ORIGINAL RESEARCH

Accommodating heterogeneity: the interaction of instructional scaffolding with student preconditions in the learning of hypothesis-based reasoning

Hanna Grimm[1] · **Peter A. Edelsbrunner[2]** · **Kornelia Möller[1]**

Abstract

Hypothesis-based reasoning with conditionals is a skill that is required for engaging in integral activities of modern elementary school science-curricula. The teaching of this skill at this early stage of education, however, is demanding, particularly in whole school classes in which it is difficult to adapt teaching to children's individual needs. We examine whether a scaffold that is static yet tailored to the context, in which the teacher explicitly models the reasoning process, manages to meet students' individual cognitive preconditions for learning this skill. Within an inquiry-based learning setting, $N = 143$ third-graders underwent either an experimental condition in which they received the explicit scaffold, or a control condition in which they did not receive this specific scaffold. Employing a latent transition analysis and a general additive model, it is examined how the additional scaffold interacted with students' prior knowledge, inhibition ability, and logical reasoning judged by their own teachers. It is found that the additional scaffolds managed to meet the needs of students with little prior knowledge; under the control condition, students with little prior knowledge showed decreased learning achievement, whereas under the experimental condition, students with differing prior knowledge learned to comparable extent and on a higher level. The scaffolds also almost fully diminished a disadvantage for students with lower teacher-judged logical reasoning, and supported students with high inhibition ability in mastering the most difficult aspect of reasoning based on irrelevant evidence. Implications for science education are discussed.

Keywords Hypothesis-based reasoning · Scaffolding · Student preconditions · Aptitude-treatment interactions · Latent transition analysis · General additive model

---

Hanna Grimm and Peter A. Edelsbrunner have equally contributed to this work.

✉ Peter A. Edelsbrunner
  peter.edelsbrunner@ifv.gess.ethz.ch

  Hanna Grimm
  hanna.grimm@uni-muenster.de

[1] University of Münster, Leonardo-Campus 11, 48149 Münster, Germany

[2] ETH Zurich, RZ H16, Clausiusstrasse 59, 8092 Zurich, Switzerland

Introduction

Conditional reasoning is one key component of logical reasoning, because much of our knowledge is conditional (Johnson-Laird & Byrne, 2002). If you are late, then the train is gone. If you put yeast in a dough, it rises. If you take a hot shower, the mirror fogs up. However, not only mundane reasoning is often conditional. Understanding conditionals is also essential in scientific reasoning and hypothesis testing (Barrouillet et al., 2008; Gauffroy & Barrouillet, 2011). Such sentences with 'if…' express hypothetical thinking, which is an essential part of scientific knowledge acquisition. Evaluating such conditionals allows drawing correct inferences from evidence.

In science education, inquiry plays a prominent role by helping students to come to understand science. One important aspect of inquiry learning is that it leads the learners through their own process of knowledge-acquisition (NGSS, 2013). Students learn about a topic through self-directed (although often to varying extent scaffolded or teacher-supported) investigations. Thereby, they do not only learn science content but also science processes, which both are included in the concept of scientific literacy (Bybee, 2002; Lazonder & Harmsen, 2016; Lederman et al., 2013). One of the process skills belonging to scientific inquiry next to for example observing, classifying, and questioning is scientific reasoning (Lederman et al., 2013). Reasoning skills such as constructing abstract models and representations, as well as the ability to inhibit prior knowledge in order to enable the entertainment of new information, are regarded as important factors in science learning (Vosniadou, 2019).

In scrutinizing, for example, physical phenomena, which are often based on conditionals (e.g. 'If the knife is made of wood, then it floats'), students need to evaluate whether evidence supports or rejects a specific hypothesis. In the following, we will use the term 'hypothesis-based reasoning' in the context of such conditionals, following the hypothetico-deductive approach to scientific reasoning (Popper, 1972). In this context, we define hypothesis-based reasoning as the ability to judge the truth-value of a conditional hypothesis given the observed outcome of an empirical test. Importantly, this kind of hypothesis-based reasoning distinguishes itself from other kinds of scientific reasoning by the special role of the conditional reasoning process. To this, affirmative, refuting, and irrelevant events must be distinguished. When testing a conditional hypothesis of the kind *If p then q*, such as "If the object is round, then it bounces," there are four combinations of possibly occurring events. The *affirmative* event, pq [round, bouncing], supports the hypothesis, and the *refuting* event, pnq [round, not bouncing], rejects the hypothesis. Whereas these two events provide information about the adequacy of the hypothesis, the two events not fulfilling the antecedence npq [not round, bouncing] and npnq [not round, not bouncing] are *irrelevant* events. These events do not convey information that would allow for a conclusive test of the hypothesis. Successful hypothesis-based reasoning with conditionals requires the reasoner to distinguish between relevant and irrelevant events (in order to decide whether the event bears on the hypothesis and neglect irrelevant events), and to distinguish whether the outcome of a relevant event affirms or refutes the hypothesis.

Research on scientific reasoning has shown that although some individuals succeed in identifying controlled experiments and conclusive tests already in earlier childhood, many children still struggle with various aspects of scientific reasoning-processes throughout childhood (e.g., Sandoval et al., 2014; Sodian et al., 1991; Piekny et al., 2013). As one source for individual differences in the ability to engage in scientific reasoning, it has been pointed out that children often struggle with differentiating between their own theories and

empirical evidence, and with coordinating these two aspects in the course of inquiry (Koerber et al., 2015).

Research in cognitive psychology has shown that children also struggle with the specific aspect of hypothesis-based reasoning with conditionals (Barrouillet et al., 2008). Gauffroy & Barrouillet (2009) explain age-related improvement of this kind of hypothesis-based reasoning by consulting the mental models-theory proposed by Johnson-Laird & Byrne (2002). According to the mental models theory, people interpret conditional sentences by constructing and manipulating mental models in working memory (Barrouillet & Lecas, 1999; Gauffroy & Barrouillet, 2011; Johnson-Laird & Byrne, 2002). The evaluation of irrelevant [np] events is a particularly demanding analytic process. Children in elementary school age do not yet possess fully developed working memory capacity, as well as the ability of inhibition (Diamond, 2013). Relations between working memory, inhibition and conditional reasoning might contribute to their limited ability in evaluating irrelevant conditionals (Vergauwe et al., 2013; Handley et al., 2004).

In the present study, we examine three student characteristics as factors influencing the learning success of hypothesis-based reasoning. In addition to inhibition ability, prior knowledge in hypothesis-based reasoning and teacher judgments of students' general ability of logical reasoning might be relevant preconditions. We examine how a scaffold in which the teacher models the underlying reasoning process interacts with these student preconditions in the acquisition of hypothesis-based reasoning with conditionals.

Training interventions to foster elementary school students' hypothesis-based reasoning

The demand of hypothesis-based reasoning ability in the context of Scientific Literacy and successful inquiry-based learning on the one hand, and the existing difficulties for young children on the other hand raise the questions whether and how hypothesis-based reasoning with conditionals can be trained and fostered in elementary school children. Whereas various studies have examined instructional support of other aspects of scientific reasoning, such as the control-of-variables strategy (for a meta-analytic overview, see Schwichow et al., 2016), to the best of our knowledge hypothesis-based reasoning with conditionals has been neglected in the science education literature. However, first approaches have been developed in this regard.

A study by Robisch et al. (2014) examined whether fostering third-graders' hypothesis-based reasoning with conditionals is possible by reducing cognitive load and facilitating the inhibition of spontaneous but often misguided reasoning processes. Following the concept of scaffolding, third-graders were instructed in one-to-one situations. Scaffolding describes the support provided for the completion of a task that learners otherwise might not be able to complete (van de Pol et al., 2010). The scaffolding in the study of Robisch et al. (2014) included modeling as well as channeling and focusing scaffolds (Pea, 2004). A wooden sorting box was used as a static scaffold (Brush & Saye, 2002) to model the proceeding of conditional reasoning. In doing so, the students dealt intensively with the respective hypothesis, focusing step-by-step on the antecedent and the consequence. Having evaluated incorrectly, the students got an adaptive verbal scaffold focusing on the relevant aspects of the hypothesis and the object. The results of Robisch et al. (2014) showed that it is possible to support third-graders' hypothesis-based reasoning in one-to-one laboratory situations by scaffolding the reasoning process. The students in their study showed significant and substantial learning gains.

In the present study, this approach to the instruction of hypothesis-based reasoning is taken to the classroom. To the best of our knowledge, this is the first study aiming at fostering this scientific reasoning-skill in school classes. It is investigated whether third graders' hypothesis-based reasoning with conditionals can be fostered in entire school classes, and how the instruction can be designed such that it best meets the heterogeneous demands of students of this age encountered in real classrooms.

Dealing with heterogeneity: instructional approaches and core ideas

Teaching in entire school classes poses the challenge to deal with learners' heterogeneous preconditions. Because of demographic changes and recently established educational policies in many countries, student heterogeneity is an increasingly common fact of classroom learning (Corno, 2008; Decristan et al., 2017; Subban, 2006). Hence, finding the best match between learners' individual preconditions and instruction is considered a central issue in educational research and practice (Kyllonen & Lajoie, 2003; Scott, 2013; Schnotz, 2010). The aspect of heterogeneity is of particular importance for the present study because it is the first study that focuses on fostering hypothesis-based reasoning with conditionals in whole classrooms of third graders. Although this kind of reasoning is demanding for third-graders, the results of Robisch et al. (2014) showed that careful one-to-one instruction can foster this skill in this age group. However, the question arises whether and how in whole classrooms, where it might be difficult to adapt instruction to children's individual needs, this skill can still be fostered.

A number of different concepts to deal with students' heterogeneous preconditions have been proposed and trialed, originating from several disciplines. It follows that they overlap in parts and are often not clearly distinguishable. These concepts encompass, to name a few central ones, differentiated instruction, individualized instruction, adaptive instruction, and personalized learning (Dumont, 2018; Suprayogi et al., 2017). Even if such different approaches pursue the same idea by coping with the diversity of students' individual needs, they differ in a variety of ways in implementation. These include for example the grouping of learners according to their level of expertise and interest, the allocation of different learning materials and adaptation of learning time, and providing additional support for struggling learners in combination with enrichment for more advanced students (Suprayogi et al., 2017).

Not only the instructional approaches, but also defining the exact aims of successfully dealing with heterogeneity is characterized by different normative positions. Hertel (2014) quotes three possible ambitions regarding the adjustment of teaching to students' individual preconditions. A first perspective aims at achieving homogeneity in learning outcomes, a second perspective at achieving comparable learning progress across students, and a third perspective at the optimal use of learners' developmental potential.

