

RESEARCH ARTICLE OPEN ACCESS

Teacher Guidance and On-the-Fly Scaffolding in Primary School Students' Inquiry Learning

Heide Sasse¹  | Anke M. Weber²  | Timo Reuter¹  | Miriam Leuchter¹ ¹RPTU, Landau, Germany | ²Department of Behavioural and Cognitive Sciences, University of Luxembourg, Luxembourg, Luxembourg**Correspondence:** Heide Sasse (heide.sasse@rptu.de)**Received:** 21 December 2023 | **Revised:** 2 September 2024 | **Accepted:** 25 September 2024**Keywords:** control of variables strategy | inquiry learning | instructional design | primary science education | responsive teaching | scaffolding | teacher guidance

ABSTRACT

In primary science education, inquiry-based science instruction stands out as an optimal learning environment for fostering domain-specific content and procedural knowledge. Recognizing the effectiveness of different forms of teacher guidance, there is an ongoing debate about the planning of high (*structured inquiry*) and low (*guided inquiry*) guidance and their optimal sequencing. This debate revolves around balancing the level of autonomy and the amount of conceptual information given to students. Furthermore, the complete understanding of the impact of responsive teaching, which encompasses a broad range of practices, including on-the-fly scaffolding such as *Promoting Participation*, *Focusing*, and *Problematizing*, remains elusive. To address this gap, this study examines the relationship between planned teacher guidance and specific instances of responsive teaching, particularly on-the-fly scaffolding in the inquiry-based science classroom. A pre-posttest design was employed, involving 164 primary school students ($M = 9.9$ years, $SD = 0.66$, 57% female) and one female experimenter. Domain-specific content knowledge contained science concepts of thermal insulation, whereas procedural knowledge comprised the application of the control-of-variables strategy. The sequential order of planned teacher guidance, *structured inquiry*, and *guided inquiry*, was systematically varied, and the experimenter was allowed to provide spontaneous on-the-fly scaffolding. The study assesses the influence of planned teacher guidance and specific instances of responsive teaching, particularly on-the-fly scaffolding on students' conceptual and procedural knowledge. Results indicate no differential learning effects based on the order of planned guidance. However, when planned *guided inquiry* was provided second, the teacher gave less on-the-fly scaffolding. Additionally, *Problematizing* had a positive effect, while *Focusing* had a negative effect on students' procedural knowledge learning.

1 | Introduction

Scientific inquiry involves students experimenting actively, that is, asking research questions and collecting, analyzing, and interpreting data (NGSS Lead States 2013). According to Kind and Osborne (2017) it includes various styles of reasoning: representing models mathematically, engaging in experimental exploration, constructing hypothetical models, categorizing, and classifying entities, predicting outcomes probabilistically, and using abductive reasoning to explain phenomena. This approach aims at facilitating students' active construction of

domain-specific content knowledge (CK), which in turn is supposed to lead to a significant transfer and long-term retention of knowledge (De Jong et al. 2023; Furtak 2006; Hume 2009; Ruzaman and Rosli 2020). Moreover, by going through a scientific inquiry process, students are supposed to acquire procedural knowledge such as the *control-of-variables strategy* (CVS). The CVS is needed to conduct a conclusive experiment in a specific content domain, with a focal variable that must be manipulated, while all other variables must be held constant (e.g. Chen and Klahr 1999; Van Vo, Csapó, and Greiff 2023). In general, studies indicate that domain-specific

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Science Education* published by Wiley Periodicals LLC.

content knowledge and procedural knowledge are related to one another and that the two can be mutually reinforcing (e.g. Koerber and Osterhaus 2019). Thus, in this study, scientific inquiry is characterized by students' active participation in experimentation, which serves as a critical component of their learning experience.

Scientific inquiry, involving active experimentation, can place a high cognitive load on students (Kaiser and Mayer 2019; Kirschner, Sweller, and Clark 2006; Sweller et al. 2024). The amount of cognitive load that primary school students experience in scientific inquiry involving experimentation may depend in part on the lack of control of variable strategy (CVS) skills, as developmental studies show (Koerber and Osterhaus 2019; Lazonder et al. 2021; Piekny and Maehler 2013). Some students might fail in applying the CVS, by manipulating the wrong variable, producing confounded results, or not even assessing the given question (Koerber et al. 2015).

Without help, primary school students may not be able to engage productively in scientific inquiry involving experimentation and may need teacher guidance (Martella, Klahr, and Li 2020). As several meta-studies have shown, scientific inquiry is effective as a teaching method, provided that students are guided by the teacher (Alfieri et al. 2011; Furtak et al. 2012; Lazonder and Harmsen 2016). With guidance, teachers might reduce the cognitive load (Corbalan, Paas, and Cuypers 2010). For scientific inquiry, teacher guidance can vary the level of inquiry by varying the extent of students' autonomy ranging from high to low and the degree of conceptual information provided (Blanchard, Southerland, and Granger 2009; Vorholzer and von Aufschnaiter 2019). These aspects of teacher guidance may be planned beforehand. However, while carrying out the planned scientific inquiry lesson, the teacher may realize that the respective goal can only be achieved by providing spontaneous scaffolding on the fly, for example, through promoting participation, focusing students' learning, and problematizing students' thoughts. Thus, many teachers adapt their planned lessons and respond in line with the responsive teaching approach. They react to their students' behavior by noticing, attending to, and responding to the substance of their students' thinking, thereby supporting their engagement in disciplinary practices (Hammer, Goldberg, and Fargason 2012).

Previous studies have either primarily compared different scopes of planned guidance: high guidance (such as *structured inquiry*) versus low guidance (such as *guided inquiry*) (Klahr and Nigam 2004; Strand-Cary and Klahr 2008), or order effects of the high-low or low-high guidance sequence (Martella, Klahr, and Li 2020; Matlen and Klahr 2013), or analyzed spontaneous guidance (e.g. Leuchter and Saalbach 2014). However, to the best of our knowledge, studies that compared different scopes of planned guidance and effects of their sequential order have not analyzed responsive teaching aspects like spontaneous scaffolding given on the fly within the planned setting, thus missing an important aspect of teaching in classrooms. In studies that have investigated guidance, it is noticeable that aspects of guidance in scientific inquiry have been investigated in laboratory settings with one-on-one interactions. In these cases, an experimenter conducts highly standardized experimental designs (Klahr and Nigam 2004), in settings with

a single experimenter and multiple children (Ashman, Kalyuga, and Sweller 2020), in real classroom settings where teachers conduct lessons with whole classes (Hammer, Goldberg, and Fargason 2012, Klika and Abels 2016), and in case studies, where one teacher's actions are analyzed in detail (Jaber 2021). As advocated by Klahr and Li (2005), optimizing research findings requires a recurring transition between investigations conducted in both laboratory and classroom settings. Our study aims to contribute to this discourse by conducting research in an ecological setting (real classrooms, with the class split in half), and implementing standardized instruction led by a trained experimenter rather than a teacher. Hence, the objective of our study was to investigate the effects of teaching scientific inquiry in a realistic classroom environment. We aimed to differentiate between the scope of planned guidance and its sequential order within a curriculum. Additionally, we examined the spontaneous guidance provided within the planned instructional setting.

1.1 | Scopes of Planned Teacher Guidance

Teachers plan the type of guidance in scientific inquiry-based learning to tailor their instruction to the learning objectives and students' abilities (Coffman 2017). Previous work in scientific inquiry has defined four types of inquiry with different levels of guidance (level 0 = verification, level 1 = structured, level 2 = guided, and level 3 = open inquiry; (see Blanchard, Southerland, and Granger 2009). In level 0, the teacher gives students a question and data-gathering methods, guiding them toward an expected conclusion. In level 1, students get a question and method but interpret the results themselves, while in level 2, they determine both method and interpretation. Level 3 involves students generating the question, and assuming full responsibility for the scientific inquiry (Settlage and Southerland 2007). This is in line with Vorholzer and von Aufschnaiter (2019) who state that the teacher must consider two dimensions when planning guidance: (1) the extent of autonomy the students are given during scientific inquiry and (2) the degree of conceptual information provided to the students. In settings with low autonomy, the students are supposed to conduct predefined experiments, similar to Blanchard's levels of inquiry 0 and 1, whereas in settings with high autonomy, the teacher (or the student) specifies a research question and leaves it to the students to design the experiments, similar to Blanchard's levels of inquiry type 2 and 3. Additionally, the teacher must decide on whether to provide explicit, implicit, or minimal conceptual information to the students (Vorholzer and von Aufschnaiter 2019). In settings with explicit conceptual information, the teacher provides direct instructions on principles such as the CVS and/or domain-specific content knowledge, corresponding to Blanchard's levels of inquiry type 0 and 1. In settings with implicit conceptual information, the teacher provides the students with hints or asks prepared questions that should help the students to discover conceptual principles without explicit explanations, corresponding to Blanchard's level of inquiry type 2. In settings with minimal conceptual information, students are supposed to discover principles such as the CVS and/or domain-specific content knowledge without receiving any conceptual information from the teacher, corresponding to Blanchard's levels of inquiry type 3 (Vorholzer and

von Aufschnaiter 2019). Thus, depending on the extent of autonomy given, combined with the degree of conceptual information provided, the scope of planned guidance and levels of inquiry will vary.

Numerous studies on primary school students' CVS learning have compared planned teacher-guided settings within *structured inquiry* to planned teacher-guided settings within *guided inquiry*. For instance, Klahr and Nigam (2004) and Strand-Cary and Klahr (2008) demonstrated that 8- to 10-year-old students who conducted guided experiments with explicit conceptual information acquired the CVS after a short time period and were able to apply it. However, primary school students might also learn the CVS in inquiry settings without explicit conceptual information, but then it takes more time (Dean Jr. and Kuhn 2007; Schalk et al. 2019). In Schalk et al.'s (2019) study with 8- to 11-year-old students, the teacher offered implicit domain-specific information, without explicit CVS instruction. After 1 year of school, the students had learned both the domain-specific content knowledge and the CVS. In contrast, Studhalter et al. (2021) found no significant learning of the CVS when 6- to 7-year-old students conducted given experiments with implicit conceptual information provided by the teacher.

Another thread of research investigated sequential order of planned teacher guidance within *structured inquiry* and *guided inquiry*, with inconsistent results. Sweller (2006) suggests that novices benefit from explicit instruction by first studying a step-by-step expert solution to a problem before attempting to solve similar problems themselves. Receiving instruction before attempting to work on a task reduces the cognitive load, enabling students to use their working memory effectively in constructing problem-solving strategies for present and future tasks (Cooper and Sweller 1987). Accordingly, an explicit-minimal sequence was found to be superior to a minimal-explicit sequence for 10-year-old's learning of domain-specific content knowledge about light energy efficiency (Ashman, Kalyuga, and Sweller 2020) and for 16-year-old's learning about the inclined plane (Hsu, Kalyuga, and Sweller 2015). In contrast, Kapur (2008) argues that it can be beneficial for students to initially tackle tasks independently, even if they encounter problems, and subsequently, are given guidance by a teacher. This approach is particularly advantageous for those who already have some level of expertise (Kalyuga et al. 2001). Correspondingly, Schwartz et al. (2011) found a minimal-explicit sequence to be superior to the reverse order for eleventh grade students' learning of kinematics and eighth grade students' learning the concept of density. However, Martella, Klahr and Li (2020) and Matlen and Klahr (2013) found that, at any stage in a sequence of inquiry activities, providing explicit conceptual information was more effective to enhance third grade students' CVS learning than giving minimal conceptual information. The contrasting findings may be due to differences in student age, prior knowledge, and subject complexity.

Nevertheless, teacher guidance is not necessarily a dichotomy where students are given exclusively either planned guidance within *guided inquiry* or *structured inquiry*, but guidance can be seen as a continuum in which the teacher spontaneously adjusts the scope of guidance in the course of the inquiry (Baur and Emden 2020; Zhang 2016). Teachers who adjust their

teaching spontaneously can be seen as having a positive belief in children's rich understanding of the world and their ability to improve it through generative activities. They encourage students to explore and apply effective scientific methods by involving them in the learning process and utilizing their existing resources (cf. Croker and Buchanan 2011; Metz 2011; Siegler 2013; Sodian, Zaitchik, and Carey 1991). Hammer, Goldberg and Fargason (2012) refer to this as a responsive teaching approach. It involves the close observation of student actions and statements, which serve as the basis for the provision of adaptive support. This support is intended to leverage students' existing knowledge and skills, with the objective of making progress towards deep conceptual understanding of scientific phenomena. Teachers may apply spontaneous, on-the-fly scaffolding (for an overview see: Van de Pol, Volman, and Beishuizen, 2010), to adapt the amount and type of planned teacher guidance according to momentary learning situations (Baur and Emden 2020; Corno 2008; Van Leeuwen and Janssen 2019). Thus, researching sequential order of planned teacher guidance within *structured inquiry* or *guided inquiry* in realistic classroom settings, spontaneous teacher guidance such as on-the-fly scaffolding has to be taken into account.

1.2 | Scopes of Spontaneous Teacher Guidance

In line with the responsive teaching approach, spontaneous teacher guidance occurs in response to students' immediate needs or questions and is often flexible in nature, addressing specific learning challenges as they arise during the learning process (Robertson, Scherr, and Hammer 2016; Van de Pol, Volman, and Beishuizen 2010). As an instructional strategy in which a teacher supports and guides students in their learning activities in real time, it can be specified as "on-the-fly scaffolding," that is, the teacher pays attention to the students' needs and intervenes in the learning process when necessary (cf. Ruiz-Primo and Furtak 2006). Such dynamic support is of particular importance when students are confronted with information that is complex, as Costley et al. (2023) and Reiser (2004) emphasize.

In situations where learners need to consider numerous elements simultaneously, explicit instruction is often required. Novice learners faced with complex tasks such as designing experiments benefit significantly from explicit instruction (Ashman, Kalyuga, and Sweller 2020). However, as learners gain expertise or engage in tasks where the complexity of the elements is inherently low, the effectiveness of explicit instruction may decrease, potentially leading to a reversal effect (Chen, Kalyuga, and Sweller 2017; Kalyuga 2007). Therefore, the provision of on-the-fly scaffolding, such as promoting participation, focusing students' learning and problematizing students' thoughts plays a crucial role in responding to the evolving learning needs of students, particularly when considering the varying complexities of learning tasks. This also applies to differently planned levels of inquiry. The less structure, the more flexibility teachers need to support students with on-the-fly scaffolding.

