DEVELOPMENT OF STUDENT UNDERSTANDING: FOCUS ON SCIENCE EDUCATION

EDITED BY: Calvin S. Kalman and Mark Lattery

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

DEVELOPMENT OF STUDENT UNDERSTANDING: FOCUS ON SCIENCE EDUCATION

Topic Editors:
Calvin S. Kalman, Concordia University, Canada
Mark Lattery, University of Wisconsin–Oshkosh, United States

Table of Contents

Editorial: Development of Student Understanding: Focus on Science Education

Calvin S. Kalman[1]* and Mark Lattery[2]

[1] Department of Physics, Concordia University, Montreal, QC, Canada, [2] Department of Physics, University of Wisconsin–Oshkosh, Oshkosh, WI, United States

Editorial on the Research Topic

Development of Student Understanding: Focus on Science Education

How can we engage a broad audience of science education researchers and practitioners to examine strategies to help students become more expert-like in their thinking? To succeed in a technologically evolving society, students must engage in critical thinking, collaborative problem solving, and evidence-based reasoning. What specific kinds of interventions are needed to assist students with varying epistemologies to attain these skills?

Many students see scientific knowledge as unconnected and conveyed by authorities, such as the instructor and the textbook; correspondingly, their own knowledge structure is fragmented and disordered—a "knowledge in pieces" (KIP) as diSessa (1983). However, many other students enter the classroom with semi-coherent and relatively stable alternative conceptions about how the world works, and also an instinct for the nature of science or scientific knowledge; e.g., students "are authentic and creative scientific modelers" (Lattery, 2017, p. 109). Whether student scientific knowledge is best characterized as a fragmented or coherent, the instructor is confronted with the difficult task of bridging student's prior knowledge with target ideas. The task is especially challenging if the student's ideas are profoundly different ("incommensurable") with target ideas. Chi (2013) noted that many concepts in student's initial flawed mental models are not transformed to the accepted scientific model despite repeated corrections or patchings of the underlying rules.

We launched this ebook to consider instructional supports that are necessary for students to examine and develop their own ideas and compare them to the ideas presented by peers, the textbook, and the instructor. This is a follow up to our previous review of three instructional strategies that show promise to address this challenge in the context of an introductory physics classroom (Kalman and Lattery, 2018). More details are also found in Kalman (2017).

In this Research Topic, ten articles touch on various aspects of helping students become more expert-like in their thinking. Four articles were submitted through *Frontiers in Education STEM Education* and six articles through *Frontiers in Psychology Educational Psychology*.

In her article, Vosniadou directly addresses the structure of students' knowledge. She cites arguments in the research literature that children start the knowledge acquisition process by forming beliefs based on their everyday experiences and lay culture. In her view "the development of science knowledge is a long and gradual process during which students use constructive learning mechanisms to assimilate new, scientific, information into their prior knowledge causing hybrid conceptions—or misconceptions. Science instruction needs to help students become aware of their experience-based beliefs that might constrain science learning causing misconceptions, provide information gradually based on students' learning progressions and develop students' scientific reasoning and executive function skills."

Hadzigeorgiou and Schulz's article is part of an extended research project investigating how to improve secondary students' motivation and engagement to learn about science. This article focuses on students' "narrative mode of thought" as a bridge to understanding science.

Seufert notes that learning with text and pictures requires learners to integrate the given information into one coherent mental representation. "Since learners often fail to integrate text and pictures, the study investigates the effects of a training for text processing strategies, picture processing strategies and strategies to map text and picture onto each other."

Kerwer and Rosman examines the dependence of epistemological change on the (un)resolvability of contradictory information, the extent to which explicit reflection on diverging information supports epistemic change, and how topic-specific diverging information affects topic- and domain-specific epistemic beliefs.

Zhao et al. show that information that displays more concrete characteristics exerted a greater cognitive inhibitory effect during the working memory task, and a greater cognitive inhibitory effect was produced when all of inhibition retrieval information clues are provided than when none of the clues are provided in the working memory task."

Kaiser and Mayer investigate the benefits of combining example-based learning with physical, hands-on investigations in inquiry-based learning for acquiring scientific reasoning skills.

Four papers concentrate on students' conceptual understanding. Nunez-Oviedo and Clement focus is on how whole class discussions can contribute to the learning of conceptual models in science. As they point out, "Science educators today still struggle with finding better ways to help students develop strong conceptual understandings as opposed to memorizing isolated facts." "It is possible to start from student-generated models that conflict with the target model in a number of ways, and still arrive at the target model for the lesson through discussion."

Han and Ellis describe how the phenomenographic method can be used to develop students' conceptual understanding of scientific concepts, to inform effective instructional design in science teaching, and to identify and improve evidence-based factors in student learning to enhance learning outcomes in science.

Munoz-Rubke et al. consider how learning formal concepts becomes more meaningful when teachers integrate what children already know and also underscore that spatial abilities have a strong and positive effect both on the motivation to learn math and on math performance itself.

Bigozzi et al. use a semi-structured interview to question faculty about their ideal teaching approach and their actual teaching approach. They also examined which component of the teaching approach is associated with students' progress in physics and critical thinking skills. The authors note that "simply going to the laboratory does not foster a constructivist learning in students, unless it is matched with reflection."

This collection of papers will hopefully engage a broad audience to extend the results presented by the authors of the articles found in this ebook to find additional ways to help students become more expert-like in their thinking.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Chi, M. T. H. (2013). "Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes," in *International Handbook of Research on Conceptual Change, 2nd Edn*, ed S. Vosniadou (New York, NY: Routledge Press), 49–70.

diSessa, A. (1983). "Phenomenology and the evolution of intuition," in *Mental Models,* eds D. Gentner and A. Stevens (New York, NY: Erlbaum), 15–34.

Kalman, C. S. (2017). *Successful Science and Engineering Teaching: Theoretical and Learning Perspectives, 2nd Edn*. New York, NY: Springer.

Kalman, C. S., and Lattery, M. J. (2018). Three active learning strategies to address mixed student epistemologies and promote conceptual change. *Front. ICT* 5:19. doi: 10.3389/fict.2018.00019

Lattery, M. J. (2017). *Deep Learning in Introductory Physics: Exploratory Studies of Modeling-Based Reasoning* Charlotte, NC: Information Age Publishing, 279.

Mechanisms of Epistemic Change—Under Which Circumstances Does Diverging Information Support Epistemic Development?

Martin Kerwer* and Tom Rosman

Leibniz Institute for Psychology Information (ZPID), Trier, Germany

Background: The number of studies on how to foster change toward advanced epistemic beliefs (i.e., beliefs about the nature of knowledge and knowing) is continuously growing because these beliefs are an important predictor of learning outcomes. In past intervention studies, presenting diverging information (e.g., descriptions of studies yielding contradictory results) reliably led to epistemic change. However, prior research insufficiently examined which aspects of diverging information affect these changes.

Aims: We investigated (1) if epistemic change differs depending on the (un)resolvability of contradictory information, (2) to what extent explicitly reflecting on diverging information supports epistemic change and (3) how topic-specific diverging information affects topic–and domain-specific epistemic beliefs. All confirmatory hypotheses were preregistered at OSF. Additionally, several exploratory analyses were conducted.

Method: To examine the research questions, we employed a simple randomized pre-post design with four experimental groups. $N = 185$ psychology students participated in the study. Experimental groups differed in the kind of diverging information included: Students either read (1) information on students applying learning strategies (control), (2) unresolvable, or (3a) resolvable controversial information on gender stereotyping. In the latter condition (3b), an additional group of participants deliberately resolved apparent contradictions in a writing task.

Results: Confirmatory latent change analyses revealed no significant group differences in epistemic change (i.e., beliefs in the control group also changed toward advanced epistemic beliefs). Using a different methodological approach, subsequent exploratory analyses nevertheless showed that presenting diverging information on gender stereotypes produced stronger topic-specific epistemic change and change in justification beliefs in the treatment groups in contrast to the control group. However, effects in the treatment groups did not differ significantly depending on the resolvability of presented controversies or for the group which was instructed explicitly to integrate controversial findings.

Conclusion: Contrary to our expectations, diverging information seems to foster epistemic change toward advanced beliefs regardless of the resolvability of presented

information, while no final conclusion concerning effects of reflection could be drawn. Moreover, our findings indicate that effects of topic-specific interventions are more pronounced on topic-specific measures. However, this relationship may vary depending on the epistemic belief dimension (e.g., justification beliefs) under investigation.

INTRODUCTION

Epistemic beliefs are conceptualized as an individual's beliefs about the nature of knowledge and knowing (Hofer and Pintrich, 1997). Even though a long tradition of interdisciplinary research on the predictors and effects of epistemic beliefs exists (Hofer and Pintrich, 1997; Greene et al., 2008, 2018; Chinn et al., 2011), interventions that aim to promote epistemic change are relatively rare (cf. Muis et al., 2016). Recently, however, interest in epistemic change surged (Kienhues et al., 2016; Muis et al., 2016; Barzilai and Chinn, 2017). This may, at least partially, be due to the fact that these beliefs have been repeatedly shown to affect how individuals deal with crucial requirements of a modern knowledge-based society, such as acquiring and evaluating knowledge (Kienhues et al., 2016; Strømsø and Kammerer, 2016). Accordingly, quasi-experimental and correlational studies point toward beneficial effects of advanced epistemic beliefs (e.g., beliefs that knowledge claims have to be weighed and evaluated) for information integration (Barzilai and Ka'adan, 2017) and sourcing (Bråten et al., 2014), while more naive types of beliefs tend to impair the performance in such tasks (e.g., Kammerer et al., 2015; Rosman et al., 2016b). In this context, the term *naive beliefs* embraces views that (1) knowledge claims can only be either true or false, or (2) the conception of knowledge as purely tentative and subjective (Kuhn et al., 2000). In line with these ideas, a recent meta-analysis by Greene et al. (2018) confirmed that epistemic beliefs are positively correlated with academic achievement, which further corroborates the importance of (fostering) those beliefs.

To allow for future intervention studies to shape individuals' epistemic development in a more efficient way, our research aims to contribute to a better understanding of the underlying mechanisms of change. In this article, we start by briefly introducing popular developmental models for epistemic beliefs, as well as established models on epistemic change and models on the domain-specificity of epistemic beliefs. Thereafter, we review recent approaches for changing epistemic beliefs in (quasi-) experimental settings, focusing on the presentation of diverging information as an especially promising method. Bringing together these theoretical perspectives, we identify three essential and unsettled research questions that relate to properties of diverging information and the domain-specificity of both the presented information and the beliefs under investigation. Subsequently, we introduce an experimental study that addresses these research questions by examining psychology students' epistemic beliefs on gender stereotyping in secondary schools. Finally, after presenting the study's results, we discuss its

implications for both future research on epistemic change and for the design of interventions that target epistemic change.

Developmental Models on Epistemic Beliefs

How are changes in epistemic beliefs thought to take place in non-experimental settings throughout an individual's lifespan? Most developmental models for describing epistemic change strongly rely on Piagetian ideas introducing *cognitive disequilibrium* as the driving force behind epistemic development (Hofer and Pintrich, 1997). More specifically, these models assume that cognitive disequilibria occur if new information contradicts previously acquired beliefs. For example, belief change may occur when math students realize that there is more than one way to solve problems in mathematics. Again typically Piagetian, almost all established developmental models postulate that epistemic development unfolds in distinct stages. In this study, we draw on the popular model of Kuhn et al. (2000), who propose a stage model that differentiates three stages of epistemic beliefs: Individuals start as *absolutists*, believing that knowledge is certain and that an objective truth exists. They then proceed to *multiplism*, whose characteristic aspect is that knowledge is seen as inherently subjective. The final and most advanced stage is called *evaluativism*, where individuals acknowledge the importance of weighing evidence and integrating contradictory knowledge claims. In our opinion, this does not imply that evaluativists deny the existence of certain knowledge. For example, an evaluativist may argue strictly in favor of vaccination if there is sufficient evidence to support its efficacy. Additionally, in a modern society with divided knowledge, advanced beliefs may also involve acknowledging one's knowledge gaps, identifying trustworthy external authorities that address these gaps (e.g., the World Health Organization for health issues), and relying on the information provided by them (Bromme et al., 2010). According to Kuhn et al. (2000), individuals successively progress from absolutism over multiplism to evaluativism in their epistemic development (although not all individuals reach the last stage). On a more fine-grained level, one may additionally characterize these rather broad stages on a set of dimensions so-called integrative models (e.g., Bendixen and Rule, 2004; Merk et al., 2018) with *certainty*, *simplicity*, *justification* and *source* of knowledge being the most prominent ones (Hofer and Pintrich, 1997). However, it should be of note that Greene et al. (2008) challenge this view by arguing that some of those dimensions, such as *simplicity* of knowledge, relate to an individual's ontological beliefs and not to their epistemic beliefs. Therefore, they suggest focusing on justification beliefs as "truly"

epistemic beliefs that determine under which circumstances individuals obtain knowledge. For this purpose, Greene et al. (2008) introduced two dimensions of justification beliefs – *justification by authority* (e.g., individuals justify knowledge claims based on experts) and *personal justification* (e.g., justification of knowledge claims based on personal experience). Subsequently, Ferguson et al. (2012) extended this framework by adding a third scale, *justification by multiple sources*, whose importance was confirmed by ensuing studies (e.g., Bråten et al., 2013).

Mechanisms of Epistemic Change—The Bendixen-Rule Model

Bendixen and Rule's (2004) *process model for personal epistemology development* describes more precisely how cognitive disequilibria presumably cause epistemic change in a certain situation. It introduces three central prerequisites of epistemic change (i.e., epistemic doubt, epistemic volition and resolution strategies), which are parts of a higher order mechanism (Bendixen, 2016). An idealized description of the proposed mechanism of change in Bendixen and Rule's model is as follows: As a starting point of epistemic change, an individual experiences epistemic doubt, a cognitive dissonance. This dissonance leads to questioning one's epistemic beliefs and may occur as a response to new information that contradicts an individual's existing beliefs (Rule and Bendixen, 2010). In order to deliberately tackle this epistemic doubt, it requires a certain amount of epistemic volition (i.e., the "will" or motivation for epistemic change), the second central component of the model (Rule and Bendixen, 2010). Thereafter, epistemic doubt is resolved by applying resolution strategies, such as reflection or social interaction, and individuals eventually adopt more advanced beliefs (Bendixen and Rule, 2004). However, proceeding to advanced beliefs is not guaranteed, even if all of these components are activated. Indeed, individuals may even regress to more naive beliefs under specific circumstances (Bendixen and Rule, 2004), which are, unfortunately, only vaguely specified in the original model. However, the notion that *epistemic doubt* may occur at any stage of an individual's epistemic development (i.e., even evaluativists are expected to question their beliefs from time to time) entails some important implications when designing intervention programs. To name only one, the interplay between prior beliefs and intervention contents has to be carefully considered (cf. Rule and Bendixen, 2010). Thus, the same instructional approach may be fruitful for absolutists, while it at the same time unintentionally evokes doubt on evaluativists' advanced beliefs. Nonetheless, this model is not uncontested, and, as Bråten (2016) stressed, the empirical validation of many assumptions of Bendixen's model, including its proposed mechanism of change, is still largely unsatisfactory.

Domain-Specificity of Epistemic Beliefs and Epistemic Change

So far, we treated epistemic beliefs in a universal way, thereby implying that beliefs on knowledge and knowing do not differ depending on the content domain they relate to. Indeed, epistemic development was initially considered to be consistent across fields or domains, and earlier research (e.g., Schommer,

1993) almost exclusively used this domain-general approach (i.e., it was assumed that individuals possess similar epistemic beliefs across content domains). Recent research has challenged this assumption by showing that epistemic beliefs encompass both domain-specific and domain-general aspects that are shared across domains (Buehl and Alexander, 2005; Muis et al., 2006). Moreover, Bråten and Strømsø (2010) argue that the same principle may also apply to specific topics, such as gender stereotyping, within certain domains or subdomains, for instance educational psychology. They further argue that the impact of epistemic beliefs on educational outcomes (such as academic achievement) should be particularly strong if beliefs and outcomes are measured on the same level of specificity. Drawing upon this thought, intervention-induced epistemic change should be particularly strong in epistemic belief measures whose specificity corresponds to the specificity of the information used to evoke epistemic doubt and subsequent changes in epistemic beliefs. Even though this assumption may sound highly plausible—especially as it is in line with findings from social psychology on the role that relevant exemplars play in behavior change (e.g., Lockwood and Kunda, 1997; Han et al., 2017), its empirical backing is certainly extendable.

Experimentally Inducing Epistemic Change

After providing this overview of the framework in which epistemic change is thought to occur, the question of how to efficiently influence individuals' epistemic development remains. As the number of research programs dedicated to achieve this aim is constantly growing, a variety of intervention approaches has been developed (see Bendixen, 2016; Muis et al., 2016). Naturally, it is theoretically sound and intuitive to evoke enduring belief change in long-term intervention programs, for example by using constructivist teaching methods (e.g., Muis and Duffy, 2013). However, short-term experimental interventions have recently become more prominent (Kienhues et al., 2016). A major advantage of this study type is that it allows for a better control of experimental circumstances and for a more specific investigation of the psychological mechanisms involved in epistemic change (even though far from all short-term interventions make use of this advantage). Moreover, those interventions have been shown to be surprisingly effective in inducing epistemic change—at least in the short term (Kienhues et al., 2008, 2011; Ferguson and Bråten, 2013). Most prominently, the presentation of *diverging information* (i.e., information that includes contradictory knowledge claims) has been shown to reliably evoke epistemic change (Kienhues et al., 2016), indicating that cognitive disequilibria (and subsequent epistemic doubt) are likely to be a driving force of epistemic development. Several interventions have been designed on this basis (Kienhues et al., 2016). For example, Kienhues et al. (2011) confronted students with conflicting knowledge claims concerning medication use for the control of cholesterol and showed that topic-specific epistemic change was more pronounced under these circumstances when compared to students that received consistent information on this topic.

Regrettably, however, most of these intervention studies fail to specify the kind of change in epistemic beliefs that is desired (Bråten, 2016); such as if they intend to reduce naive beliefs or

foster advanced beliefs. Especially studies that are not strongly based on Kuhn's framework often seem to strive to simply reduce absolute beliefs and tend to neglect possible adverse effects of strong multiplistic beliefs. More precisely, frequently proposed adverse effects of multiplism encompass impaired viewpoint and text comprehension (Bråten et al., 2013; Barzilai and Eshet-Alkalai, 2015) as well as impeded sourcing (Barzilai et al., 2015). Thus, even though the mere presentation of conflicting (or diverging) information has been shown to efficiently reduce absolutism, such interventions do not ensure that evaluativistic beliefs prosper. In fact, it is much more likely that an individual will simply "replace" absolute beliefs with multiplistic beliefs or that already existing multiplistic views are strengthened when he or she is confronted with inconsistent evidence on a specific topic. Furthermore, from a theoretical point of view, one may suggest that backward transitions from evaluativism to multiplism might occur if individuals are repeatedly confronted with diverging information including controversies that are more difficult to integrate (e.g., the conflicting intervention condition of Kienhues et al., 2011). As outlined above, this kind of epistemic change is, in our view, not worth striving for. Therefore, we need interventions that make individuals avert both absolute and multiplistic beliefs, while at the same time supporting a change toward evaluativistic beliefs.

The Resolvable Controversies Intervention

To address this need, Rosman et al. (2016a) developed an intervention approach, which—by drawing on so-called *resolvable controversies*—aims to reduce both absolutism *and* multiplism simultaneously, as well as to foster evaluativism. On a global level, it illustrates, based on apparently conflicting findings of studies on gender stereotyping at secondary schools, how to identify contextual factors that help to explain controversies when evidence seems to be ill-structured—or, more strictly speaking, it exemplifies how to weigh knowledge claims (Rosman et al., 2016a).