In the present study, we take the position that instruction successfully manages to deal with heterogeneity if it leads to optimal learning gains across a broad range of student preconditions. Students should be supported in utilizing their individual learning potential as far as possible. Within the context of the present study, we pose the aim to promote the learning of hypothesis-based reasoning of students with varying levels of prior knowledge, inhibition ability, and teacher-judged logical reasoning. In the following, these three preconditions are discussed, followed by a discussion of how the scaffolding-approach that is implemented here is aimed at meeting children's heterogeneity in these preconditions.

Heterogeneity in student preconditions

Heterogeneity in student preconditions can refer to a broad range of different characteristics, such as developmental level, cognitive/intellectual ability, motivation, gender, different mother tongue than spoken in class, ethnicity, and sociocultural background (Randi & Corno, 2005; Tomlinson et al., 2003). With regard to the learning of hypothesis-based reasoning, we focus on heterogeneity in learners' cognitive abilities, because promoting hypothesis-based reasoning is based on a strong cognitive component (Barrouillet et al., 2008; Gauffroy & Barrouillet, 2011).

In the context of the present study, we consider three student characteristics as cognitive preconditions that might affect their learning during the instruction of hypothesis-based reasoning: (1) prior knowledge in hypothesis-based reasoning, (2) inhibition ability, and (3) students' logical reasoning measured by teachers' judgment. Although these three variables and approaches to their measurement do not cover the whole range of heterogeneity in student preconditions, they do cover three diverse and important aspects of cognitive abilities.

The role of the first precondition, students' prior knowledge, differs between contexts and learning contents; however, prior knowledge is generally considered a major determinant of learning (Schroeders et al., 2016; Simonsmeier et al., 2022). In the present study, prior knowledge amounts to students' prior levels of hypothesis-based reasoning based on affirmative, refuting, and irrelevant events. If students already show skilled reasoning on any of these three aspects before instruction, this implies that they will not be able to gain any further from the respective part of instruction. Therefore, controlling for prior knowledge is important, in order to control for this situation in the examination of students' development from before to after the instruction. In addition, from the perspective of mental models theory, prior knowledge implies that students have schemata available that they can build upon. This allows connecting the new information to their prior knowledge and freeing capacities for extending their available schemata to new types of events that they have not yet mastered (Barrouillet & Lecas, 1999; Gauffroy & Barrouillet, 2011; Johnson-Laird & Byrne, 2002). However, prior schemata might not always be correct and might have to be restructured. In hypothesis-based reasoning with conditionals, children might for example possess schemata in which evidence that does not fulfill the antecedence is incorrectly used for the evaluation of a hypothesis (Barrouillet & Lecas, 1999).

The second student precondition, inhibition ability, is generally considered an important predictor of hypothesis-based reasoning. Inhibition is the ability to tune out task-irrelevant processes (Tiego et al., 2018). In domain-specific learning, inhibition ability allows reasoning independently of one's own goals or beliefs (Handley et al., 2004) Robisch et al. (2014) found that inhibitory control is a predictor for reasoning success. In a study with 61 ten-year-olds, they were able to assert that children's reasoning in belief-based problems is dependent on the ability to inhibit belief-based responses. We assume that inhibition ability plays a role in hypothesis-based reasoning because it enables students to suppress domain-specific intuitions that are linked to their prior knowledge. For example, focusing on the prior knowledge that objects filled with air bounce can distract students from concentrating on the underlying reasoning process in which they have to focus on the relation between antecedent ("if objects filled with air bounce…") and consequence ("…then this object [which is filled with air] should bounce"), and only afterwards connect these to domain-specific inferences. In addition, we include this aspect because in comparison to

domain-specific prior knowledge, inhibition ability is a domain-general and therefore more context-independent student precondition.

The third student precondition, logical reasoning, covers a more general representation of cognitive ability. More general reasoning abilities can contribute to individual differences in the acquisition of scientific reasoning skills during inquiry-based instruction (Wagensveld et al., 2015). In the present study, we assess teacher judgments of students' logical reasoning. We define teacher-judged logical reasoning as teachers' overall subjective evaluation of students' verbal and non-verbal skills related to reasoning and problem solving. We provide teachers with descriptions of multiple areas encompassing non-verbal and verbal skills, asking them to provide an overall judgment for each student. Even if teacher judgments are far from perfect accuracy (Südkamp et al., 2012), they are often the primary source of information on students' educational progress. The results of meta-analyses about teacher judgment accuracy indicate positive and fairly high correlations of judgment and test performance regarding academic achievement (Südkamp et al., 2012) and moderate judgment accuracy for cognitive abilities (Machts et al., 2016). These findings indicate that although they cannot be considered a directly and objectively measured criterion, teachers' judgments can provide an informative indicator of students' academic potential and cognitive abilities. We include this subjective measure as a predictor for students' acquisition of hypothesis-based reasoning because in addition to the two other predictors, prior knowledge and inhibition, it is a less specific, more general indicator of students' cognitive abilities. In addition, it is closely related to the skills that students are expected to acquire, because hypothesis-based reasoning is a type of logical reasoning.

Scaffolding as an instructional approach to meet students' heterogeneous preconditions

The concept of scaffolding connects two approaches of dealing with heterogeneity, which is why we consider scaffolding as an appropriate approach for fostering hypothesis-based reasoning in the classroom. On the one hand, one way of dealing with students' heterogeneous preconditions is to adapt support individually, realized by changing so-called sight structures of classroom instruction (Seiz et al., 2016). These encompass organizational adaptations such as group-building, adaptations of material, and learning time (Tomlinson et al., 2003). In addition, in the idea of scaffolding, a teacher's support has to be adapted to the learners' needs, regardless of whether a single student or a group of students is involved (van de Pol et al., 2010). This especially concerns the principle of fading instructional support, which is based on the need of individual adaptation (van de Pol et al., 2010; Wood et al., 1976).

On the other hand, meeting students' heterogeneous preconditions can be pursued by improving deep structures rather than sight structures (Seiz et al., 2016). In contrast to the idea that different learners require different kinds of instruction (Cronbach & Snow, 1977; Schnotz, 2010), adapting deep structures emphasizes teaching quality as an overall strategy. Deep structures refer to the process of teaching and they describe the quality of interactions between teacher and student (Seiz et al., 2016). In this regard, several authors identified three dimensions of teaching quality: Cognitive activation/ instructional support, supportive climate, and classroom management (Decristan et al., 2017; Pianta & Hamre, 2009). Including the first two dimensions, scaffolding aims at implementing deep structures as well.

Educational research indicates that in comparison to sight structures, the quality of deep structures is more important for students' learning outcomes (Hattie, 2009; Seidel & Shavelson, 2007). Instructional settings focusing on the implementation of supportive deep structures can particularly impact the influence of heterogeneity in the classroom on students' learning outcomes (Vorholzer et al., 2018; Wagensveld et al., 2015). For elementary school classes with large student heterogeneity, particularly cognitive activation and supportive climate can have positive effects on learning outcomes (Decristan et al., 2017).

The prototypical scaffolding (Wood et al., 1976) takes more time than is generally available in day-to-day science instruction (Hogan & Pressley, 1997). For this reason, similar support is often provided for all students, instead of tailoring instructional support to each student's needs (Martin et al., 2018). However, it is important that essential elements of scaffolding like consistent diagnosis and fading of support are not being neglected (Puntambeker & Hübscher, 2005).

Scaffolding is characterized by two central techniques. Reiser (2004) differentiates between the complementary mechanisms of structuring the task and problematizing the subject matter. Structuring helps students to decompose complex tasks, to focus effort, and to monitor their own progress. Structuring tasks can disencumber working memory. Problematizing, on the other hand, poses a challenge to the learner, by eliciting articulation and decisions, surfacing gaps and disagreements (Reiser, 2004). Drawing attention to refuting evidence can help learners to recognize a cognitive conflict, which can support conceptual change (Posner et al., 1982; Kang et al., 2004).

Although these and further approaches for dealing with heterogeneity have been suggested and partially undergone trials, meeting the different needs of individual students remains a challenging task for teachers. Roll et al. (2018) point out that little is known about the effect of instructional guidance on the interaction with student preconditions. Thus, the question of who benefits from which kind and amount of support, but also the question of the interaction of teaching quality with different student abilities remain to be further investigated. Similarly, Decristan et al. (2017) emphasize that until now, "the interplay between heterogeneity in the classroom and teaching quality has not received much attention in empirical research".

The present study

The present study presents a classroom-based training of hypothesis-based reasoning with conditionals. Within this context, we investigate whether adapting deep structures in an experimental condition by implementing a particular approach to scaffolding manages to support third-graders with heterogeneous preconditions in acquiring this skill. We employ a recently developed scaffolding method that has been carefully adapted to this learning context. We compare an active control condition of third-graders engaging in structured inquiry on hypothesis-based reasoning to an experimental condition in which students receive additional instructional support through two kinds of scaffolds.

In the control condition, the students undergo inquiry activities that are structured regarding the cognitive difficulty of the contents. Specifically, embedded within the context of elasticity, they receive inquiry materials to test a number of conditional hypotheses regarding the bouncing propensity of different objects. In this active control group, the task is structured such that the hypotheses that the students test are of increasing difficulty with regard to the underlying hypothesis-based reasoning process.

In the experimental condition, the students receive two kinds of scaffolds. First, they receive a structuring scaffold in which the teacher, supported by a visual aid (i.e., the sorting box), models the hypothesis-based reasoning process. With the visual aid, the teachers explicate the structure of the reasoning process. The students then receive the same inquiry materials and examine the same conditional hypotheses in the same structured order as the students in the control condition. As a second scaffold, in this condition the teacher provides cognitively activating, problematizing verbal scaffolds during students' inquiry to react adaptively to students' difficulties with the reasoning process. Thus, the focal interest of the present study is to examine how providing these two scaffolds (the structuring scaffold provided through modeling with the visual aid, and teachers' adaptive problematizing scaffolds) serves students' preconditions for learning hypothesis-based reasoning with conditionals.

We expect that these additional scaffolds increase the range of preconditions under which the students can successfully acquire hypothesis-based reasoning. This can mean either that the scaffolding decreases the limiting influence of lower student preconditions on the learning outcome, or that it enables students with higher preconditions to achieve higher learning gains than without the additional scaffolds. With our aim of dealing with heterogeneity in mind, within this context we examine three specific questions.