Accordingly, there is an inclination among teachers to provide on-the-fly scaffolding, planned or intuitively, in the discretionary scope of the lesson, as research shows. Hammer, Goldberg

and Fargason (2012) studied a third grade “energy” lesson where one teacher in her class prioritized recognizing and promoting productive scientific thinking over following a strict lesson plan. The data showed the teacher used spontaneous scaffolds like asking for explanations, encouraging comparisons, and categorizing students’ statements to engage their conceptual and epistemological resources. Ainley and Luntley (2007) studied the teaching practices of six mathematics teachers in their own classes in England. They found that some teachers used spontaneous scaffolding effectively, while others lacked the subject expertise to do so. Some teachers, despite having strong subject knowledge, missed opportunities to engage responsively with students. The study highlights how responsive teaching impacts instructional outcomes. Schwarz et al. (2021) examined how 16 elementary science teachers in their own classes influence sense-making by analyzing their responsiveness to students’ thinking. They identified various on-the-fly scaffolding strategies, for example, asking questions, and pressing for reasoning. Klika and Abels (2016) compared the instruction of one teacher in his own eighth grade class using *structured inquiry* (level 1) and *guided inquiry* (level 2) over the course of a year. They found that in the *structured inquiry* setting, only a few students, often the same ones, actively participated in the discourse. In contrast, students who were less involved in the *structured inquiry* lessons showed more engagement and even took on leadership roles in the *guided inquiry* lessons. However, the students mainly focused on the practical aspects of the experiments rather than the underlying scientific concepts, which led to higher participation but not necessarily better academic learning outcomes. The study also identifies challenges and dilemmas in the teaching process, such as the limited time frame while adhering to the curriculum, which restricts the opportunities for responsive teaching with on-the-fly scaffolding during *guided inquiry*. Overall, the studies indicate that while on-the-fly scaffolding is one strategy teachers can employ to respond to students’ thinking and enhance engagement, its effectiveness depends on the teacher’s expertise and the instructional context. This highlights both the advantages and challenges of applying responsive teaching strategies.

2 | Specific On-the-Fly Scaffolding Strategies

In addition to studies on the general use of on-the-fly scaffolding, there are various studies that have delved deeper into different strategies for on-the-fly scaffolding. To begin with, there are studies that have examined strategies which essentially involve teachers ensuring that students understand the assigned tasks and remain engaged (Doyle 2013). There are several basic aspects of typical teaching practices. Teachers may consistently monitor their students’ work progress and provide on-the-fly scaffolding to promote their active participation in a task when students encounter difficulties (Marzano et al. 2000). According to Ruiz-Primo and Brookhart (2017), basic support measures to promote students’ participation include repetition of tasks, specifying tasks and pointing out missing aspects of processing the task. Regarding the role of teachers’ appreciation of students’ participation, the results seem to be inconsistent. Chiu (2004) found that ninth graders needed less explicit explanation when their task processing was regularly

acknowledged by teachers. Shanab et al. (1981) showed that teachers’ encouraging statements such as “That’s much better than average” supported college students’ engagement in a task. However, Amato-Zech, Hoff, and Doepke (2006) found that encouraging statements had no effect on primary students’ on-task behavior. Meta-analytic studies examining the effects of encouraging statements on motivation have shown that such statements tend to increase intrinsic motivation (Cameron and Pierce 1994; Deci, Koestner, and Ryan 1999). Specifically, Deci, Koestner, and Ryan (1999) found that the exact wording of encouraging statements matters. Statements such as “Excellent, you should continue to work so well” might even be perceived as pressure and thus have a counterproductive effect on intrinsic motivation, especially in primary school students. Moreover, Drews et al. (2020) found that for primary school students, encouraging statements had no effect on learning when the task itself was inherently motivating. In addition to these studies on rather typical teacher practices, developmental studies indicate that students’ concentration can wane over time (Betts et al. 2006). Therefore, teachers should employ cues and highlight strategies to improve their concentration, to draw attention to important aspects of a task (Calleja, Foster, and Hodgen 2023).

In a study with sixth grade students, Munshi et al. (2022) found that focused, individual teacher advice for undergraduate students improved both high and low achievers’ task performance, whereas generalized advice only had an effect on high achievers. Mulder, Lazonder, and de Jong (2014) showed that high school students who received focusing prompts on the CVS developed more conclusive experiments than students without these prompts. In addition, Soncini, Matteucci, and Butera (2021) found that primary school students’ mathematics learning improved when teachers pointed out students’ mistakes while also encouraging them to stay focused on the task. Moreover, Simons and Klein (2007) showed that seventh grade students performed significantly better in a problem-based learning environment when they were given on-the-fly scaffolding in the form of provided help cards that limited the students’ tasks to focus their attention on essential aspects, compared to students who were not given scaffolding.

Finally, studies have shown that spontaneous on-the-fly scaffolding can also encourage students to think further by problematizing, pointing out discrepancies in students’ thoughts and providing cognitive conflicts (Reiser 2004). English and King (2019) found that when teachers encouraged students to reflect on their strategy, sixth graders improved their problem solving. Similarly, Heemsoth and Heinze (2016) found that seventh and eighth grade students’ procedural knowledge improved when they were prompted to reflect on their solution strategy rather than just the solution. Furthermore, teachers may explicitly ask students to compare their solutions with each other, prompting them to actively engage in their knowledge construction (Kleickmann 2012). For instance, Rittle-Johnson and Star (2007) studied the effects of comparing different student solutions among seventh graders regarding linear equations. The experimental condition was explicitly asked to compare different ways of problem solving. The control condition was given similar tasks and solution examples. However, the solutions were presented without the explicit request to be

compared. The results showed that the experimental condition was superior in terms of flexible procedural knowledge. Moreover, a study by Miller-Cotto, Booth, and Newcombe (2022) showed that 11-year-old students who were prompted to solve science tasks using self-explanations, acquired more accurate understanding compared to a control condition. Furthermore, Siverling et al. (2021) showed that scientific reasoning in fifth to eighth graders can be significantly enhanced by teachers asking students to justify the set-up of their experiments. However, Studhalter et al. (2021) found that problematizing may have no effect or even a negative effect on preschool students' knowledge construction if it is not individualized.

By reviewing the studies above, we find that there are numerous scaffolding strategies that responsive teachers can employ in the classroom. In this study, we identify three scaffolding categories that can guide us to design strategies to enhance young children's scientific understanding:

- *Promoting participation*: Summarizing the findings of for example, Chiu (2004), Deci, Koestner, and Ryan (1999), or Shanab et al. (1981) this includes common methods like explaining tasks, keeping students engaged, and encouraging their engagement.
- *Focusing*: Summarizing the findings of for example, Mulder, Lazonder, and de Jong (2014), Munshi et al. (2022), or Soncini, Matteucci, and Butera (2021) this includes strategies to make tasks easier by guiding students' focus on important parts through cues and pointing out strategies.
- *Problematizing*: Summarizing the findings of for example, English and King (2019), Miller-Cotto, Booth, and Newcombe (2022), or Siverling et al. (2021), this includes strategies to deepen students' understanding by prompting them to think more deeply, highlighting inconsistencies in their thoughts, and providing cognitive conflicts.

Yet, in general, research shows that while most teachers agree that being responsive to students' learning needs by providing on-the-fly scaffolding has significant benefits, their initial response when questioned about it is often to express concern about students persisting with wrong ideas or deviating from prescribed standards-based curricula (Maskiewicz 2015; Radoff et al. 2018; Richards et al. 2020). Thus, teachers often fail to adapt spontaneously to the learning situation with on-the-fly scaffolding, but rather stick to their planned instruction, giving all students the same support (Corno 2008; Hoffman and Duffy 2016; Martin et al. 2019; Wittwer and Renkl 2008). Consequently, further research is needed to explore the use of scaffolding strategies and its impact on students' learning. In particular, we must delve into the impact of diverse strategies on students' learning, aiming to gather evidence that lays the groundwork for equipping teachers with effective methodologies.

2.1 | Challenges in Implementing Research-Based Teaching

Sweller, Kirschner, and Clark (2007) emphasize that translating research findings into effective teaching practices poses significant challenges. Particularly in the realm of inquiry-based

learning, research often encounters methodological hurdles, with key variables insufficiently controlled and confounding factors overlooked. Zhang et al. (2022) state that there is a lack of empirical evidence supporting research-based approaches in teaching. They point out that existing studies largely rely on program-based investigations, complicating the assessment of specific teaching methods' effectiveness. Consequently, conducting studies on effective teaching practices presents significant challenges and requires a balance between laboratory research and classroom interventions (Klahr and Li 2005). Research efforts such as those by Chase and Klahr (2017), Leuchter and Naber (2019), or Weber and Leuchter (2022) have attempted to address this issue by implementing an intermediate step that integrates elements of both laboratory and classroom research. This approach aims to preserve as much of the authentic classroom environment as possible while ensuring a high degree of standardization. We tried to achieve this by conducting a study in the students' original classroom, with each class halved, but by having a trained experimenter conducting the lessons. This methodology aims at providing insights into aspects of responsive teaching practices involving on-the-fly scaffolding in the context of planned guidance, within *guided inquiry* compared to *structured inquiry*.

3 | The Present Study

To sum up, findings from studies on sequential order of planned teacher guidance within *structured inquiry* and within *guided inquiry* during scientific inquiry for students' learning are inconsistent: Whereas some studies indicate superior learning for *structured inquiry—guided inquiry* sequences (Ashman, Kalyuga, and Sweller 2020; Hsu, Kalyuga, and Sweller 2015), other studies suggest the reverse order to be superior (e.g. Kalyuga et al. 2001; Kapur 2008) or showed that the order did not matter (Martella, Klahr, and Li 2020; Matlen and Klahr 2013). However, these studies were conducted in the laboratory, providing highly controlled conditions. Yet, realistic classroom settings, in which teaching for scientific inquiry take place, might look different. In realistic classrooms, the teacher can compensate possible disadvantages of planned guidance within either *guided inquiry* or *structured inquiry* with on-the-fly scaffolding. To the best of our knowledge, previous studies on the sequential order of planned guidance within *structured inquiry* and planned guidance within *guided inquiry* have not yet addressed the influence of on-the-fly scaffolding (Berthold, Eysink, and Renkl 2009; Limón 2001).

In the present study, we put into practice scientific inquiry in realistic classrooms with one experimenter in two ways: One way was *Planned Guidance within Structured Inquiry (S)* where students had to follow a step-by-step instruction to set up a given experiment with given material to investigate a given research question. By following the step-by-step instruction, students received implicit conceptual information on the CVS. The other way was *Planned Guidance within Guided Inquiry (G)*, where students had to set up their own experiments with self-chosen material to investigate a given research question. The selection of materials also included unusable materials. By setting up their own experiments, students received minimal conceptual information on the CVS. In *both settings*, we allowed

the experimenter to provide spontaneous on-the-fly scaffolding to students who, in her estimation, required it to complete the task. We compared the sequential order of *Planned Guidance within Structured Inquiry (S)* and *Planned Guidance within Guided Inquiry (G)* in two experimental conditions: *S followed by G* versus *G followed by S*. In particular, we are interested in how the experimenter uses on-the-fly scaffolding, depending on the sequential order. Possibly, if *G follows S*, the experimenter might decide that less on-the-fly scaffolding is needed, because the students already are supposed to have built up the CVS principle before in *S*. By investigating the occurrence of spontaneous on-the-fly-scaffolding, we aim at shedding light on possible teacher strategies in a realistic classroom environment.

We addressed the following research questions (RQ):

RQ1 *How do the different types of on-the-fly scaffolding (Promoting Participation, Focusing and Problematizing) contribute to students' CVS learning when controlled for students' prior domain-specific content knowledge and science self-concept regardless of the experimental condition?*

RQ2 *Does (a) the amount and (b) the type of on-the-fly scaffolding differ for the sequential order S followed by G versus G followed by S?*

RQ3 *Does the order S followed by G versus G followed by S influence students' CVS and domain-specific content knowledge learning gains?*

4 | Method

4.1 | Sample

In total, 164 students from 10 classes participated in the study. Each class was randomly divided in half, resulting in a total of 20 half classes. The participants were drawn from six primary schools located in rural German towns, aged 8–11 years of age, $M = 9.9$, $SD = 0.66$, including 93 female students. This sample reflects typical demographic characteristics in terms of age and sex. The racial background of the students was not documented, following the common practice in German research, where samples typically lack explicit information on this aspect.

Each half class was assigned to one of the experimental conditions (10 half classes for the group *S followed by G*, 10 half classes for the group *G followed by S*). Standard procedures were used for recruiting participants: Written consent was obtained from the parents of each participant. This consent was approved by the university institutional review board and the school board, including ethics approval. Participation in the study was voluntary, and the students were able to revoke their consent at any time during the course. The school board allowed for a total of 4 h of study in each class, resulting in two lessons, each 60 min long, one *Planned Guidance within Structured Inquiry (S)* and one *Planned Guidance within Guided Inquiry (G)*. The remaining time was used for testing. The instruction took place in the respective classrooms of the students, to maintain their familiarity with the learning environment.

4.2 | General Procedure

Students in the participating groups were given two lessons on “thermal insulation” with one task each lesson. Students' domain-specific content knowledge of thermal insulation and their CVS skills were assessed before the first lesson (pretest) started, immediately after the second lesson (posttest), and about 6 weeks after the second lesson (follow-up test) using an identical test. Students' self-concept regarding science was assessed in the pretest (see Table 1).

Each lesson included one task (Task 1: *Why do desert foxes have bigger ears than arctic foxes?*, Task 2: *Why do we wear a hat in winter?*). The tasks were rated by five experts as equally difficult for this age group. One lesson was *S* and the other lesson was *G* (work sheets, Figures 1–4). The order of the tasks and the order of *S* and *G* were counterbalanced (see Table 2). To investigate order effects (i.e. the order in which a guidance type was applied), groups were assigned to one of two experimental conditions: *S followed by G* or *G followed by S*.