Recently, Rosman and Mayer (2018) used the following procedures for implementing the intervention: First, 18 short abstracts of conflicting studies on gender stereotyping and gender-specific discrimination in schools are presented. A crucial component of the resolvable controversies intervention is that apparent contradictions in these texts can be resolved (or integrated) by identifying the context in which a certain type of discrimination (favoring either boys or girls) occurs. To support this process, participants are additionally asked in adjunct questions who is discriminated against according to the present study. For example, intervention contents imply that girls are discriminated against in physics while boys are discriminated against in languages and literature. In this case, participants are thought to identify the factor "subject matter" as a contextual factor that explains apparent inconsistencies between the studies. This resolvability of apparent contradictions is thought to induce epistemic doubt concerning both absolutism and multiplism because a variation in findings exists but is explainable (Rosman et al., 2016a). According to Rosman, Mayer and Merk (under review), this insight should subsequently be generalized to higher-level domains

(e.g., educational psychology). Unfortunately, on an empirical level, prior studies did not explicitly confirm this assumption—for example, by introducing a control condition drawing on inexplicable discrepancies in findings (i.e., "unresolvable" controversies)—but focused on the overall efficacy of the intervention instead.

In the second part of Rosman and Mayer's (2018) intervention, subjects proceeded by integrating conflicting findings in a writing task. In the resolution instruction of this writing task (i.e., the most prolific instruction for eliciting epistemic change), subjects were required to complete a scientific essay which illustrates conditions of gender-specific discrimination based on the presented studies. Because of the didactical properties of the presented controversies, subjects are expected to identify the aforementioned contextual factors under these circumstances. As the effects of both parts of the intervention (i.e., the reading and writing tasks) have never been disentangled, it remains unclear to what extent the intervention's efficacy can be attributed to either one of both of those distinct intervention contents. Examining these reading and writing tasks separately would be particularly insightful for clarifying how deeply diverging information has to be processed in order to affect epistemic beliefs. For example, drawing upon Bendixen and Rule's model of epistemic change, the writing task might trigger the resolution of epistemic doubt that was evoked by the presentation of diverging information. The underlying mechanism would be that a reflection on conflicting information in presented texts (during the writing task) prompts a reflection on one's own epistemic doubt that has been evoked by the respective texts. Although some studies investigated links between explicit reflection on epistemic beliefs and subsequent changes in those beliefs (see Lunn Brownlee et al., 2016), prior research failed to address the distinct relationship between receiving diverging information, reflecting on it, and epistemic change.

Research Questions

Based on these considerations, the purpose of our study is to shed some light onto how exactly diverging information may foster change toward advanced epistemic beliefs. Our first research question aims at identifying specific circumstances and characteristics of diverging information that trigger change toward certain types of epistemic beliefs.

(1) Under which circumstances does diverging information evoke epistemic change toward advanced belief types (i.e., no simple reduction of absolutism at the cost of rising multiplistic beliefs, but a reduction of both absolutism and multiplism, and a simultaneous change toward evaluativism)?

Moreover, we want to examine the effects of a deep processing of diverging information by separating effects of the *presentation* of diverging information (which should be closely related to the occurrence of epistemic doubt) from effects of *reflecting* on this information (which is possibly connected to the resolution of this doubt). Thus, our second research question is:

(2) Will interventions based on resolvable controversies still be able to induce epistemic change toward advanced epistemic

beliefs after removing all components that are linked to reflecting on how to integrate conflicting information?

As described above, it is plausible to assume that changes in epistemic beliefs depend on the level of specificity of both the administered intervention (i.e., presented diverging information) and the epistemic belief measure used. More specifically, intervention effects may be stronger if both levels of specificity correspond to each other. In our last research question, we will empirically scrutinize this assumption and examine to what extent changes in topic-specific beliefs (e.g., beliefs regarding the topic of gender stereotypes) carry over to higher-level domains (e.g., beliefs regarding educational psychology).

(3) Are the effects of topic-specific epistemic change interventions more pronounced in topic-specific epistemic belief measures?

In the next section, materials and methods of our study designed specifically to answer these questions are described.

MATERIALS AND METHODS

All planned procedures and hypotheses of our confirmatory analyses have been preregistered at the Open Science Framework (https://osf.io/te7wk/). For the reader's convenience, they are re-iterated here. Moreover, this section also includes information on actually collected data, exploratory outcomes and exploratory analyses. All study measures and methods were in compliance with the Declaration of Helsinki and the APA Ethics Code (American Psychological Association, 2002). Ethical approval was obtained from the Ethics Committee of the German Psychological Association and prior to their participation, all students gave their informed consent. Since study inclusion and pre-intervention measurements were conducted online, no written informed consent could be obtained at study inclusion. However, we provided an information sheet and consent form (for download) and subjects were only allowed to enter the study if they confirmed (by checking a box) that they agreed to the conditions specified in these documents. As all other study measures, these procedures for online data collection and study inclusion were approved by the Ethics Committee of the German Psychological Association.

Participants and Study Timeline

Our research questions were investigated with data from an experimental study employing a 4×2 pre-post design with one between-subjects factor (intervention type with four levels) and one within-subjects factor (repeated measurement factor with two levels). In total, $N = 201$ psychology students (minor and major), who were recruited at Trier University by means of flyers and mailing lists, partook in the online pre-intervention measurement. At least 1 week after this measurement, the second measurement occasion took place in group sessions at a university lab. In the second measurement occasion— that included the intervention as well as the post-intervention measurement—$N = 185$ students participated (92.04% of participants who had enrolled at the first measurement occasion) and received 20 Euro upon study completion. For one

participant, pre-intervention and post-intervention data could not be matched and, thus, data of the first measurement occasion had to be treated as missing data. Thus, our dataset contains $N = 184$ students whose demographic data is known. These participants (89.67% females) had a mean age of $M = 23.21$ ($SD = 3.13$). 95.65% of our participants studied psychology as their major subject (59.78% Bachelor and 35.87% Master students), while 4.35% took a minor in psychology. The median study duration was six semesters ($M = 5.85$, $SD = 2.97$).

Procedures and Materials

Intervention

We modified Rosman and Mayer's (2018) resolvable controversies intervention that has been described above to address our research questions. We pursued two aims with this modification: (1) to inspect how the resolvable nature of presented controversies affects epistemic change, and (2) to examine the distinct effects of presenting diverging information (i.e., evoking epistemic doubt) on epistemic change by separating effects of doubt from effects that are possibly related to deeper level processing (i.e., the resolution strategy *reflection*).

To clarify if epistemic advancement does indeed depend on the resolvability of the controversies, we "masked" the resolvable nature of these controversies by distorting the effects of contextual factors that explain diverging findings (see **Figure 1** for an illustrative example). For example, if the original intervention text states that boys are consistently discriminated against in languages and literature, the modified version stated that some studies find that boys are discriminated against in languages and literature while others find that girls are disadvantaged in these subjects. Thus, we eliminated the pattern that underlies the presented conflicting information and, hence, the intervention should induce doubt concerning absolutism only because diverging findings cannot be integrated anymore. Multiplism, in contrast, might even be fostered since the abundance of conflicting information is likely to convey views of the knowledge body in question as extremely tentative and inconsistent.

Considering the second aim, that is singling out effects of epistemic doubt, we shortened the original resolvable controversies intervention of Rosman and Mayer (2018). The original paradigm uses both reading and writing about resolvable controversies. By means of specific writing instructions, participants are invited to integrate conflicting information and, thus, reflect on this information. It cannot be finally ruled out that this higher level processing of diverging information also causes reflection on participants' epistemic doubt. Thus, we separated effects of inducing epistemic doubt by the mere presentation of diverging information from effects of reflecting on this information by comparing a shortened version of the intervention, where the writing task is left out, to the original intervention that includes this writing task.

In order to test the overall efficacy of our intervention, we compared changes in epistemic beliefs in these three treatment conditions[1] to changes in a control group. Participants in

---

[1] In the following, we will refer to all experimental groups that received any kind of diverging information on gender stereotyping as *treatment groups* or *treatment*

Picault et al. (2003) compared grades in <u>German</u> of 113 high school students to their results in a standardized reading comprehension test (the so-called ELFE 1-6). Although boys/girls scored slightly higher than girls/boys in the reading comprehension test, the average grade boys/girls received was significantly lower. Consequently, boys/girls got poorer grades than girls/boys despite similar aptitude.

☐ **Boys** are disadvantaged in secondary schools.

☐ **Girls** are disadvantaged in secondary schools.

☐ **Neither boys nor girls are disadvantaged** in secondary schools.

☐ The study is **not related** to gender stereotyping in secondary schools.

In their study, Klerus et al. (2005) had 132 secondary school students take a standardized <u>physics</u> test. Moreover, students were also asked to state their grade in physics. Even though boys and girls did not show any differences in the standardized test, girls/boys had significantly poorer grades in physics. Thus, girls/boys received worse grades than boys/girls despite similar aptitude.

☐ **Boys** are disadvantaged in secondary schools.

☐ **Girls** are disadvantaged in secondary schools.

☐ **Neither boys nor girls are disadvantaged** in secondary schools.

☐ The study is **not related** to gender stereotyping in secondary schools.

Sperber et al. (2001) compared 126 secondary school students' grades in <u>history</u> to their results in a standardized school performance test (the so-called 'Hamburger Schulleistungstest'). Grades were assessed by means of self-reports. Analyses revealed no gender differences, neither with regard to the school performance test nor with regard to grades in history.

☐ **Boys** are disadvantaged in secondary schools.

☐ **Girls** are disadvantaged in secondary schools.

☐ **Neither boys nor girls are disadvantaged** in secondary schools.

☐ The study is **not related** to gender stereotyping in secondary schools.

FIGURE 1 | Three resolvable controversies sample texts. Cues allowing to resolve the controversies are underlined and red marks illustrate how texts were modified in the "Unresolvable Read" group (please note that only half of the cues were changed resulting in an overall random pattern of discrimination). The complete German version of the texts is available on request.

the control group read texts on students employing learning strategies. To design this task as similar as possible to the gender stereotypes reading task—which required participants to rate for each presented study if boys or girls were discriminated against (adjunct questions)—each text snippet of the control task contained two descriptions of students employing different learning strategies that were compared to each other. For example, participants learned that two students applied different approaches concerning the length and distribution of their learning units. While one student learned from 9 a.m. to 6 p.m. and only took a short lunch break of 20 min, the other student only learned for 2 h at a time and took extensive breaks in between. After reading both descriptions, participants were asked to assess the characteristics of these learning strategies on a

set of scales, such as required effort or generation of detailed knowledge.

To sum up, intervention conditions or "experimental groups" in our study differed in the kind of intervention that participants received:

- *Control (learning strategies).* Group 1 read texts on students employing different learning strategies,
- *Unresolvable Read.* Group 2 read conflicting materials which cannot be resolved by identifying moderator variables (i.e., a modified version of the conflicting materials that are used in groups 3a and 3b),
- *Resolvable Read.* Group 3a read conflicting materials whose contradictions could be resolved (i.e., the original reading task of the resolvable controversies intervention),
- *Resolvable Read and Write.* Group 3b read conflicting materials whose contradictions could be resolved (the same task that group 3a received) and was additionally subjected

conditions (i.e., irrespectively of the (un)resolvable nature of these information or if subjects had to write an integrating text on these controversies).

to the resolution writing task of the resolvable controversies intervention.

The following time limits applied to respective tasks: Participants were allowed a maximum of 15 min for the reading task and 45 min for the writing task (in group 3b).
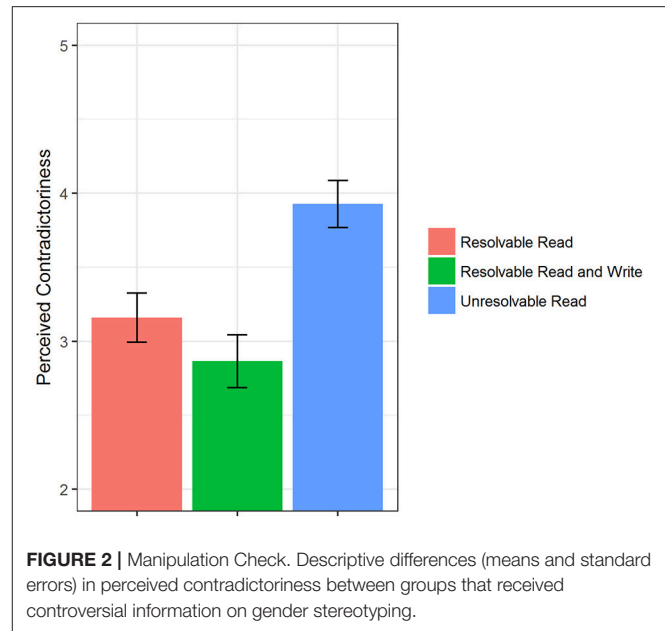
Assignment to Groups

Upon the start of the second measurement occasion, randomized assignment of participants to experimental groups was carried out using the respective function of the survey software Unipark. The study was single-blind (i.e., study staff could become aware of the assigned experimental group during the intervention). However, since all instructions that differed between groups and that were related to experimental manipulations were given in computerized form, this could not affect data quality. As expected, experimental groups did not differ significantly (all $p > 0.10$) in any demographic variables we assessed (i.e., age, gender, study semester, study subject, secondary school grades), nor in any pre-test scores on our dependent variables.

Manipulation Check

To evaluate whether our manipulation worked as intended, we checked if presented information on gender stereotypes were perceived as more controversial and contradictory in the "Unresolvable Read" group when compared to the "Resolvable Read" and "Resolvable Read and Write" groups. The underlying rationale is that—since we intended to thwart the integration of conflicting results by our modification of Rosman's intervention—higher scores on perceived contradictoriness indicate that diverging information has been recognized as non-resolvable in this group.

In order to test whether the expected differences occurred, we employed a self-report questionnaire that assessed to what extent subjects perceived presented information on gender stereotyping to be controversial or conflicting. A sample item is "Upon reading the texts…findings seemed to be very contradictory." The reliability on this scale was good (Omega total ranging from 0.80 to 0.81 in the three treatment groups). As a statistical technique, we used multiple regression analyses with the "Unresolvable Read" group as reference category and dummy-coded variables for group membership as predictors. It should be of note that the contradictoriness was only assessed for the "Unresolvable Read," "Resolvable Read" and "Resolvable Read and Write" group because of its topic-specific focus. Assessment took place after the intervention was finished in respective groups (i.e., after reading the controversies in the "Unresolvable Read" and "Resolvable Read" group and after writing a text on these controversies in the "Resolvable Read and Write" group). **Figure 2** provides a graphical overview of reported contradictoriness' mean scores separated by intervention group.

Results of these multiple regression analyses revealed that the perceived overall contradictoriness of presented information differed significantly between groups, $R^2 = 0.13$, $F_{(2, 136)} = 10.61$, $p < 0.001$. More precisely, estimates for dummy-coded regression coefficients indicate that subjects in the "Unresolvable Read"

FIGURE 2 | Manipulation Check. Descriptive differences (means and standard errors) in perceived contradictoriness between groups that received controversial information on gender stereotyping.

group rated presented information to be more inconsistent than subjects in both the "Resolvable Read" group ($b = -0.77$, $t_{(136)} = -3.213$, $p < 0.01$) and the "Resolvable Read and Write" group ($b = -1.06$, $t_{(136)} = -4.468$, $p < 0.001$).

Thus, our manipulation succeeded in "masking" the resolvability of inconsistent findings which is an integral part of the original intervention. Participants in the "Unresolvable Read group judged information concerning gender stereotypes to be more controversial than subjects in the "Resolvable Read" and "Resolvable Read and Write" groups.

Dependent Variables

Confirmatory dependent measures are the FREE-GST, a topic-specific measure of epistemic beliefs and the FREE-EDPSY, a domain-specific measure of epistemic beliefs. Both measures are based on Kuhn et al. (2000) framework and were initially developed and validated in a recent study of Rosman, Mayer and Merk (under review).

Primary outcome: topic-specific epistemic beliefs (FREE-GST)

The FREE-GST measures topic-specific epistemic beliefs on gender-stereotype discrimination in secondary schools. The questionnaire starts with the presentation of three controversial positions on gender stereotype discrimination (i.e., boys are disadvantaged, girls are disadvantaged, neither boys nor girls are disadvantaged). Thereafter, 15 statements on this controversy, which represent either absolute, multiplistic, or evaluativistic beliefs, are to be rated on a 6-point Likert scale (5 statements per belief type). A sample item for evaluativism is "Gender specific discrimination can be diverse. Accordingly, depending on certain contextual factors, rather one or the other view is correct."

Secondary outcome: domain-specific epistemic beliefs (FREE-EDPSY)

The FREE-EDPSY applies the same procedure to domain-specific epistemic beliefs in educational psychology. It introduces controversial scientific positions relating to the domain of educational psychology (i.e., an argument about the efficacy of an unspecified method of this field, such as a learning strategy or a teaching method). Subsequently, just like in the FREE-GST, 15 statements relating to either absolute, multiplistic, or evaluativistic beliefs are presented. A sample item for multiplism is "In educational research, scientists interpret their findings based on their personal opinion. Actually, nobody can know for sure whether specific methods are beneficial for learning or not."

Computation of scales and indices for the FREE-GST and FREE-EDPSY

Absolutism, multiplism and evaluativism scores were computed as mean scores of the respective items for the FREE-GST and FREE-EDPSY, exactly as has been done in prior research (e.g., Rosman and Mayer, 2018). After inspecting psychometric properties of these scales, we decided to drop one item of the multiplism scale because reliabilities increased for both the FREE-GST and the FREE-EDPSY if this item was excluded.

Furthermore, we combined absolutism, multiplism and evaluativism scores to the so-called D-index, which Krettenauer (2005) proposed as an overall measure of advanced epistemic beliefs. Applying Krettenauer's formula to our questionnaires, the D-index was computed as *Evaluativism –.5 x (Absolutism + Multiplism)* for the FREE-GST and the FREE-EDPSY. Because the D-Index condenses changes across absolutism, multiplism and evaluativism, we expected the power to detect such overall changes toward advanced beliefs to be higher in analyses using the D-Index. However, as the D-index was not part of our preregistration, analyses including this index are exploratory.

Exploratory outcome: psychology-specific justification beliefs

We assessed psychology-specific justification beliefs by a domain-specific adaptation of a domain-general German questionnaire (Klopp and Stark, 2016). Klopp and Stark's questionnaire builds on items originally developed by Ferguson et al. (e.g., Bråten et al., 2013; Ferguson and Bråten, 2013). The questionnaire differentiates the three types of justification beliefs that were introduced above: (1) personal justification, (2) justification by authority, (3) justification by multiple sources. All scores were computed as mean scores.

Covariates

To control for influences of third variables, we measured a set of potential covariates. Need for cognitive closure was assessed by Schlink and Walther's (2007) questionnaire as connections to epistemic change have already been empirically shown for this construct (Rosman et al., 2016a). Additionally, (Bendixen and Rule, 2004) repeatedly emphasized the (theoretical) importance of environmental factors. In order to account for this, we employed Schiefele and Jacob-Ebbinghaus (2006) study satisfaction questionnaire. Moreover, as Bendixen and Rule's model on epistemic change is closely connected to conceptual

change theory (Bendixen and Rule, 2004), covariates that are proposed in the conceptual change literature, i.e. need for cognition, task value, prior topic interest and self-reported prior knowledge (Dole and Sinatra, 1998; Sinatra and Mason, 2013), were included as well. Therefore, we employed an established measurement instrument by Bless et al. (1994) for need for cognition and a questionnaire that proved to reliably assess task value dimensions in prior research (Gaspard et al., 2017). Since these variables were only included in exploratory analyses if they differed at least marginally significantly between groups (see below), further details are only provided for control variables that are relevant for the present paper in **Tables 2**, **3**.

Hypotheses

Based on the research questions that were introduced above, we derived the following hypotheses:

H1. Epistemic belief change can be induced by text-based interventions that evoke epistemic doubt. The predicted patterns of epistemic change regarding the three developmental stages of epistemic beliefs (absolutism, multiplism, evaluativism) can be found in **Table 1**.