1) How does structural and problematizing scaffolding, that is provided in addition to an inquiry-based learning sequence, interact with students' prior knowledge in the acquisition of hypothesis-based reasoning?

Our hypothesis regarding this research question is that we expect the scaffolding to decrease the limiting impact of students' prior knowledge, because instructional support can be particularly supportive when prior knowledge is rather low (Schnotz, 2010). We expect that both of the scaffolding measures will contribute to this effect, because the structuring scaffold is assumed to make students aware of their mental schemata, and the problematizing scaffold is assumed to make students aware of problematic aspects of those schemata.

Furthermore, a limiting impact of individual differences in students' inhibition ability on their potential to improve their hypothesis-based reasoning is to be expected, because inhibition ability counts as one of the main predictors of learning hypothesis-based reasoning (Handley et al., 2004; Robisch et al., 2014). Regarding inhibition ability, we examine the following research question:

2) How does the additional scaffolding interact with students' inhibition ability in the acquisition of hypothesis-based reasoning?

We expect that the scaffolding will allow more students to exploit their inhibition ability, because additional instructional support might relieve students' working memory and support them in inhibiting misleading intuitions, such as overriding the influence of domain-specific prior knowledge on their inquiry activities (Schauble, 1996). As a result, the scaffolding should support students with varying levels of inhibition ability in improving their hypothesis-based reasoning. Specifically, the structuring scaffold is assumed to help students in comprehending crucial steps of the reasoning process in which they have to neglect their prior knowledge, which might be difficult to figure out with insufficient inhibition.

Finally, we examine the following research question regarding students' logical reasoning:

3) How does the additional scaffolding interact with students' teacher-judged logical reasoning?

In the experimental condition, teachers provide scaffolds through a stepwise modeling of the reasoning process. We expect these scaffolds to support students with lower logical reasoning, because they help the students to follow the process in a systematic manner without having to delineate the whole process purely on their own. Therefore, we expect the scaffolding to decrease the impact of students' logical reasoning on their learning achievement. We expect in particular the problematizing scaffolds to support students in spotting and correcting errors in their logical reasoning during the inquiry phase.

In addition to a decrease of the limiting impact of students' preconditions, we also expect that under the experimental condition, students' general level of learning outcomes will be higher than in the control condition. If these two criteria were met, this would indicate that in accordance with our intention, the scaffolding successfully managed to meet students' individual requirements.

Method

Sample

Four teachers from four schools in the federal state of North Rhine-Westphalia in Germany were approached to participate in the present study. These four teachers and participating schools agreed that two school classes from the forthcoming wave of third-graders in each of the schools would participate in the study. This resulted in an initial sample of $N = 198$ third-graders from eight school classes in four schools. These students underwent one of two intervention conditions (experimental condition—inquiry based learning with additional scaffolds; control condition—inquiry-based learning without additional scaffolds) delivered by the four different teachers. Each teacher instructed one school class in the experimental condition and another school class in the control condition, in order to balance teacher effects across intervention conditions. After the interventions took place, an implementation check for was conducted by checking on videographies of all lessons whether the teachers stuck to the two different intervention conditions. This dealt as validation of teacher fidelity by condition, and of valid implementation of the teaching scripts within each of the conditions. One of the four teachers delivered the scaffolds meant exclusively for the experimental condition also in the control condition. Consequently, we excluded data from all students who underwent instruction by this teacher ($n = 55$) in order to maintain balanced teacher effects across conditions, leading to a final sample of $N = 143$ students whose data could be included in the analyses.

The 143 third-graders had a mean age of 9.11 ($SD = 0.50$) years, including 81 girls and 62 boys, with $n = 73$ students in the experimental condition, and $n = 70$ students in the control condition. The students stemmed from urban and rural areas. The socioeconomic background in the area of four participating school classes can be described as generally rather high, with a low percentage of residents who are unemployed or receive welfare benefits. Two further school classes stemmed from a school in a middle-sized town with an

immigrant population of about 20%. The teachers indicated that overall, 60 of the 143 participating students had German as their second language. However, they also indicated that only five of the students were not fluent in speaking German and flawless in understanding German. Therefore, language background was not considered in particular in the design of the intervention and in the data analyses.

Assessment instruments

Assessment materials translated from German into English by the study authors are available from https://osf.io/3uf6b.

Hypothesis-based reasoning

To assess students' ability of hypothesis-based reasoning with conditionals, we administered truth-testing tasks based on Barrouillet et al. (2008). Barrouillet et al. (2008) administered truth-testing tasks on a computer, with four trials for each of the four different kinds of conditionals. We adapted these tasks into a paper-pencil-test with physical materials in order to resemble a typical inquiry situation in which children have to interpret an event's evidential value. Trained assistants conducted a material-based survey with the entire class at the same time. The students were confronted with four different assumptions, in the form of relative clauses such as 'Objects that are filled with air bounce.' To address the needs of third-graders, the assumptions were integrated into a small background story, in which other children presented their thoughts. Different objects were presented and it was demonstrated whether the objects bounce, or not. For example, the assistant dropped a tennis ball, stating: "This tennis ball is filled with air and bounces. What does the tennis ball tell you about Tim's [story character] assumption?" The students had to decide and note down whether the event verified or disproved the proposed assumption, or whether it was irrelevant for that assumption. The students were required to reason about the truth values of each of the four presented assumptions, putting a cross on a sheet on which the different answer options were figurally indicated in a very simple manner. In each case, all four of the possible events (pq, p¬q, ¬pq, ¬p¬q) were presented once. To avoid order effects, the sequence of the event possibilities was varied across the four assumptions. The test did not differ before and after the unit.

Inhibition ability

Students' inhibition ability was assessed by an author of the study with a fruit-stroop task (for details and a validation in German-speaking children, see Jansen et al., 1999; Archibald & Kerns, 1999). In this task, seven lines were presented to the students. Each line showed four incorrectly colored fruits such as lemons and plums. The students had to name the correct colors of the fruits as quickly as possible. The study author conducting the assessment recorded the overall time needed to solve all tasks as an indicator of students' inhibition ability. Students' reaction time in seconds was reverse-coded, such that higher inhibition ability is indicated by higher inhibition scores in all analyses. The test was conducted in 1:1-situations in a separate room with each single student.

Logical reasoning

Teacher judgments were requested to assess students' logical reasoning. In accordance with the regional school grading system, the teachers were asked to provide a grading on the school grades scale from 1 (very good) to 6 (insufficient), with the opportunity to specify the grading in more detail by adding a plus or minus to the grade. We asked the teachers to evaluate the students in this way because it is in accordance with the regular regional grading system that all of them use on a daily basis and are well-acquainted with. Specifically, the teachers were asked to provide an overall judgment of each student's abilities, based on short descriptions of logical thinking (reasoning, detecting relations, organizing & classifying), creative and self-guided thinking aimed at problem solving (developing ideas, finding solutions, gathering and developing rationales), expressive powers in oral communication, and comprehension of complex verbal expressions.

Design and procedure

The children underwent an instructional sequence of seven units, each of which lasted 45 min, of which only the second to fourth unit are in focus of this study. This part of the intervention was aimed at fostering students' hypothesis-based reasoning with conditionals. In these three learning units, the students had to apply hypothesis-based reasoning to investigate assumptions. Only in those three units, the experimental condition differed from the control condition, whereby the teacher implemented additional scaffolds to foster hypothesis-based reasoning. In the other four units, which did not deal with the students' alternative assumptions, the students were introduced to the study context and developed research questions together with the teacher (first unit) and in the end, they received instruction on the scientifically correct concepts of elasticity and plasticity (fifth to seventh unit).

All units were distributed over three weeks. The pretest for assessing students' hypothesis-based reasoning, inhibition ability, and teacher judgments of logical reasoning was conducted two weeks before the first unit. The posttest on students' hypothesis-based reasoning was conducted after the fourth learning unit, the last unit that differed between the control condition and the experimental condition. The students were assessed again in a follow-up test, the data from which are not considered here. The follow-up test took place three months after the intervention, and we do not believe that after such a long delay we can still reliably examine effects related to the research question in focus here, that is, interactions between the conditions and students' preconditions, because statistical power would be far too low."

The topic of the learning units was "the bouncing ball" (see Thiel, 1987), with the background of elasticity and plasticity. This topic was chosen to contextualize the intervention within a science topic that is appropriate for third-graders. However, the focus of the intervention was on fostering students' hypothesis-based reasoning with conditionals. All units included teacher-guidance and inquiry-based student activities in which the students worked in pairs to examine the truth value of conditional hypotheses. They first stated their own assumptions regarding each respective hypothesis, then tested the assumption empirically with objects that the hypotheses referred to, and then they were asked to draw inferences by referring back from the observed evidence to the initial hypothesis. In both conditions, the students received the same amount of time, the same set of research questions (conditional hypotheses), and materials to test these questions. Also, in both conditions, this task was structured by asking the students first to test a rather easy hypothesis

to start with, then all hypotheses that did not include negations, and finally hypotheses that included negations. In a pilot study, it was found that predetermining this general order of hypotheses for the students avoided that students got lost and did not manage to work adequately on any of the hypotheses. Thus, this structuring element was also part of the control condition, which makes it an active control condition with this structuring element of guidance.

The only difference between the two conditions were two kinds of scaffolds that the students in the experimental condition received in addition to the structuring element of the hypothesis order. These comprised structuring scaffolds and problematizing scaffolds (Reiser, 2004). The central element of the scaffolds was a wooden sorting box (Fig. 1). This wooden box was used as a static scaffold serving two aims. First, the wooden sorting box was aimed at helping the students to structure and visualize the complex hypothesis-based reasoning process (structuring scaffolds). Second, the teachers used the wooden box as a lead for adaptive scaffolds, asking cognitively activating questions. The box was aimed used to visualize and verbalize the reasoning process, relieving students' working memory by reducing cognitive load and supporting the inhibition of irrelevant information.