The lessons were conducted by one trained experimenter, who is also one of the authors of this paper (a 31-year-old female with 4 years of teaching expertise). In line with the responsive teaching approach, the experimenter was allowed to use on-the-fly scaffolding during the lessons. She was instructed to provide as little support as possible, only providing on-the-fly scaffolding when she deemed it necessary, to maintain the course of the lesson. Thus, she was allowed to give different types and amounts of on-the-fly scaffolding based on her own assessment of the students' needs

TABLE 1 | General procedure.

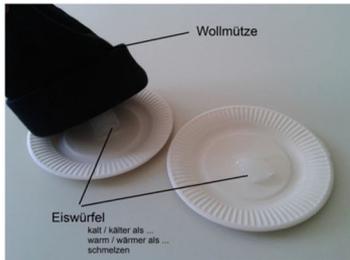
Step 1 Pretest	Step 2 Task	Step 3 Task	Step 4 Posttest	Step 5 Follow-up test
Questionnaires: • Content knowledge • CVS skills • Science self-concept	Lesson 1	Lesson 2	Questionnaires: • Content knowledge • CVS skills	Questionnaires: • Content knowledge • CVS skills

Warum tragen wir im Winter eine Mütze?

1. **Forschfrage:** Warum tragen wir im Winter eine Mütze?

2. **Vermutet**, warum wir im Winter eine Mütze aufsetzen.

3. **Führt folgendes Experiment durch:**



- Nehmt euch 2 Eiswürfel.
- Legt einen Eiswürfel auf einen Teller.
- Legt einen Eiswürfel in die Mütze.
- Legt die Mütze mit dem Eiswürfel auf den anderen Teller
- Setzt den Timer auf 10 Minuten und startet ihn.
- Berührt die Eiswürfel nicht mehr.

4. **Beobachtet**, was passiert!

5. **Dokumentiert** eure Beobachtungen.

6. **Ergebnis:** Schreibt auf, was ihr beobachten konntet.

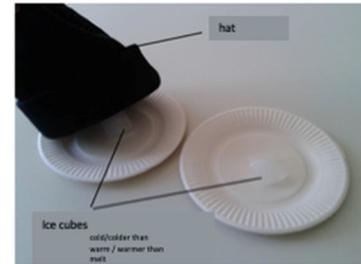
7. Könt ihr eure Beobachtungen erklären? Schreibt auf!

Why do we wear a hat in winter?

1. **Research question:** Why do we wear a hat in winter?

2. **Why do we wear a hat in winter? What is your hypothesis?**

3. **Carry out the following experiment:**



- Take 2 ice cubes.
- Place one ice cube on a plate.
- Place the other ice cube in the hat.
- Place the hat with the ice cube on the other plate.
- Set the timer to 10 minutes and start it.
- Do not touch the ice cubes again.

4. **Observe** what happens!

5. **Document** your observations.

6. **Results:** Write down what you observed.

7. **Can you explain your observations? Write down!**

FIGURE 1 | Work sheet high guidance: Why do we wear a hat in winter? (German and English Version).

Warum tragen wir im Winter eine Mütze?

Forschfrage: Warum tragen wir im Winter eine Mütze?

Was denkt ihr? Schreibt eure **Vermutungen** auf.

Wie könnt ihr die Frage beantworten?

Überlegt euch gemeinsam ein geeignetes Experiment und wählt Materialien dazu aus.

Führt das Experiment durch.

Dokumentiert eure Ergebnisse.

Könt Ihr eure **Ergebnisse** erklären? Schreibt auf!

Why do we wear a hat in winter?

Research question: Why do we wear a hat in winter?

What do you think? Write down your **hypotheses**.

How can you answer the question?

Think about a suitable experiment and select materials for it.

Carry out the experiment.

Document your observations

Can you explain your **results**? Write down!

FIGURE 2 | Work sheet low guidance: Why do we wear a hat in winter? (German and English Version).

Warum haben Wüstenfüchse größere Ohren als Polarfüchse?



1. **Forscherfrage:** Warum haben Wüstenfüchse größere Ohren als Polarfüchse?

2. **Vermute**, warum die Ohren unterschiedlich groß sind. Schreibt eure **Vermutungen** auf.

3. **Führt folgendes Experiment** durch:

- Füllt in den großen und in den kleinen Messbecher je 100 ml 40 °C warmes Wasser.
- Stellt den Timer auf 2 Minuten.
- Messt die Temperatur im Wasser alle zwei Minuten.

Das Thermometer muss die ganze Zeit im Wasser bleiben.
Achtung: Das Thermometer kann ausgehen!

4. **Beobachtet** was passiert!

5. **Dokumentiert** eure Beobachtungen: Tragt die Temperaturen in die Tabelle ein.

6. **Ergebnis:** Schreibt auf, was ihr beobachten konntet.

7. **Könnt ihr eure Beobachtungen erklären?** Schreibt auf!

Why do desert foxes have bigger ears than Arctic foxes?



1. **Research question:** Why do desert foxes have bigger ears than Arctic foxes?

2. Why do the ears have different sizes? **What is your hypothesis?**

3. **Carry out the following experiment:**

- Pour 100 ml of 40 °C water into the large measuring jug and 100 ml into the small measuring jug.
- Set the timer to 2 minutes.
- Measure the temperature in the water every two minutes.

The thermometer must remain in the water at all times.
Caution: The thermometer may turn off!

4. **Observe** what happens!

5. **Document** your observations.

6. **Results:** Write down what you were able to observe. Write the temperatures in the table.

7. **Can you explain your observations?** Write them down!

FIGURE 3 | Work sheet high guidance: Why do desert foxes have bigger ears than arctic foxes? (German and English Version).

spontaneously. However, to create an ecological setting that is as valid as possible the experimenter received theoretical training on the three types of scaffolding beforehand (*Promoting Participation*, *Focusing*, and *Problematization*). The training included the application of these three types as verbal on-the-fly scaffolding (Table 3, in the *scaffolding* section) as well as the use of prepared help-cards (Figure 5 and 6). In addition, she was also trained in the practical implementation of the three types of on-the-fly scaffolding in a 2-h session with two groups of ten children. She was unfamiliar with the students who took part in the training or study. The experimenter was supported by an assistant (male, 25 years old) who handed out teaching materials. The intervention was videotaped and coded.

4.3 | Classroom Procedure

The structure of the two lessons was identical: The lesson started with an introduction to the task, with the entire half class seated in a circle (15 min). Afterwards, the students were asked to state their assumptions, which were collected by the experimenter and shown on the board (5 min). This was followed by the work phase according to the experimental condition (30 min). In each half class, two to four students ($M = 2.96$) were randomly assigned to teams in which they carried out the experiments. At the end of the lesson, the

students presented the results of the work phase, with the entire half class seated in a circle (10 min).

4.4 | Measures

4.4.1 | Content Knowledge Test

The students' domain-specific content knowledge of thermal insulation was assessed by a researcher-developed questionnaire. The questions were aligned with the learning objectives of the lessons. Initially, the test consisted of 18 items, however, one item depicting an oven was excluded because it did not fit into the internal consistency. The final version of the test is shown in Figure 7. The test consisted of a total of 17 items, whereby 16 claims were to be answered with true/false/unknown. In one item, students had to complete a drawing. Items were rated as 0 (false) or 1 (correct). A mean score was calculated for the domain-specific content knowledge variable, ranging from 0 to 1.

4.4.2 | Procedural Knowledge (CVS) Test

The questionnaire assessed students' procedural knowledge of the CVS with 5 items focusing on the development and

Warum haben Wüstenfüchse größere Ohren als Polarfüchse?



Forscherfrage: Warum haben Wüstenfüchse größere Ohren als Polarfüchse?

Was denkt ihr? Schreibt eure **Vermutungen** auf.

Wie könnt ihr die Frage beantworten?

Überlegt euch **gemeinsam** ein geeignetes Experiment und wählt **Materialien** dazu aus.

Führt das Experiment durch und notiert eure **Beobachtungen**.

Dokumentiert eure **Ergebnisse**.

Könnt Ihr eure **Ergebnisse** erklären? Schreibt auf!

Why do desert foxes have bigger ears than Arctic foxes?



Research question: Why do desert foxes have bigger ears than Arctic foxes?

Research question: Why do desert foxes have bigger ears than Arctic foxes?

What do you think? Write down your **hypotheses**.

How can you answer the question?

Think about a suitable experiment and select materials for it.

Carry out the experiment.

Document your observations.

Can you explain your **results**? Write it down!

FIGURE 4 | Work sheet low guidance: Why do desert foxes have bigger ears than arctic foxes? (German and English Version).

TABLE 2 | Counterbalanced manipulation: Sequence of the task, structured inquiry and guided inquiry.

	Experimental condition A		Experimental condition B	
Task	1	2	1	2
Teacher guidance	Structured inquiry		Guided inquiry	
Task	2	1	2	1
Teacher guidance	Guided inquiry		Structured inquiry	

evaluation of experiments from a validated test (see Edelsbrunner et al. 2018). The students had to find out and explain which experimental set up was fair, considering whether the focal variable was manipulated, and other variables were correctly held constant. All items treated domains unrelated to the intervention of the study (force, aviation, and mechanics) (Figure 8).

4.4.3 | Science Self-Concept

The students' academic self-concept in relation to science learning was assessed using 8 items (Figure 9). The items were taken from the questionnaires TIMMS "The Trends in

International Mathematics and Science Study" (Bos 2012), "Science Interest Survey" (Lamb et al. 2012) (Curricular Engagement and Self-Concept) and "The Test of Science Related Attitudes—TOSRA" (Robinson and Fraser 2013) (Extracurricular Engagement). The science self-concept was surveyed using a 4-point Likert scale (0 = not at all, 3 = very much).

The measures were tested and recorded on tablet PCs. There were no time restrictions. No time constraints were imposed, with the typical completion time being 20–30 min.

4.4.4 | Scaffolding

The experimenter's on-the-fly scaffolding was coded by two trained raters, both research assistants, and one reference rater using videos and transcripts of the lessons. The raters' training began with an introduction to the different categories of scaffolding. Categories were classified using a scale based on Hsin and Wu (2011): (1) *Promoting Participation*, (2) *Focusing*, or (3) *Problematizing*. Next, 10 transcripts/videos of the experimenters' actions were coded deductively alongside the reference rater. Afterward, each rater independently coded the scaffolds inductively. For every five independently coded transcripts/videos, two were jointly evaluated to ensure consistency. Deviations in the coding of the experimenter's scaffolds were rare, as the experimenter was trained in advance on the scaffolds to be

TABLE 3 | Overview of on-the-fly scaffolding strategies used in the study.

Scaffolding code	Scaffolding aim	Example
Promoting participation	Clarification of task and procedure Maintaining students' involvement	Teacher monitors work status and <ul style="list-style-type: none"> • Repeats task • Points out missing aspects of the task • Specifies task • Acknowledges progress of work • Gives encouraging statements
Focusing	Focusing students' attention	<ul style="list-style-type: none"> • Gives specific, individual advice • Gives strategy prompts • Points out mistakes • Encourages to use help cards
Problematizing	Stimulating higher order thinking	<ul style="list-style-type: none"> • Prompting reflection • Prompting comparison • Prompting explanations • Prompting reasoning

given. The inter-rater reliability between the two raters and the reference rater was $\kappa > 0.85$. All experimenter statements were categorized and scored as events (0 = not appearing, 1 = appearing).

4.5 | Data Analysis

The statistics program R, version 4.0.3 (R Core Team 2023), was used to analyze the data. First, we computed descriptive statistics of all variables for the whole data set. All variables were normally distributed with skew $< |2|$ and kurtosis $< |7|$ (West, Finch, and Curran 1995).

Research question 1 on the effect of on-the-fly scaffolding to students' CVS learning was addressed by running additional mixed-level models with children on level-2 and time of measurement on level-1 with procedural knowledge as the outcome variable and time of measurement as an indicator for rate of change as the independent variable. In a second step, science self-concept and content knowledge were added to the model as control variables. Last the three types of scaffolds, *Problematizing*, *Focusing*, and *Promoting Participation* were added to the model as predictors.

To address the following research questions, we divided the sample into two groups according to the order they had received guidance in (*S followed by G* or *G followed by S*). We addressed research question 2a on differences in the amount of scaffolds between group order, we compared the number of scaffolds received between the two order (*structured inquiry* followed by *guided inquiry* or *guided inquiry* followed by *structured inquiry*) with *z*-tests of proportions. For research question 2b on differences in the type of scaffolds between group order, two proportions *z*-tests were conducted.

Last, to test research question 3 whether the order *S followed by G* versus *G followed by S* influences changes in students' CVS and domain-specific content knowledge, we used mixed-level

models with children on level-2 and time of measurement on level-1, content or procedural knowledge as the dependent variable, and time of measurement as an indicator for rate of change and order of guidance received as independent variables. Missing data occurred when children missed the test dates that they had agreed upon with their schools (e.g. because of illness). A total of 153 children had complete data sets. Therefore, the number of participants varied between different analyses. Some students dropped out of the study, for example, because of illness. Others had missing values on some of the items. Pairwise deletion was used to include the largest amount of data possible. However, to ensure that missingness would not affect the results, we ran all the analyses only with the children who had a complete data set, that is, used listwise deletion. The results are the same for the pairwise and listwise deletion.

5 | Results

The descriptive statistics for the measures can be found in Table 4.

RQ1 *How do the different types of on-the-fly scaffolding (Promoting Participation, Focusing, and Problematizing) contribute to students' CVS learning when controlled for students' prior domain-specific content knowledge and science self-concept regardless of the experimental condition?*

To answer RQ1, we specified hierarchical mixed-effects models to investigate whether on-the-fly scaffolding affected changes in students' CVS skills, controlled for interindividual covariates of domain-specific content knowledge and science self-concept. We investigated the normal distribution of residuals. The assumption of normal distribution was slightly violated for Model 1, $W = 0.99$, $p = 0.034$. Therefore, a robust mixed-effects model was calculated instead, which yielded similar results as the nonrobust model. Shapiro–Wilk tests showed no significant deviation from normal distribution for the other two models,

 <p>Hilfekarte 1</p> <p><i>Impuls: Überlegt gemeinsam, welche Materialien ihr auswählt.</i></p>	<ul style="list-style-type: none"> • ihr braucht eine Mütze und zwei Eiswürfel • legt einen Eiswürfel in die Mütze 	 <p>Help Card 1</p> <p><i>Hint: Think together about which materials to choose.</i></p>	<ul style="list-style-type: none"> • you need a hat and two ice cubes • put one ice cube in the hat
 <p>Hilfekarte 2</p> <p><i>Impuls: Vergleicht die Eiswürfel</i></p>	<ul style="list-style-type: none"> • schaut in regelmäßigen Abständen nach den Eiswürfeln • nutzt den Timer 	 <p>Help Card 2</p> <p><i>Hint: Compare the ice cubes</i></p>	<ul style="list-style-type: none"> • check the ice cubes at regular intervals • use the timer

FIGURE 5 | Help cards: Why do we wear a hat in winter? (German and English Version).