More specifically, we expect small to moderate effects for the following differences between intervention conditions:

- *H1a*. Reading multiple texts presenting conflicting scientific evidence will induce epistemic change, whereas reading texts on students employing different learning strategies will not induce epistemic change.
- *H1b*. Evaluativism will increase if the conflicts between the texts may be resolved by identifying moderator variables ('resolvable controversies') compared to a condition including texts in which the conflicts cannot be resolved.
- *H1c*. The 'resolvable controversies' intervention reliably induces epistemic change even if it is shortened by leaving out the writing task. Incremental effects of the writing task will be small to moderate.

H2. All effects on epistemic change will be more pronounced in the topic-specific measure FREE-GST compared to the domain-specific FREE-EDPSY questionnaire.

In the following, statistical procedures for testing these hypotheses are described.

Statistical Analyses

All statistical analyses were conducted in R 3.5.0 (R Core Team, 2018). The package lavaan 0.6-1 (Rosseel, 2012) was used for latent variable analyses.

TABLE 1 | Predicted pattern of effects (FREE-GST, FREE-EDPSY).

	Absolutism	Multiplism	Evaluativism
Resolvable read and write	–	–	++
Resolvable read	–	–	+
Unresolvable read	–	+	0
Control (learning strategies)	0	0	0

+, increase in epistemic beliefs; –, decrease in epistemic beliefs; 0, no change in epistemic beliefs.

TABLE 2 | Intercorrelations and reliabilities of study variables at the pre-intervention measurement occasion (t1).

Correlation \ *p*-values	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Absolutism (topic-specific)	**0.73**	0.047	<0.001	<0.001	0.017	<0.001	0.014	0.007	0.327	0.132	0.479	<0.001	<0.001
2 Multiplism (topic-specific)	0.147	**0.69**	0.004	0.764	<0.001	0.014	<0.001	0.605	0.707	0.088	0.338	<0.001	<0.001
3 Evaluativism (topic-specific)	−0.313	−0.212	**0.71**	0.002	0.007	<0.001	<0.001	0.845	0.019	0.002	0.056	<0.001	<0.001
4 Absolutism (domain-specific)	0.662	0.022	−0.224	**0.76**	0.512	0.001	0.167	<0.001	0.554	0.390	0.922	<0.001	<0.001
5 Multiplism (domain-specific)	0.176	0.701	−0.199	0.049	**0.79**	0.017	<0.001	0.184	0.504	0.793	0.710	<0.001	<0.001
6 Evaluativism (domain-specific)	−0.274	−0.180	0.757	−0.253	−0.176	**0.69**	0.002	0.738	0.036	0.019	0.299	<0.001	<0.001
7 Personal Justification	0.181	0.623	−0.262	0.102	0.711	−0.229	**0.77**	0.198	0.131	0.953	0.714	<0.001	<0.001
8 Justification by Authority	0.198	−0.038	−0.014	0.254	−0.098	−0.025	−0.095	**0.76**	0.056	0.364	0.505	0.336	0.324
9 Justification by Multiple Sources	−0.073	−0.028	0.172	−0.044	0.050	0.154	0.112	−0.141	**0.73**	0.604	0.404	0.043	0.201
10 Task Value	−0.112	−0.126	0.226	−0.064	−0.020	0.173	0.004	0.068	0.039	**0.90**	0.053	0.001	0.051
11 Prior Interest Gender Stereotypes	0.053	−0.071	0.141	0.007	0.028	0.077	−0.027	−0.050	0.062	0.143	**0.77**	0.188	0.641
12 D–Index (topic–specific)	−0.641	−0.562	0.845	−0.407	−0.455	0.661	−0.469	−0.071	0.150	0.236	0.098	–	<0.001
13 D–Index (domain–specific)	−0.514	−0.430	0.661	−0.577	−0.568	0.818	−0.497	−0.073	0.095	0.145	0.035	0.783	–

N = 184 and 183 (for correlations involving prior interest in gender stereotypes or task value); values in bold on the diagonal = Omega Total; the lower triangle contains correlation estimates while the upper triangle represents corresponding p-values (two-tailed tests).

Statistical Model

Confirmatory analyses

We used latent difference score modeling (McArdle, 2009) to analyze our data. The main outcome variables of our analyses were changes in epistemic beliefs (i.e., absolutism, multiplism and evaluativism scores of the FREE-GST and FREE-EDPSY), which were operationalized as latent change scores (see **Figure 3** for more details). These latent change scores were predicted by dummy-coded intervention group variables. In order to investigate group differences not related to the reference group, we defined these effects as new parameters of the structural equation model. The same procedure holds for comparisons between topic–and domain-related measures (H2). Analyses concerning H1 were conducted separately for absolutism, multiplism and evaluativism (for FREE-GST and FREE-EDPSY, respectively) resulting in a total number of six target models. A logical precondition of H2 (more pronounced effects on epistemic change for the topic-specific FREE-GST) is that group differences in epistemic change exist. Therefore, H2 was only to be tested if any significant group differences were found in analyses that are related to H1. However, H2-analyses were performed even if the revealed pattern of effects contradicted the hypothesized pattern of effects. H2-analyses were conducted separately for absolutism, multiplism and evaluativism resulting in a maximum possible number of three target models.

The following procedure was employed for testing our hypotheses: First, intervention group was dummy-coded with the control group as reference category[2]. Thereafter, we estimated a null model that fixed differences in epistemic change between groups ($b_1 = b_2 = b_3 = 0$) [H1] or between topic-specific and domain-specific measures ($b_{0GST} = b_{0EDPSY}$, $b_{1GST} = b_{1EDPSY}$, $b_{2GST} = b_{2EDPSY}$, $b_{3GST} = b_{3EDPSY}$) [H2] to zero. Subsequently, we compared this null model to a target model that imposed no restrictions on differences in epistemic change between groups ($b_1 = x_1$, $b_2 = x_2$, $b_3 = x_3$) [H1] or topic–and domain-specific measures ($b_{0GST} = x_4$, $b_{0EDPSY} = x_5$, $b_{1GST} = x_6$, $b_{1EDPSY} = x_7$, $b_{2GST} = x_8$, $b_{2EDPSY} = x_9$, $b_{3GST} = x_{10}$, $b_{3EDPSY} = x_{11}$) [H2]. If the corresponding likelihood ratio test (LRT) revealed that epistemic change differed significantly between groups [H1] or measures [H2], we inspected the estimated model parameters in order to examine group [H1] or measure differences [H2] in epistemic change. We used the standard $p < 0.05$ criteria for likelihood ratio tests and for determining if the estimated effects of (dummy-coded) intervention group variables were significantly different from those expected if the null hypothesis was correct. As the expected direction of effects as well as the expected order of effects is explicitly predicted, we used one-tailed tests whenever appropriate.

---

[2]This is a minor modification to the planned procedure in our preregistration which suggested using the "Resolvable Read and Write" group as reference. However, this modification does not substantially affect our confirmatory analyses as it only changes how the model is parameterized (and not if effects become significant or not). We chose this procedure as it allowed us a more convenient interpretation of results (i.e., in terms of consistency with exploratory analyses).

TABLE 3 | Means and standard deviations of all study variables separated by intervention group.

	Pre-intervention (t1)								Post-intervention (t2)							
	M_{LS}	SD_{LS}	M_{UR}	SD_{UR}	M_{RR}	SD_{RR}	M_{RW}	SD_{RW}	M_{LS}	SD_{LS}	M_{UR}	SD_{UR}	M_{RR}	SD_{RR}	M_{RW}	SD_{RW}
Absolutism (topic-specific)	2.671	0.850	2.813	0.751	2.822	0.694	2.847	0.740	2.383	0.832	2.396	0.777	2.509	0.714	2.238	0.884
Multiplism (topic-specific)	2.978	0.738	3.076	0.758	2.924	0.673	3.059	0.747	2.924	0.859	2.777	0.738	2.614	0.737	2.777	0.911
Evaluativism (topic-specific)	4.871	0.606	4.870	0.558	4.813	0.666	4.749	0.720	4.891	0.682	5.048	0.530	4.935	0.595	5.026	0.673
Absolutism (domain-specific)	2.636	0.739	2.752	0.683	2.891	0.802	2.872	0.696	2.474	0.784	2.522	0.762	2.652	0.756	2.481	0.881
Multiplism (domain-specific)	3.083	0.885	3.011	0.756	2.832	0.825	3.229	0.746	2.799	0.843	2.777	0.834	2.614	0.881	2.899	0.817
Evaluativism(domain-specific)	4.987	0.602	4.922	0.537	4.917	0.563	4.877	0.687	5.052	0.583	5.074	0.534	4.991	0.537	5.111	0.575
D-Index (topic-specific)	2.047	1.082	1.925	0.908	1.940	0.932	1.796	1.030	2.238	1.141	2.461	0.895	2.373	0.738	2.518	0.953
D-Index (domain-specific)	2.127	0.936	2.040	0.886	2.056	0.930	1.826	0.968	2.416	0.919	2.424	0.814	2.358	0.804	2.421	0.917
Personal Justification	2.607	1.035	2.630	1.012	2.420	0.765	2.901	0.914	2.341	0.868	2.457	0.898	2.326	0.899	2.660	0.936
Justification by Authority	3.785	0.838	3.543	0.946	3.543	0.884	3.418	0.844	3.761	0.746	3.370	0.856	3.580	0.859	3.156	1.056
Justification by Multiple Sources	5.022	0.796	5.080	0.689	4.935	0.848	5.043	0.680	5.043	0.729	5.203	0.638	5.203	0.573	5.262	0.637
Task Value	2.920	0.675	2.707	0.618	2.549	0.680	2.660	0.681	–	–	–	–	–	–	–	–
Prior Interest Gender Stereotypes	4.341	1.131	4.399	1.104	4.072	0.883	4.809	1.056	–	–	–	–	–	–	–	–
Perceived Contradictoriness	–	–	–	–	–	–	–	–	–	–	3.928	1.082	3.159	1.133	2.865	1.219
N	45*	45*	46	46	46	46	47	47	46	46	46	46	46	46	47	47

M, arithmetic mean; SD, standard deviation; indices specify the intervention group LS, Learning Strategies (Control); UR, Unresolvable Read; RR, Resolvable Read; RW, Resolvable Read and Write. *Due to missing values the sample size for prior interest in gender stereotypes and task value was 44.

Exploratory analyses

In addition to this preregistered procedure, we introduced an alternating model which proposed that the presentation of topic-specific diverging information had an overall effect on epistemic beliefs that was invariant across treatment groups (i.e., in the "Resolvable Read," "Resolvable Read and Write" and the "Unresolvable Read" group). Strictly speaking, this "equal group effects" model thereby suggests that neither the writing task nor the resolvable or unresolvable nature of the intervention materials mattered, but that the mere presentation of diverging information may trigger epistemic change. In order to specify this model, we restricted effects of dummy-coded variables to be equal across treatment conditions ($b_1 = b_2 = b_3$) and repeated our analyses for the FREE-GST and FREE-EDPSY. Furthermore, we analyzed the five additional exploratory outcomes introduced above: justification beliefs (personal justification, justification by authority, justification by multiple sources), and the D-Indices of the FREE-GST respectively the FREE-EDPSY.

As a consequence, we extended our model comparison procedure for choosing a target model as follows: In a first step, we compared the equal group effects model ($b_1 = b_2 = b_3$) to the null model ($b_1 = b_2 = b_3 = 0$) based on a likelihood ratio test.

The selected model of the first step was subsequently compared to our target model from the confirmatory analyses ($b_1 = x_1$, $b_2 = x_2$, $b_3 = x_3$). Otherwise, we applied the same procedures as for confirmatory hypothesis testing.

We also checked for pre-test differences on covariates that were measured before group assignment took place by means of ANOVAs with group as factor. If any marginally significant or significant differences between groups on covariates existed, we conducted additional analyses that introduced these covariates as predictors of both pre-intervention beliefs and epistemic change in our latent change model.

Finally, we investigated if the intervention was especially beneficial for subjects that held more naive epistemic beliefs (i.e., prior beliefs as indicated by pre-intervention values). For this purpose, we divided our sample into groups with more naive or more advanced epistemic beliefs—as has been done in prior research on epistemic change (e.g., Kienhues et al., 2008). More precisely, we repeated all prior exploratory analyses that yielded significant intervention effects and used multiple group modeling to test if these intervention effects differed between naive and advanced groups. For each multiple group model, we split our sample into a naive and an advanced group based on the median

FIGURE 3 | Exemplary latent change model for testing H1. Latent change in epistemic beliefs Δ GST_A (i.e., latent change in absolutism on the FREE-GST) is predicted by dummy-coded variables indicating group membership (i.e., *RRW* for "Resolvable Read and Write", *RR* for "Resolvable Read" and *UR* for "Unresolvable Read"). Latent change itself is operationalized as the part of an observed outcome variable GST_{A2} (i.e., absolutism on the FREE-GST post-intervention) that differs from its pre-intervention measurement GST_{A1} (i.e., absolutism on the FREE-GST pre-intervention).

score of pre-intervention values of the outcome variable under investigation and tested if intervention effects differed between these groups based on LRTs.

Statistical Power and Sample Size Calculation
Our a priori determined target sample size was 212 participants (i.e., 53 for each experimental group). In order to calculate this target sample size, we conducted a simulation study in R. For each condition of this simulation study (i.e., tested sample size), we generated 1,000 datasets and, subsequently, analyzed the data using the statistical model described above. The expected effect size in the population model of this simulation study was derived from a previous study by Rosman, Mayer and Merk (under review), who examined epistemic change using the resolvable controversies intervention and employed a similar design to our current study. In this study, the authors showed that modifying the resolvable controversies intervention by introducing alternating writing tasks caused significant differences in epistemic change between conditions (i.e., a standardized regression coefficient of 0.276 for change in evaluativism). As we assumed that dropping the writing task or changing the resolvable nature of the presented controversies were much stronger modifications of the established resolvable controversies intervention, we expected larger effects in the current study. Our simulation study revealed that such effects would be detectable for a sample size of $n = 53$ subjects per group:

The power for detecting small to moderate effects (i.e., beta = 0.40), which range above the practical significance criterion introduced by Ferguson (2009), surpassed 85%. Moreover, the power for detecting moderate effects (i.e., beta = 0.50) was above 96% for this sample size. A reanalysis with our actual sample size (46 subjects per group) showed that the power for detecting small to moderate effects still approximated 80% and was therefore acceptable.

RESULTS

Reliabilities and intercorrelations of all study variables for the first measurement occasion are given in **Table 2**, while means and standard deviations (separated by group) are given in **Table 3**. Moreover, considerable ceiling effects existed for the *justification by multiple sources* scale (pre 15.22% and post 20.00% of all subjects showed values at the upper limit of the scale), as well as small ceiling effects for *evaluativism* on both the FREE-GST (2.72% pre and 8.11% post) and the FREE-EDPSY (6.52% pre and 7.57% post). Floor effects for all other measures were neglectable (<5.00% pre respectively 6.50% post), while the D-Index was completely unaffected by ceiling effects. There were no univariate or multivariate outliers on dependent variables according to the criteria of our preregistration (i.e., based on z-scores with $p(z) < 0.001$ for univariate outliers and a mahalanobis distance with $p(\chi^2, df = 6) < 0.001$ for multivariate outliers). Thus, no outlier-corrected analyses were performed.

Confirmatory Analyses
A graphical overview of mean changes in epistemic beliefs on primary and secondary outcomes divided by experimental groups is given in **Figure 4**.

Hypothesis 1
None of the likelihood ratio tests that were planned in our preregistration reached significance (all $p > 0.05$ see **Tables 4, 5** for more details). Thus, we found no significant group differences in epistemic change according to the preregistered criterion. For topic-specific beliefs, as measured by the FREE-GST, we observed, across experimental groups, significant declines in absolutism ($b_0 = -0.407$, $p < 0.001$) and multiplism ($b_0 = -0.242$, $p < 0.001$), while evaluativism increased significantly ($b_0 = 0.153$, $p < 0.01$). The same pattern was observed for domain-specific beliefs that were assessed by the FREE-EDPSY with regard to absolutism ($b_0 = -0.254$, $p < 0.001$) and multiplism ($b_0 = -0.271$, $p < 0.001$) and evaluativism ($b_0 = 0.134$, $p < 0.001$, see **Table 6** for more details).

Hypothesis 2
As prespecified in our statistical analysis plan, Hypothesis 2 was not tested because confirmatory analyses concerning Hypothesis 1 revealed no significant differences between groups.
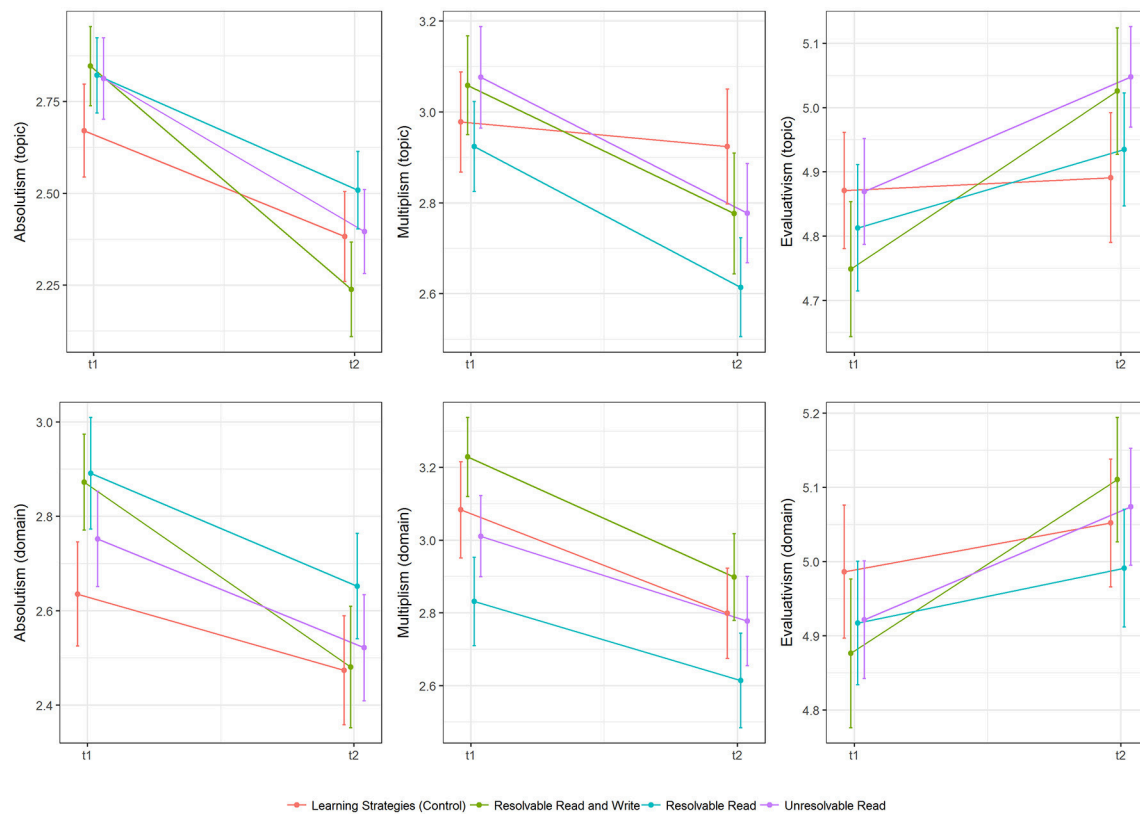
FIGURE 4 | Descriptive differences (means and standard errors) in epistemic change in topic-specific (FREE-GST) and domain-specific (FREE-EDPSY) epistemic beliefs for absolutism, multiplism and evaluativism.

Exploratory Analyses
Equal Group Effects Model
When repeating our analyses with the equal group effects model ($b_1 = b_2 = b_3$), all likelihood ratio tests on primary and secondary outcomes still failed to reach statistical significance when comparing the equal group effects model to the null model (all $p > 0.05$ see **Tables 4**, **5** for more details).

D-Index
Descriptive changes in the D-Index are depicted in **Figure 5**, while more information on descriptive statistics is available in **Table 3**.