Both teacher and the students used such sorting boxes. In the beginning of the second unit, the teacher used such a wooden box to model the reasoning process, starting with focusing the premise before exploring the consequence of the examined assumption. In the following, the modeling procedure is described with the exemplary hypothesis "objects

Fig. 1 Wooden sorting box
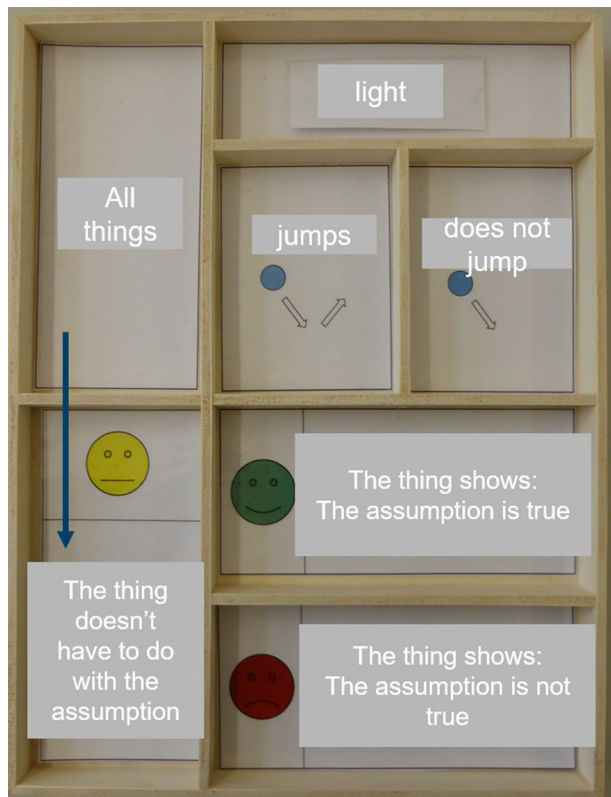
that are filled with air bounce." For this hypothesis, successful hypothesis-based reasoning would require a student to distinguish between relevant and irrelevant events (i.e., whether the object is filled with air; neglecting the event in case this premise is not fulfilled), and to distinguish between events that affirm (an air-filled object bounces) or refute (an air-filled object does not bounce) the hypothesis, and then based on this information to draw the correct inference. In the first step of the teacher modeling, all objects that should be tested were placed in the top-left field. A table tennis ball is chosen as the first object, and the teacher poses the question whether the ball fulfills the premise or not. In this case, the object is filled with air, which is why the table tennis ball runs through the right part of the box. The ball drops onto a field with a sign labeled "filled with air." Now, it is tried whether the ball bounces. It does, therefore, it is placed in the box with the green smiley, where it is indicated that the object verifies the assumption. Objects that refute the assumption of 'bouncing' are placed in the box with the red smiley. If the premise is not fulfilled (for example a wooden cube not filled with air), the object is sorted out in the box with the yellow smiley, indicating that the object can contribute no information about the truth of the assumption. The teacher demonstrates the process for using the sorting box, focusing students' attention toward essential thoughts such as "Is the object with air or not?", "What do you expect the object will do following your assumption?", "What do you see? Does the object bounce or not?", and "What does the object tell you about your assumption?". Following the teacher's demonstration, the students test their assumptions on their own using one sorting box for each pair of students. The teachers further support students in this process by continuing with structuring and problematizing questions in an adaptive manner. Specifically, the teacher observed all student pairs and when noting uncertainties or mistakes in the reasoning procedure, the teacher would support the students by posing such questions, fading this instructional support when students did not require it anymore to proceed further.

   To support the teachers in implementing the instruction in the control condition and in the experimental condition, three training courses lasting three hours each were supplied. Detailed teaching scripts as well as all teaching material were given to the teachers. The study leader went through the teaching script together with the teachers in detail. The courses and accompanying teaching scripts included instructions on how to deliver the training sessions, and how to work with the wooden sorting box and implement the cognitively activating hints and questions with school classes in the experimental condition. They were asked to align their lessons closely to the guidelines.

Statistical approach

In order to analyze the interaction of the two different intervention conditions with students' preconditions across these three scores representing hypothesis-based reasoning, we used two statistical modeling approaches. First, we estimated a latent transition analysis. This analysis allows extracting ability-profiles across the three scores, both at pretest and at posttest. These profiles represent subgroups of students that differ systematically from each other in their patterns of hypothesis-based reasoning ability (for a conceptual explanation of this analytic approach and its significance for educational science, see Hickendorff et al., 2018). This analytic approach was chosen because it offers a comprehensive depiction of students' learning patterns across the three outcome variables, and it is able to deal with ceiling effects stemming from high prior knowledge or instructional success (i.e., high mean scores after the instruction); in regular models such as analysis of variance, it would

not be possible to take these characteristics of the data into account in an informative manner. By modeling how students with different preconditions move between the ability profiles from before to after the instruction, this analysis can also capture linear and non-linear relations between the three hypothesis-based reasoning scores and students' preconditions.

In addition, from prior research it is known that hypothesis-based reasoning from irrelevant events (npq and npnq-conditionals) is a particularly challenging aspect for third-graders (Grimm et al., 2018). We estimated a second statistical model that provides readily interpretable visualizations regarding whether and how the scaffolding manages to promote students' learning of this key-aspect independently of their preconditions. Specifically, we estimated a general additive model (GAM; Vaci & Bilalić, 2017), a regression technique able to capture non-linear relations. These models can capture non-linear interactions between the intervention conditions and students' preconditions, thereby also handling potential ceiling effects. Adding to the information gained from the first analytic approach, the GAM provides informative visualizations regarding how students' preconditions affect the challenging learning of reasoning based on irrelevant conditionals (npq, npnq) in both intervention conditions. In this model, we included all relevant predictor variables at the same time. These included prior knowledge in reasoning based on irrelevant conditionals, inhibition ability, teacher-judged logical reasoning, intervention condition, and interactions of condition with the preconditions. Prior knowledge was only represented by students' pretest scores regarding irrelevant conditionals, since this was the same construct as the dependent variable in the analysis. The effects of the different student preconditions estimated in this model are controlled for the respective other variables.

In the following, we describe the relevant results from these two analytic approaches for our research questions, that is, how the scaffolding interacted (1) with students' prior knowledge, (2) with their inhibition ability, and (3) with their logical reasoning. After summarizing a preparatory psychometric evaluation of our items and presenting descriptive statistics, we describe the model fitting process of the latent transition analysis and the resulting student profiles, followed by general information on the model fitting results of the general additive model. Then, we present the focal results for our three research questions successively. We report the results from the latent transition analysis and the general additive model regarding (1) the effect of the scaffolding on the impact of students' prior knowledge on their learning patterns, then (2) the effect of scaffolding on the impact of students' inhibition ability, and finally (3) on that of students' logical reasoning.

Results

Psychometric evaluation and descriptive statistics

We analyzed six scores representing students' solution rates on the items representing hypothesis-based reasoning based on affirmative evidence (2 of the pq-items, 2 further items excluded due to ceiling effects; internal consistency estimated after Dunn et al., 2014: $Omega = 0.90$ at pretest, 0.82 at posttest), refuting evidence (the 4 pnq-items; $Omega = 0.68$ at pretest, 0.74 at posttest), and irrelevant evidence (the 8 npq- and npnq-items; $Omega = 0.94$ at pretest, 0.96 at posttest) at pretest and at posttest. The combining of the npq- and npnq-items into a single mean score was backed by a confirmatory factor analysis indicating that students' ability on these two types of items correlated perfectly ($r = 1.00$; model fit after combining both factors: $\chi 2(321) = 391.24$, $p < .001$,

Table 1 Descriptive statistics of main study variables per condition

	Experimental condition		Control condition		Condition difference
	M	SD	M	SD	Cohen's d +− 95%CI
Inhibition ability	51.60	17.33	51.64	18.64	0.00 [− 0.33; 0.34]
Logical reasoning	8.68	2.62	9.14	2.82	− 0.17 [− 0.50; 0.16]
Pretest affirmative	0.63	0.41	0.58	0.44	0.13 [−0.20; 0.48]
Posttest affirmative	0.84	0.84	0.74	0.40	0.29 [− 0.06; 0.62]
Pretest refuting	0.71	0.25	0.72	0.21	− 0.02 [− 0.35; 0.32]
Posttest refuting	0.75	0.25	0.73	0.27	0.05 [− 0.30; 0.40]
Pretest irrelevant	0.27	0.28	0.34	0.37	− 0.20 [− 0.54; 0.13]
Posttest irrelevant	0.71	0.32	0.54	0.38	0.49 [0.12; 0.85]
Affirmative gain	0.19	0.41	0.17	0.45	0.04 [− 0.31; 0.41]
Refuting gain	0.02	0.30	0.01	0.30	0.04 [− 0.34; 0.38]
Irrelevant gain	0.44	0.37	0.18	0.33	0.73 [0.34; 1.13]

Mean-scores on hypothesis-based reasoning variables represent percentage of solved items, on inhibition ability time in seconds (min. = 25, max. = 121), logical reasoning was rated on an adapted grading scale (min. = 4, max. = 14). Cohen's d estimates are indicated together with bootstrapped confidence intervals to correct for non-normality

Table 2 Intercorrelations of main study variables

	Inhibition	Reasoning	Pr_aff	Po_aff	Pr_ref	Po_ref	Pr_irr
Reasoning	0.26						
Pretest affirmative	0.09	0.18					
Posttest affirmative	0.27	0.27	0.39				
Pretest refuting	0.15	0.27	0.41	0.15			
Posttest refuting	0.16	0.42	0.28	0.39	0.24		
Pretest irrelevant	0.05	0.27	0.27	0.31	0.08	0.05	
Posttest irrelevant	0.22	0.14	0.27	0.37	0.06	0.10	0.44

Pr_aff Pretest affirmative, *Po_aff* Posttest affirmative, *Pr_ref* Pretest refuting, *Po_ref* Posttest refuting, *Pr_irr* Pretest irrelevant

$RMSEA = 0.04$, $CFI = 0.97$, no salient residual associations, see Greiff & Heene, 2017). Descriptive statistics of students' preconditions and the three mean scores representing hypothesis-based reasoning ability at pretest and posttest are provided in Table 1, and their intercorrelations in Table 2. The descriptive statistics, including effect sizes indicate that students in the control condition had slightly higher scores on reasoning based on irrelevant events at pretest, and that the experimental condition increased more strongly than the control condition on this aspect.