 <p>Hilfekarte 1</p> <p><i>Impuls: Überlegt gemeinsam, welche Materialien ihr auswählt.</i></p>	<ul style="list-style-type: none"> • ihr sollt „große“ und „kleine Ohren“ miteinander vergleichen • dafür braucht ihr einen großen und einen kleinen Gegenstand 	 <p>Help card 1</p> <p><i>Hint: Think together about which materials to choose.</i></p>	<ul style="list-style-type: none"> • you should compare "big" and "small ears" with each other • for this you need a large and a small object
 <p>Hilfekarte 2</p> <p><i>Impuls: Was fließt in den Ohren?</i></p>	<ul style="list-style-type: none"> • in den Ohren der Füchse fließt warmes Blut • in eurem Experiment könnt ihr stattdessen warmes Wasser verwenden 	 <p>Help card 2</p> <p><i>Hint: What is flowing in ears?</i></p>	<ul style="list-style-type: none"> • warm blood flows in the ears of the foxes • in your experiment you can use warm water instead
 <p>Hilfekarte 3</p> <p><i>Impuls: Vergleicht die Ohren!</i></p>	<ul style="list-style-type: none"> • gebt die <u>gleiche Menge</u> warmes Wasser in die beiden unterschiedlich großen „Ohren“ (Messbecher) • misst die Temperatur des Wassers immer in gleichen Abständen (z.B. alle 2 Minuten) • tragt eure Ergebnisse in eine Tabelle ein 	 <p>Help card 3</p> <p><i>Hint: Compare the ears!</i></p>	<ul style="list-style-type: none"> • pour the <u>same amount</u> of warm water into the two differently sized "ears" (measuring cup) • always measure the temperature of the water at equal intervals (e.g. every 2 minutes) • write down your results in a table

FIGURE 6 | Help cards: Why do desert foxes have bigger ears than arctic foxes? (German and English Version).

Model 2, $W = 0.99$, $p = 0.291$, and Model 3, $W = 0.99$, $p = 0.159$. The results of the three models are presented in Table 5. Model 1 shows that time affects change in CVS skills, thus confirming the results of RQ1. Model 2 integrates domain-specific content knowledge and science self-concept and shows that domain-specific content knowledge is positively related to change in CVS skills and that the effect of time vanishes when integrating the covariates into the model. In a last step, on-the-fly scaffolding was included into Model 3. *Promoting Participation* was

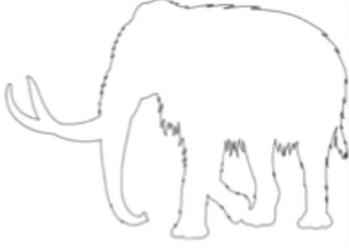
not related to students' change in CVS skills, while *Focusing* was negatively related, and *Problematising* was positively related. A deviance test shows that Model 2 explains the data better than Model 1, $\Delta\chi^2 = 47.76$, $df = 2$, $p < 0.001$ and that Model 3 explains the data better than Model 2, $\Delta\chi^2 = 12.24$, $df = 3$, $p = 0.007$.

RQ2a Does the amount of on-the-fly scaffolding differ for the sequential order S followed by G versus G followed by S?



Anna und Tim sind im Zoo. Sie sehen Wüstenfüchse, die in der warmen Wüste leben, und Polarfüchse, die in kalten Gebieten leben. Dabei stellen sie fest, dass Wüstenfüchse viel größere Ohren haben, als Polarfüchse. Sie überlegen, warum das so ist. Gemeinsam stellen sie Vermutungen an.

Kreuze an, ob die Aussagen von Anna und Tim stimmen oder nicht:

	Richtig	Falsch	Weiß nicht
Der Polarfuchs friert nicht, weil er den Schnee liebt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Füchse mit längeren Körperteilen frieren schneller als Füchse mit kürzeren Körperteilen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Der Polarfuchs hat kleine Ohren, damit er im Schnee gut getarnt ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weil der Polarfuchs kleine Ohren hat, bleibt ihm warm.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je weiter man sich vom Äquator den Polen nähert, desto größer werden die Ohren vieler Tiere.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Der Wüstenfuchs braucht seine großen Ohren nur um gut hören zu können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zeichne die Ohren ein:			
Der Elefant lebt im warmen Afrika		Das Mammut lebte im eiskalten Sibirien	
			

Lisa und Max wollen mit ihrem Hund Bello draußen spielen. Es ist Winter und sehr kalt. Sie setzen eine Mütze auf und ziehen Handschuhe an. Sie überlegen warum sie durch die Mütze und die Handschuhe weniger frieren. Gemeinsam stellen sie Vermutungen an.

Kreuze an, ob die Aussagen von Lisa und Max stimmen oder nicht:

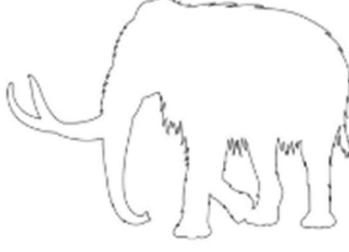
	Richtig	Falsch	Weiß nicht
Der Pullover erzeugt Wärme.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dem Hund ist warm, weil sein Fell Wärme erzeugt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ein Eiswürfel mit Mütze, schmilzt schneller als ein Eiswürfel ohne Mütze.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kreise die Dinge ein, die Wärme erzeugen können:			
			

FIGURE 7 | Questionnaire: Conceptual knowledge (German and English Version).



Anna and Tim are at the zoo. They see desert foxes, which live in the warm desert, and Arctic foxes, which live in cold regions. They realise that desert foxes have much bigger ears than Arctic foxes. They ponder why this is the case. Together they make assumptions.

Tick whether Anna and Tim's statements are true or not:

	true	false	I don't know
The Arctic fox doesn't freeze because it loves the snow.			
Foxes with longer body parts freeze faster than foxes with shorter body parts.			
The Arctic fox has small ears so that it is well hidden in the snow.			
Because the Arctic fox has small ears, it stays warm.			
The closer you get to the poles from the equator, the larger the ears of many animals become.			
The desert fox only needs its large ears to be able to hear well.			
Draw the ears:			
The elephant lives in warm Africa		The mammoth lived in icy Siberia	
			

Lisa and Max want to play outside with their dog Bello. It's winter and very cold. They put on a hat and gloves. They think about why the hat and gloves make them less cold. Together they make assumptions.

Tick whether Lisa and Max's statements are true or not:

	true	false	I don't know
The pullover produces heat.			
The dog is warm because its fur produces heat.			
An ice cube with a hat melts faster than an ice cube without a hat.			
Circle the things that can produce heat:			
			

FIGURE 7 | (Continued)

Welches Flugzeug verbraucht am wenigsten Treibstoff?

Herr Müller baut Flugzeuge und möchte, dass sie möglichst wenig Treibstoff verbrauchen. Er hat verschiedene Ideen, wovon der Treibstoffverbrauch abhängen könnte:

Er überlegt sich, dass ein Flugzeug eine spitze oder eine runde Nase haben kann.		
Er überlegt sich, dass die Höhenruder unten oder oben angebracht werden können.		
Er überlegt sich, dass ein Flugzeug doppelte oder einfache Flügel haben kann.		

Herr Müller vermutet, dass ein Flugzeug mit einer spitzen Nase weniger Treibstoff verbraucht als ein Flugzeug mit einer runden Nase.

Was soll Herr Müller tun, um herauszufinden, ob die Form der Flugzeugnase für den Treibstoffverbrauch wichtig ist?

Kreuze die richtige Antwort an:

- Herr Müller muss ein paar Flugzeuge bauen und vergleichen, wie viel Treibstoff sie verbrauchen.
- Herr Müller muss zwei Flugzeuge bauen, eines mit runder Nase und eines mit spitzer Nase. Sie müssen aber sonst ganz gleich sein. Dann muss er vergleichen, wie viel Treibstoff sie verbrauchen.
- Herr Müller muss zwei ganz unterschiedliche Flugzeuge bauen, bei denen er die Nase, die Flügel und die Höhenruder unterschiedlich macht. Dann muss er vergleichen, wie viel Treibstoff sie verbrauchen.

FIGURE 8 | Questionnaire: CVS; selected items (German and English Version).

In total, we found 488 incidences of on-the-fly scaffolding given by the experimenter. The sequential order *S* followed by *G* comprised 212 incidents, whereas *G* followed by *S* comprised 276 incidents.

To get a more detailed picture on how the scaffolding incidences were distributed among the students, we grouped the students according to the number of scaffolding incidents they were given by the experimenter. Forty-four students (29%) were given three or more on-the-fly scaffolds, referred to in the following as *High Scaffolding*, 53 students (34%) were given two on-the-fly scaffolds, referred to in the following as *Medium Scaffolding*, and 57 students (37% of the sample) were given zero to one on-the-fly scaffolds, referred to in the following as *Low Scaffolding*.

In the *structured inquiry (S)* condition, we found 35 incidences of on-the-fly scaffolding. The sequential order *S* followed by *G* comprised 12 incidents, whereas *G* followed by *S* comprised 23 incidents.

Again, we grouped the students according to the number of scaffolding incidents they were given by the experimenter. No student was given three or more on-the-fly scaffolds, 27 students (17%) were given two on-the-fly scaffolds, and 130 students (83% of the sample) were given zero to one on-the-fly scaffolds. We compared the number of students classified as *High*, *Medium*, or *Low Scaffolding* for the two sequential orders. The results are presented in Tables 6 and 7.

To compare the probability of on-the-fly scaffolding for the two experimental conditions with *guided inquiry (G)*, *z*-tests of

Welche Kugel rollt am weitesten?

Robert baut verschiedene Rampen, von denen er Kugeln herunterrollen lassen kann. Robert baut ganz verschiedene Rampen. Sie unterscheiden sich in der Steigung, der Oberfläche und der Länge. Er hat auch verschiedene Kugeln – nämlich eine leichte und eine schwere Kugel.

Steigung: Er kann die Rampe steil oder flach machen.	
Oberfläche: Er kann die Oberfläche der Rampe glatt oder rau machen.	
Länge: Er kann die Rampe kurz oder lang machen.	
Gewicht der Kugel: Er kann eine leichte oder schwere Kugel herunterrollen lassen.	

Robert vermutet, dass eine Kugel weiter rollt, wenn die Rampe steil ist.

Um seine Vermutung zu überprüfen, baut er die beiden unten abgebildeten Rampen, die sich in der Steigung, der Oberfläche und der Länge voneinander unterscheiden. Dann lässt er auf der einen Rampe die leichte und auf der anderen Rampe die schwere Kugel herunterrollen. Anschließend vergleicht er, wie weit die beiden Kugeln bei den verschiedenen Rampen gerollt sind.

Rampe 1: 	Rampe 2: 
---	---

Ist das ein gutes Experiment, um herauszufinden, ob eine Kugel bei einer steilen Rampe weiter rollt als bei einer flachen?

Kreuze die richtige Antwort an:

Ja, das ist ein gutes Experiment.
Warum ist das ein gutes Experiment? Begründe deine Antwort.

Nein, das ist kein gutes Experiment.
Warum ist das kein gutes Experiment? Begründe deine Antwort.

Which airplane needs the least fuel?

Mr. Miller builds airplanes. His airplanes should use as little fuel as possible. He has a couple of ideas about what might be important for how much fuel an airplane needs.

The form of the nose can be pointed or rounded.



The rudder can be on the bottom or on the top.



The wings can be double or single.



Mr. Miller believes that airplanes with a pointed nose need less fuel than airplanes with a rounded nose. What should Mr. Miller do to find out whether the form of the nose is important for how much fuel the airplane needs?

Select the correct answer:

- Mr. Miller should build a couple of planes. Then he can compare how much fuel each one needs.
- Mr. Miller should build two planes: one with a pointed and one with a round nose, but with the same rudder and wings. Then, he can compare how much fuel each one needs.
- Mr. Miller should build two very different planes. They should have different noses, rudders, and wings. Then, he can compare how much fuel each one needs.

Which ball rolls the farthest?

Robert builds several ramps to let balls roll down. They differ in steepness, surface, and length. He also has two different balls, a light one and a heavy one.

Steepness:
The ramp can be steep or flat.



Surface:
The ramp can be smooth or rough.



Length:
The ramp can be short or long.



Ball:
The ball can be light or heavy.



FIGURE 8 | (Continued)

Robert believes that a ball rolls farther when the ramp is steep. To test his belief, he builds two ramps. The ramps differ in steepness, surface, and length.

Ramp 1:



Ramp 2:



On one ramp, he rolls the light ball. On the other ramp, he rolls the heavy ball. He measures how far the balls roll.

Is this a good experiment to find out whether a ball rolls farther on a steep ramp than on a flat ramp?

Select the correct answer:

Yes, this is a good experiment.
Why is this a good experiment? Please justify:

No, this is not a good experiment.
Why is this not a good experiment? Please justify:

FIGURE 8 | (Continued)

proportions were calculated. They revealed that when *S* was followed by *G*, the experimenter was less likely to give a high amount of on-the-fly scaffolding, $z = 9.53$, $df = 1$, $p = 0.002$, and more likely to give a low amount of on-the-fly scaffolding, $z = 5.19$, $df = 1$, $p = 0.023$ than when *G* was followed by *S*. The probability of giving a medium amount of on-the-fly scaffolding did not differ for the experimental conditions, $z = 0.39$, $df = 1$, $p = 0.535$.

Similarly, to compare the probability of on-the-fly scaffolding for the two experimental conditions with *structured inquiry* (*S*), *z*-tests of proportions were calculated. In both groups, the experimenter was equally likely to give a low amount of on-the-fly

scaffolding, $z(1) = 0.12$, $p = 0.727$, and equally likely to provide a medium amount of on-the-fly scaffolding, $z(1) = 0.12$, $p = 0.727$. As there was no high amount of on-the-fly scaffolding, the probability of it was not calculated.