For topic-specific advanced epistemic beliefs, LRTs indicated that the equal group effects model fitted our data best. In other words, effects on epistemic change for the control group and the three topic-specific intervention groups (i.e., the "Resolvable Read and Write," "Resolvable Read," "Unresolvable Read" groups) differed significantly ($\Delta \chi^2 = 6.413$, $df = 1$, $p < 0.05$), while differences in effect estimates between experimental conditions did not reach statistical significance ($\Delta \chi^2 = 2.830$, $df = 2$, $p = 0.243$). When analyzing parameter estimates of the model, we obtained the following pattern of effects: Even though D-index scores (an indicator of advanced epistemic beliefs) increased significantly in the control group ($b_0 = 0.253$, $p < 0.05$), this increase was significantly

larger across topic-specific intervention groups ($b_1 = 0.300$, $p < 0.05$).

For the respective measure on domain-specific beliefs, LRTs indicated that neither for the equal group effects model, nor for a model with unrestricted group effects, model fit improved significantly. Across groups, we observed a significant increase in the D-Index for domain-specific beliefs ($b_0 = 0.397$, $p < 0.001$). **Tables 4**, **5** provide more details on model fit difference tests and overall model fit, while **Table 6** presents parameter estimates.

As epistemic change differed between groups, we tested Hypothesis 2 for the D-Index. Concerning Hypothesis 2, we selected (again based on LRTs) a model that restricted effects on topic-specific and domain-specific measures to be equal across topic-specific intervention groups ($b_1 = b_2 = b_3$) but allowed these effects (and the intercept in the control group) to differ between topic- and domain-specific measures (see **Table 7** for more details on model difference tests). Model inspection showed that intervention effects on epistemic change were indeed significantly more pronounced in the topic-specific D-Index than in the domain-specific D-index (b_{1GST}-$b_{1EDPSY} = 0.237$, $p < 0.05$), while effects in the control group did not differ significantly (b_{0GST}-$b_{0EDPSY} = -0.100$, $p = 0.396$). Again, **Table 6** provides further details on parameter estimates.

TABLE 4 | Fit indices and model difference tests for the FREE-GST.

	Without covariates				$\Delta\chi^2$	Δdf	p	With covariates				$\Delta\chi^2$	Δdf	p
	χ^2	df	p					χ^2	df	p				
ABSOLUTISM														
M0. No intervention effect	**7.905**	**6**	**0.245**	M0 vs. M1	1.096	1	0.295	**7.062**	**6**	**0.315**	M0 vs. M1	1.015	1	0.314
M1. Equal group effects	6.809	5	0.235	M1 vs. M2	–	–	–	6.047	5	0.302	M1 vs. M2	–	–	–
M2: Varying intervention effects	1.673	3	0.643	M0 vs. M2	6.232	3	0.101	0.944	3	0.815	M0 vs. M2	6.118	3	0.106
MULTIPLISM														
M0. No intervention effect	**5.197**	**6**	**0.519**	M0 vs. M1	3.597	1	0.058	6.403	6	0.380	M0 vs. M1	3.854	1	0.0496
M1. Equal group effects	1.600	5	0.901	M1 vs. M2	–	–	–	**2.549**	**5**	**0.769**	M1 vs. M2	0.508	2	0.776
M2: Varying intervention effects	1.242	3	0.743	M0 vs. M2	3.955	3	0.266	2.041	3	0.564	M0 vs. M2	–	–	–
EVALUATIVISM														
M0. No intervention effect	**4.904**	**6**	**0.556**	M0 vs. M1	2.440	1	0.118	6.219	6	0.399	M0 vs. M1	4.143	1	0.042
M1. Equal group effects	2.464	5	0.782	M1 vs. M2	–	–	–	**2.076**	**5**	**0.839**	M1 vs. M2	0.455	2	0.797
M2: Varying intervention effects	1.047	3	0.790	M0 vs. M2	3.856	3	0.277	1.621	3	0.655	M0 vs. M2	–	–	–
D-INDEX														
M0. No intervention effect	10.502	6	0.105	M0 vs. M1	6.413	1	0.011	11.405	6	0.077	M0 vs. M1	8.079	1	0.004
M1. Equal group effects	**4.088**	**5**	**0.537**	M1 vs. M2	2.830	2	0.243	**3.326**	**5**	**0.650**	M1 vs. M2	1.800	2	0.407
M2: Varying intervention effects	1.259	3	0.739	M0 vs. M2	–	–	–	1.526	3	0.676	M0 vs. M2	–	–	–

Boldface = target model.

TABLE 5 | Fit indices and model difference tests for the FREE-EDPSY.

	Without covariates				$\Delta\chi^2$	Δdf	p	With covariates				$\Delta\chi^2$	Δdf	p
	χ^2	df	p					χ^2	df	p				
ABSOLUTISM														
M0. No intervention effect	**6.004**	**6**	**0.423**	M0 vs. M1	0.375	1	0.540	**5.430**	**6**	**0.490**	M0 vs. M1	0.235	1	0.628
M1. Equal group effects	5.629	5	0.344	M1 vs. M2	–	–	–	5.195	5	0.393	M1 vs. M2	–	–	–
M2: Varying intervention effects	3.968	3	0.265	M0 vs. M2	2.037	3	0.565	3.457	3	0.326	M0 vs. M2	1.973	3	0.578
MULTIPLISM														
M0. No intervention effect	**6.141**	**6**	**0.408**	M0 vs. M1	0.033	1	0.855	**6.335**	**6**	**0.387**	M0 vs. M1	0.176	1	0.675
M1. Equal group effects	6.108	5	0.296	M1 vs. M2	–	–	–	6.159	5	0.291	M1 vs. M2	–	–	–
M2: Varying intervention effects	6.010	3	0.111	M0 vs. M2	0.131	3	0.988	6.046	3	0.109	M0 vs. M2	0.289	3	0.962
EVALUATIVISM														
M0. No intervention effect	**3.859**	**6**	**0.696**	M0 vs. M1	0.486	1	0.486	**3.138**	**6**	**0.791**	M0 vs. M1	1.165	1	0.280
M1. Equal group effects	3.374	5	0.643	M1 vs. M2	–	–	–	1.973	5	0.853	M1 vs. M2	–	–	–
M2: Varying intervention effects	0.704	3	0.872	M0 vs. M2	3.156	3	0.368	0.606	3	0.895	M0 vs. M2	2.532	3	0.470
D-INDEX														
M0. No intervention effect	**5.637**	**6**	**0.465**	M0 vs. M1	0.467	1	0.494	**4.634**	**6**	**0.592**	M0 vs. M1	0.571	1	0.450
M1. Equal group effects	5.170	5	0.396	M1 vs. M2	–	–	–	4.063	5	0.540	M1 vs. M2	–	–	–
M2: Varying intervention effects	2.640	3	0.450	M0 vs. M2	2.996	3	0.392	2.646	3	0.449	M0 vs. M2	1.988	3	0.575

Boldface = target model.

Justification Beliefs

Observed changes in justification beliefs are depicted in **Figure 6**, while **Table 8** details overall model fit and model difference tests. Finally, information on parameter estimates of the target models can be retrieved from **Table 9**.

Personal justification

For *personal justification*, we found no group differences in epistemic change ($p > 0.05$ for all LRTs). Overall, personal justification beliefs decreased significantly ($b_0 = -0.201$, $p < 0.001$) across groups.

Justification by authority

Regarding the next scale of the justification beliefs questionnaire, *justification by authority*, LRTs indicated that a model with varying (freely estimated) effects between experimental conditions fitted our data best ($\Delta\chi^2 = 9.708$, $df = 3$, $p < 0.05$, see **Table 8** for more details). According to this model, beliefs in justification by authority decreased significantly in the "Resolvable Read and Write" group ($b_1 = -0.378$, $p < 0.05$) and the "Unresolvable Read" group ($b_3 = -0.247$, $p < 0.05$) when compared to epistemic change in the control group. The corresponding effect in the "Resolvable Read"

TABLE 6 | Regression coefficients of target models predicting epistemic change in absolutism, multiplism, evaluativism and the D-Index (measured by FREE-GST and FREE-EDPSY).

	Absolutism				Multiplism				Evaluativism				D-Index			
	No covariates		Covariates		No covariates		Covariates		No covariates		Covariates		No covariates		Covariates	
	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
FREE-GST																
Intercept	**−0.407****	0.050	**−0.407****	0.049	**−0.242****	0.053	−0.072	0.099	**0.153****	0.044	0.019	0.072	**0.253***	0.114	**0.225***	0.113
Intervention	0.000	–	0.000	–	0.000	–	**−0.226***	0.113	0.000	–	**0.178***	0.085	**0.300***	0.127	**0.337****	0.125
Task Value		–	−0.091	0.057			0.001	0.058			0.045	0.045			0.027	0.061
Prior Interest			0.053	0.060			0.006	0.052			0.016	0.047			0.029	0.057
FREE–EDPSY																
Intercept	**−0.254****	0.049	**−0.254****	0.049	**−0.271****	0.048	**−0.271****	0.048	**0.134****	0.036	**0.134****	0.036	**0.397****	0.054	**0.397****	0.053
Intervention	0.000	–	0.000	–	0.000	–	0.000	–	0.000	–	0.000	–	0.000	–	0.000	–
Task Value			0.050	0.054			0.049	0.053			0.031	0.037			−0.018	0.058
Prior Interest			−0.013	0.051			−0.061	0.054			0.035	0.035			0.072	0.061
FREE–GST and FREE-EDPSY																
Intercept$_{GST}$							−0.097	0.097			0.053	0.051			**0.224***	0.107	**0.199+**	0.105
Intercept$_{EDPSY}$							**−0.298****	0.096			0.053	0.051			**0.323****	0.091	**0.334****	0.092
Intervention$_{GST}$							**−0.192+**	0.106			**0.117+**	0.062			**0.337****	0.118	**0.370****	0.114
Intervention$_{EDPSY}$							0.035	0.105			**0.117+**	0.062			0.099	0.102	0.085	0.102
Task Value$_{GST}$							0.004	0.057			0.040	0.045					0.030	0.061
Prior Interest$_{GST}$							0.005	0.052			0.018	0.047					0.028	0.057
Task Value$_{EDPSY}$							0.051	0.053			0.040	0.038					0.069	0.060
Prior Interest$_{EDPSY}$							−0.062	0.053			0.032	0.035					−0.012	0.058
ΔControl							**0.201***	0.093							−0.100	0.117	−0.135	0.117
ΔIntervention							**−0.227***	0.094							**0.237***	0.118	**0.284***	0.116

*N = 185; reference group (0/0/0 dummy coding) = control (learning strategies); EST, unstandardized regression weight; SE, standard error; boldface scores = two-tailed significance test; +p < 0.10; *p < 0.05; **p < 0.01.*

FIGURE 5 | Descriptive differences (means and standard errors) in epistemic change in topic-specific (FREE-GST) and domain-specific epistemic beliefs (FREE-EDPSY) for the D-index.
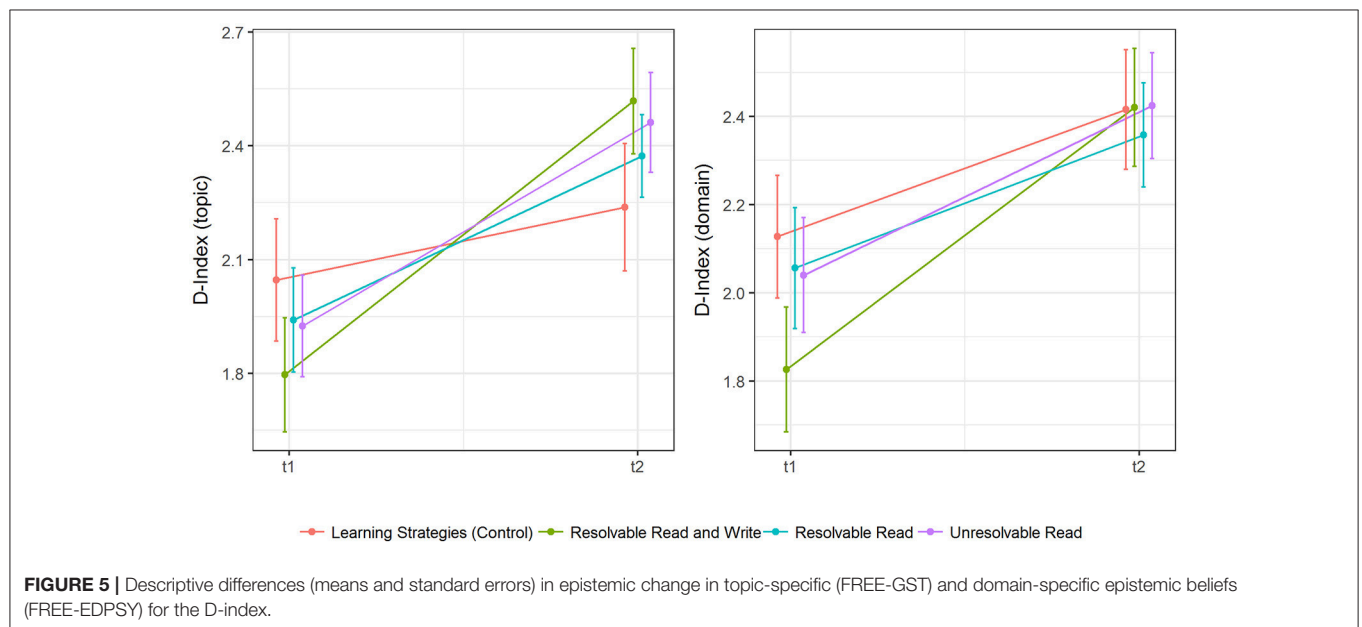
group ($b_2 = -0.037$, $p = 0.771$) and overall change in the control group ($b_0 = 0.066$, $p = 0.477$) did not reach statistical significance.
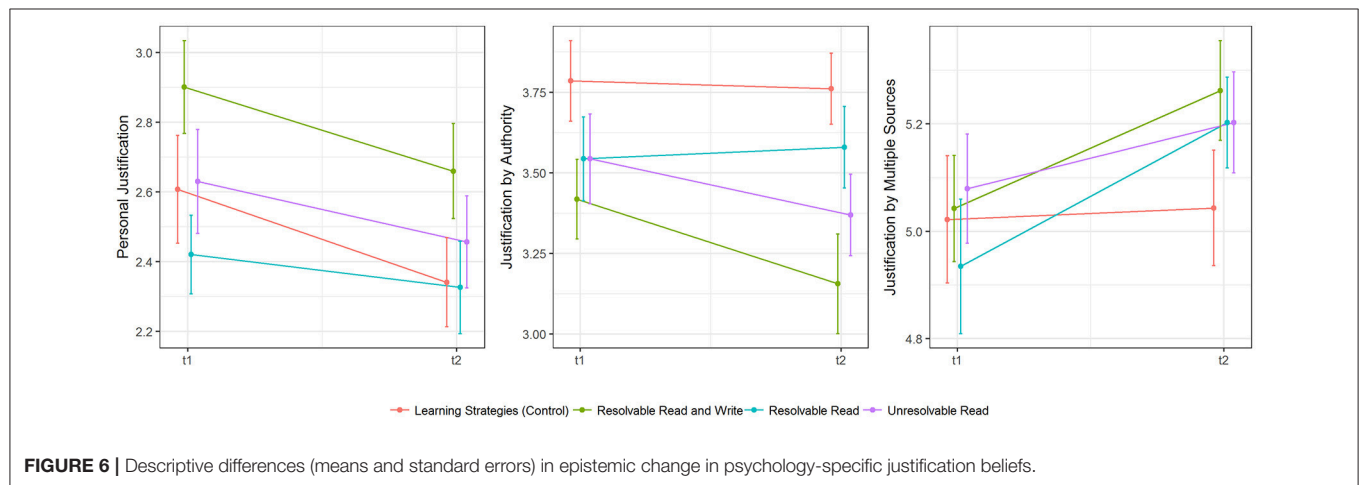
Justification by multiple sources

Finally, we selected a model with effects that were fixed to be equal for all groups that received a topic-specific intervention on

TABLE 7 | Fit indices and model difference tests for Hypothesis 2.

	Without covariates							With covariates						
	χ^2	df	p		$\Delta\chi^2$	Δdf	p	χ^2	df	p		$\Delta\chi^2$	Δdf	p
ABSOLUTISM														
M0. No difference between measures	–	–	–		–	–	–	–	–	–		–	–	–
M1. Equal group effects for each measure	–	–	–		–	–	–	–	–	–		–	–	–
MULTIPLISM														
M0. No difference between measures	–	–	–		–	–	–	20.849	14	0.106	M0 vs. M1	6.329	2	0.042
M1. Equal group effects for each measure	–	–	–		–	–	–	**14.520**	**12**	**0.269**				
EVALUATIVISM														
M0. No difference between measures	–	–	–		–	–	–	**31.567**	**14**	**0.005**	M0 vs. M1	2.973	2	0.226
M1. Equal group effects for each measure	–	–	–		–	–	–	28.594	12	0.005				
D-INDEX														
M0. No difference between measures	26.781	14	0.021	M0 vs. M1	6.405	2	0.041	28.844	14	0.011	M0 vs. M1	8.550	2	0.014
M1. Equal group effects for each measure	**20.376**	**12**	**0.060**					**20.293**	**12**	**0.062**				

Boldface = target model.

FIGURE 6 | Descriptive differences (means and standard errors) in epistemic change in psychology-specific justification beliefs.

gender-stereotypes for *justification by multiple sources* ($\Delta \chi^2 = 4.010$, $df = 1$, $p < 0.05$, see **Table 8** for more details). Participants of the treatment groups showed a change toward stronger beliefs in *justification by multiple sources* ($b_1 = 0.185$, $p < 0.05$) when compared to participants in the control group whose beliefs remained unchanged ($b_0 = 0.017$, $p = 0.836$).

Controlling for Pre-test Differences on Covariates

Analyses on pre-intervention differences on covariates revealed that groups differed at least marginally significant on self-reported intrinsic task value, i.e. a positive attitude toward dealing with psychological science, $F_{(3, 179)} = 2.47$, $p < 0.10$, $\eta^2 = 0.04$, and prior topic interest, i.e. self-reported interest in the topic

gender stereotyping, $F_{(3, 179)} = 3.93$, $p < 0.01$, $\eta^2 = 0.06$, at the first measurement occasion (and therefore prior to group assignment). More specifically, Tukey-*post-hoc*-tests indicated that participants who were later assigned to the "Resolvable Read" group had significantly lower values ($p < 0.05$) on the intrinsic task value scale when compared to the control group and on prior topic interest when compared to the "Resolvable Read and Write" group. Apart from that, no *post-hoc* comparisons yielded significant results. Due to the randomized assignment of participants to intervention conditions, these differences can only be attributed to mere chance. To deal with the issue, however, we included these variables as covariates that predicted pre-intervention differences in epistemic beliefs and epistemic

TABLE 8 | Fit indices and model difference tests for psychology-specific justification beliefs.

	Without covariates							With covariates						
	χ^2	df	p		$\Delta\chi^2$	Δdf	p	χ^2	df	p		$\Delta\chi^2$	Δdf	p
PERSONAL JUSTIFICATION														
M0. No intervention effect	**7.572**	**6**	**0.271**	M0 vs. M1	1.386	1	0.239	**8.204**	**6**	**0.224**	M0 vs. M1	1.108	1	0.293
M1. Equal group effects	6.187	5	0.288	M1 vs. M2	–	–	–	7.096	5	0.214	M1 vs. M2	–	–	–
M2: Varying intervention effects	6.154	3	0.104	M0 vs. M2	1.418	3	0.701	7.085	3	0.069	M0 vs. M2	1.119	3	0.772
JUSTIFICATION BY AUTHORITY														
M0. No intervention effect	13.431	6	0.037	M0 vs. M1	3.563	1	0.059	13.301	6	0.038	M0 vs. M1	2.168	1	0.141
M1. Equal group effects	9.868	5	0.079	M1 vs. M2	–	–	–	11.133	5	0.049	M1 vs. M2	–	–	–
M2: Varying intervention effects	**3.723**	**3**	**0.293**	M0 vs. M2	9.708	3	0.021	**2.940**	**3**	**0.401**	M0 vs. M2	10.361	3	0.016
JUSTIFICATION BY MULTIPLE SOURCES														
M0. No intervention effect	5.507	6	0.481	M0 vs. M1	4.010	1	0.045	7.050	6	0.316	M0 vs. M1	5.564	1	0.018
M1. Equal group effects	**1.497**	**5**	**0.913**	M1 vs. M2	0.557	2	0.757	**1.487**	**5**	**0.915**	M1 vs. M2	0.808	2	0.668
M2: Varying intervention effects	0.940	3	0.816	M0 vs. M2	–	–	–	0.679	3	0.878	M0 vs. M2	–	–	–

Boldface = target model.