Table 3 Fit indices for the latent transition analysis with varying number of ability-profiles

Number of profiles	AIC	BIC	aBIC	Entropy
1	1073	1126	1069	NA
2	769	819	766	0.88
3	560	655	554	0.84
4	460	611	450	0.86

AIC Akaike information criterion, *BIC* Bayesian information criterion, *aBIC* sample-size adjusted Bayesian information criterion, *Entropy* reliability of profile separation
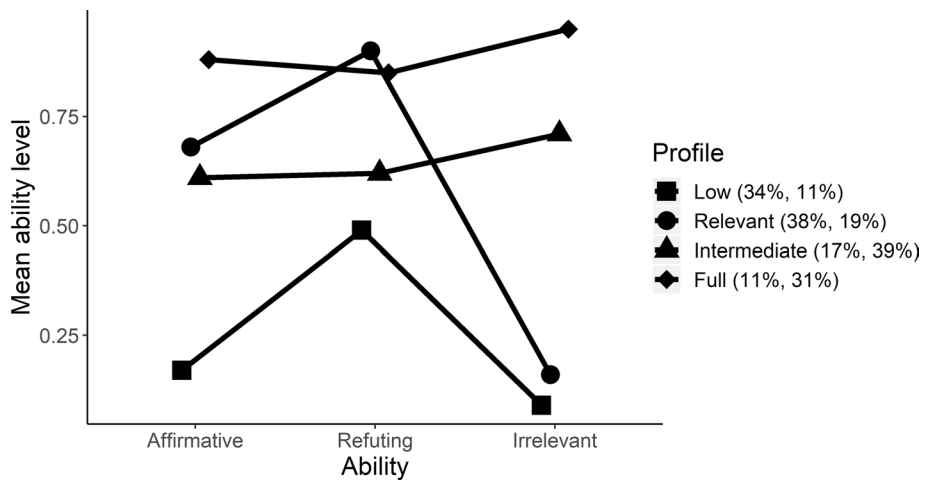

Fig. 2 The four student ability-profiles extracted in the latent transition analysis. Numbers in brackets indicate percentage of students showing the respective ability-profile at pretest and at posttest

Latent transition analysis: model fitting and resulting student profiles of hypothesis-based reasoning

In the latent transition analysis, the best model fit (Table 3) was achieved by extracting four different ability-profiles based on the data of students in both intervention conditions. Models with higher numbers of ability-profiles did not converge, and this model showed the lowest BIC, which is a reliable indicator of the correct number of systematic patterns (Nylund et al., 2007). After identifying the four student ability-profiles, the intervention condition, students' inhibition ability and logical reasoning were added to the model in order to examine the impact of these variables on students' learning patterns.

The four extracted ability-profiles are depicted in Fig. 2. The biggest profile at pretest described 38% of students (19% at posttest), who could successfully solve all items involving evidence from events relevant to the hypothesis; that is, these students had high ability levels on the affirmative and refuting items, but low ability levels on the irrelevant items. Consequently, this profile was labeled the *relevant* profile. The second biggest profile at pretest described 34% of students (11% at posttest). These students had the lowest ability levels on all three abilities and therefore this profile was labelled the *low* profile. A third

profile described 17% of students (39% at posttest) who showed moderate ability across all three variables (*intermediate* profile), and the fourth profile described 11% of students (39% at posttest) who could successfully solve items of all three kinds and therefore was labelled the *full* profile. To summarize, the profiles differed from each other as follows (see Fig. 2): In the *low* profile, students showed the lowest solution rates (below 50% of items) on all three kinds of items; in the *intermediate* profile, they showed moderate solution rates (above 50% but still far-from-perfect) on all three kinds of items; in the *relevant* profile, they showed moderate solution rates on items involving affirmative events (similar to the intermediate profile) and high solution rates (almost 100% solved) on items involving refuting events, but low solution rates (similar to the low profile) on items involving irrelevant events; and in the *full* profile, they showed the highest solution rates on affirmative and refuting (similar to the relevant profile) items, and this was also the only profile with high solution rates on the items involving irrelevant events.

General additive model: numerical results and relations of student preconditions with posttest-scores

The numerical results from the general additive model are presented in Table 4. A first noteworthy result, which however is not in the focus of this study, is the main effect of condition. Students in the experimental condition were estimated to reach a mean score of 0.2 points more (i.e., to solve on average 20% more items) than students in the experimental condition. In addition, Fig. 3 presents scatter plots between students' preconditions and their posttest scores on reasoning based on irrelevant events.

Table 4 Results of the general additive model

Model term	*Estimate*	*t*	*p*
Condition	− 0.2	− 3.9	<0.001
	edf	*F*	
Prior knowledge (experimental condition)	1.1.0	0.4	0.054
Prior knowledge (control condition)	1.93	4.39	<0.001
Inhibition (experimental condition)	0.95	1.99	<0.001
Inhibition (control condition)	0.00	0.00	0.429
Logical reasoning (experimental condition)	0.00	0.00	0.473
Logical reasoning (control condition)	0.00	0.00	0.653
Prior knowledge*inhibition (experimental condition)	0.97	0.09	0.176
Prior knowledge*inhibition (control condition)	0.38	0.03	0.309
Prior knowledge*logical reasoning (experimental condition)	0.41	0.06	0.190
Prior knowledge*logical reasoning (control condition)	3.07	0.96	0.001

Condition is the only linear parameter and therefore has a linear parameter estimate and *t*-value, while the other parameters are non-linear (*smooth*) parameters that have estimated degrees of freedom (*edf*; close to 0 indicates lack of effect, above 1 non-linear effect) and associated *F*-values. Condition was coded with a 0 for the experimental condition and a 1 for the control condition
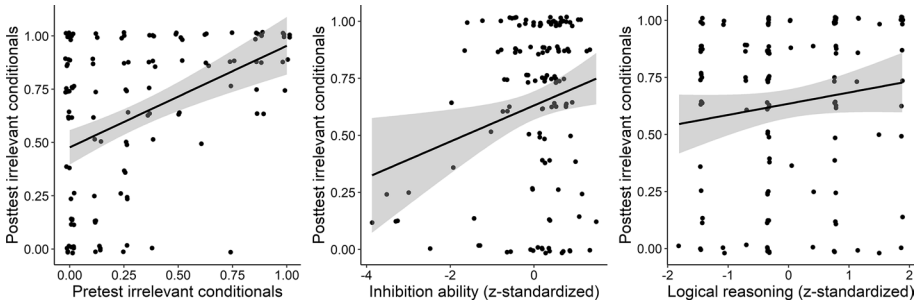
Fig. 3 Scatter plots between the three student preconditions (prior knowledge, left; inhibition ability, middle; logical reasoning, right) and their posttest scores on reasoning based on irrelevant events. Fit line indicates linear association with a 95% confidence interval band

The interaction of scaffolding with students' prior knowledge

In order to examine the first research question, concerned with the effect of scaffolding on the impact of students' prior knowledge on their learning patterns, we added the intervention condition to the latent transition analysis. This allowed examining whether and how the probability of transitioning from an initial profile at pretest into the same or another profile at posttest differed between the two intervention conditions. In other words, this analysis showed commonalities and differences between the two intervention conditions in students' learning patterns. In this analysis, students' prior knowledge is represented in their profiles at pretest. The result from this analysis is shown in Fig. 4.
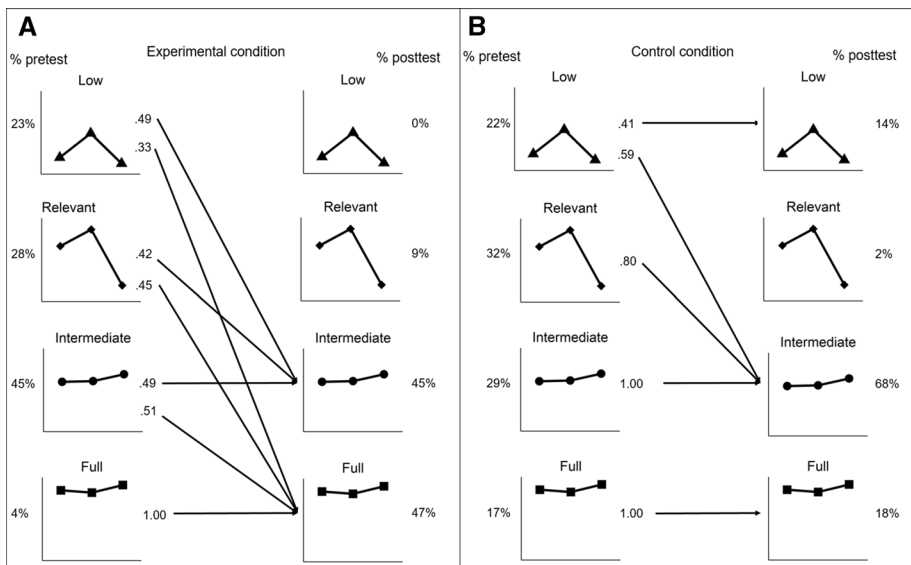
Fig. 4 Transition probabilities between four reasoning ability-profiles from pre- to posttest, for students in experimental condition **A** and in control condition **B**

Among the overall 70 students in the control condition (Fig. 4B), about half of those who started in the *low* profile moved into the *intermediate* profile after instruction (13%, $n=9$), whereas the remainder of these stayed within the *low* profile (9%, $n=6$). Also of those starting in the *relevant* profile, most (26%, $n=18$) moved into the *intermediate* profile. To summarize, in the control condition, students with low prior knowledge and those with prior knowledge about reasoning based on relevant events managed to acquire some ability in reasoning based on irrelevant events. Finally, all students (46%; $n=32$) who already showed the *intermediate* profile or the *full* profile before the instruction stayed in their initial profiles. Overall, of the 70 students in the control condition most (68%; $n=48$) students after the instruction reached the *intermediate* profile, whereas none reached a systematic transition from any other profile into the *full* profile.

Among the overall 73 students in the experimental condition (Fig. 4A), about half of those who started in the *low* profile moved into the intermediate profile after instruction (11%, $n=8$), whereas a third of these moved into the full profile (8%, $n=5$). Also of those starting in the *relevant* profile, almost half moved into the *intermediate* profile (12%, $n=9$), whereas the remainder of these mostly moved into the *full* profile (13%, $n=9$). To summarize, in the experimental condition, about half of the students with low prior knowledge or prior knowledge about reasoning based on relevant events managed to acquire some ability in reasoning based on irrelevant events, and most of the remaining students managed to acquire full understanding with all kinds of events. Finally, of the students who showed the intermediate profile at pretest, half stayed within this profile ( ) (46%; $n=32$) who already showed the *intermediate* profile or the *full* profile before the instruction stayed in their initial profiles. half of the students who started with rather high prior knowledge in the *intermediate* profile (16 out of 33) managed to acquire full ability on all aspects and move into the *full* profile. Students starting in the *low* ability- or *relevant*-profiles in similar parts managed to either acquire some ability on all aspects and move into the *intermediate* profile (8 resp. 9 students), or to master all aspects and move into the *full* profile (6 resp. 9 students). Overall, this comparison of the learning patterns in both conditions reveals that in the experimental condition, more students with little prior knowledge managed to learn, and only within the experimental condition students made the difficult step of mastering reasoning based on irrelevant events.