RQ2b Does the type of on-the-fly scaffolding differ for the sequential order *S* followed by *G* versus *G* followed by *S*?

Concerning the type of scaffolding in the experimental condition with *guided inquiry* (*G*), two proportions *z*-tests were conducted. Results showed no differences for *Promoting Participation*, $z(1) = 0.30$, $p = 0.582$ between the experimental conditions. However, when *S* was followed by *G*, the

experimenter used more *Problematising* than when starting with *G*, $z(1) = 8.89$, $p = 0.003$. In contrast, when *G* was followed by *S*, the experimenter used more *Focusing* than when starting with *S*, $z(1) = 11.70$, $p < 0.001$ (Table 8).

Similarly, in the experimental conditions with *structured inquiry* (*S*), results showed no order effects for *Focusing*, $z(1) = 3.19$, $p = 0.074$, and *Problematising*, $z(1) = 0.33$, $p = 0.568$. However, when *S* was followed by *G*, the experimenter used more *Promoting Participation* than when starting with *G*, $z(1) = 4.67$, $p = 0.031$ (Table 9).

RQ3 Does the order *S* followed by *G* versus *G* followed by *S* influence students' CVS and domain-specific content knowledge learning gains?

We examined the change in students' CVS skills and domain-specific content knowledge depending on the sequence of *S* and *G*, that is, whether the students engaged in *S* or *G* first, from pretest to follow-up test.

The descriptive results show that CVS mean scores and domain-specific content knowledge mean scores increased from pretest to follow-up test in both experimental conditions (Table 10). We then analyzed this increase using inferential statistics. We investigated the intraclass correlation and found that differences between students explained 68% of the variance in changes in CVS skills and 37% of the variance in changes in domain-specific content knowledge. This indicates that the points of measurement were nested in students and therefore a two-level structure was necessary.

Thus, we specified two mixed effects model with students on level-2 and point of measurement on level-1. To check the assumptions of normal distribution of the residuals, we ran a

Shapiro–Wilk normality test on the models. For the model with CVS skills as the outcome, the Shapiro–Wilk test showed a deviation from the normal distribution, $W = 0.99$, $p = 0.008$. Therefore, a robust mixed effect model was run instead, which showed similar results to the nonrobust model. For the model with content knowledge as the outcome, the Shapiro–Wilk test did not show a deviation from normal distribution, $W = 0.99$, $p = 0.053$. The results are presented in Table 11. Results suggest that the increase in CVS and domain-specific content knowledge did not differ for the experimental conditions, indicating that the sequence of *S* and *G* did not affect the learning of CVS nor of domain-specific content knowledge.

6 | Discussion

In the present study, we manipulated the level of planned teacher guidance (*structured inquiry* vs. *guided inquiry*) during primary school students' scientific inquiry on the topic of thermal insulation. In both settings, the students conducted given experiments without explicit teaching of the CVS principle. In the setting with *Planned Guidance within Structured Inquiry*, the students were given low autonomy (predefined experiments) including a written step-by-step instruction on how to conduct the experiments and implicit conceptual information on the CVS. In the setting with *Planned Guidance within Guided Inquiry*, the students were given high autonomy (developing their own experiments) and minimal conceptual information on the CVS. Moreover, in the setting with *Planned Guidance within Guided Inquiry*, the experimenter provided implicit conceptual information on the CVS principle via spontaneous on-the-fly scaffolding. Each student went through both settings, but in a different order: One half of the students started with conducting given experiments (*Planned Guidance*

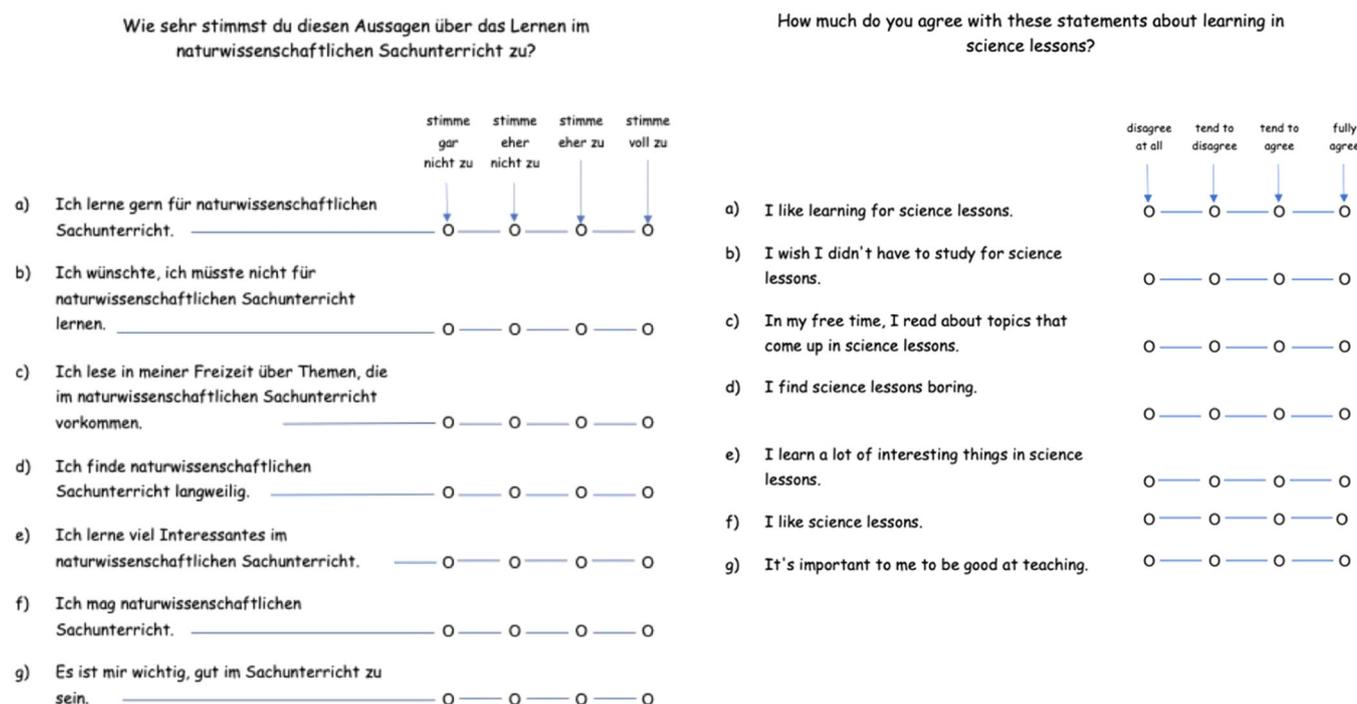


FIGURE 9 | Questionnaire: Science self-concept (German and English Version).

Wie sehr stimmst du diesen Aussagen über naturwissenschaftlichen Sachunterricht zu?

	stimme gar nicht zu	stimme eher nicht zu	stimme eher zu	stimme voll zu
a) Normalerweise bin ich gut im naturwissenschaftlichen Sachunterricht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Naturwissenschaftlicher Sachunterricht fällt mir schwerer als vielen meiner Mitschüler.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) Ich bin einfach nicht gut im naturwissenschaftlichen Sachunterricht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) Ich lerne schnell im naturwissenschaftlichen Sachunterricht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) Meine Lehrerin sagt, ich bin gut in naturwissenschaftlichen Sachunterricht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) Mir fällt naturwissenschaftlicher Sachunterricht schwerer als alle anderen Fächer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g) Ich bin mit meinen Leistungen im naturwissenschaftlichen Sachunterricht zufrieden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h) Für meine Zukunft ist es notwendig gut über Themen des naturwissenschaftlichen Sachunterrichts Bescheid zu wissen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How much do you agree with these statements about learning in science lessons?

	disagree at all	tend to disagree	tend to agree	fully agree
a) I'm normally good at science.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Science lessons are more difficult for me than for many of my classmates.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) I'm just not good at science.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) I learn quickly in science lessons.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) My teacher says I'm good at science.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) I find science lessons more difficult than any other subject.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g) I am satisfied with my performance in science lessons.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h) For my future, I need to have a good understanding of scientific subjects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wie sehr stimmst du diesen Aussagen über Naturwissenschaften zu?

	stimme gar nicht zu	stimme eher nicht zu	stimme eher zu	stimme voll zu
a) Ich wäre gerne in einer Naturwissenschaften - AG.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Mir wird langweilig, wenn ich zu Hause naturwissenschaftliche Sendungen im Fernsehen sehe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) Ich finde es blöd naturwissenschaftliche Bücher oder Zeitschriften geschenkt zu bekommen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) Ich mag es, während meiner Ferien Bücher über Naturwissenschaften zu lesen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) Ich würde gerne zu Hause naturwissenschaftliche Experimente durchführen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) Ich finde es langweilig, nach der Schule mit meinen Freunden über Naturwissenschaften zu sprechen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g) Ich würde mich freuen später in einem naturwissenschaftlichen Labor zu arbeiten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h) Ich finde es langweilig eine Sendung über Naturwissenschaften im Radio anzuhören.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i) Ich würde gerne am Wochenende ein naturwissenschaftliches Museum besuchen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j) Ich finde es langweilig, Zeitungsartikel über Naturwissenschaften zu lesen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k) Ich finde es blöd, naturwissenschaftliches Spielzeug geschenkt zu bekommen (zum Beispiel einen Experimentierkoffer oder ein Mikroskop).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How much do you agree with these statements about learning in science lessons?

	disagree at all	tend to disagree	tend to agree	fully agree
a) I would like to be in a science club.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) I get bored when I watch science programs on TV at home.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) I think it's stupid to receive science books or magazines as gifts.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) I like to read books about science during my vacations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) I would like to carry out scientific experiments at home.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) I find it boring to talk about science with my friends after school.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g) I would love to work in a science laboratory later on.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h) I find it boring to listen to a program about science on the radio.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i) I would like to visit a science museum at the weekend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j) I find it boring to read newspaper articles about science.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k) I think it's stupid to get science toys as presents (for example an experiment kit or a microscope).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE 9 | (Continued)

TABLE 4 | Descriptive statistics.

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	α
CK pretest	130	0.31	0.17	0–1	0.71
CK posttest	152	0.57	0.23	0–1	0.78
CK follow-up	130	0.54	0.23	0–1	0.79
PK pretest	149	0.29	0.26	0–1	0.49
PK posttest	135	0.35	0.30	0–1	0.74
PK follow-up	146	0.37	0.31	0–1	0.75
Science SC	139	2.18	0.53	0–3	0.83
<i>Promoting participation</i>	164	1.84	1.31	0–5	
<i>Focusing</i>	164	4.03	2.07	0–9	
<i>Problematizing</i>	164	3.45	1.87	0–5	

Abbreviations: CK, content knowledge; PK, procedural knowledge (CVS); science SC, science self-concept.

within Structured Inquiry) followed by a lesson in which they had to design their own experiments without any material scaffolding (*Planned Guidance within Guided Inquiry*). The other half of the students worked in reverse order. The lessons' topic was counterbalanced.

Our primary research aim was to investigate the effects of different types of guidance, with on-the-fly scaffolding, on learning during scientific inquiry. We were particularly interested in the role of responsive teaching with on-the-fly scaffolding during *structured inquiry* (following Blanchard's level 1 of scientific inquiry), and *guided inquiry* as (following Blanchard's level 2 of scientific inquiry), when these are sequenced in different order. We examined (1) what type of on-the-fly scaffolding contributed to students' CVS learning after controlling for their prior domain-specific content knowledge and science self-concept, (2) whether the amount and type of on-the-fly scaffolding during the lessons differed based on the order used, and (3) whether students' CVS and domain-specific content knowledge learning differed based on the respective order used.

We found that (1) *Problematizing* on-the-fly scaffolding contributed to students' CVS learning, (2) the experimenter provided more *Focusing* on-the-fly scaffolding when *Planned Guidance within Guided Inquiry* was first, while providing more *Problematizing* on-the-fly scaffolding when *Planned Guidance within Guided Inquiry* was the second lesson, as well as the experimenter provided more *Promoting Participation* on-the-fly scaffolding when *Planned Guidance within Structured Inquiry* was first, and (3) students' CVS and domain-specific content knowledge learning did not differ, independently of the respective order used. We will discuss these findings in the following sections.

The experimenter was allowed to promote students' participation, focus students' attention, and problematize students' thinking whenever she felt it was appropriate. The different types of on-the-fly scaffolding had different effects on students' CVS learning and were also given differently by the experimenter depending on the sequential order of the planned guidance. Experimenter's *Problematizing* had a positive effect

on students' CVS learning which is in line with other studies on the benefits of engaging students in active and reflective learning processes (Aleven and Koedinger 2002; English and King 2019; Heemsoth and Heinze 2016; Miller-Cotto, Booth, and Newcombe 2022; Rittle-Johnson and Star 2009). We found that *Problematizing* was more likely to be given when *Planned Guidance within Guided Inquiry* was second. In this sequential order, students had already been introduced to experimentation in the preceding *Planned Guidance within Structured Inquiry* lesson. Thus, students who received more *Problematizing* received it in addition to the knowledge they had already built up beforehand. Moreover, this sequential order, with the *Planned Guidance within Structured Inquiry* lesson preceding *Planned Guidance within Guided Inquiry*, may have led the experimenter to believe that the students' skills were sufficient for more challenging scaffolding that would provide additional cognitive stimulation. The experimenter might have thought that these students would be able to cope with the challenges of provoked cognitive conflicts.

In contrast to other studies (Chiu 2004; Gettinger and Kohler 2013; Mirkovic et al. 2020; Shanab et al. 1981), we found no effects in relation to the experimenter's promotion of participation on students' CVS learning. This result is consistent with the findings of Amato-Zech, Hoff, and Doepke (2006), who also found no effect of teacher's encouraging statements on students' participation. However, on the one hand, during *Planned Guidance within Guided Inquiry*, the amount of *Promoting Participation* did not differ between the sequential order, resulting in a lack of variance. The experimenter might have taken for granted that all students, regardless of the sequential order, could profit equally from clarifying goals of a task, supporting procedures, and keeping students involved. On the other hand, during *Planned Guidance within Structured Inquiry*, we found that the experimenter gave more *Promoting Participation* for the sequential order *S followed by G*. The experimenter might have felt that providing more *Promoting Participation* scaffolds during the first lesson was necessary to help students adjust to the new learning environment and to ensure they understood the task structure.