TABLE 9 | Regression coefficients of target models predicting epistemic change in justification beliefs.

	Personal justification				Justification by authority				Justification by multiple sources			
	No covariates		Covariates		No covariates		Covariates		No covariates		Covariates	
	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
Intercept	**−0.201****	0.053	**−0.201****	0.053	0.066	0.092	0.029	0.089	0.017	0.084	0.010	0.085
Resolvable read and write	0.000	–	0.000	–	**−0.378***	0.146	**−0.352***	0.148	**0.185***	0.093	**0.220***	0.095
Resolvable read	0.000	–	0.000	–	−0.037	0.128	0.044	0.123	**0.185***	0.093	**0.220***	0.095
Unresolvable read	0.000	–	0.000	–	**−0.247***	0.126	**−0.211**$^+$	0.119	**0.185***	0.093	**0.220***	0.095
Task value			−0.037	0.053			**0.107***	0.050			0.064	0.050
Prior interest			0.035	0.050			0.066	0.051			−0.051	0.059

N = 185; reference group (0/0/0 dummy coding) = control (learning strategies); EST, unstandardized regression weight; SE, standard error; boldface scores = two-tailed significance test; $^+p < 0.10$; *$p < 0.05$; **$p < 0.01$.

change in our analyses and repeated all analyses specified above. To facilitate interpreting results of these analyses, both covariates were z-standardized prior to inclusion.

Results of the controlled analyses differed for topic-specific beliefs on multiplism and evaluativism. For both multiplism and evaluativism, as measured by the FREE-GST, we chose an equal group effects model ($b_1 = b_2 = b_3$) based on LRTs (see **Table 4** for more details). Parameter estimates of these models indicate that epistemic beliefs in the control group did not change significantly (multiplism: $b_0 = -0.072$, evaluativism: $b_0 = 0.019$, both $p > 0.05$). When compared to these effects, we observed a significantly more pronounced decline in multiplism ($b_1 = -0.226$, $p < 0.05$) and increase in evaluativism ($b_1 = 0.178$, $p < 0.05$) across topic-specific intervention groups.

Subsequently, we also tested Hypothesis 2 on multiplism and evaluativism while controlling for pre-test differences. For multiplism, an equal group effects model was chosen based on

LRTs (see **Table 7** for more details). Inspection of parameter estimates revealed that treatment effects were significantly more pronounced in topic-specific measures ($b_{1GST} - b_{1EDPSY} = -0.227$, $p < 0.05$) while epistemic change toward advanced beliefs in the control group was significantly more prominent in domain-specific measures ($b_{0GST} - b_{0EDPSY} = 0.201$, $p < 0.05$). For evaluativism, model fit did not significantly increase upon allowing effects to differ between domain-specific and topic-specific measures (see **Table 7** for more details) and therefore, a model that restricted intercept and slope to be equal across topic– and domain-specific measures was chosen. Parameter estimates for this model imply that evaluativism scores in the control group did not change significantly over time ($b_0 = 0.053$, $p = 0.298$) while in comparison a significant increase of evaluativism was detected across measures for the treatment groups ($b_1 = 0.117$, $p < 0.05$; one-tailed). In other words, epistemic change in evaluativism does not differ between topic–and domain-specific beliefs (and H2 is therefore rejected), while an overall increase

in topic–and domain-specific evaluativistic beliefs is observed for the treatment groups. Apart from these findings, results did not differ for any other previously reported analyses with respect to the significance of results or selected target model (see **Tables 4–9** for further details).

Prior Beliefs and Epistemic Change

Exploring the relationship between pre-intervention values, instruction (i.e., treatment groups) and latent change scores, we found that treatment effects were descriptively stronger in the more naive group but that these differences failed to reach significance for all outcome measures (all $p > 0.05$).

DISCUSSION

Effects of Diverging Information on Epistemic Change

Hypothesis 1

Surprisingly, confirmatory analyses revealed no significant group differences between experimental groups. Results suggest that this lack of significant findings is largely due to a profound decrease in topic-specific and domain-specific absolutism and multiplism that takes place in our control group. Overall, this trend toward advanced beliefs in the control group and a decrease in multiplism as well as an increase in evaluativism in the "Unresolvable Read" group are the most important deviations from our a priori expected pattern of results concerning Hypothesis 1 (see **Table 1**). Applying these results to our specific hypotheses H1a, H1b, and H1c, we draw the following conclusions.

Hypothesis 1a

The second part of H1a assumed that the learning strategies task in the control group would not induce epistemic change. As stated above, our data clearly point toward a rejection of this hypothesis as advanced beliefs concerning absolutism and multiplism thrive in the control group. How can we explain this unexpected trajectory? After re-inspecting the materials from our control group, we tend to reframe the learning strategies task, i.e., reading texts on students employing different learning strategies, as a presentation of diverging information on the topic of learning strategies. More specifically, participants may interpret each description of a student employing a learning strategy as a "case study" that introduces a new knowledge claim regarding the efficacy of a certain learning strategy. Hence, this presentation of conflicting knowledge claims might engender a decline of absolute beliefs, while the subsequent task that requires participants to compare these knowledge claims on a set of predefined criteria (the adjunct questions) may trigger an integration of diverging information and, therefore, thwart a change toward multiplistic beliefs. Along these lines, selecting the topic "learning strategies" and this kind of control task may have been ill-fated choices with regard to obtaining significant differences between treatment and control groups because both the gender stereotypes interventions and the learning strategies task are settled in the educational psychology domain. Possibly, our subjects perceived learning strategies to be even more

prototypical for this domain. Therefore, crossover-effects may exist for beliefs on different topics that are settled within the same domain (i.e., learning strategies and gender stereotyping within educational psychology). On the other hand, these "ill-fated choices" opened up a highly interesting new perspective for examining the diverging information paradigm. Based on our control group, we are actually able to compare effects of the mere presentation of any kind of diverging information, to science-based diverging information that was explicitly designed to evoke epistemic doubt and change toward advanced beliefs.

Nonetheless, as a consequence, the actual effect size of examined effects (and thus the power of our tests) that compared effects of gender stereotype interventions to control groups might be lower than expected for H1a. At least the non-significant effects in confirmatory analyses substantiate this theory. In spite of this fact, exploratory analyses introduce some evidence in favor of H1a as they revealed that topic-specific interventions fostered topic-specific epistemic change toward advanced beliefs when compared to the control group (an increase in the D-Index, a decrease in multiplism and an increase in evaluativism). Interestingly, this finding also holds for psychology-specific justification beliefs (a decreased belief in justification by authority in the "Resolvable Read and Write" and the "Unresolvable Read" group, as well as an increased belief in justification by multiple sources across treatment groups).

In conclusion, H1a can be partially confirmed as we observed some kind of treatment effect on five out of eleven outcome variables. Unexpectedly, the control task induced epistemic change toward advanced beliefs but exploratory analyses revealed that change toward advanced beliefs was more prominent for the treatment groups (in particular, evaluativism did only change in these groups). Additionally, treatment group interventions promoted the development of advanced justification beliefs more efficiently, which indicates that the mere presentation of any kind of diverging information does not equally affect all dimensions of epistemic beliefs.

Hypothesis 1b

Contrary to our expectations, changes in evaluativism in the "Unresolvable Read" group were similar to changes in the "Resolvable Read and Write" and "Resolvable Read" groups. Therefore, no significant differences were found for evaluativism between treatment groups. Even more importantly, non-significant effects do not seem to be due to power issues as the "Unresolvable Read" tended to outperform the "Resolvable Read" group—at least on a descriptive level. In a nutshell, our results indicated that epistemic change differed between treatment groups only on one out of eleven outcomes and in this case the observed effect even contradicted the expected pattern of effects (i.e., beneficial effects occurred in the "Unresolvable Read" group). Thus, H1b is completely rejected; the consequences of this will be discussed in the implications section.

Hypothesis 1c

The first part of this hypothesis (efficacy in the "Resolvable Read" group) is strongly connected to H1a and, thus, can be regarded as partially confirmed. A precondition for testing the

second part of this hypothesis ("difference in effects in the "Resolvable Read" and "Resolvable Read and Write" group is small to moderate") in a statistically sound way was that the corresponding target model would have been chosen by LRTs. Unfortunately, this was not the case as chosen target models restricted effects to be equal across groups. Therefore, they did not allow to introduce model constraints on effect parameters of dummy-coded intervention groups or to include differences between those effects as additional parameters in our model (i.e., for testing the hypothesis "difference smaller than value x").

On the other hand, the fact that differences between groups did not become significant based on LRTs implies that overall differences in efficacy cannot be very large because otherwise they would have been detected (as our power analyses indicate). Still, these LRT did not explicitly test the null hypothesis for H1c and descriptive statistics indicate that (small) differences might exist for some outcome measures. In other words, we cannot say for sure if the writing instruction supported epistemic change in our study but we can rule out with some certainty that it was a prerequisite for change. In conclusion, our data tend to confirm the first part of H1c (overall efficacy of the reading task), but are not able to fully test the second part of H1c that pertains to incremental effects of reflecting on diverging information.

Hypothesis 2

Our statistical analysis plan prescribed that H2 (i.e., differences in the efficacy concerning domain–and topic-specific measures) was only examined if differences between experimental groups occurred. Due to the fact that no differences between experimental groups (H1) were found in confirmatory analyses, Hypothesis 2 was not tested in our confirmatory analyses.

However, evidence in favor of this hypothesis stems from exploratory analyses, where significantly stronger effects in topic-specific measures were found for the D-Index and for multiplism (when controlling for covariates). Although findings for evaluativism descriptively confirmed this trend, the corresponding effects failed to reach significance. All in all, we found the hypothesized relationship between effects on topic–and domain-specific measures in two out of three cases, in which it could be meaningfully tested, and, therefore, Hypotheses 2 can be regarded as partially confirmed.

Then again, extrapolating from this notion, we would expect to find even weaker differences between effects in our topic-specific intervention groups and our control group for justification beliefs in psychological science, as this is the highest level-domain investigated by our study (i.e., gender stereotypes are a topic within educational psychology, which represents a subdomain of psychological science). Interestingly, this was not the case. On the contrary, we found effects for justification beliefs that would have been significant according to the criteria of our confirmatory analyses. Hence, different dimensions of epistemic beliefs seem to respond in very distinct ways to various aspects of administered interventions. Possibly, the learning strategies control task is only generalized to educational psychology (as a method within this domain), while the resolvable controversies intervention is generalized to both the topic of gender stereotyping and psychological science as

a whole (because it deals with research findings on gender stereotypes).

Implications and Further Directions

With our first research question, we aimed to create a better understanding of how exactly diverging information affect epistemic change. The findings that we obtained for subjects that received unresolvable controversial information tell a very interesting story in this regard and offer promising starting points for future research. To our surprise, advanced epistemic beliefs (especially justification beliefs) prospered under these circumstances. This is even more remarkable as manipulation check analyses indicated that subjects actually perceived the presented information to be more inconsistent than subjects in the other groups. Why do subjects not regress to simpler multiplistic beliefs when facing this entirely inconsistent information but instead progress to advanced beliefs? Various explanations are conceivable: Possibly, our subjects found some way to integrate conflicting findings and went to great lengths in order to integrate conflicting findings (e.g., by identifying an alternating pattern). Alternatively, they may attribute inconsistencies of presented information solely on the limited amount of information that was offered by our intervention. Especially evaluativists could readily align new information to their existing beliefs by arguing that contextual factors exist but that prior research has, up to now, failed to identify those factors. In accordance with this notion, Rule and Bendixen (2010) argued that schema theory (Anderson et al., 1977) might offer a fruitful framework for understanding the role of prior beliefs in epistemic change. Furthermore, applying our findings to the current situation in psychology (e.g., the replication crisis), one could suggest that ill-structured knowledge does not necessarily hinder individuals' epistemic development after all. Indeed, our results suggest that advanced justification beliefs might prosper under this "climate of contradictoriness." On the other hand, this also implies that our population's prior competence in integrating conflicting knowledge claims might have been distinctively high. Therefore, it may be questionable if our results can be generalized beyond higher education students in psychology— even though existing research on beneficial effects of "standard" diverging information interventions (Kienhues et al., 2016) possibly corroborates our findings. This body of research also includes quasi-experimental studies from other disciplines whose findings are consistent with our observations in the "Unresolvable Read" group. For example, Han and Jeong (2014) showed that epistemic beliefs of (gifted) high school students who planned to major or majored in science and engineering prospered when they attended a Science-Technology-Society education program. In this education program, they were (among others) confronted with dilemmas in engineering and natural science that—just like the unresolvable controversies in our study—could not be resolved within the course. Nevertheless, these unresolvable dilemmas fostered advanced beliefs and moral judgment (Han and Jeong, 2014). As a consequence, future research should examine, which degree of inconsistency fosters epistemic development and from when on it hinders

progress, while paying close attention to the role of prior beliefs and educational background. Conceptual change research on "dissonance producing approaches" (e.g., contrasting common misconceptions to scientists' views) for teaching and their limitations (c.f. Clement, 2013) should provide some valuable input for this purpose.

Concerning our second research question, which aimed at investigating effects of reflecting on diverging information, results are harder to interpret. However, the concept of "epistemic reflexivity" that was introduced by Feucht et al. (2017) as an internal dialog that is focused on "personal epistemologies leading to action for transformative practices in the classroom" (p. 234) might be able to shed some light on the observed pattern of effects. The effects of *reflection* may not be very large because reflecting on diverging information lacks goal-orientation (i.e., the goal of epistemic change was not explicitly given in the writing task instructions). Hence, Lunn Brownlee et al.'s (2017) framework for epistemic reflexivity might be applied when designing future epistemic change interventions in order to ensure that reflection leads to reflexive thinking. Framing the same argument in Bendixen and Rule's model (Bendixen and Rule, 2004; Rule and Bendixen, 2010), one could also reason that subjects' "will" to resolve epistemic doubt (i.e., epistemic volition) may have been insufficient. Since epistemic doubt, epistemic volition and resolution strategies are thought to be part of higher order mechanisms in their model (Rule and Bendixen, 2010), larger effects of reflecting on diverging information might become apparent if subjects' epistemic volition is simultaneously targeted by interventions. Therefore, even though this is somewhat speculative, our results could point to the importance of epistemic volition in epistemic change, an aspect that should be investigated in future research. One way to do so would be the design of intervention components that are tailored specifically to affect *epistemic doubt*, *epistemic volition* or *reflection* and to investigate their incremental effects on epistemic change.

Moreover, our study gave some interesting insights into how effects of topic-specific interventions are generalized—a pressing issue in epistemic change research (cf. Bråten, 2016). In fact, experimental studies often possess a narrow topic-specific scope (cf. Muis et al., 2016) and, therefore, their overall impact on an individual's more general epistemic development may be questionable (cf. Bråten, 2016). With regard to this concern, Kienhues et al. (2008) have argued that topic-related epistemic cognitions can be used to exemplify notions beyond this topic. Thus, their so-called *exemplary principle* predicts that a certain way of dealing with epistemic problems can be transferred when approaching problems in related areas. Our research corroborates to this notion. As could have been predicted by the *exemplary principle*, we found carry over effects within the domain of educational psychology: Topic-specific intervention effects of our gender stereotyping intervention were transferred to domain-specific beliefs and even to higher-level justification beliefs.

Furthermore, the presentation of diverging information on the topic of learning strategies caused an unexpected decrease in absolute beliefs regarding another topic within the same domain (i.e., gender stereotyping within educational psychology). However, not all topic-specific beliefs were equally affected. More specifically, diverging information on learning strategies did not result in significant changes in evaluativism (topic–or domain-specific) nor in justification beliefs. This yields two important implications which pertain to both our first and last research question: First, the generalization of epistemic beliefs seems to depend on the dimension of epistemic beliefs under investigation. Possibly, it is comparatively easy to change beliefs on the structure of knowledge (i.e., certainty and simplicity) by presenting (any kind of) diverging information that is settled within a certain domain. In contrast, changing other belief dimensions (e.g., justification beliefs) might require interventions that are specifically tailored to modify epistemic beliefs. Future research should address this question, where Greene et al. (2008, 2010) distinction between ontological beliefs and epistemic beliefs may prove to be a valuable starting point for this endeavor. Secondly, we saw that evoking doubt regarding absolute beliefs was comparatively easy as we required no didactical concept in order to change those beliefs. Our learning strategies task efficiently reduced topic–and domain-specific absolute beliefs—at least in the short term—even though it was actually designed as a control task. Drawing upon this thought, epistemic change interventions that aim at a simple reduction of absolutism might lack in ambition because individuals are likely to encounter a vast amount of diverging information in their everyday life (in particular in softer disciplines and/or in higher education). Additionally, our findings suggest that these insights might be readily conferred to adjacent domains. However, once more, specific characteristics of our sample have to be taken into account when interpreting these findings and future research should examine if our observed pattern of effects holds in confirmatory studies for other populations.

Limitations

First, one may criticize that findings and conclusions of our study are largely based on exploratory analyses. However, our exploratory analyses modified confirmatory analyses in no substantial way as we derived exploratory analyses and outcomes from our prespecified theory and did not alter our research questions or hypotheses. Instead, we investigated the same questions on a more basic level in order to meaningfully examine if the overall paradigm had worked as intended. Nonetheless, as for all exploratory research, it is the task of future confirmatory studies to validate our findings. Until then, these findings should be cautiously interpreted.

Secondly, the duration of our intervention was rather short. This is particularly true considering the mismatch between intervention duration and length of normative development process that the intervention aims at. However, this is not uncommon for this kind of intervention (cf. Muis et al., 2016) and is indeed well-founded, as this experimental setting allows to disentangle the mechanism of change in the first place. Moreover, to settle the issue of targeting a long-term process by short-term interventions, Ferguson et al. (2012) referred to Vygotsky (1978). Based on his framework, they argued that short-term interventions in an experimental setting might be able

to accelerate or compress development processes that normally require longer periods of time. Nonetheless, long-term effects of those short-term interventions should be investigated in future studies by including follow-up measurements.

Concerning the power of our analyses, the significance criteria might have been chosen too restrictive for some exploratory analyses. We used the standard $p < 0.05$ criteria for likelihood ratio tests although we wanted to inspect one-sided effects in some cases. This procedure was designed to avoid an increased Type I error rate because of multiple testing when comparing effects for multiple treatment groups simultaneously. Unfortunately, the power in the equal group effects model of our exploratory analyses may have been diminished because only one intervention effect is estimated within this model and, thus, multiple testing is not an issue here. As a consequence, in some analyses, we obtained no significant LRT while the (single) parameter estimate would have been significant according to our criteria. Ceiling effects may further contribute to these power issues. However, exploratory analyses revealed that the intervention efficacy did not vary depending on the developmental level of epistemic beliefs. This possibly indicates that all groups were equally affected by ceiling effects (if at all). On the other hand, the existence of those ceiling effects further justifies our choice of the D-Index as exploratory outcome which does not suffer from this issue.