The estimates from the non-linear regression model regarding this research question are depicted in Fig. 5. Here, only students' showing in more detail how the intervention
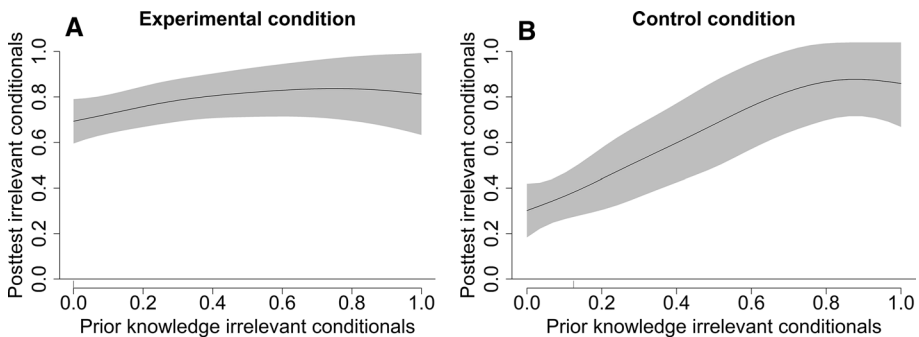

Fig. 5 The influence of students' prior knowledge about reasoning based on irrelevant events on their posttest ability, after they have undergone the experimental **A** or control **B** condition. 95% confidence intervals in grey

interacted with students' prior knowledge in the learning of the demanding aspect of reasoning based on irrelevant events.

The estimates from the non-linear regression illustrate that the two intervention conditions differently impacted the influence of students' prior knowledge on their learning of reasoning based on irrelevant events. In the control condition (Fig. 5B), a linear and positive association is visible, indicating that students' learning achievement depended quite strongly on their prior knowledge. In the experimental condition (Fig. 5A), the association is very weak; even students with little prior knowledge (below 0.5) managed to acquire substantial ability and achieved mean scores above 0.6 at posttest. This analysis together with the latent transition analysis illustrates that the experimental condition managed to decrease the limiting influence of students' prior knowledge on the challenging learning of reasoning based on irrelevant events.

The interaction of scaffolding with students' inhibition ability

To examine the second research question, how the intervention conditions affected the impact of students' inhibition ability on their learning patterns, inhibition was added to the latent transition analysis in addition to prior knowledge. The result from this analysis is shown in Fig. 6. There was clear effect of the intervention condition on the impact of students' inhibition ability for a very limited number of learning patterns. For learning patterns not included in Fig. 6, no relation between inhibition ability and transition probabilities was visible.

In the experimental condition (Fig. 6A), with higher inhibition ability students were less likely to stay in, or transition from the *low* into the *relevant* profile. Instead, students with higher inhibition ability more likely transitioned into the *full* profile. In the control condition (Fig. 6B), the opposite effect of inhibition ability was observed: Students with higher inhibition ability were more likely to stay in, or transition from the *low* into the *relevant* profile, instead of the *intermediate* profile. Integrating these results, in the control condition inhibition ability was a positive predictor of students' learning of reasoning based on relevant events but not based on irrelevant events, while in the experimental condition it was a positive predictor of acquiring ability on both types of events.

The estimates from the non-linear regression model (Fig. 7) further illustrate this result regarding the difficult aspect of reasoning based on irrelevant events: While in the control
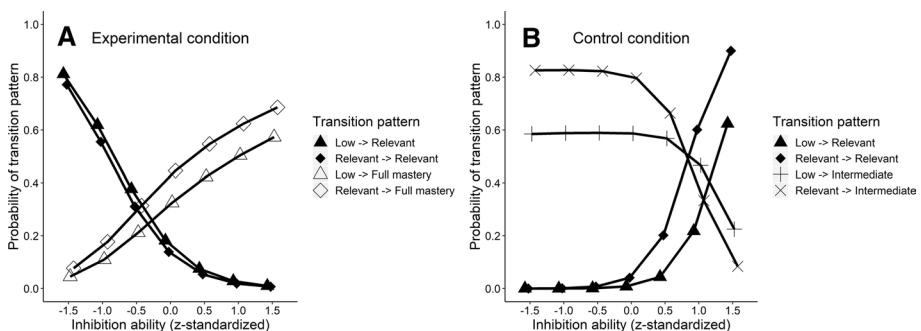

Fig. 6 The relation between students' inhibition ability and the probability to show a certain learning transition in the experimental **A** and control condition **B**
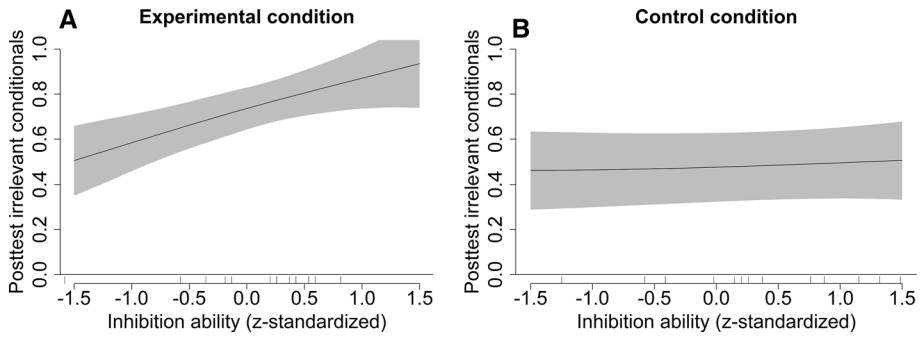
Fig. 7 The influence of students' inhibition ability (x-axis) on their posttest ability in reasoning based on irrelevant events (y-axis), after they have undergone the experimental **A** or control **B** condition. 95% confidence intervals in grey

condition (Fig. 7B) students' inhibition ability had no visible impact on the acquisition of this aspect, in the experimental condition (Fig. 7A) it was visible that inhibition ability had a positive impact on students' learning.

The interaction of scaffolding with students' logical reasoning

Teacher-judged logical reasoning was added to the latent transition analysis in addition to prior knowledge to examine the third research question, how the intervention conditions affected the impact of this variable on students' learning patterns. The result from this analysis is shown in Fig. 8. Again, for learning patterns not included in Fig. 8, no relation between logical reasoning and transition probabilities was visible.

In the experimental condition (Fig. 8A), only students with exceptionally low logical reasoning were likely to stay in, or transition from the *relevant* into the *low* profile. Instead, students with better logical reasoning stayed in, or transitioned from the *low* into the *relevant* profile. In the control condition (Fig. 8B), partially similar effects were observed: Students with lower logical reasoning were likely to stay in, or transition from the *relevant* into the *low* profile. Instead, students with high logical reasoning stayed in, or transitioned from the *low* into the *relevant* profile. In addition, students with average logical reasoning
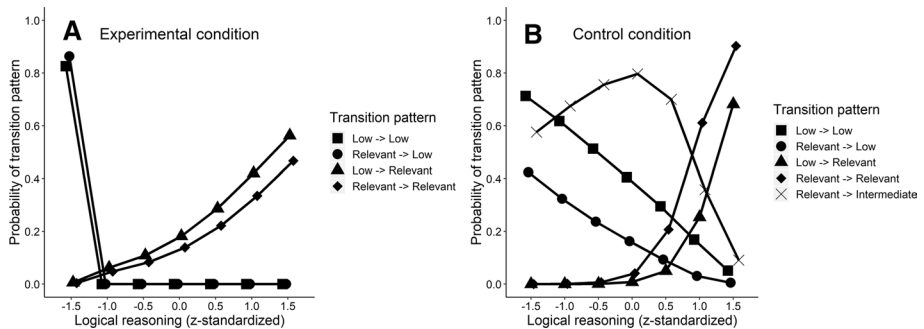

Fig. 8 The relation between students' logical reasoning and the probability to show a certain learning transition in the experimental **A** and control condition **B**

most likely transitioned from the *relevant* into the *intermediate* profile. Another visible difference was that in the experimental condition, the positive impact of students' logical reasoning was visible already below the mean, while students in the control condition managed to harness their logical reasoning if it was above average. Integrating these results, in both conditions logical reasoning was a positive predictor of students' learning of reasoning based on relevant events, in the experimental condition starting at low levels, and in the control condition especially at levels above average.

The estimates from the non-linear regression model (Fig. 9) further illustrate these results: In the experimental condition (Fig. 9A), logical reasoning had no substantial association with students' acquisition of reasoning ability based on irrelevant events. In the control condition (Fig. 9B), however, there was an interaction between students' prior knowledge and their logical reasoning (which required adding the additional dimension of color to this figure, in order to be able to represent this higher-order interaction visually): Among students with higher teacher-rated logical reasoning, only those with high prior knowledge managed to learn reasoning from the irrelevant events. Among students with lower teacher-rated logical reasoning, however, also those with little prior knowledge managed to learn reasoning from the irrelevant conditionals. This result indicates that teachers' ratings of students' logical reasoning have some slight negative predictive value regarding how well students with varying prior knowledge manage to learn reasoning based on the difficult irrelevant events. Contrary to teachers' ratings, even those with lower-rated logical reasoning managed to acquire newfound ability on this aspect under the control condition.

