abell (Routledge), 593–613.

Osborne, J., Borko, H., Fishman, E., Gomez Zaccarelli, F., Berson, E., Busch, K. C., et al. (2019). Impacts of a practice-based professional development program on elementary teachers' facilitation of and student engagement with scientific argumentation. *Am. Educ. Res. J.* 56, 1067–1112. doi: 10.3102/0002831218812059

Parsons, S. A. (2012). Adaptive teaching in literacy instruction: case studies of two teachers. *J. Lit. Res.* 44, 149–170. doi: 10.1177/1086296X12440261

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., et al. (2018). Teachers' instructional adaptations: a research synthesis. *Rev. Educ. Res.* 88, 205–242. doi: 10.3102/0034654317743198

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *J. Learn. Sci.* 13, 423–451. doi: 10.1207/s15327809jls1303_6

Petermann, F., Metz, D., and Fröhlich, L. P. (2010). *SET 5–10. Sprachstanderhebungstest für kinder im Alter zwischen 5 und 10 Jahren [test of language comprehension of 5- to 10-year-old children]*. Göttingen, Germany: Hogrefe.

Pino-Pasternak, D., Whitebread, D., and Tolmie, A. (2010). A multidimensional analysis of parent-child interactions during academic tasks and their relationships with children's self-regulated learning. *Cogn. Instr.* 28, 219–272. doi: 10.1080/07370008.2010.490494

Praetorius, A.-K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic dimensions of teaching quality: the german framework of three basic dimensions. *ZDM-Math. Educ.* 50, 407–426. doi: 10.1007/s11858-018-0918-4

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., and Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learn. Instr.* 31, 2–12. doi: 10.1016/j.learninstruc.2013.12.002

Pratt, M. W., and Savoy-Levine, K. M. (1998). Contingent tutoring of long-division skills in fourth and fifth graders: experimental tests of some hypotheses about scaffolding (no. 2). Elsevier science. *J. Appl. Dev. Psychol.* 19, 287–304.

Puntambekar, S., and Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: what have we gained and what have we missed? *Educ. Psychol.* 40, 1–12. doi: 10.1207/s15326985ep4001_1

Reiser, B. J. (2004). Scaffolding complex learning: the mechanisms of structuring and problematizing student work. *J. Learn. Sci.* 13, 273–304. doi: 10.1207/s15327809jls1303_2

Ruiz-Primo, M. A., and Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *J. Res. Sci. Teach.* 44, 57–84. doi: 10.1002/tea.20163

Schneider, M., Vamvakoussi, X., and van Dooren, W. (2012). "Conceptual change" in *Encyclopedia of the Sciences of Learning*. ed. N. M. Seel (New York, NY: Springer), 735–738.

Seidel, T., and Prenzel, M. (2006). Stability of teaching patterns in physics instruction: findings from a video study. *Learn. Instr.* 16, 228–240. doi: 10.1016/j.learninstruc.2006.03.002

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., et al. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Appl. Meas. Educ.* 21, 295–314. doi: 10.1080/08957340802347647

Smit, J., van Eerde, H. A. A., and Bakker, A. (2013). A conceptualisation of whole-class scaffolding. *Br. Educ. Res. J.* 39, 817–834. doi: 10.1002/berj.3007

Smith, F., Hardman, F., Wall, K., and Mroz, M. (2004). Interactive whole class teaching in the national literacy and numercy strategies. *Br. Educ. Res. J.* 30, 395–411. doi: 10.1080/01411920410001689706

van de Pol, J. (2012). Scaffolding in teacher-student interaction: exploring, measuring, promoting and evaluating scaffolding. doctoral dissertation. Faculty of Social and Behavioural Sciences Institute, research institute child development and education (CDE). Available at: http://hdl.handle.net/11245/1.392689

van de Pol, J., Volman, M., and Beishuizen, J. (2010). Scaffolding in teacher–student interaction: a decade of research. *Educ. Psychol. Rev.* 22, 271–296. doi: 10.1007/s10648-010-9127-6

van de Pol, J., Volman, M., and Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learn. Instr.* 21, 46–57. doi: 10.1016/j.learninstruc.2009.10.004

van de Pol, J., Volman, M., and Beishuizen, J. (2012). Promoting teacher scaffolding in small-group work: a contingency perspective. *Teach. Teach. Educ.* 28, 193–205. doi: 10.1016/j.tate.2011.09.009

van de Pol, J., Volman, M., Oort, F., and Beishuizen, J. (2014). Teacher scaffolding in small-group work: an intervention study. *J. Learn. Sci.* 23, 600–650. Available at: http://www.jstor.org/stable/43828358

van de Pol, J., Volman, M., Oort, F., and Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instr. Sci.* 43, 615–641. doi: 10.1007/s11251-015-9351-z

Vorholzer, A., and von Aufschnaiter, C. (2019). Guidance in inquiry-based instruction – an attempt to disentangle a manifold construct. *Int. J. Sci. Educ.* 41, 1562–1577. doi: 10.1080/09500693.2019.1616124

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Weiß, R. H. (2006). *CFT 20-R. Grundintelligenztest Skala 2 – Revision [Culture Fair Test]*. Göttingen, Germany: Hogrefe.

Wischgoll, A., Pauli, C., and Reusser, K. (2015). Scaffolding—how can contingency lead to successful learning when dealing with errors? *ZDM-Math. Educ.* 47, 1147–1159. doi: 10.1007/s11858-015-0714-3

Wischgoll, A., Pauli, C., and Reusser, K. (2019). High levels of cognitive and motivational contingency with increasing task complexity results in higher performance. *Instr. Sci.* 47, 319–352. doi: 10.1007/s11251-019-09485-2

Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *J. Child Psychol. Psychiatry* 17, 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wood, D., Wood, H., and Middleton, D. (1978). An experimental evaluation of four face-to-face teaching strategies. *Int. J. Behav. Dev.* 1, 131–147. doi: 10.1177/016502547800100203