Focusing, in our study giving individual advice, giving prompts on strategies, pointing out mistakes and encouraging to use help cards had a negative effect on students' CVS learning, which is in contrast to many studies (Kapur 2016; Mulder, Lazonder, and de Jong 2014; Munshi et al. 2022; Simons and Klein 2007; Soncini et al. 2022). However, the experimenter gave more *Focusing* when the *Planned Guidance within Guided Inquiry* lesson took place first. Thus, *Focusing* was given to students who did not have the chance to build up knowledge beforehand. This may have led the experimenter to consider that the students' skills were insufficient for independent scientific inquiry learning and that *Focusing* might reduce students' cognitive load (Sweller 1988), directing the students' attention to essential aspects of the task. However, according to Reiser (2004), too much structure may prevent students from engaging actively in a task. Thus, the students in our study may have found the *Focusing* scaffolds rigid, and therefore could not fully engage in the learning process. In addition, the *Focusing* scaffolds may have been too prescriptive or direct, not allowing for students to engage in their own ways of problem solving and

TABLE 5 | Mixed-effects models to investigate whether on-the-fly scaffolding affected students' CVS skills, controlling for other interindividual covariates.

Fixed effects	Model 1			Model 2				Model 3			
	γ	SE	<i>t</i>	γ	β	SE	<i>t</i>	γ	β	SE	<i>t</i>
Intercept	0.31***	0.02	15.27	0.11	—	0.09	1.27	0.12	—	0.11	1.06
Time	0.04**	0.01	4.36	−0.02	—	0.01	−1.20	−0.02	—	0.01	−1.19
Content knowledge				0.43***	0.36	0.06	7.00	0.43***	0.36	0.06	6.90
Science SC				0.01	0.02	0.04	0.35	0.00	0.02	0.04	0.94
<i>Promoting participation</i>								0.01	0.01	0.02	0.71
<i>Focusing</i>								−0.02*	−0.01	0.01	−2.18
<i>Problematizing</i>								0.03*	0.01	0.01	2.53
Random effects	Var		SD	Var		SD		Var		SD	
Person	0.06		0.24	0.04		0.21		0.04		0.20	
Level-1-residuum	0.02		0.14	0.02		0.15		0.02		0.15	
Marginal R^2 /conditional R^2	0.01/0.74			0.14/0.69				0.20/0.70			

Note: Marginal R^2 and conditional R^2 are used as effect size measures (Rights and Sterba 2021).

Abbreviations: Science SC, science self concept; SE, standard error; time, rate of change.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

TABLE 6 | Distribution of on-the-fly scaffolding during guided inquiry depending on the sequence of S and G.

Groupings	<i>S</i> followed by <i>G</i>	<i>G</i> followed by <i>S</i>
High scaffolding	13 (28%) students	31 (41%) students
Medium scaffolding	29 (33%) students	24 (32%) students
Low scaffolding	36 (39%) students	21 (28%) students

Note: In the column *S* followed by *G*, the percentages add up to 101% due to rounding principles.

Abbreviations: G, planned guidance within guided inquiry with on-the-fly scaffolding; S, planned guidance within structured inquiry.

TABLE 7 | Distribution of on-the-fly scaffolding during structured inquiry depending on the sequence of S and G.

Groupings	<i>S</i> followed by <i>G</i>	<i>G</i> followed by <i>S</i>
High scaffolding	0 (0%) students	0 (0%) students
Medium scaffolding	15 (19%) students	12 (16%) students
Low scaffolding	65 (81%) students	65 (84%) students

Abbreviations: G, planned guidance within guided inquiry with on-the-fly scaffolding; S, planned guidance within structured inquiry.

TABLE 8 | Amount and type of on-the-fly scaffolding incidences during guided inquiry for the sequential orders S followed by G versus G followed by S.

	Total <i>n</i>	<i>Promoting participation</i> (%)	<i>Focusing</i> (%)	<i>Problematizing</i> (%)
S followed by G	276	(21%)	(37%)	(42%)
G followed by S	212	(19%)	(47%)	(34%)

Abbreviations: G, planned guidance within guided inquiry with on-the-fly scaffolding; S, planned guidance within structured inquiry.

thus, hindering intrinsic motivation (Ryan and Deci 2000). Moreover, the timing of *Focusing* may have been inappropriate, resulting in the experimenter pointing out the students' mistakes too early, potentially causing the students to not deal productively with their own mistakes in alignment with an error-friendly classroom (Soncini et al. 2022). Instead, they may have been discouraged from continuing to work on the task. Otherwise, *Focusing* may have been given too late, so that the

students did not have enough time to sufficiently reflect on their task completion.

We also analyzed the students' test scores and found improvements for both the CVS and domain-specific content knowledge from pretest to follow-up-test. This is in line with the findings from Schalk et al. (2019) who showed that 8- to 11-year-old primary school students improved their CVS skills and domain-

TABLE 9 | Amount and type of on-the-fly scaffolding incidences during structured inquiry for the sequential orders S followed by G versus G followed by S.

	Total <i>n</i>	Promoting participation (%)	Focusing (%)	Problematising (%)
<i>S</i> followed by <i>G</i>	12	68	15	17
<i>G</i> followed by <i>S</i>	23	46	31	23

TABLE 10 | Means and standard deviations of the change in students' CVS skills depending on the sequence of S and G.

	<i>S</i> followed by <i>G</i>		<i>G</i> followed by <i>S</i>	
	<i>n</i>	<i>M</i> (SD)	<i>n</i>	<i>M</i> (SD)
T1 CK	72	0.30 (0.17)	73	0.28 (0.14)
T2 CK	72	0.58 (0.23)	73	0.56 (0.24)
T3 CK	63	0.55 (0.23)	71	0.57 (0.23)
T1 PK	72	0.31 (0.27)	73	0.28 (0.25)
T2 PK	63	0.36 (0.32)	64	0.34 (0.28)
T3 PK	64	0.36 (0.33)	71	0.36 (0.30)

Note: The differences in *n* compared to the calculations in the "Measures" section are due to missing data.

Abbreviations: T1, test 1; T2, test 2; T3, test 3; CK, content knowledge; G, planned guidance within guided inquiry; PK, procedural knowledge (CVS); S, planned guidance within structured inquiry.

specific content knowledge by conducting given experiments over the course of a school year without getting explicit teaching of the CVS principles. Even more so, our findings suggest that an improvement in CVS skills without explicit teaching on the CVS principle can occur over the course of only two lessons. However, this contrasts with the findings of Studhalter et al. (2021), who found that in realistic classroom settings, most 6- to 7-year-old students failed in acquiring the CVS principle by conducting given experiments without explicit CVS teaching for 1 h. Yet, students in the study by Studhalter et al. (2021) were younger than the students in the study of Schalk et al. (2019) and in the present study (8- to 11-year-olds), indicating that learning the CVS without explicit teaching might work for older but not for younger children.

In our study, learning gains in CVS and domain-specific content knowledge did not differ for the order of *Planned Guidance within Guided Inquiry* and *Planned Guidance within Structured Inquiry*. This is in line with results from Martella, Klahr and Li (2020) and Matlen and Klahr (2013), but contrary to previous studies that found order effects. Some of these studies concluded that a sequence of *structured inquiry—guided inquiry* guidance was superior for learning than a sequence of *guided inquiry—structured inquiry* guidance (Ashman, Kalyuga, and Sweller 2020; Hsu, Kalyuga, and Sweller 2015). Other studies found that a sequence of *guided inquiry—structured inquiry* guidance was superior to learning than a sequence of *structured inquiry—guided inquiry* guidance (Kapur 2008; Schwartz et al. 2011). The results of our study, conducted in a controlled but realistic classroom setting, expand these findings by suggesting that both orders of *Planned Guidance within Guided Inquiry* and *Planned Guidance within Structured Inquiry* can be applied in case of 8- to 11-year-old students' learning the CVS and domain-specific content knowledge.

However, this outcome could be attributed to the specific design of our study, as we defined *Planned Guidance within Structured Inquiry* as low autonomy with implicit conceptual information on the CVS through written step-by-step instructions on how to conduct the given experiments. *Planned Guidance within Guided Inquiry* was defined as allowing for high autonomy with minimal conceptual information. In our study, the experimenter was able to provide on-the-fly scaffolding in accordance with a responsive teaching approach. Consequently, the fact that we did not find any order effects could be attributed to the higher amount of spontaneous on-the-fly scaffolding in the *Planned Guidance within Guided Inquiry* condition. In our specific context, when we analyze an experimenter's behavior, it is conceivable that professional self-conception leads the experimenter to take responsibility for her students' learning. Consequently, the experimenter may use on-the-fly scaffolding to mitigate the potential disadvantages of the initial instructional environment.

7 | Limitations

First, our study design was conducted in a realistic classroom setting. Hence, the students' behavior was authentic, and the scaffolds given by the experimenter were spontaneous. Our results can therefore be interpreted in light of an ecologically valid setting. Nevertheless, they may not be as reliable, and objective compared to a highly standardized laboratory study. In addition, there could be the potential for experimenter bias, despite receiving theoretical training on different types of scaffolding. The experimenter's personal teaching style and biases could still influence the type and amount of on-the-fly scaffolding provided during the lessons. Furthermore, the fact that the experimenter was one of the authors of the paper might introduce a conflict of interest, as she may have vested interests in the success of the study.

Second, the study was conducted in two lessons 60-min each with two different topics, which is in line with other study designs (Bihler et al. 2018; Guerrero-Rosada et al. 2021; Slot et al. 2018). However, instructional quality often varies across lessons and topics (Hill, Charalambous, and Kraft 2012; Praetorius et al. 2014). Thus, increasing the number of lessons in a broader range of science topics, would have allowed us to better control for these influences, thereby increasing the validity of our results.

Third, only one experimenter gave the lessons. Hence, the decision of which scaffolds to give was based on one subjective assessment of one person. Therefore, our results may be interpreted with caution due to limited objectivity. Thus, future studies should involve more experimenters or teachers (Eid, Gollwitzer, and Schmitt 2015). Then, experimenters or teachers should be able to give insight into their professional self-conception to be able to explain variance in their scaffolding (Buehl and Beck 2015).

TABLE 11 | Change in CVS skills and domain-specific content knowledge depending on the sequence of S and G.

Fixed effects	CVS skills			Content knowledge		
	γ	SE	t	γ	SE	t
Intercept <i>S</i> first	0.33***	0.03	10.03	0.50***	0.02	22.75
Intercept <i>G</i> first	0.30***	0.03	10.73	0.47***	0.02	21.57
Δ Intercept <i>S</i> first— <i>G</i> first	0.03	0.04	0.62	0.03	0.03	1.01
Time* <i>S</i> first	0.03**	0.01	2.90	0.11***	0.01	7.68
Time* <i>G</i> first	0.04**	0.01	2.33	0.13***	0.01	8.76
Δ Time* <i>S</i> first—Time* <i>G</i> first	0.01	0.02	0.31	0.02	0.02	0.41
Random effects	Var	SD		Var	SD	
Person	0.06	0.24		0.03	0.16	
Level-1-residuum	0.02	0.14		0.02	0.16	
Marginal R^2 /conditional R^2	0.01/0.75			0.16/0.59		

Note: Marginal R^2 considers only the variance explained by the fixed effects, conditional R^2 considers the variance explained by the total model according to the approach by Nakagawa and Schielzeth (2013); marginal R^2 and conditional R^2 are used as effect size measures (Rights and Sterba 2021).

Abbreviations: G, planned guidance within guided inquiry with on-the-fly scaffolding; SE, standard error; S, planned guidance within structured inquiry, time, rate of change.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Fourth, this study did not consider control variables, such as students' science curiosity or the socioeconomic status of parents. Yet, research has shown that these variables significantly influence the learning process of children (Van Schijndel, Jansen, and Raijmakers 2018; Freitag and Schulz 2018). Thus, in future studies such variables should be controlled to further explain learning gains in inquiry-based science education.

Fifth, it would have been desirable to conduct an interview with the experimenter immediately after the lessons to better understand her perspective on the decision-making processes involved in providing scaffolds. According to the study design, the experimenter appeared to adhere to offering only the scaffolds she deemed necessary. An interview would have provided direct insights into her considerations, pedagogical intentions, and the specific reasons behind her decisions. This would have helped refine the interpretation of the results and enhance understanding of the potential implications of her actions.

Finally, regarding the effects of the scaffolding given, we showed that *Problematizing* had a positive effect on learning. However, the experimenter mainly gave these scaffolds to students who had the chance to build up knowledge in the prior lesson, having received *Planned Guidance within Structured Inquiry* first. It is possible that the experimenter unintentionally disadvantaged students who received *Planned Guidance within Guided Inquiry* first, misjudging the needs of these students to also receive *Problematizing*. Therefore, in a future study, it would be valuable to train experimenters or teachers on how to accurately assess students' needs and provide adaptive scaffolding regardless of the sequential order of lessons.

8 | Conclusions

Our investigation into the impact of *Planned Guidance within Structured Inquiry* and *Planned Guidance within Guided Inquiry* sequencing on students' CVS learning has revealed noteworthy

insights. While the order of guidance did not yield significant effects on overall CVS learning, a closer examination illuminated distinctive patterns in the provision of on-the-fly scaffolds. In practical terms, our findings underscore the potential for teachers in real classrooms to optimize students' development of scientific inquiry skills through thoughtful lesson planning and the preparation and use of on-the-fly scaffolding. Our findings contribute to the ongoing discourse on optimal methodologies for science education, a discussion that De Jong et al. (2023) and Sweller et al. (2024) have highlighted recently. Our study demonstrates that a standardized instructional approach is not necessarily required for science learning. Instead, teachers may adapt their methods by offering varying degrees of support, based on their perception of the individual needs of students, in accordance with a responsive teaching approach. This aligns with the consensus of both research groups, emphasizing that a successful pedagogical strategy is likely to involve a synthesis of inquiry-based learning, as highlighted by De Jong et al. (2023), and direct instruction, as recommended by Sweller et al. (2024). In conclusion, our study contributes to the understanding of the nuanced interplay between guidance and the provision of on-the-fly scaffolding in shaping primary school students' scientific inquiry skills.