We conducted an additional descriptive analysis to further examine the surprising finding that some students with lower logical reasoning in the control condition apparently exhibited substantial learning gains on the irrelevant conditionals (orange area in Fig. 9B). We examined a histogram of posttest scores on the irrelevant conditionals for this subgroup of students with lower logical reasoning (more than $1.0SD$ below mean) and lower prior knowledge (mean score below 0.5 at pretest). This subgroup encompassed overall 25 students stemming from the two intervention conditions. The histogram confirms the surprising finding and illustrates its source: As visible in Fig. 10, there seems to be a clear distinction between students with lower preconditions who managed to master the irrelevant
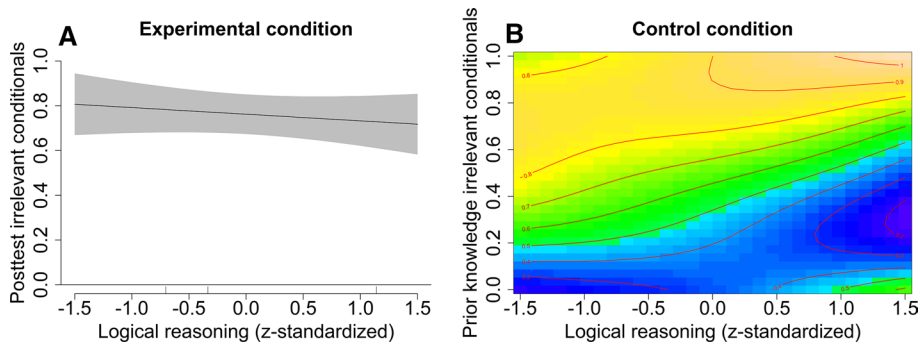

Fig. 9 The influence of students' logical reasoning on their ability in reasoning based on irrelevant events after they have undergone the experimental **A** or control **B** condition. For the experimental condition, students' posttest ability estimate is on the y-axis; in the control condition, the impact of logical reasoning depended on students' prior knowledge (y-axis); estimates of the outcome variable, students' ability on irrelevant events at posttest, is indicated in different colors (blue: low ability 0.0−0.4; green: moderate ability 0.4–0.7; orange: high ability 0.7−1.0)
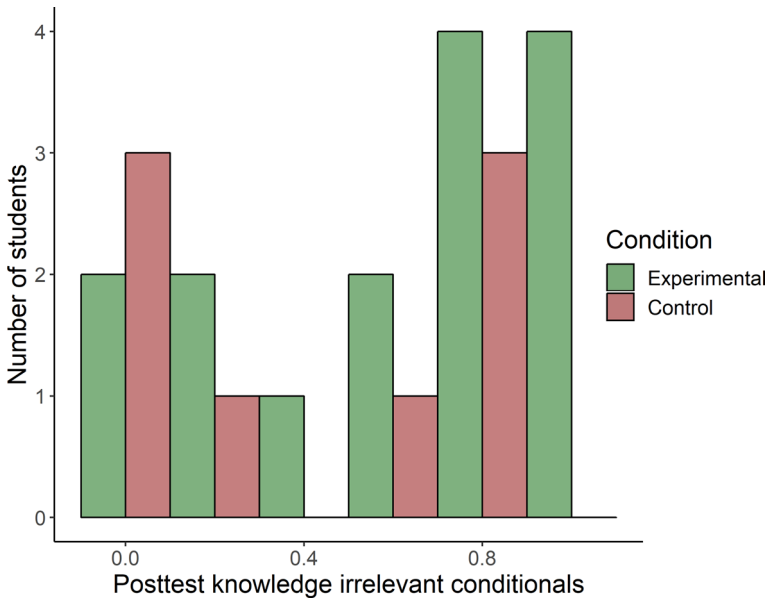
Fig. 10 A histogram of posttest knowledge on irrelevant conditionals (x-axis) for students with low prior knowledge and low logical reasoning. Y-axis indicates number of students with respective mean posttest knowledge, while different colors indicate the respective students' condition

conditionals, and those who did not. Half of the students in the control condition, despite low prior knowledge and low logical reasoning, managed to make this step, yielding strong learning gains on the irrelevant conditionals (Fig. 10: red bins with lower posttest-scores on left of x-axis vs. red bins with higher posttest-scores on right of x-axis). In the experimental condition, the amount of students making this difficult step even appeared to be bigger (Fig. 10: green bins on left vs. right of x-axis). The students in the control group making this difficult step apparently caused the non-linear regression model to produce the estimates depicted in Fig. 9B.

Discussion

We examined whether additional scaffolds manage to meet heterogeneity in third-graders' individual preconditions for acquiring hypothesis-based reasoning ability in an inquiry-based classroom-setting. Positive relations between students' levels of preconditions and probabilities to transition into proficient profiles of hypothesis-based reasoning indicate that in comparison to the control condition without additional scaffolds, in the experimental condition students managed to benefit from the instruction across a broader range of prior knowledge, inhibition ability, and teacher-judged logical reasoning. As visible from the latent transition analysis, under the experimental condition even students with lower preconditions managed to learn substantially. In addition, also students with higher preconditions benefitted from the scaffolding, particularly in making the most difficult step of understanding reasoning based on irrelevant conditionals. This was, for example, visible in the positive relation of inhibition ability with the probability to change into the *full*

profile from before to after the instruction. However, we also found that the scaffolding did not support the learning of all students, for example regarding teacher-judged logical reasoning. Following, we discuss these findings as well as the general learning patterns in the scaffolding-supported experimental condition and in the control condition. Then, we discuss the potential of our analytic approach, its limitations and more general limitations of our study, and implications for science education.

Scaffolding and the impact of students' prior knowledge

The results regarding the impact of students' prior knowledge confirm our hypothesis that the additional scaffolds in the experimental condition would diminish the impact of this precondition. Both analyses showed that particularly the students with low prior knowledge profited from the additional scaffolds.

There were some students that did not improve at all and stayed in the profile with the lowest overall ability level. This finding confirms the challenging nature of hypothesis-based reasoning for third-graders (Robisch et al., 2014). This reasoning skill is not trivial to learn, stressing the need for promoting this ability in inquiry-based science learning. But the results also show that the instructional support provided for acquiring the ability of hypothesis-based reasoning can be successful. Comparing the movements between the profiles, a substantial number of students in the experimental condition moved into the *full* profile, managing to master all facets of hypothesis-based reasoning, rather independently of their profile at pretest. In the control condition, students only reached the *intermediate* profile, with partial ability on all facets. Only those from the control condition who already had been in the *full* profile stayed there. This finding proves the success of implementing additional scaffolds in the experimental condition and confirms the findings of Robisch et al. (2014) regarding the possibility of promoting hypothesis-based reasoning in third-graders. We can conclude that the additional scaffolds in the experimental condition removed the disadvantage of students with lower prior knowledge. By this we can assume that the influence of students' prior knowledge as a major determinant of learning (Simonsmeier et al., 2022) on the learning outcome is decreased by the implementation of scaffolding.

Another aspect concerning prior knowledge that we would like to mention is that the estimates of internal consistencies of the assessment of hypothesis-based reasoning were rather high. This indicates little contextual dependence: Regardless whether students have to apply a specific hypothesis-based reasoning-skill in situations involving for example balloons or balls, they tend to achieve similarly across contexts. We believe that this does not necessarily indicate the absence of influence on the context on children's reasoning, which would be in contrast to the general finding that scientific reasoning-skills function context-dependent (e.g., Opitz et al., 2021). Rather, we assume that children who managed to overcome their own content knowledge and instead stick to the aim of the task (i.e., evaluating the evidential value independently of their own content knowledge) managed doing so similarly across contexts.

Scaffolding and the impact of students' inhibition ability

Regarding the interaction of the additional scaffolds with students' inhibition ability, the students of the control condition only achieved partial acquisition of hypothesis-based reasoning, whereas the students of the experimental condition achieved full acquisition

of hypothesis-based reasoning. The implementation of scaffolding seems to increase the impact of inhibition ability on students' potential to learn hypothesis-based reasoning, by enabling the optimal use of each student's potential. This means that in the experimental condition, there was no disadvantage for students with lower inhibition ability, but a special benefit for students with higher inhibition ability.

In both conditions, mastering reasoning based on relevant events was learned quite independently of students' inhibition ability. However, in the experimental condition, students with lower inhibition ability managed to master reasoning based on relevant events, while those with high inhibition ability in addition mastered the difficult aspect of reasoning based on irrelevant events. Hence, implementing additional scaffolds supported all students in learning by exploiting their inhibition ability. That high inhibition ability enables to reach the *full* profile confirms the findings of Handley et al. (2004), claiming inhibition ability to be a strong predictor for reasoning on belief-based problems. Inhibition ability supports the process of decontextualizing the reasoning from students' own beliefs (Handley et al., 2004). The findings of the present study fit well in this context, because in the intervention the students had to reason about their own alternative concepts and in the end had to refute their initial, potentially wrong assumptions.

In the experimental condition, the students with high inhibition ability did not only learn the refuting conditionals particularly well, but with transitioning into the *full* profile, they also learned to reason about the irrelevant conditionals. It is important to find such influential factors, because reasoning about irrelevant conditionals is especially difficult to learn for children (Gauffroy & Barrouillet, 2011). Inhibition ability might be helpful for learning to reason about irrelevant conditionals because the underlying cognitive process requires full focus on the antecedence first, and then full neglect of the consequence. In the experimental condition, the teacher modelled the process to focus the antecedence first, and to decide whether an event carries relevant information for the hypothesis, or not. Learning from this model, the students of the experimental condition might have been able to utilize their inhibition ability by realizing the modelled stepwise approach.

In the control condition, inhibition ability helped students to learn reasoning based on relevant events and thus reach the *relevant* profile. One explanation for this finding is that inhibition ability is especially necessary for recognizing refuting evidence, because in this case one's own beliefs have to be inhibited (Handley et al., 2004).

Overall, the students in the experimental condition profited to a higher extent from their inhibition ability than the students in the control condition, particularly in the demanding learning of reasoning based on irrelevant events. This can be explained by looking at the two core functions of reasoning proposed by Handley et al. (2004), which are the capacity of working memory and inhibition ability. In the experimental condition, the explicit scaffolding aims at relieving working memory, which enables to exploit inhibition ability. In the control condition, children's working memory might be at its limits, thwarting utilization of inhibition ability. The implemented scaffolding in the experimental condition seems to provide a kick for those students with high inhibition ability to reason about irrelevant conditionals. Without receiving the stepwise modelling of identifying irrelevant conditionals in the control condition, inhibition ability seems not to be able to impact the learning achievement regarding reasoning based on irrelevant events.

Scaffolding and the impact of students' logical reasoning

In both groups, an effect of scaffolding on the impact of students' logical reasoning on their learning was observed. Students with higher logical reasoning were more likely to master reasoning based on relevant events and thus reach the *relevant* profile. In the experimental condition, this effect appeared for all students except for those with very low logical reasoning, whereas in the control condition this effect appeared only for students with above-average logical reasoning. These findings suggest that the implementation of scaffolding diminishes the disadvantage of students with lower logical reasoning. In the experimental condition, students with lower logical reasoning could also reach the *relevant* profile, which was not the case in the control condition. Hence, in the experimental group students with a broader range of logical reasoning could profit from the intervention. A small group of students with very little logical reasoning who ended up in the *low* profile indicates again the high challenge of learning hypothesis-based reasoning in elementary school (Gauffroy & Barrouillet, 2011; Robisch et al., 2014). The question, however, remains whether those students are still too young to learn hypothesis-based reasoning, or whether they would be able to profit from instruction with increased adaptive support.