Conclusion

In sum, this study illustrates that many questions remain unanswered when it comes to understanding the relationship between (properties of) diverging information, epistemic doubt and subsequent changes on different dimensions of epistemic beliefs. It shows that evoking doubt regarding absolute beliefs is relatively easy because individuals seem to be skillful in recognizing varying knowledge claims and subsequently averting absolute beliefs. Additionally, we found evidence for the existence of carry-over effects from topic-specific interventions for both higher-level domain-specific beliefs (i.e., beliefs regarding educational psychology and psychological science as a whole) and beliefs pertaining to other topics within the same domain (i.e., effects of the learning strategies task on beliefs on gender stereotyping). In this context and for epistemic change in general, the role of reflecting on presented conflicting information should be thoroughly addressed by future research. Finally, we may need to reconsider our understanding on how individuals acquire and retain evaluativistic beliefs and the role that non-resolvable controversial information play in this development.

DATA AVAILABILITY STATEMENT

The data generated and analyzed for this study can be found in PsychArchives: doi: 10.23668/psycharchives.930.

AUTHOR CONTRIBUTIONS

TR and MK conceived, planned and preregistered the experiment based on a project proposal written by TR. MK conducted the study, analyzed the data and prepared the first draft of the manuscript. TR reviewed critically, revised the article and supervised the project.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *Am. Psychol.* 57, 1060–1073. doi: 10.1037//0003-066X.57.12.1060

Anderson, R. C., Reynolds, R. E., Schallert, D. L., and Goetz, E. T. (1977). Frameworks for comprehending discourse. *Am. Educ. Res. J.* 14, 367–381. doi: 10.3102/00028312014004367

Barzilai, S., and Chinn, C. A. (2017). On the goals of epistemic education. Promoting apt epistemic performance. *J. Learn. Sci.* 1, 1–37. doi: 10.1080/10508406.2017.1392968

Barzilai, S., and Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learn. Instruct.* 36, 86–103. doi: 10.1016/j.learninstruc.2014.12.003

Barzilai, S., and Ka'adan, I. (2017). Learning to integrate divergent information sources. The interplay of epistemic cognition and epistemic metacognition. *Metacogn. Learn.* 12, 193–232. doi: 10.1007/s11409-016-9165-7

Barzilai, S., Tzadok, E., and Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts. Pathways from views of knowing to written argumentation. *Instruct. Sci.* 43, 737–766. doi: 10.1007/s11251-015-9359-4

Bendixen, L. D. (2016). "Teaching for epistemic change in elementary classrooms," in *Handbook of Epistemic Cognition*, eds. J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge), 281–299.

Bendixen, L. D., and Rule, D. C. (2004). An integrative approach to personal epistemology. A guiding model. *Educ. Psychol.* 39, 69–80. doi: 10.1207/s15326985ep3901_7

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., and Schwarz, N. (1994). Need for cognition. Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben [Need for cognition: a scale measuring engagement and happiness in cognitive tasks]. *Zeitschr. für Sozialpsychol.* 25, 147–154.

Bråten, I. (2016). "Epistemic cognition interventions: Issues, challenges, and directions," in *Handbook of Epistemic Cognition*, eds. J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge), 360–371

Bråten, I., Ferguson, L. E., Strømsø, H. I., and Anmarkrud, Ø. (2013). Justification beliefs and multiple-documents comprehension. *Eur. J. Psychol. Educ.* 28, 879–902. doi: 10.1007/s10212-012-0145-2

Bråten, I., Ferguson, L. E., Strømsø, H. I., and Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue. Relations between epistemic cognition while reading and sourcing and argumentation in essays. *Br. J. Educ. Psychol.* 84, 58–85. doi: 10.1111/bjep.12005

Bråten, I., and Strømsø, H. I. (2010). When law students read multiple documents about global warming. Examining the role of topic-specific beliefs about the nature of knowledge and knowing. *Instr. Sci.* 38, 635–657. doi: 10.1007/s11251-008-9091-4

Bromme, R., Kienhues, D., and Porsch, T. (2010). "Who knows what and who can we believe? Epistemological beliefs are beliefs about knowledge (mostly) to be attained from others," in *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice*, eds L. D. Bendixen and F. C. Feucht (Cambridge, UK: Cambridge University Press), 163–193

Buehl, M. M., and Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief Profiles. *Am. Educ. Res. J.* 42, 697–726. doi: 10.3102/00028312042004697

Chinn, C. A., Buckland, L. A., and Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition. Arguments from philosophy and psychology. *Educ. Psychol.* 46, 141–167. doi: 10.1080/00461520.2011.587722

Clement, J. J. (2013). "Roles for explanatory models and analogies in conceptual change," in *International Handbook of Research on Conceptual Change*, *2nd Edn*, ed S. Vosniadou (Hoboken: Taylor and Francis), 412–446.

Dole, J. A., and Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educ. Psychol.* 33, 109–128. doi: 10.1080/00461520.1998.9653294

Ferguson, C. J. (2009). An effect size primer. A guide for clinicians and researchers. *Professional Psychol.* 40, 532–538. doi: 10.1037/a0015808

Ferguson, L. E., and Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs. Changes and relations to multiple-text comprehension. *Learn. Instruct.* 25, 49–61. doi: 10.1016/j.learninstruc.2012.11.003

Ferguson, L. E., Bråten, I., and Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence. A think-aloud study. *Learn. Instruct.* 22, 103–120. doi: 10.1016/j.learninstruc.2011.08.002

Feucht, F. C., Lunn Brownlee, J., and Schraw, G. (2017). Moving beyond reflection. Reflexivity and epistemic cognition in teaching and teacher education. *Educ. Psychol.* 52, 234–241. doi: 10.1080/00461520.2017.1350180

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., and Nagengast, B. (2017). Assessing task values in five subjects during secondary school. Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemp. Educ. Psychol.* 48, 67–84. doi: 10.1016/j.cedpsych.2016.09.003

Greene, J. A., Azevedo, R., and Torney-Purta, J. (2008). Modeling epistemic and ontological cognition. Philosophical perspectives and methodological directions. *Educ. Psychol.* 43, 142–160. doi: 10.1080/00461520802178458

Greene, J. A., Cartiff, B. M., and Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. *J. Educ. Psychol.* 110, 1084–1111. doi: 10.1037/edu0000263

Greene, J. A., Torney-Purta, J., and Azevedo, R. (2010). Empirical evidence regarding relations among a model of epistemic and ontological cognition, academic performance, and educational level. *J. Educ. Psychol.* 102, 234–255. doi: 10.1037/a0017998

Han, H., and Jeong, C. (2014). Improving epistemological beliefs and moral judgment through an STS-based science ethics education program. *Sci. Eng. Ethics* 20, 197–220. doi: 10.1007/s11948-013-9429-4

Han, H., Kim, J., Jeong, C., and Cohen, G. L. (2017). Attainable and relevant moral exemplars are more effective than extraordinary exemplars in promoting voluntary service engagement. *Front. Psychol.* 8:283. doi: 10.3389/fpsyg.2017.00283

Hofer, B. K., and Pintrich, P. R. (1997). The development of epistemological theories. Beliefs about knowledge and knowing and their relation to learning. *Rev. Educ. Res.* 67, 88–140. doi: 10.3102/00346543067001088

Kammerer, Y., Amann, D. G., and Gerjets, P. (2015). When adults without university education search the Internet for health information. The roles of Internet-specific epistemic beliefs and a source evaluation intervention. *Comput. Hum. Behav.* 48, 297–309. doi: 10.1016/j.chb.2015.01.045

Kienhues, D., Bromme, R., and Stahl, E. (2008). Changing epistemological beliefs. The unexpected impact of a short-term intervention. *Br. J. Educ. Psychol.* 78, 545–565. doi: 10.1348/000709907X268589

Kienhues, D., Ferguson, L. E., and Stahl, E. (2016). "Diverging information and epistemic change," in *Handbook of Epistemic Cognition*, eds. J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge), 318–330.

Kienhues, D., Stadtler, M., and Bromme, R. (2011). Dealing with conflicting or consistent medical information on the web. When expert information breeds laypersons' doubts about experts. *Learn. Instruct.* 21, 193–204. doi: 10.1016/j.learninstruc.2010.02.004

Klopp, E., and Stark, R. (2016). *Entwicklung eines Fragebogens zur Erfassung Domänenübergreifender Epistemologischer Überzeugungen [Development of a Domain-General Epistemological Beliefs Questionnaire]*. Unpublished manuscript, Department of Educational Science, Saarland University, Saarbrücken.

Krettenauer, T. (2005). Die Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen und das Problem der Übertragbarkeit von Interviewverfahren in standardisierte Fragebogenmethoden [Measuring the developmental level of epistemological beliefs and the problem of transfering interview procedures to standardized questionnaire methods]. *Zeitschr. für Entwicklungspsychol. und Pädagog. Psychol.* 37, 69–79. doi: 10.1026/0049-8637.37.2.69

Kuhn, D., Cheney, R., and Weinstock, M. P. (2000). The development of epistemological understanding. *Cogn. Dev.* 15, 309–328. doi: 10.1016/s0885-2014(00)00030-7

Lockwood, P., and Kunda, Z. (1997). Superstars and me. Predicting the impact of role models on the self. *J. Personal. Soc. Psychol.* 73, 91–103. doi: 10.1037//0022-3514.73.1.91

Lunn Brownlee, J., Ferguson, L. E., and Ryan, M. (2017). Changing teachers' epistemic cognition. A new conceptual framework for epistemic reflexivity. *Educ. Psychol.* 52, 242–252. doi: 10.1080/00461520.2017.1333430

Lunn Brownlee, J., Schraw, G., Walker, S., and Ryan, M. (2016). "Changes in preservice teachers' personal epistemologies," in *Handbook of Epistemic Cognition*, eds J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge), 300–317.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annu. Rev. Psychol.* 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612

Merk, S., Rosman, T., Muis, K. R., Kelava, A., and Bohl, T. (2018). Topic specific epistemic beliefs. Extending the theory of integrated domains in personal epistemology. *Learn. Instruc.* 56, 84–97. doi: 10.1016/j.learninstruc.2018.04.008

Muis, K. R., Bendixen, L. D., and Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research. Philosophical and empirical reflections in the development of a theoretical framework. *Educ. Psychol. Rev.* 18, 3–54. doi: 10.1007/s10648-006-9003-6

Muis, K. R., and Duffy, M. C. (2013). Epistemic climate and epistemic change. Instruction designed to change students' beliefs and learning strategies and improve achievement. *J. Educ. Psychol.* 105, 213–225. doi: 10.1037/a0029690

Muis, K. R., Trevors, G. J., and Chevrier, M. (2016). "Epistemic climate for epistemic change," in *Handbook of Epistemic Cognition*, eds J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge,), 331–359.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rosman, T., and Mayer, A.-K. (2018). Epistemic beliefs as predictors of epistemic emotions. Extending a theoretical model. *Br. J. Educ. Psychol.* 88, 410–427. doi: 10.1111/bjep.12191

Rosman, T., Mayer, A.-K., Peter, J., and Krampen, G. (2016a). Need for cognitive closure may impede the effectiveness of epistemic belief instruction. *Learn. Individ. Differ.* 49, 406–413. doi: 10.1016/j.lindif.2016.05.017

Rosman, T., Peter, J., Mayer, A.-K., and Krampen, G. (2016b). Conceptions of scientific knowledge influence learning of academic skills. Epistemic beliefs and the efficacy of information literacy instruction. *Stud. Higher Educ.* 43, 96–113. doi: 10.1080/03075079.2016.1156666

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Rule, D. C., and Bendixen, L. D. (2010). "The integrative model of personal epistemology development: Theoretical underpinnings and implications for education," in *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice*, eds L. D. Bendixen and F. C. Feucht (Cambridge, UK: Cambridge University Press), 94–123.

Schiefele, U., and Jacob-Ebbinghaus, L. (2006). Lernermerkmale und Lehrqualität als Bedingungen der Studienzufriedenheit [Student characteristics and perceived teaching quality as conditions of study satisfaction]. *Zeitschr. für Pädagogische Psychol.* 20, 199–212. doi: 10.1024/1010-0652.20.3.199

Schlink, S., and Walther, E. (2007). Kurz und gut. Eine deutsche Kurzskala zur Erfassung des Bedürfnisses nach kognitiver Geschlossenheit [Short and sweet: a German short scale to measure need for cognitive closure]. *Zeitschr. für Sozialpsychol.* 38, 153–161. doi: 10.1024/0044-3514.38.3.153

Schommer, M. (1993). Epistemological development and academic performance among secondary students. *J. Educ. Psychol.* 85, 406–411. doi: 10.1037/0022-0663.85.3.406

Sinatra, G. M., and Mason, L. (2013). "Beyond knowledge: learner characteristics influencing conceptual change," in *International Handbook of Research on Conceptual Change, 2nd Edn,* ed. S. Vosniadou. (Hoboken: Taylor and Francis), 377–394

Strømsø, H., and Kammerer, Y. (2016). "Epistemic cognition and reading for understanding in the internet age," in *Handbook of Epistemic Cognition*, ed. J. A. Greene, W. A. Sandoval, and I. Bråten (London: Routledge), 231–246.

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

Check for
updates

The Influence of Teaching Approach on Students' Conceptual Learning in Physics

Lucia Bigozzi[1], Christian Tarchi[1]*, Carlo Fiorentini[2], Paola Falsini[2] and Federica Stefanelli[1]

[1] *Department of Education and Psychology, University of Florence, Florence, Italy,* [2] *Centro di Iniziativa Democratica Insegnanti (Center of Teachers' Democratic Initiative), Florence, Italy*

Physics is fundamental to secure future needs for scientific and technological competence (Angell et al., 2004), but many countries experience a drop in students' performances in international assessments (Organisation for Economic Co-operation Development [OECD], 2018), as well as in rates of enrolment in undergraduate programs in scientific disciplines (STEM). Socio-constructivist theories have produced a reforming movement in several educational systems, in particular in the area of sciences, but teacher often consider them an idealistic view of education and do not consider themselves metacognitively competent enough to foster thinking in the classroom. In this study, we investigated the efficacy of different teaching methods on high-school students' conceptual knowledge of physics, after the effect of science-related beliefs and critical thinking skills was controlled. We adopted a mixed-method with sequential design, in which quantitative and qualitative data flow are inter-mixed. In specific, we interviewed four high school physics teachers to identify teaching approaches (qualitative approach) and compared them in terms of efficacy on students' performances (quantitative approach). Four teachers and 77 10th grade students participated. Teachers were interviewed during the school years and asked questions about their teaching experience, their teaching approach (Kang and Wallace, 2005) and their epistemic beliefs (Tsai, 2002). Students performances in Science-related beliefs (Conley et al., 2004), critical thinking (Cornell Critical Thinking Test Level X, Millman et al., 2005), and conceptual knowledge in physics (The Force and Motion Conceptual Evaluation, Ramlo, 2002) were evaluated twice, at the beginning and at the end of the school year. The independent-sample t-tests on pre-test variables did not reveal any statistically significant difference between groups. Results from the complex samples GLM revealed statistically significant differences on post-test scores in conceptual knowledge in physics, after the effect of covariates was controlled. Overall, the study contributes to our understanding on current teaching practices in school, and their effect on students' conceptual understanding of physics concepts.

Keywords: physics, conceptual learning, critical thinking, teaching approach, epistemic beliefs

INTRODUCTION

In several countries there is great concern about students' performances in science. International assessments, such as PISA (Program for International Student Assessment) or TIMSS (Trends in International Mathematics and Science Study) have revealed a high percentage of underachieving students, and a low percentage of excellent performance in science. According to PISA, although students express interest in science topics and recognize that science plays an important role in the world, their performances are not excellent, and greatly depend on how science is taught in their schools (Organisation for Economic Cooperation Development [OECD], 2018). According to TIMSS, no countries show a significant increase in performances in Physics from 1995 to 2015 in students, and only a small percentage reach the high benchmark (Stephens et al., 2016). Not surprisingly, few students enroll in undergraduate programs in scientific disciplines (STEM) compared to other domains, and even adults fail at understanding science-related topics when they are brought to their attention, affecting their decision-making processes.

Thus, it is crucial for research to focus on the influence of how science is taught in schools on students' conceptual learning of science (Bigozzi et al., 2002). In this study, we will focus on Physics, and in specific on high school students' conceptual learning of force and motion. There are several reasons why high school students struggle in learning physics concepts. In specific, in this study, we investigated the effect of students' pre-instructional conceptions of physics, science-related beliefs, and critical thinking.

Conceptual Understanding of Physics in High School

Physics was one of the first areas in which students' pre-instructional conceptions were studied (e.g., McCloskey, 1983; Aretz et al., 2016). Students' pre-instructional conceptions that are deeply rooted in daily life experiences have been defined in several ways (e.g., misconceptions, alternative conceptions, intuitive conceptions, naive conceptions, and the like), and there is a plethora of studies showing that they impair their conceptual understanding of science topics (Vygotsky, 1978; Ramlo, 2008; Bigozzi et al., 2011, 2014; Vosniadou, 2013). Rather than being blank slates, students begin physics with a well-established set of theories grounded on their common-sense beliefs about how the physical world works (Hestenes et al., 1992). If instruction does not take students' pre-instructional conceptions into consideration, it will be almost totally ineffective (Hestenes et al., 1992). Of notice is that conceptual change is domain-specific, that is new information obtained through experience and/or instruction can lead to a specific restructuring in a delimited area of our knowledge. Two fundamental types of conceptual change have been hypothesized: weak and radical restructuring (Carey, 1985). In weak structuring, new information is integrated in pre-existing schemes, causing an increase in the relationships among concepts, but without altering the fundamental attributes; in radical restructuring, new information determines a change in the structure of the individual's concepts and relationships

between concepts. For what concerns the topic of force and motion, prior studies have established that common-sense beliefs are incompatible with Newtonian concepts (Hestenes et al., 1992), calling for radical restructuring as an aim of instruction. An example of radical restructuring in the physical domain of force and motion would be a shift from thinking of force as an entity to thinking of it as even a process (Ramlo, 2002). In the next paragraphs, we will discuss two individual difference variables (i.e., science-related beliefs and critical thinking) and a contextual variable (teaching approach) that have been found to be associated with students' conceptual learning in physics.

Science-Related Beliefs

Students' epistemic beliefs, that is beliefs about the nature of knowledge and knowing, are receiving increased research interest in several domain of knowledge (see Greene et al. (2016)). Since Schommer's (1990) seminal studies, students' epistemic beliefs have been repeatedly associated with conceptual learning in science. Hofer and Pintrich (1997) suggested that four dimensions represent students' beliefs about the nature of knowledge and knowing. The former ones are reflected by the certainty and development dimensions: students vary in the degree to which they believe that there is always a right answer or, conversely, whether there may be more than one answer to complex problems; and in the degree to which they think that theories can evolve and change or not. Students' beliefs about the nature of knowing are reflected by the source and justification dimensions: students vary in the degree in which they believe whether knowledge originates from external authorities or is internally constructed; and in the ways in which they cite evidence and evaluate claims.

Several studies have found a relationship between epistemic beliefs and conceptual knowledge of physics (Stathopoulou and Vosniadou, 2007; Franco et al., 2012). For instance, Stathopoulou and Vosniadou (2007) studied this association in Greek secondary school students in two studies, and found that students with a high epistemological sophistication in physics reported a higher conceptual understanding of physics, as assessed by the Force and Motion Conceptual Evaluation instrument (Thornton and Sokoloff, 1998) than students with a low epistemological sophistication in physics. The authors concluded that sophisticated physics-related epistemological beliefs are necessary but not sufficient for conceptual understanding of physics. Franco et al. (2012) found that when undergraduate students' epistemic beliefs were consistent with the knowledge representation of a physics text about Newtonian laws, they showed better learning than when their epistemic beliefs were inconsistent. Science-related beliefs and conceptual understanding of physics are associated, although the direction of this association is unclear. Mason et al. (2013) investigated the relationships between epistemic beliefs and achievement in science in three age groups (5th, 8th, and 11th graders). They found that for 11th graders the hypothesized model explained a smaller portion of variance in achievement in science as compared to the other age groups. Epistemic beliefs had a direct effect on knowledge in science, which in turn has a direct effect on achievement in science.