Acknowledgments

We would like to thank all children who participated in our study. Special thanks go to our research assistants Assistant-A and Assistant-B. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

For the purpose of transparency and scientific integrity, the data supporting the results of this research are available upon request from the

corresponding author, Heide Sasse. We are committed to sharing our findings and facilitating further scholarly dialogue.

Ethics Statement

In conducting this study, we affirm that we obtained informed consent from all participants involved. The research procedures strictly adhered to the ethical standards outlined by the American Psychological Association (APA). The well-being, rights, and confidentiality of the participants were safeguarded throughout the course of the study.

References

- Ainley, J., and M. Luntley. 2007. "The Role of Attention in Expert Classroom Practice." *Journal of Mathematics Teacher Education* 10: 3–22.
- Aleven, V. A., and K. R. Koedinger. 2002. "An Effective Metacognitive Strategy: Learning By Doing and Explaining With a Computer-Based Cognitive Tutor." *Cognitive Science* 262: 147–179. https://doi.org/10.1207/s15516709cog2602_1.
- Alfieri, L., P. J. Brooks, N. J. Aldrich, and H. R. Tenenbaum. 2011. "Does Discovery-Based Instruction Enhance Learning?." *Journal of Educational Psychology* 103: 1–18. <https://doi.org/10.1037/a0021017>.
- Amato-Zech, N. A., K. E. Hoff, and K. J. Doepke. 2006. "Increasing On-Task Behavior in the Classroom: Extension of Self-Monitoring Strategies." *Psychology in the Schools* 43, no. 2: 211–221. <https://doi.org/10.1002/pits.20137>.
- Ashman, G., S. Kalyuga, and J. Sweller. 2020. "Problem-Solving or Explicit Instruction: Which Should Go First When Element Interactivity Is High." *Educational Psychology Review* 32: 229–247.
- Baur, A., and M. Emden. 2020. "How to Open Inquiry Teaching? An Alternative Teaching Scaffold to Foster Students' Inquiry Skills." *Chemistry Teacher International* 3, no. 1: 20190013.
- Berthold, K., T. H. S. Eysink, and A. Renkl. 2009. "Assisting Self-Explanation Prompts Are More Effective Than Open Prompts When Learning With Multiple Representations." *Instructional Science* 37: 345–363.
- Betts, J., J. McKay, P. Maruff, and V. Anderson. 2006. "The Development of Sustained Attention in Children: The Effect of Age and Task Load." *Child Neuropsychology* 12, no. 3: 205–221.
- Bihler, L. M., A. Agache, K. Schneller, J. A. Willard, and B. Leyendecker. 2018. "Expressive Morphological Skills of Dual Language Learning and Monolingual German Children: Exploring Links to Duration of Preschool Attendance, Classroom Quality, and Classroom Composition." *Frontiers in Psychology* 9: 888. <https://doi.org/10.3389/fpsyg.2018.00888>.
- Blanchard, M. R., S. A. Southerland, and E. M. Granger. 2009. "No Silver Bullet for Inquiry: Making Sense of Teacher Change Following an Inquiry-Based Research Experience for Teachers." *Science Education* 93, no. 2: 322–360.
- Bos, W. 2012. *TIMSS 2011 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [TIMSS 2011 Mathematical and Scientific Competencies of Primary School Children in Germany in International Comparison]*. Waxmann Verlag.
- Buehl, M. M., and J. S. Beck. 2015. "The Relationship between Teachers' Beliefs and Teachers' Practices." In *International Handbook of Research on Teachers' Beliefs*, edited by In. H. Fives and M. G. Gill, 67–84. Routledge.
- Calleja, J., C. Foster, and J. Hodgen. 2023. "Teachers' Structuring of Mathematical Inquiry Lessons: Shifting From "Task-First" to "Scaffolded Inquiry." *Research in Mathematics Education*. Published ahead of print, February 28, 2023. <https://doi.org/10.1080/14794802.2023.2176915>.
- Cameron, J., and W. D. Pierce. 1994. "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis." *Review of Educational Research* 64, no. 3: 363–423. <https://doi.org/10.3102/00346543064003363>.
- Chase, C. C., and D. Klahr. 2017. "Invention Versus Direct Instruction: For Some Content, It's a Tie." *Journal of Science Education and Technology* 26: 582–596.
- Chen, O., S. Kalyuga, and J. Sweller. 2017. "The Expertise Reversal Effect Is a Variant of the More General Element Interactivity Effect." *Educational Psychology Review* 29: 393–405.
- Chen, Z., and D. Klahr. 1999. "All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy." *Child Development* 70, no. 5: 1098–1120.
- Chiu, M. M. 2004. "Adapting Teacher Interventions to Student Needs During Cooperative Learning: How to Improve Student Problem Solving and Time On-Task." *American Educational Research Journal* 41, no. 2: 365–399. <https://doi.org/10.3102/00028312041002365>.
- Coffman, T. 2017. *Inquiry-Based Learning: Designing Instruction to Promote Higher Level Thinking*. Rowman & Littlefield.
- Cooper, G., and J. Sweller. 1987. "Effects of Schema Acquisition and Rule Automation on Mathematical Problem-Solving Transfer." *Journal of Educational Psychology* 79, no. 4: 347–362.
- Corbalan, G., F. Paas, and H. Cuypers. 2010. "Computer-Based Feedback in Linear Algebra: Effects on Transfer Performance and Motivation." *Computers & Education* 55, no. 2: 692–703. <https://doi.org/10.1016/j.compedu.2010.03.002>.
- Corno, L. 2008. "On Teaching Adaptively." *Educational Psychologist* 43, no. 3: 161–173.
- Costley, J., A. Gorbunova, M. Courtney, O. Chen, and C. Lange. 2023. "Problem-Solving Support and Instructional Sequence: Impact on Cognitive Load and Student Performance." *European Journal of Psychology of Education* 39: 1817–1840. <https://doi.org/10.1007/s10212-023-00757-7>.
- Crocker, S., and H. Buchanan. 2011. "Scientific Reasoning in a Real-World Context: The Effect of Prior Belief and Outcome on Children's Hypothesis-Testing Strategies." *British Journal of Developmental Psychology* 29, no. 3: 409–424.
- Dean, D. Jr., and D. Kuhn. 2007. "Direct Instruction vs. Discovery: The Long View." *Science Education* 91: 384–397. <https://doi.org/10.1002/sce.20194>.
- Deci, E. L., R. Koestner, and R. M. Ryan. 1999. "The Undermining Effect Is a Reality After All—Extrinsic Rewards, Task Interest, and Self-Determination: Reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999)." *Psychological Bulletin* 125, no. 6: 692–700. <https://doi.org/10.1037/0033-2909.125.6.692>.
- De Jong, T., A. W. Lazonder, C. A. Chinn, et al. 2023. "Let's Talk Evidence—The Case for Combining Inquiry-Based and Direct Instruction." *Educational Research Review* 39: 100536. <https://doi.org/10.1016/j.edurev.2023.100536>.
- Doyle, W. 2013. "Ecological Approaches to Classroom Management." In *Handbook of Classroom Management*, edited by In. C. M. Carolyn and C. S. Weinstein, 107–136. New York: Routledge.
- Draws, R., G. Tani, P. Lopes Cardozo, and S. Chiviawsky. 2020. "Positive Feedback Praising Good Performance Does Not Alter the Learning of an Intrinsically Motivating Task in 10-Year-Old Children." *European Journal of Human Movement* 45: 46–54. <https://doi.org/10.21134/eurjhm.2020.45.5>.
- Edelsbrunner, P. A., L. Schalk, R. Schumacher, and E. Stern. 2018. "Variable Control and Conceptual Change: A Large-Scale Quantitative Study in Elementary School." *Learning and Individual Differences* 66: 38–53.
- Eid, M., M. Gollwitzer, and M. Schmitt. 2015. *Statistik und Forschungsmethoden*. Weinheim, Basel: Beltz.
- English, L. D., and D. King. 2019. "Stem Integration in Sixth Grade: Designing and Constructing Paper Bridges." *International Journal of Science and Mathematics Education* 17, no. 5: 863–884. <https://doi.org/10.1007/s10763-018-9912-0>.

- Freitag, H.-W., and A. Schulz. 2018. *Der sozioökonomische Status der Schülerinnen und Schüler [The socio-economic status of pupils]*. Berlin: Data report 2018. <https://www.bpb.de/nachschlagen/datenreport-2018/bildung/277991/der-soziooekonomische-status-der-schuelerinnen-und-schueler>.
- Furtak, E. M. 2006. "The Problem With Answers: An Exploration of Guided Scientific Inquiry Teaching." *Science Education* 90, no. 3: 453–467.
- Furtak, E. M., T. Seidel, H. Iverson, and D. C. Briggs. 2012. "Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis." *Review of Educational Research* 82, no. 3: 300–329.
- Gettinger, M., and K. M. Kohler. 2013. "Process-Outcome Approaches to Classroom Management and Effective Teaching." In *Handbook of Classroom Management*, edited by C. M. Evertson and C. S. Weinstein, 83–106. New York: Routledge.
- Guerrero-Rosada, P., C. Weiland, M. McCormick, et al. 2021. "Null Relations Between Class Scores and Gains in Children's Language, Math, and Executive Function Skills: A Replication and Extension Study." *Early Childhood Research Quarterly* 54: 1–12. <https://doi.org/10.1016/j.ecresq.2020.07.009>.
- Hammer, D., F. Goldberg, and S. Fargason. 2012. "Responsive Teaching and the Beginnings of Energy in a Third Grade Classroom." *Review of Science, Mathematics and ICT Education* 6, no. 1: 51–72. <https://doi.org/10.26220/rev.1694>.
- Heemsoth, T., and A. Heinze. 2016. "Secondary School Students Learning From Reflections on the Rationale Behind Self-Made Errors: A Field Experiment." *The Journal of Experimental Education* 84, no. 1: 98–118. <https://doi.org/10.1080/00220973.2014.963215>.
- Hill, H. C., C. Y. Charalambous, and M. A. Kraft. 2012. "When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study." *Educational Researcher* 41, no. 2: 56–64. <https://doi.org/10.3102/0013189X12437203>.
- Hoffman, J. V., and G. G. Duffy. 2016. "Does Thoughtfully Adaptive Teaching Actually Exist? A Challenge to Teacher Educators." *Theory Into Practice* 55, no. 3: 172–179.
- Hsin, C.-T., and H.-K. Wu. 2011. Using Scaffolding Strategies to Promote Young Children's Scientific Understandings of Floating and Sinking. *Journal of Science Education and Technology* 20, no. 5: 656–666. <https://doi.org/10.1007/s10956-011-9310-7>.
- Hsu, C.-Y., S. Kalyuga, and J. Sweller. 2015. "When Should Guidance Be Presented in Physics Instruction?." *Archives of Scientific Psychology* 3, no. 1: 37–53.
- Hume, A. 2009. "Authentic Scientific Inquiry and School Science." *Teaching Science* 55, no. 2: 35–41.
- Jaber, L. Z. 2021. "He Got a Glimpse of the Joys of Understanding"—The Role of Epistemic Empathy in Teacher Learning." *Journal of the Learning Sciences* 30, no. 3: 433–465. <https://doi.org/10.1080/10508406.2021>.
- Kaiser, I., and J. Mayer. 2019. "The Long-Term Benefit of Video Modeling Examples for Guided Inquiry." *Frontiers in Education* 4: 104. <https://doi.org/10.3389/educ.2019.00104>.
- Kalyuga, S. 2007. "Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction." *Educational Psychology Review* 19, no. 4: 509–539. <https://doi.org/10.1007/s10648-007-9054-3>.
- Kalyuga, S., P. Chandler, J. Tuovinen, and J. Sweller. 2001. "When Problem Solving Is Superior to Studying Worked Examples." *Journal of Educational Psychology* 93, no. 3: 579–588. <https://doi.org/10.1037/0022-0663.93.3.579>.
- Kapur, M. 2008. "Productive Failure." *Cognition and Instruction* 26, no. 3: 379–424.
- Kapur, M. 2016. "Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning." *Educational Psychologist* 51, no. 2: 289–299. <https://doi.org/10.1080/00461520.2016.1155457>.
- Kind, P., and J. Osborne. 2017. "Styles of Scientific Reasoning: A Cultural Rationale for Science Education?." *Science Education* 101, no. 1: 8–31.
- Kirschner, P. A., J. Sweller, and R. E. Clark. 2006. "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching." *Educational Psychologist* 41, no. 2: 75–86. https://doi.org/10.1207/s15326985ep4102_1.
- Klahr, D., and J. Li. 2005. "Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back." *Journal of Science Education and Technology* 14: 217–238.
- Klahr, D., and M. Nigam. 2004. "The Equivalence of Learning Paths in Early Science Instruction: Effects of Direct Instruction and Discovery Learning." *Psychological Science* 15, no. 10: 661–667.
- Kleickmann, T. 2012. *Kognitiv aktivieren und inhaltlich strukturieren im naturwissenschaftlichen Sachunterricht*. Kiel: IPN Leibniz-Institut f. d. Pädagogik d. Naturwissenschaften an d. Universität Kiel.
- Klika, D., and S. Abels. 2016. "Scaffolding Guided Inquiry-Based Chemistry Education at an Inclusive School." In *Science Education Research: Engaging Learners From a Sustainable Future, Proceedings of ESERA 2015, Helsinki*, edited by J. Lavonen, K. Juuti, J. Lampiselkä, A. Uitto and K. Hahl (Hrsg.), 1030–1037.
- Koerber, S., D. Mayer, C. Osterhaus, K. Schwippert, and B. Sodian. 2015. "The Development of Scientific Thinking in Elementary School: A Comprehensive Inventory." *Child Development* 86, no. 1: 327–336. <https://doi.org/10.1111/cdev.12298>.
- Koerber, S., and C. Osterhaus. 2019. "Individual Differences in Early Scientific Thinking: Assessment, Cognitive Influences, and Their Relevance for Science Learning." *Journal of Cognition and Development* 20, no. 4: 510–533. <https://doi.org/10.1080/15248372.2019.1620232>.
- Lamb, R. L., L. Annetta, J. Meldrum, and D. Vallett. 2012. "Measuring Science Interest: Rasch Validation of the Science Interest Survey." *International Journal of Science and Mathematics Education* 10: 643–668.
- Lazonder, A. W., and R. Harmsen. 2016. "Meta-Analysis of Inquiry-Based Learning: Effects of Guidance." *Review of Educational Research* 86, no. 3: 681–718.
- Lazonder, A. W., N. Janssen, H. Gijlers, and A. Walraven. 2021. "Patterns of Development in Children's Scientific Reasoning: Results From a Three-Year Longitudinal Study." *Journal of Cognition and Development* 22, no. 1: 108–124. <https://doi.org/10.1080/15248372.2020.1814293>.
- Leuchter, M., and B. Naber. 2019. "Studying Children's Knowledge Base of One-Sided Levers As Force Amplifiers." *Journal of Research in Science Teaching* 56, no. 1: 91–112.
- Leuchter, M., and H. Saalbach. 2014. "Verbale Unterstützungsmaßnahmen im Rahmen eines naturwissenschaftlichen Lernangebots in Kindergarten und Grundschule [Verbal Support Measures in the Context of a Science Learning Program in Kindergarten and Elementary School]." *Unterrichtswissenschaft* 42, no. 2: 117–131.
- Limón, M. 2001. "On the Cognitive Conflict As an Instructional Strategy for Conceptual Change: A Critical Appraisal." *Learning and Instruction* 11, no. 4–5: 357–380.
- Martella, A. M., D. Klahr, and W. Li. 2020. "The Relative Effectiveness of Different Active Learning Implementations in Teaching Elementary School Students How to Design Simple Experiments." *Journal of Educational Psychology* 112, no. 8: 1582–1596.
- Martin, N. D., C. Dornfeld Tissenbaum, D. Gnesdilow, and S. Puntambekar. 2019. "Fading Distributed Scaffolds: The Importance of Complementarity Between Teacher and Material Scaffolds." *Instructional Science* 47: 69–98.