The non-linear regression model showed that in the control condition, some students with lower logical reasoning tended to show higher learning gains on irrelevant conditionals than those with higher logical reasoning. We were able to explain this result by finding in a histogram that half of these students in the control condition, despite lower logical reasoning, managed to make the difficult step of learning the irrelevant conditionals. This is surprising, because we assumed that logical reasoning would be especially important for learning the difficult irrelevant conditionals. On the other hand, supporting our hypotheses, the histogram shows that in the experimental condition, even more students from this group with lower logical reasoning managed to make this difficult step. Thus, it seems that although against our expectations some students with lower logical reasoning managed to make this difficult step, in accordance with our expectations this was the case for more students who received the additional scaffolds. Due to the limited sample size within this subgroup, this finding should be interpreted with care and replicated and further explored in future research.

Potential of the analytic approach

In this study, we selected a combination of two rather rarely seen analytic approaches in educational research: A latent transition analysis, and a general additive model. In our case, these approaches had advantages that we believe allowed insights which would be hard to yield with more common approaches such as analysis of variance and other general linear models. The latent transition analysis allowed modeling multivariate profiles of students' abilities in hypothesis-based reasoning. In more common univariate and purely linear models, relations between the three variables (e.g., students proficient with both situations involving relevant evidence, but not with irrelevant evidence) would be lost, and they would also be hardly visible in multivariate extensions. In addition, there were some students who already showed high levels of hypothesis-based reasoning before the instruction, and more students with high levels after the instruction. In regular linear models, this would lead to ceiling effects, causing biased parameter estimates and difficulties with their interpretation. In the latent transition analysis, the very proficient students received their own profile, which elegantly handled this issue: The model separated these students from

the rest, allowing informative interpretations of these and the other students' learning pathways from before to after the instruction. The general additive model similarly handled this issue; by capturing non-linear relations, ceiling effects were covered by a kink in the regression line for students approaching the maximum score.

Limitations

Despite the advantages of our analytic approaches, one limitation is the limited sample size. Although latent transition analyses have delivered informative insights based on similar sample sizes in educational research before (e.g., Schneider & Hardy, 2013), the present results should be interpreted with caution, particularly those based on smaller numbers of students (e.g., concerning the smaller student profiles).

Another limitation is the small number of teachers who provided the instruction in the present study. We tried to balance teacher effects by having each teacher instruct under both conditions. However, the teachers might have been influenced in their teaching by their own hypotheses regarding the benefits and drawbacks of the two conditions.

Finally, in the present study we looked into specific aspects of cognitive preconditions that we regarded of central importance for acquiring hypothesis-based reasoning. Further factors that we suggest to take up in future studies are additional student experiences beyond prior knowledge in hypothesis-based reasoning, such as their experience with inquiry-based learning settings and their expertise in further inquiry activities. In addition, the teacher judgments of students' logical reasoning might be replaced or extended with objective measures. Objective measures lack the subjective component of a teacher judgment and take more time for implementation, but they would improve validity from the perspective of yielding a pure measure of students' cognitive abilities.

Implications for science education

Commonly, inductive reasoning is taught in scientific inquiry teaching; deductive reasoning seems to be more often neglected. Despite accumulating research indicating that even young children can reason scientifically (Sandoval et al., 2014; Sodian et al., 1991; Piekny et al., 2013), in accordance with Barrouillet et al. (2008) our findings indicate that third-graders commonly fail on rather simple tasks requiring hypothesis-based reasoning. This was particularly the case on tasks involving irrelevant events, before they had received the related instruction. At the same time, the results after our intervention indicate that students can also substantially improve in hypothesis-based reasoning with conditionals. This result encourages the fostering of not only inductive but also deductive reasoning strategies, in order to support the holistic implementation of scientific literacy.

The instructional scaffolds developed in this study may serve as an example for fostering hypothesis-based reasoning also in other scientific contexts, for example within the topic of floating and sinking. We would like to emphasize the importance of taking care to relieve children's working memory and to support inhibition. This can be achieved by structuring the mental process in substeps: First focusing the premise, and then considering the consequence. As our example shows, a sorting box can provide a helpful visualizing tool. It might be helpful to increase the level of difficulty by first working with statements without negation. Instructional support can be provided by modelling by the teacher and does not necessarily require individual adaptation. Fading of the support might be useful for students with high cognitive preconditions.

Overall, we can conclude that the implementation of such scaffolding measures leads to a higher benefit for all students independently of their preconditions than without these scaffolds that were provided in addition to structuring of the difficulty regarding the inquiry process. The extent of individual adaptations in the present study was kept low against the background of several different needs in teaching heterogeneous classes. Still, the cognitive activation and structuring of the tasks seemed to be able to meet students' heterogeneous preconditions. While it remains important that the essential elements of scaffolding like individual fading of support are kept upright (Puntambeker & Hübscher, 2005), such individual adaptations might further increase learning outcomes, but they are very costly to implement (Tomlinson et al., 2003). The present study shows that implementing scaffolding carefully adapted to the teaching context as an overall strategy of high teaching quality already contributes substantially to the optimal use of students' developmental potential.

References

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*, 115–129

Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*, 760–772

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*, 289–302.

Brush, T. A., & Saye, J. W. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning, 1*, 1–13.

Bybee, R. (2002). Scientific literacy - mythos oder realität? In W. Gräber, P. Nentwig, T. Koballa, & R. Evans (Eds.), *Scientific literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (pp. 21–43). Wiesbaden: VS Verlag für Sozialwissenschaften.

Corno, L. (2008). On teaching adaptively. *Educational Psychologist, 43*, 161–173.

Cronbach, L. J., & Snow, R. (1977). *Aptitudes and instructional methods*. New York: Irvington.

Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, *86*, 109–121

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168

Dumont, H. (2018). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *103*, 1–29

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review*, *29*, 249–282

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental psychology*, *47*, 1000–1011

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, *33*, 313–317

Grimm, H., Robisch, C., & Möller, K. (2018). Förderung hypothesenbezogener Schlussfolgerungen im naturwissenschaftlichen Sachunterricht durch gezieltes Scaffolding—Gelingt dies unter Feldbedingungen? *Zeitschrift für Grundschulforschung, 11*, 349–363.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, *10*, 175–195

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge

Hertel, S. (2014). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. In B. Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E. M. Kirschhock, & G. Ranger (Eds.), *Individuelle Förderung und Lernen in der Gemeinschaft* (pp. 19–34). Wiesbaden: Springer VS

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, *66*, 4–15

Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan (Ed.), *Scaffolding student learning. Instructional approaches and issues* (pp. 75–107). Cambridge: Brookline Books.

Jansen, S., Mannhaupt, G., Marx, H., & Skowronek, H. (1999). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678

Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, *34*, 71–96

Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, *86*(1), 327–336

Kyllonen, P. C., & Lajoie, S. P. (2003). Reassessing aptitude: Introduction to a special issue in honor of Richard E. snow. *Educational Psychologist, 83*, 79–83.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning. *Review of Educational Research, 86*, 681–718.

Lederman, N., Lederman, J., & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *International Journal of Education in Mathematics Science and Technology, 1*, 138–147.

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, *19*, 85–103

Martin, N. D., Tissenbaum, D., Gnesdilow, C., & Puntambekar, S. (2018). Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science, 53*, 1–30.

NGSS (2013). Next generation science standards for today's students and tomorrow's workforce (second draft; Appendix F). Washington: Archieve, Inc. with NRC, NISTA, AAAS. Retrieved from https://www.nextgenscience.org/ Accessed 16 April 2020

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*, 535–569

Opitz, A., Heene, M., & Fischer, F. (2021). Using differential item functioning to analyze the domain generality of a common scientific reasoning test. *European Journal of Psychological Assessment, 38*(4), 251–260. https://doi.org/10.1027/1015-5759/a000662

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*, 423–451

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109–119.

Piekny, J., Grube, D., & Maehler, C. (2013). The relation between preschool children's false-belief understanding and domain-general experimentation skills. *Metacognition and Learning, 8*, 103–119.

Popper, K. R. (1972). *Objective knowledge*. Oxford: Oxford University Press

Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211–227

Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*, 1–12

Randi, J., & Corno, L. (2005). Teaching and learner variation. *British Journal of Educational Psychology*, *3*, 47–69

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13*, 273–304.

Robisch, C., Tröbst, S., & Möller, K. (2014). Hypothesenbezogene Schlussfolgerungen im Grundschulalter fördern. *Zeitschrift für Grundschulforschung*, *2*, 88–101

Roll, I., Butler, D., Yee, N., Welsh, A., Perez, S., & Briseno, A. (2018). Understanding the impact of guiding inquiry: the relationship between directive support, student attributes, and transfer of knowledge, attitudes, and behaviours in inquiry learning. *Instructional Science*, *46*, 77–104

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist, 49*(2), 139–152. https://doi.org/10.1080/00461520.2014.917589

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*, 102–119

Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, *49*, 1639–1649

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science*, *38*, 315–323

Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, *59*, 84–95

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, *39*, 37–63

Scott, C. (2013). The search for the key for individualized instruction. In J. Hattie & E. M. Andermann (Eds.), *International guide to student achievement. Educational psychology handbook series* (pp. 385–388). New York, NY: Routledge.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454–499.

Seiz, J., Decristan, J., Kunter, M., & Baumert, J. (2016). Differenzielle Effekte von Klassenführung und Unterstützung für Schülerinnen und Schüler mit Migrationshintergrund. *Zeitschrift für Pädagogische Psychologie*, *30*, 237–249

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, *57*, 31–54

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, *62*, 753–766

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal*, *7*, 935–947

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*, 743–762

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education*, *67*, 291–301

Thiel, S. (1987). Wie springt ein Ball? *Grundschule*, *1*, 18–23

Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology, 9*, 1339.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., & Brimijoin, K. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted, 27*, 119–145.

Vaci, N., & Bilalić, M. (2017). Chess databases as a research vehicle in psychology: Modeling large data. *Behavior research methods*, *49*, 1227–1240

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction. A decade of research. *Educational Psychology Review, 22*, 271–296.

Vergauwe, E., Gauffroy, C., Morsanyi, K., Dagry, I., & Barrouillet, P. (2013). Chronometric evidence for the dual-process mental model theory of conditional. *Journal of Cognitive Psychology*, *25*, 174–182

Vorholzer, A., Aufschnaiter, C., & Boone, W. J. (2018). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research Science Education, 103*, 1–27.

Vosniadou, S. (2019). The development of students' understanding of science. *Frontiers in Education, 4*, 32.

Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, *43*, 365–379

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17*, 89–100

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.