Results from a developmental perspective suggested that in 11th grade only mastery goals directly influence domain knowledge. Thus, given the existence of an association between teaching approach and achievement goals (Urdan and Schoenfelder, 2006), it could be expected that in high school a constructivist learning environment may enhance students' knowledge in science by inducing mastery goals, rather than performance goals (e.g., by creating a learning environment in which all ideas are equally useful, rather than asking students about their ideas with the purpose of correcting them).

Critical Thinking

Critical thinking is considered as a necessary component of a 21st-century active citizenship that participates in a pluralistic and democratic society (Angeli and Valanides, 2009). For this reason, the development of this kind of thinking is considered the primary goal of science education (Tiruneh et al., 2017). Critical thinking is a type of reflective thinking, focused on deciding what we should believe or do (Ennis, 1989). A critical thinker needs the skills to identify what is implicit in reasoning and to judge if the basis of an inference is solid or not. According to Ennis (1989), it is possible to decide what to believe through different processes, namely induction, deduction, and value judgment. Each of these processes taps on several critical thinking skills: identifying the source of information, analyzing the credibility of information, comparing new information with prior knowledge, and drawing conclusions based on their critical thinking (Linn, 2000).

Initially, critical thinking was taught as a separate track from other subjects, whereas more recently efforts have been made to embed critical thinking skills within subject matter instruction (Niu et al., 2013; Tiruneh et al., 2014). The relationship between critical thinking and conceptual understanding in science is bidirectional: students need critical thinking skills to understand scientific concepts, but science learning might enhance their critical thinking skills, if the latter are targeted by the teacher and embedded in the curriculum. Successful teaching of CT skills in within the teaching of domain-specific knowledge should result in both, deeper conceptual understanding of the subject and development of critical thinking skills (Tiruneh et al., 2017). Miri et al. (2007) compared a group of high school students who were exposed to teaching strategies designed to enhance critical thinking in science classes to two other control groups, a science one and a non-science one. A mixed method research model was applied: critical thinking was measured at the beginning and at the end of the school year, and teaching strategies promoting critical thinking were identified through semi-structured interviews. According to the results, the experimental group showed a statistically significant improvement in critical thinking skills compared with the control groups. Teaching approach plays a fundamental role in mediating critical thinking improvement over time. The next paragraph will discuss the influence of teaching approach on conceptual understanding in science in general, and physics in specific.

Teaching Approach

In the past half-century, literature on teaching has largely disputed whether students learn more in an unguided or minimally guided environment in which they must discover and construct knowledge, or, conversely, whether they should be provided with direct instructional guidance on discipline-specific concepts and procedures (Kirschner et al., 2006). The debate was initiated by the influence of constructivism on learning, which also produced several minimally guided approaches (e.g., discovery learning, inquiry learning, constructivist learning, and the like, Kirschner et al., 2006). Most of these approaches are implemented in science courses, in which students are asked to discover science laws and principles by acting as scientists (van Joolingen et al., 2005). However, there are several reasons why constructivism, interpreted in this way, is not widely used in educational systems. First, minimally guided environments may induce teachers to reduce the use of important aspects of learning, such as providing feedback. For instance, Zhang (2018) investigated the detrimental effect of withholding answers from students and found that students involved in hands-on activities with feedback provided achieved better science learning performances than students in hands-on only condition, with answers withheld, and students in the direct instruction condition. The author concluded that withholding answers during inquiry-based learning had hindered students' understandings of concepts, development in reasoning skills, and ability to transfer knowledge to real-life situations. Some authors consider minimally guided constructivist approaches as theoretically incompatible with human cognitive architecture (Kirschner et al., 2006). For instance, working memory is limited, and problem-solving, a central component of constructivist approaches, places a huge load on working memory, which is not available to be used to learn (Kirschner et al., 2006). Moreover, assuming that the way an expert works in a domain is equivalent to the way in which a novice learns in the same domain might be a fundamental error, and research has consistently shown that guided instruction leads to better learning results than unguided instruction does (Kirschner et al., 2006). In a recent study, the beliefs of 87 science teacher about the differences between students' experiments and scientific experiments were collected. According to the results, they considered all experimentation as a kind of science practice; however, these two types of experimentation are also characterized by differences. The three largest dimensions involved in students' experimentation were pedagogical, procedural, and epistemic whereas, for scientific experiments, the major dimensions involved were procedural, epistemic, and materials (Wei and Li, 2017). These results demonstrate how the practice of experimentation should be substantially different when the students are involved in it. Indeed, in students-led experiments it is important that the teachers take into account the pedagogical dimension.

Rather than claiming that constructivist teaching approaches are ineffective, we propose to investigate how constructivist principles can be included in approaches in which the teacher is assigned a fundamental designing and managing role (e.g., guided instruction). As discussed earlier, constructivist theories recognize that students bring to science class pre-instructional conceptions on world phenomena derived from their everyday experiences, and they are not going to revise them if simply exposed to new theories, unless they are provided with reflective

experiences (Boddy et al., 2003). Students' pre-instructional concepts are often viewed by teachers as obstacles to science learning, but they may serve as resources if teachers increase their understanding about the range of possible ideas that students hold about science topics (Larkin, 2012). Socio-constructivist theories encourage teachers to focus more on inquiry (Mortimer and Scott, 2003) and student-centered instructional practices (Schneider et al., 2005). However, inquiry-based activities should be integrated with classroom talk (Mortimer and Scott, 2003). All science teachers recognize the importance of experimentation in teaching, but they often fail at introducing scientific discourse in their classes. Classroom discourse should not just be used as preparation for the experiment or as after-experiment analysis, but rather should be used to foster learning progression, to search for new knowledge or to answer new questions (Bereiter, 1994). Physics classrooms based on progressive discourse greatly increase students' conceptual understanding of physics more than content-centered classrooms do (Bigozzi et al., 2014). Some studies have focused on science teachers' use of laboratory in their teaching practices (Hofstein and Lunetta, 1982) and found that typically lab activities are used more as "a frill" rather than an integrated component of their course (Tobin, 1986; Kang and Wallace, 2005). The laboratory can be used in several ways: to verify a law, in this case, the laboratory is organized as a sequence of hands-on activities, in which students follow guidelines describing each step (how to mount the instrument, how to measure, and the like); or alternatively, students are let free in their inquiry, without any specific instruction; or, finally, the laboratory is perceived as a "break" from classroom lectures.

Finally, science teachers vary in the extent to which they aim at teaching content only or, conversely whether they integrate higher-order skills, such as critical thinking, in their program. Critical thinking is certainly a core component of teachers' professional development, but only a very few teachers succeed at implementing teaching strategies that enhance students' critical thinking skills (Miri et al., 2007). However, the laboratory is not the only hands-on activity that can be used in the classroom. Effective science teaching approaches should include for the science several types of practical activities (or Making). Some examples of Making could be found in the use of ICT in the classroom, in the production of scale models and in the organization of trips with formative aims. Making may augment other forms of learning activities, as the traditional transmission lessons; or it may provide a context for assessing students' understanding of the scientific practices, such as experimental design in laboratory activity (Bevan, 2017).

Constructivism suggests that people's actions are influenced by ideas and theories constructed earlier based on everyday experiences, and this applies to both students and teachers. Thus, teachers' epistemological beliefs have been hypothesized as a central variable influencing teaching approach (Hewson and Hewson, 1987; Tsai, 2002; Kang and Wallace, 2005). There are several aspects of teachers' beliefs that might influence their teaching approach: beliefs about the nature of science, beliefs about how to teach science, and beliefs about how students learn science. A previous study conducted on teachers' epistemological beliefs suggested that most teachers hold a traditional view of

teaching and science according to which science is best taught by transferring knowledge, giving clear and firm concepts to students, and presenting scientific truths and facts (Tsai, 2002).

Hestenes et al. (1992) assembled a large database of test results from a standardized test on force (FCI, Force Concept Inventory), demonstrating two trends: traditional teaching methods, based on lecture and homework, did not lead to substantial improvements in learning the laws regarding force and the motion as measured by the FCI, and interactive engagement generated much more substantial learning gains on the FCI. Many studies have focused on identifying aspects of the teaching approach that influence students' conceptual understanding of physics, but focusing mostly on epistemological beliefs (e.g., Lederman et al., 2002) or testing the efficacy of instructional components (e.g., use of laboratory, Kang and Wallace, 2005), rather than focusing on the overall teaching approach, including epistemological approach, activities implemented, views on the nature of learning, use of laboratory or classroom discussion, and the like.

This Study

In this study, we investigated the association between teaching approach and high school students' conceptual understanding of a physical topic (i.e., force and motion). Rather than testing the efficacy of a research-designed intervention, we wanted to analyze difference between teaching practices influenced by real-life teaching approaches. We also included science-related beliefs and critical thinking as control variables for two main reasons: they are associated with conceptual understanding of physics, thus representing a potentially confounding variable; and a growth in these skills is desirable and an expected effect of a science course. We applied a mixed-method with sequential design (Johnson and Onwuegbuzie, 2004; Creswell and Plano Clark, 2011), according to which a research question is explored with a quantitative and a qualitative method. Data streams are intermixed to benefit from the strengths of both approaches. High school students' conceptual knowledge of physics (i.e., force and motion), science-related beliefs, and critical thinking skills were assessed twice, at the beginning and at the end of the school year (quantitative approach). Physics teachers were interviewed to find similarities and differences in teaching methods (qualitative approach). Finally, the influence of the teaching method on students' growth from the beginning to the end of the school year was investigated (intermixture between quantitative and qualitative approach). We expected to identify two main approaches of teaching physics, one more content-centred and one more student-centered, and we expected the latter approach to foster a higher increase in students' conceptual understanding of physics, science-related skills and critical thinking skills than the former one.

MATERIALS AND METHODS

Participants

The participants of this study were 84 high school students, enrolled in Grade 10 (Age = 15.80 ± 0.43; 59 males and 25 females). Students came from four different classes, from two

different high schools located in a mid-size city in Central Italy. All students spoke Italian as their mother-tongue language. At the time of the study, no participant was diagnosed with a physical or mental disability, was included in a diagnostic process, or identified by the teachers as having special educational needs, thus all participants could be defined as typically developing. The two schools were located in areas characterized by a middle-high socio-economic level. The participating schools were not following any specific program to empower relevant variables for this study and adhered to the national curriculum. This study was carried out in accordance with the recommendations of AIP (*Associazione Italiana Psicologi*, Italian Association of Psychologists) and of the University of Florence, Italy. Ethics approval was not required at the time the research was conducted by the University of Florence. Participants' parents subjects gave written informed consent in accordance with the Declaration of Helsinki (World Medical Association, 2013).

Procedure and Research Design

All physics teachers working in the territory were contacted for a meeting with the researchers, in which the aims of the study were explained. Teachers were eligible to participate in the study if they had a minimum of five years' teaching experience; were teaching Grade 10 at the time of the study; were teaching the concept of force and motion; and were not following any experimentation or specific program at the time of the study, nor had their grade 10 students followed any specific program the year before (grade 9, which in Italy is the first year of high school). Five teachers were considered eligible and expressed interest in participating in the study. During the research, one teacher had to take leave for personal reasons, and consequently was excluded from the data analysis.

At the beginning of the school year, in October, we measured students' performances in conceptual understanding of physics, science-related beliefs, and critical thinking. In the middle of the school year (i.e., March), after the topic of force and motion had already been introduced and concluded in each class, participating teachers were interviewed (interviews were audio-recorded and transcribed for qualitative analysis). At the end of the school year, in May, students' performances in conceptual understanding of physics, science-related beliefs, and critical thinking were measured again. All steps were conducted by a researcher trained by the first and second author of the study.

As a result of the analysis of teachers' interviews, two groups were identified, one applying a student-centered approach (two teachers, 39 students) and one applying a content-centered approach (two teachers, 45 students). All teachers used the laboratory, lectures, and classroom discussion several times during the last school year. However, the order of these teaching components differed, as we will discuss further on in the manuscript. For instance, while the two student-centered teachers claimed to use the laboratory as a starting point of a teaching unit, the content-centered teachers took students to the laboratory after lectures, to apply the theoretical principles addressed in. Another source of differences between the two groups was the use of hands-on activities other than the laboratory. Only the student-centered teachers claimed to use

alternative hands-on activities to replace the laboratory, such as reading of original texts written by past scientists or field-trips. For all the other teaching components (type of exams, material available in the laboratories, syllabus, time allotted to a teaching unit within the course) the four classrooms were equivalent.

Measures
Conceptual Understanding of Physics

This variable was measured through the Force and Motion Conceptual Evaluation (FMCE, Thornton and Sokoloff, 1998), a multiple-choice test of students' conceptual understanding of Newton's Laws of Motion. Scores on the FMCE are strongly related to students' score in the force concept inventory (FCI, Hestenes et al., 1992), a widely applied test measuring students' understanding of one-dimensional kinematics and Newton's laws; two-dimensional motion with constant acceleration; impulsive forces; vector sums; cancellation of forces; and identification of forces. In this study, we opted for the FMCE to measure students' conceptual understanding of physics, as it provides a detailed measure of their understanding of one-dimensional forces and motion (Thornton et al., 2009), the unit of study chosen as a reference to ask teachers about their teaching method. Previous studies had used the FMCE with high school and college students, and proved its validity and reliability (Ramlo, 2008). The FMCE consists of 43 questions, and multiple choices range from five to nine answers. Overall, questions aim at assessing whether students are able to adopt a Newtonian framework or, conversely, rely on everyday experience-based conceptions. Questions use a natural language and graphical representations (e.g., [Questions] 8–10 refer to a toy car which is given a quick push so that it rolls up an inclined ramp. After it is released, it rolls up, reaches its highest point and rolls back down again. *Friction is so small that it can be ignored.*[1] The text is followed by a graphical representation of a car on an inclined ramp facing downwards (Thornton and Sokoloff, 1998, p. 347). The test was translated into Italian by a bilingual researcher, and back-translated by another bilingual researcher. The two versions were compared and no significant differences were found. The Italian version was also expert-validated by two Physics teachers with more than 20 years of experience in teaching high-school students. Minor revisions in wording were suggested, with no semantically or conceptually significant departures from the original version. Students' scores could range between 0 and 43, and reliability scores were $\omega^2 = 0.72$ at the pre-test and $\omega = 0.88$ at the post-test.

Science-Related Beliefs

Students' science-related beliefs were assessed through a self-report instrument (developed by Conley et al., 2004; Italian version by Mason et al., 2010, 2013). The instrument taps four dimensions of science-related epistemological beliefs: source (e.g., "Whatever the teacher says in science class is

---

[1] Italics are used in the original version.

[2] Reliability scores for the conceptual understanding of physics measure were calculated through McDonald's ω, because of the differences in nature between questions in the same instruments (i.e., multiple choices ranging from five to nine answers).

true"), certainty (e.g., "All questions in science have one right answer"), development (e.g., "Sometimes scientists change their minds about what is true in science"), and justification (e.g., "Good answers are based on evidence from many different experiments") through 26 items on a 5-point Likert scale (from strongly disagree to strongly agree). The instrument was originally developed for elementary-school students (Conley et al., 2004), but has also been successfully implemented with high-school students, and proved to be valid and reliable (Tsai et al., 2011). As the main focus of the present study was not epistemological beliefs, we calculated a total score to assess students' overall sophistication in science-related beliefs. Students' scores could range between 26 and 130, and reliability scores were $\alpha = 0.75$ at the pre-test and $\alpha = 0.83$ at the post-test.

Critical Thinking

Students' critical thinking was assessed through the Cornell Critical Thinking Test – Level X (CCTT; Millman et al., 2005). The test includes 71 multiple-choice items (three alternatives) and assesses the following skills: hypothesis-testing skills, credibility of source and observation skills, deduction skills, and assumption identification skills. The test is delivered in a narrative context, in which students follow the events of a group of explorers that landed on a planet to find out what happened to the first group of explorers. An example of item was as follows: "You are given two reports, you have to read them both and decide whether one of them is more credible than the other. (A) The mechanic analyses the rivers around the village and reports, the water is not drinkable; (B) the medical officer says, we still cannot tell whether the water is drinkable; (C) A and B are equally credible." In this case, the right answer is B, since the medical officer should have more expertise on drinkable water than the mechanic has). The test was translated into Italian by a bilingual researcher and back-translated by another bilingual researcher. The two versions were compared, and no significant differences were found. The Italian version was also expert-validated by two teachers with more than 20 years of experience in teaching to high school students. Students' scores could range between 0 and 70, and reliability scores were $\alpha = 0.76$ at the pre-test and $\alpha = 0.82$ at the post-test.

Semi-Structured Interview

Teachers were interviewed at a time agreed with them by a trained researcher, with no prior relationship with the teachers. At the time, when the interviews were conducted during the school-year, thus before post-test. Thus, at the time in which they were conducted, students' gains in target variables were still unknown. Interview duration ranged between 45 min and 1 h. Interviews were audio-recorded and transcribed. The semi-structured interview included three sections: teaching experience and program, teaching method, and epistemological beliefs (see **Supplementary Material** for the full semi-structured interview).

In the first section, we asked questions to establish the teacher's expertise (e.g., "how long have you been teaching for?") and to establish equivalence in the Physics program delivered during the school year ("What was the Physics program this year, and in specific what topic related to force and motion did you discuss with the students?"). We used the first part of the semi-structured interview in order to collect objective data about the teaching practices implemented by the participating teachers.

The second section aimed at identifying teaching approaches, and questions were derived from Kang and Wallace's (2005) study. Teachers were invited to think about how they taught physics in general, and the topic of force and motion in specific, and to describe a typical lesson. Then teachers were asked questions on the use of laboratory (e.g., "What roles do you believe lab activities play in your teaching?"); the use of group work and individual activities, and how often they used discussion to stimulate learning. Finally, teachers were asked which technique was more effective and which one was less effective in promoting conceptual understanding.

The last section aimed at investigating teachers' epistemological beliefs of science, teaching science and learning science. Questions about beliefs of science [e.g., "After scientists have developed a scientific theory (e.g., atomic theory, evolution theory), does the theory ever change?"] were derived from Lederman et al.'s study (2002), whereas questions about beliefs of teaching (e.g., "Could you describe what an ideal science teaching environment would look like?") and learning science (e.g., "What do you think about the responsibilities of students when learning science?") were derived from Tsai's study (2002).

Transcripts of teachers' semi-structured interviews were investigated through thematic analysis, "a method for identifying, analyzing, and reporting patterns (themes) within data" (Braun and Clarke, 2006, p. 6). Thematic analysis allows one to search for themes across the entire data set, rather than within a data item (Braun and Clarke, 2006). In this study, we searched themes across interviews, rather than, for example, counting the frequency of specific aspects within each interview. Transcripts were analyzed after the post-test stage, but before students' score were analyzed, thus coders were blind towards post-test group differences.

RESULTS

Descriptive results are presented in **Table 1**. An analysis of students' conceptual understanding of physics reveals very low scores at the pre-test, and an increase at the post-test, although students' performances are still far from full mastery of Newtonian perspective.

The analysis of correlational scores showed that initial levels of critical thinking were associated with conceptual understanding of physics as assessed at both time points. Science-related beliefs were associated with critical thinking skills at both time points, but they were not associated with conceptual understanding of physics. Each variable at the post-test was associated with the initial performance as assessed at the pre-test.

Teaching Approach

The qualitative analysis of teachers' interviews, carried out with the method of thematic analysis, revealed the existence of two main teaching approaches: one defined as "guided-constructivism approach" (GCA) characterized by a focus

TABLE 1 | Descriptive results (minimum and maximum scores, mean, and standard deviation) for the total sample ($n = 84$) and divided by group (student-centered $N = 39$, teacher-centered $N = 45$), and correlation among variables for the total group.