- Marzano, R. J., B. B. Gaddy, and C. Dean. 2000. *What Works in Classroom Instruction*. Aurora: Mid-continent Research for Education and Learning (McREL).
- Maskiewicz, A. C. 2015. "Navigating the Challenges of Teaching Responsively: An Insider's Perspective." In *Responsive Teaching in Science and Mathematics*, edited by A. D. Robertson, R. E. Scherr and D. Hammer, 123–143. New York: Routledge.
- Matlen, B. J., and D. Klahr. 2013. "Sequential Effects of High and Low Instructional Guidance on Children's Acquisition of Experimentation Skills: Is It All in the Timing." *Instructional Science* 41: 621–634.
- Metz, K. E. 2011. "Disentangling Robust Developmental Constraints From the Instructionally Mutable: Young Children's Epistemic Reasoning About a Study of Their Own Design." *Journal of the Learning Sciences* 20, no. 1: 50–110.
- Miller-Cotto, D., J. L. Booth, and N. S. Newcombe. 2022. "Sketching and Verbal Self-Explanation: Do They Help Middle School Children Solve Science Problems." *Applied Cognitive Psychology* 36, no. 4: 919–935. <https://doi.org/10.1002/acp.3980>.
- Mirkovic, J., A. Aggarwal, D. Weinman, P. Lepe, J. Mache, and R. Weiss. 2020. "Using Terminal Histories to Monitor Student Progress on Hands-on Exercises." *Proceedings of the 51st ACM Technical Symposium on Computer Science Education, Portland, OR, USA*. 866–872. <https://doi.org/10.1145/3328778.3366935>.
- Mulder, Y. G., A. W. Lazonder, and T. de Jong. 2014 February. "Using Heuristic Worked Examples to Promote Inquiry-Based Learning." *Learning and Instruction* 29: 56–64. <https://doi.org/10.1016/j.learninstruc.2013.08.001>.
- Munshi, A., G. Biswas, E. Davalos, O. Logan, and M. Rushdy. 2022. "Adaptive Scaffolding to Support Strategic Learning in an Open-Ended Learning Environment [Full Paper]." In *The 30th International Conference on Computers in Education (ICCE), Kuala Lumpur, Malaysia*. 150–156.
- Nakagawa, S., and H. Schielzeth. 2013. "A General and Simple Method for Obtaining R^2 From Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4, no. 2: 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>.
- NGSS Lead States. 2013. *Next Generation Science Standards: For States, by States*. National Washington, DC: Academies Press.
- Piekny, J., and C. Maehler. 2013. "Scientific Reasoning in Early and Middle Childhood: The Development of Domain-General Evidence Evaluation, Experimentation, and Hypothesis Generation Skills." *British Journal of Developmental Psychology* 31, no. 2: 153–179. <https://doi.org/10.1111/j.2044-835X.2012.02082.x>.
- Praetorius, A.-K., C. Pauli, K. Reusser, K. Rakoczy, and E. Klieme. 2014. "One Lesson Is All You Need? Stability of Instructional Quality Across Lessons." *Learning and Instruction* 31: 2–12. <https://doi.org/10.1016/j.learninstruc.2013.12.002>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Radoff, J., A. D. Robertson, S. Fargason, and F. Goldberg. 2018. "Responsive Teaching and High-Stakes Testing." *Science and Children* 55, no. 9: 88–91.
- Reiser, B. J. 2004. "Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work." *Journal of the Learning Sciences* 13, no. 3: 273–304. https://doi.org/10.1207/s15327809jls1303_2.
- Richards, J., A. Elby, M. J. Luna, A. D. Robertson, D. M. Levin, and C. G. Nyeggen. 2020. "Reframing the Responsiveness Challenge: A Framing-Anchored Explanatory Framework to Account for Irregularity in Novice Teachers' Attention and Responsiveness to Student Thinking." *Cognition and Instruction* 38, no. 2: 116–152.
- Rights, J. D., and S. K. Sterba. 2021. "Effect Size Measures for Longitudinal Growth Analyses: Extending a Framework of Multilevel Model R-Squareds to Accommodate Heteroscedasticity, Autocorrelation, Nonlinearity, and Alternative Centering Strategies." *New Directions for Child and Adolescent Development* 2021, no. 175: 65–110.
- Rittle-Johnson, B., and J. R. Star. 2007. "Does Comparing Solution Methods Facilitate Conceptual and Procedural Knowledge? An Experimental Study on Learning to Solve Equations." *Journal of Educational Psychology* 99, no. 3: 561–574. <https://doi.org/10.1037/0022-0663.99.3.561>.
- Rittle-Johnson, B., and J. R. Star. 2009. "Compared With What? The Effects of Different Comparisons on Conceptual Knowledge and Procedural Flexibility for Equation Solving." *Journal of Educational Psychology* 101, no. 3: 529–544. <https://doi.org/10.1037/a0014224>.
- Robertson, A. D., Scherr, R. E., and Hammer, D., eds. 2016. *Responsive Teaching in Science and Mathematics* (1–35). New York: Routledge.
- Robinson, E., and B. J. Fraser. 2013. "Kindergarten Students' and Parents' Perceptions of Science Classroom Environments: Achievement and Attitudes." *Learning Environments Research* 16: 151–167.
- Ruiz-Primo, M. A., and S. M. Brookhart. 2017. *Using Feedback to Improve Learning*. Routledge.
- Ruiz-Primo, M. A., and E. M. Furtak. 2006. "Informal Formative Assessment and Scientific Inquiry: Exploring Teachers' Practices and Student Learning." *Educational Assessment* 11, no. (3–4): 237–263. <https://doi.org/10.1080/10627197.2006.9652991>.
- Ruzaman, N. K., and D. I. Rosli. 2020. "Inquiry-Based Education: Innovation in Participatory Inquiry Paradigm." *International Journal of Emerging Technologies in Learning (IJET)* 15, no. 10: 4–15.
- Ryan, R. M., and E. L. Deci. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25, no. 1: 54–67. <https://doi.org/10.1006/ceps.1999.1020>.
- Schalk, L., P. A. Edelsbrunner, A. Deiglmayr, R. Schumacher, and E. Stern. 2019. "Improved Application of the Control-of-Variables Strategy As a Collateral Benefit of Inquiry-Based Physics Education in Elementary School." *Learning and Instruction* 59: 34–45.
- Schwartz, D. L., C. C. Chase, M. A. Oppezzo, and D. B. Chin. 2011. "Practicing Versus Inventing With Contrasting Cases: The Effects of Telling First on Learning and Transfer." *Journal of Educational Psychology* 103: 759–775. <https://doi.org/10.1037/a0025140>.
- Schwarz, C. V., M. Braaten, C. Haverly, and E. X. de los Santos. 2021. "Using Sense-Making Moments to Understand How Elementary Teachers' Interactions Expand, Maintain, or Shut Down Sense-Making in Science." *Cognition and Instruction* 39, no. 2: 113–148.
- Settlage, J., and S. A. Southerland. 2007. *Teaching Science to all Children: Using Culture as a Starting Point*. New York: Routledge.
- Shanab, M. E., D. Peterson, S. Dargahi, and P. Deroian. 1981. "The Effects of Positive and Negative Verbal Feedback on the Intrinsic Motivation of Male and Female Subjects." *The Journal of Social Psychology* 115, no. 2: 195–205. <https://doi.org/10.1080/00224545.1981.9711659>.
- Siegler, R. S., ed. 2013. "The Origins of Scientific Reasoning." In *Children's Thinking*, 109–150. New York: Psychology Press.
- Simons, K. D., and J. D. Klein. 2007. "The Impact of Scaffolding and Student Achievement Levels in a Problem-Based Learning Environment." *Instructional Science* 35, no. 1: 41–72. <https://doi.org/10.1007/s11251-006-9002-5>.
- Siverling, E. A., T. J. Moore, E. Suazo-Flores, C. A. Mathis, and S. S. Guzey. 2021. "What Initiates Evidence-Based Reasoning?: Situations That Prompt Students to Support Their Design Ideas and Decisions." *Journal of Engineering Education* 110, no. 2: 294–317.
- Slot, P. L., D. Bleses, L. M. Justice, J. Markussen-Brown, and A. Højen. 2018. "Structural and Process Quality of Danish Preschools: Direct and Indirect Associations With Children's Growth in Language and

- Preliteracy Skills." *Early Education and Development* 29, no. 4: 581–602. <https://doi.org/10.1080/10409289.2018.1452494>.
- Sodian, B., D. Zaitchik, and S. Carey. 1991. "Young Children's Differentiation of Hypothetical Beliefs From Evidence." *Child Development* 62, no. 4: 753–766.
- Soncini, A., M. C. Matteucci, and F. Butera. 2021. "Error Handling in the Classroom: An Experimental Study of Teachers' Strategies to Foster Positive Error Climate." *European Journal of Psychology of Education* 36, no. 3: 719–738. <https://doi.org/10.1007/s10212-020-00494-1>.
- Soncini, A., E. P. Visintin, M. C. Matteucci, C. Tomasetto, and F. Butera. 2022. "Positive Error Climate Promotes Learning Outcomes Through Students' Adaptive Reactions Towards Errors." *Learning and Instruction* 80: 101627. <https://doi.org/10.1016/j.learninstruc.2022.101627>.
- Strand-Cary, M., and D. Klahr. 2008. "Developing Elementary Science Skills: Instructional Effectiveness and Path Independence." *Cognitive Development* 23, no. 4: 488–511.
- Studhalter, U. T., M. Leuchter, A. Tettenborn, A. Elmer, P. A. Edelsbrunner, and H. Saalbach. 2021. "Early Science Learning: The Effects of Teacher Talk." *Learning and Instruction* 71: 101371.
- Sweller, J. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12, no. 2: 257–285.
- Sweller, J. 2006. "The Worked Example Effect and Human Cognition." *Learning and Instruction* 16, no. 2: 165–169. <https://doi.org/10.1016/j.learninstruc.2006.02.005>.
- Sweller, J., P. A. Kirschner, and R. E. Clark. 2007. "Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries." *Educational Psychologist* 42, no. 2: 115–121. <https://doi.org/10.1080/00461520701263426>.
- Sweller, J., L. Zhang, G. Ashman, W. Cobern, and P. A. Kirschner. 2024. "Response to De Jong et al.'s (2023) Paper "Let's Talk Evidence—The Case for Combining Inquiry-based and Direct Instruction"." *Educational Research Review* 42: 100584. <https://doi.org/10.1016/j.edurev.2023.100584>.
- Van de Pol, J., M. Volman, and J. Beishuizen. 2010. "Scaffolding in Teacher–Student Interaction: A Decade of Research." *Educational Psychology Review* 22: 271–296.
- Van Leeuwen, A., and J. Janssen. 2019. "A Systematic Review of Teacher Guidance During Collaborative Learning in Primary and Secondary Education." *Educational Research Review* 27: 71–89.
- Van Schijndel, T. J. P., B. R. J. Jansen, and M. E. J. Raijmakers. 2018. "Do Individual Differences in Children's Curiosity Relate to Their Inquiry-Based Learning?." *International Journal of Science Education* 40, no. 9: 996–1015.
- Van Vo, D., B. Csapó, and S. Greiff. 2023. "Development of the Control of Variables Strategy in Physics Among Secondary School Students." *Thinking Skills and Creativity* 49: 101371. <https://doi.org/10.1016/j.tsc.2023.101371>.
- Vorholzer, A., and C. von Aufschnaiter. 2019. "Guidance in Inquiry-Based Instruction—An Attempt to Disentangle a Manifold Construct." *International Journal of Science Education* 41, no. 11: 1562–1577.
- Weber, A. M., and M. Leuchter. 2022. "Fostering Children's Block Building Self-Concepts and Stability Knowledge Through Construction Play." *Journal of Applied Developmental Psychology* 80: 101400. <https://doi.org/10.1016/j.appdev.2022.101400>.
- West, S. G., J. F. Finch, and P. J. Curran (1995). *Structural Equation Models With Nonnormal Variables: Problems and Remedies*.
- Wittwer, J., and A. Renkl. 2008. "Why Instructional Explanations Often Do Not Work: A Framework for Understanding the Effectiveness of Instructional Explanations." *Educational Psychologist* 43, no. 1: 49–64.
- Zhang, L. 2016. "Is Inquiry-Based Science Teaching Worth the Effort? Some Thoughts Worth Considering." *Science & Education* 25: 897–915.
- Zhang, L., P. A. Kirschner, W. W. Cobern, and J. Sweller. 2022. "There Is an Evidence Crisis in Science Educational Policy." *Educational Psychology Review* 34, no. 2: 1157–1176.