		Total					GCA		CCA							
		N	Min	Max	M	SD	M	SD	M	SD	1	2	3	4	5	6
1	Beliefs – 1	84	60	98	77.38	7.24	77.08	8.04	77.63	6.56	1					
2	Critical thinking – 1	84	33	56	46.15	5.24	45.26	5.30	46.85	5.15	0.32**	1				
3	Physics – 1	84	2	23	9.29	4.66	9.35	4.66	9.30	4.84	0.11	0.30**	1			
4	Beliefs – 2	84	86	125	108.03	8.43	107.84	10.16	108.17	7.02	0.57**	0.27*	0.05	1		
5	Critical thinking – 2	84	22	62	46.07	7.08	44.74	5.71	47.00	7.84	0.11	0.41**	0.19	0.35**	1	
6	Physics – 2	84	2	38	12.33	7.24	14.91	7.41	10.70	6.82	0.09	0.24*	0.41**	0.16	0.10	1

$**p < 0.01$; $*p < 0.05$. GCA, guided-constructivism approach; CCA, content-centered approach.

TABLE 2 | A comparison chart between significant dimensions derived by interviews to GCA and CCA teachers.

Dimension	Guided-constructivism approach (GCA)	Content-centered approach (CCA)
Teachers' epistemological beliefs of science	Sophisticated beliefs	Sophisticated beliefs
Goal of instruction	Explaining phenomena that students can observe in everyday life	Transmitting the skills necessary for a theoretical and abstract understanding of physical phenomena
Start of teaching unit	Laboratory (and students' direct experience) as the start of the teaching unit	Lecture as the start of the teaching unit
Laboratory	It is the ideal place to make students dissatisfied with their naive theories and needs to be coupled with progressive discussion. Actually, they go beyond the distinction between laboratory and classroom teaching, with these two becoming mere physical places, rather than methods. One could have a laboratory in a classroom, or lecture in the laboratory.	It is a place suitable for group work, in which students can work together on experiments.
Students' pre-instructional conceptions	Pre-instructional conceptions made by the students represent the starting point of a lesson	Instruction should aim at correcting students' pre-instructional conceptions
Students' participation in the classroom discourse	Laboratory is not always effective in eliciting students' conceptions about physics topics, and field-trips as well as videos or reading of original writings by past scientists should be used to relate physics concepts to real life experiences.	CCA teachers simply rely on the laboratory and do not implement alternative participatory activities in their teaching of physics.
Students' role	Students should have an active role to students in the deconstruction of naive schemes and conceptual understanding of physics concepts	Students should be motivated to learn.
Teacher's role and guidance	Lectures, laboratory activity and classroom discussion should be guided by the teacher. Overall, students need to be guided throughout all the steps of science learning. Teachers need to be aware of students' pre-instructional conceptions and guide them through scientifically valid conceptions.	Teachers should explain the theory and set up experiments for students

on students' conceptualization and guided by teachers; and another one defined as "content-centered approach" (CCA) and characterized by a traditional teaching approach (see **Table 2** for a comparative chart). Interestingly, the two teaching methods differed for characteristics of the teaching method used, but not for educators' science-related beliefs. Indeed, from the analysis of the questions derived from the teachers' epistemological beliefs of science (Lederman et al., 2002), no substantial difference emerged, and teachers expressed similar views on the nature of science. Teachers' epistemological beliefs are a central variable influencing teaching approach (Hewson and Hewson, 1987; Tsai, 2002; Kang and Wallace, 2005), and a potentially confounding variable in this study. Use of laboratory, classroom discussion,

attribution role to students, and teachers may depend on their beliefs about the nature of science, but in this study all teachers are considered equivalent. Moreover, teachers' sophisticated epistemological beliefs are probably associated to their students' growth in science-related beliefs.

Conversely, from the qualitative analysis of the questions on the teaching approach (derived from Tsai, 2002; Kang and Wallace, 2005), significant differences emerged.

Although all four teachers showed the use of some typical constructivist teaching techniques, the two GCA teachers claimed to have a more substantial focus on the conceptual construction of concrete meanings of physics. In other words, the teachers aimed at explaining phenomena that students can observe in

everyday life. The two CCA teachers stated that in the classroom they aimed at transmitting the skills necessary for a theoretical and abstract understanding of physical phenomena, without explicitly mentioning the importance for the students to become able to understand the real phenomena (see Q1 in **Table 3**). Importantly, no teacher mentioned the explicit teaching of higher-order skills, which is probably associated to the lack of growth in critical thinking skills over the school year in the students of our sample (Miri et al., 2007).

Most teachers would say that the laboratory is important in science teaching, but they might differ in the role attributed to it in their lesson plan. Moreover, their actual use of laboratory in their teaching practices might depend on availability of instruments and thus, change from school to school. Therefore, we asked teachers to describe their ideal teaching approach. As a consequence of this focus on concrete concepts, GCA teachers affirmed that the ideal teaching method should be to start from the laboratory and from students' direct experience of the physical phenomena that are going to be discussed in classroom. CCA teachers consider the laboratory as an important aspect of physics teaching too, but they believe that the starting point of teaching should be the lecture. Whereas students' pre-instructional concepts are often viewed by teachers as obstacles to science learning, GCA teachers consider them as resources to successfully begin a teaching unit (Larkin, 2012). CCA teachers see the laboratory as a place suitable for group work, whereas GCA teachers believe that the laboratory is the ideal place to make students dissatisfied with their naive theories, thus provoking in them the cognitive dissonance necessary to motivate them towards change and learning of correct concepts on the phenomena of physics (see Q2 in **Table 3**). GCA teachers organize the laboratory activity in brief, qualitative observations, always fostering individual reflection and collective discussion, acknowledging the importance of integrating inquiry-based activities with classroom talk (Mortimer and Scott, 2003). In this way, they go beyond the distinction between laboratory and classroom teaching, with these two becoming mere physical places, rather than methods. The laboratory, as well as the classroom discussion, are seen by constructivist teachers as moments where it is possible to start from students' mistakes to help them construct correct theories on empirical phenomena. In this perspective, GCA teachers, unlike CCA ones, believe that the mistakes made by the students represent the starting point of a lesson. Pre-instructional concepts, should be valorized, rather than corrected, stimulating in the student the cognitive reasoning preliminary to conceptual change and learning (see Q3 in **Table 3**). Of notice, students derive their pre-instructional concepts from everyday experiences and will not revise them if simply exposed to new theories, unless they are provided with reflective experiences (Boddy et al., 2003).

GCA teachers also use several means to foster students' participation in the classroom discourse on science. Laboratory is not always effective in eliciting students' conceptions about physics topics, and field-trips as well as videos or reading of original writings by past scientists (who sometimes struggled with the same pre-instructional conceptions our students hold) might support teachers in this step (see Q4 in **Table 3**).

Most of the time, CCA teachers simply rely on the laboratory and do not implement alternative participatory activities in their teaching of physics, although prior studies have shown that physics classrooms based on progressive discourse greatly increase students' conceptual understanding of physics more than content-centered classrooms do (Bigozzi et al., 2014).

Another important source of differences between the two teaching approaches identified in this study lies in what role teachers attribute to students in science learning. Whereas CCA teachers attribute a motivational role to students, GCA teachers emphasize the importance of attributing an active role to students in the deconstruction of naive schemes and conceptual understanding of physics concepts (see Q5 in **Table 3**).

GCA teachers are aware that students need to be guided throughout all the steps of science learning. Teachers need to be aware of students' pre-instructional conceptions and guide them through scientifically valid conceptions. As such, also the laboratory activity and classroom discussion should be guided by the teacher (see Q6 in **Table 3**). Past studies have shown that constructivist approaches with minimally guidance by the teacher are not effective in promoting conceptual understanding of scientific concepts (Kirschner et al., 2006; Zhang, 2018).

Groups' Equivalency at the Beginning of the School Year

To determine the equivalence between groups at the beginning of the school year, we conducted a series of t-tests for independent samples, with group as independent variable and pre-test scores in science-related beliefs, critical thinking and conceptual understanding of physics as dependent variables. We used the two one-sided tests (TOST) procedure for testing equivalence (Lakens, 2017). Whereas traditional t-tests for independent samples allow to refuse the null hypothesis, the TOST procedure allows to verify equivalence between means. No significant differences emerged, so we could conclude that the two groups were equivalent (see **Table 4**).

Effect of Teaching Approach on Students' Gains

Research hypotheses were explored through a generalized linear model for complex samples (complex-samples GLM) conducted with the software IBM SPSS version 19. Complex-samples GLM allows to control the effect of data nested within clusters (in our case, classrooms), and thus to test group differences with adjustment for clustering by classrooms (Aerts et al., 2002). Educational studies have often do deal with clustered data. Clustered data arise when the data from the whole study can be classified into a number of different groups, referred to as clusters, and observations within a cluster are more alike than observations from different clusters (Galbraith et al., 2010). Each cluster contains multiple observations, giving the data a "nested" or "hierarchical" structure, with individual observations (i.e., students) nested within the cluster (i.e., classrooms). Modeling approaches are particularly useful when there are other covariates that need to be included in the analysis (Galbraith et al., 2010), such as in the case of the present study.

TABLE 3 | Excerpts from participating teachers' interviews representing differences in teaching approach.

Questions	GCA teacher #1	GCA teacher #2	CCA teacher #1	CCA teacher #2
(Q1) How would you describe your teaching approach in Physics?	Contextualizing the concepts and showing how the theories have developed in history to understand everyday aspects of life.	Showing students how things work through knowledge of Physics concepts.	I try to get students accustomed to a theoretical dissertation a bit abstract	General dissertations, without being too specific
(Q2) What is the ideal teaching approach to Physics	Having the class in the laboratory, because facts and phenomena, the things you observe and perceive should be the starting point; because if we use book pages or notes as the starting point, students simply repeat the lesson and they do not even know what they are talking about.	What I noticed is that if you match a theoretical class with a practical one, learning is more enduring. The ideal method is to be in the laboratory, a laboratory fully equipped, rather than having a whiteboard class, and start a topic from a practical situation that students can perceive.	Matching individual studying with a theoretical discussion in the classroom	Lecture and subsequent discussion in classroom aiming at correcting errors
(Q3) When is classroom discussion productive	It is productive when we try to emphasize the most important claims, but also the incorrect ones. We actually linger on the incorrect claims, we make them emerge and become more explicit. Large part of classroom discussion consists of starting from making students aware of their common-sense beliefs, and transforming them into scientific conceptualization.	It is productive when there is participation, including also the students who generally are more resistant to participate. It is productive when I ask some questions and students try to answer them, even if their answers are wrong I still consider the discussion productive, as it opens up an issue, and represents a starting point.	It is productive when students take turns in speaking and are all involved.	In a classroom discussion, I am the one asking questions. I let students answer them, stressing the right one, and correcting the wrong ones.
(Q4) What activities do you consider most successful in your teaching approach to elicit students' understanding of physics of force and motion?	Reading of original texts written by past scientists, collective discussion of everyday physics-related aspects of life, images or videos.	Sometimes I take students out of the classroom into contexts in which physics concepts can be observed. Once I took them to D'Azeglio square to examine the oscillating motion of swings. Once, a student shot a video of angular motion (with him sitting on a stool) and we watched it and discussed it in class.	I do not implement any particular activity with the topic of force. With other topics we might do some field-trips, but not very often.	99% of the time I teach through lectures.
(Q5) What is the students' role in science learning?	Students should synthesize what we discuss together in the classroom. They should elaborate on the topics and reach a deep understanding of concepts. A big challenge is to remove common-sense beliefs about force and motion, although they are persistent. I let the knowledge be constructed by them; I think that learning is effective when they reach a conclusion on their own. Once they reach a conclusion and revise a common-sense belief, the teacher's job is to underline what has happened and make sure that students are aware about the conceptual change that took place.	Students' responsibility is to elaborate on what they have empirically experienced, regardless of whether they experienced it incidentally or in school. The teachers should promote this type of reflection.	Students need to feel involved in the classroom	Students should find the motivation to get engaged.

(Continued)

TABLE 3 | Continued

Questions	GCA teacher #1	GCA teacher #2	CCA teacher #1	CCA teacher #2
(Q6) What is the teachers' role in science learning?	An important task is to remove students' common-sense conceptions, and transform them into scientifically correct Newtonian conceptions. In the laboratory, it is important that the teacher stimulates certain reflections. Sometimes I conduct the experiment, or I have the students conduct it in small groups, and I go around to be sure that they are reflecting During classroom discussion, the teacher should capture the most significant aspects, and synthesize them. My role is that of a guide. Students can briefly discuss among themselves, but they should also be led towards a collective discussion, which needs to be guided to be effective.	The teacher needs to put the students in condition to practice concepts; the teacher should not just talk about them. The laboratory activity means asking a question to students. Students need to be guided as they should always be aware of what they are doing and why they are doing certain things. My role is to communicate what we do and the main problem we are inquiring. I participate in their work, but I also observe, overall, they need to be masters of the activity.	My main responsibility is that they feel involved. I ask them to stop me when they do not understand something. Then, they need to proceed on their own, and elaborate knowledge rather than learning concepts by heart.	Three functions, (1) explaining what they need to do; (2) guide them through theory; and (3) be sure that all the safety norms in the laboratory are respected

TABLE 4 | Pre-test differences between groups: Results from TOST Independent Samples *t*-test.

	t	*Df*	*p*	95% LCI	95% UCI
Beliefs – 1	−0.20	74.9	0.84	−3.51	2.76
Critical Thinking – 1	−1.28	63.7	0.21	−3.68	0.50
Physics – 1^{+}	0.05	69.4	0.73	−0.07	0.10

$^{+}$*Variable normalized through monotonic transformation.*

TABLE 5 | Results from the complex samples GLM.

Dependent variables	Parameters	*df*	Wald's *F*	*p*
Physics – 2^{+} [$R^2 = 0.39$]	Group	1, 3	15.50	0.03
	Physics – 1^{+}	1, 3	7.95	0.07
	Beliefs – 1	1, 3	0.35	0.60
	Critical thinking – 1	1, 3	8.72	0.06
Beliefs – 2 [$R^2 = 0.43$]	Group	1, 3	8.40	0.06
	Physics – 1^{+}	1, 3	0.15	0.73
	Beliefs – 1	1, 3	15.61	0.03
	Critical thinking – 1	1, 3	6.67	0.08
Critical thinking – 2 [$R^2 = 0.32$]	Group	1, 3	0.31	0.62
	Physics – 1^{+}	1, 3	1.25	0.35
	Beliefs – 1	1, 3	0.17	0.71
	Critical thinking – 1	1, 3	18.91	0.02

$^{+}$*Variable normalized through monotonic transformation.*

Classroom was included as cluster variable to account for random effects. Group was included as factor to analyze differences between teaching approaches in post-test scores. Outcome variables were post-test scores in science-related beliefs, critical thinking, and conceptual understanding of physics. Pre-test scores in science-related beliefs, critical thinking, and conceptual understanding of physics as covariates, to account for initial differences.

The group variable explained post-test performances in conceptual understanding of, but not in science-related beliefs. Post-test scores in science-related beliefs, critical thinking, and conceptual understanding of physics were all associated to their respective pre-test scores. Post-test scores in conceptual understanding of physics were also associated to pre-test scores in critical thinking (see **Table 5**).

Overall, both teaching approaches are effective in promoting growth in science-related beliefs, probably because of teachers' sophisticated epistemological beliefs. Neither of the two teaching approaches are effective in promoting students' critical thinking skills, probably because they fail at embedding explicit teaching of higher order skills in their teaching practices (Miri et al., 2007). Whereas both teaching approaches may be effective in promoting a learning of theoretical principles and laws, the GCA approach is more successful in promoting conceptual understanding of physics concepts.

DISCUSSION

This study contributes to our understanding of teaching approaches to physics in high school, and how they are associated

with students' conceptual learning of force and motion. To evaluate the teaching method and in order to understand which characteristics of it could predict students' performance, we implemented a qualitative thematic analysis of the teachers' interviews. The semi-structured interview investigated teachers' teaching approach about physics in general and the topic of force and motion in specific (i.e., use discussion, laboratory, individual and group work in their teaching practices, their epistemological beliefs about science, and their epistemological beliefs about teaching science. The questions were derived from past studies (Lederman et al., 2002; Tsai, 2002; Kang and Wallace, 2005).

Past studies have suggested that teachers' epistemological beliefs about science play an important role in their teaching practices in the classroom (Hewson and Hewson, 1987; Lederman et al., 2002; Tsai, 2002; Kang and Wallace, 2005), but in our study teachers held similar views on the nature of science, allowing us to consider them equivalent in epistemological beliefs on science, and focus our analysis on the teaching approach only. Teachers might hold sophisticated beliefs about the nature of science, but these do not automatically transfer to their practices (Yoon and Kim, 2016). Moreover, teachers reported similar teaching practices, which are generally associated to general principles of constructivism (use of laboratory, importance of discussion, assigning an active role to students, and the like). For example, all teachers affirm that when they teach physics to students they start to explain to them the real events that each student knows. In other words, all teachers explained the physics concepts starting from students' experiences, and this is an important aspect in the constructivist method (Mortimer and Scott, 2003; Bigozzi et al., 2014). Thus, on surface, all teachers believed that they were teaching according to constructivist principles. Differences emerged when teachers were asked about their practices when teaching about force and motion (Mansour, 2009), that is, when their teaching approach was inquired more in depth. The thematic analysis revealed the presence of two main teaching approaches, one defined as guided-constructivism approach, and the other one as content-centered approach. The two GCA teachers attributed a specific role to the laboratory, an integrated component in the teaching practice in which students can have experience of their own beliefs, rather than using it as "a frill" (Tobin, 1986; Kang and Wallace, 2005). In this study, GCA teachers assigned a seminal role to the laboratory, as it gives rise to the whole teaching module. Moreover, the laboratory setting allows teachers to guide also the moment in which students become aware of their own and each other's pre-instructional conceptions. Some teachers interpret constructivist teaching as unguided teaching, but this approach has been demonstrated to be ineffective (Kirschner et al., 2006). Rather than having students discover laws and principles by improvising as scientists (van Joolingen et al., 2005), they should be guided in each step of the inquiry, and supported to become aware of their own conceptions and provided with a reflective experience (e.g., experiment in the laboratory, field-trip, video, historical readings, and the like) on the perceived phenomena (Boddy et al., 2003). For instance, a laboratory activity should also be integrated with classroom talk (Bereiter, 1994; Mortimer and Scott, 2003). The laboratory

activity should produce knowledge in students. Laboratory activity and classroom lectures should not be considered as distinct moments. In choosing which experiment to engage students with, the teacher needs to choose one that might create cognitive dissonance in the students, make them ask questions, and foster desire of knowing. GCA teachers asked students a disposition towards conceptual change in a guided environment, rather than being the only agent of such a conceptual change. CCA teachers tend to value students' performance in terms of conformity to the criteria of the discipline, and consider the evaluation and correction of the learner's conceptualization as the main teacher's task (Mansour, 2009).

The analysis of post-test scores revealed that the GCA teachers' students outperformed the CCA teachers' students in conceptual understanding of force and motion at the end of the school year, after checking on the effects of conceptual understanding of force and motion, critical thinking and science-related beliefs at the beginning of the school-year. No group differences were found for post-test science-related beliefs or critical thinking. An analysis of descriptive scores shows that critical thinking does not increase from pre-test to post-test, whereas science-related beliefs appear to improve in both groups. Thus, the reason why the teaching approach does not influence these variables might differ. Critical thinking might not improve as neither of these two approaches explicitly targets it. One might expect an improvement in critical thinking skills as an ancillary effect of GCA, but teachers might need to increase guidance in this direction. For instance, exposing students to the original writings of famous physicists might improve their conceptual understanding of physics, but unless these writings are compared with non-authoritative writings, students do not reflect on the differences between scientists' approach to a problem versus laypeople's approach to it, and do not practice (or improve) observation and credibility of sources skills. For what concerns science-related beliefs, they appear to improve in both groups over the school year, so the two teaching approaches might be equivalent in their efficacy.

In conclusion, secondary school teaching can be meaningful if a balance between experimentation and observation, historical contextualization, use of videos and simulation, is achieved. Of course, such a balance must take into consideration the school resources. The simultaneous and balanced use of all these methodological instruments allow one to create classrooms based on scientific knowledge construction, within which textbooks are just one (and not the main) of the several learning aids. Rather than transmission of knowledge by the teacher, physics teaching should be characterized as a shared construction of knowledge, which is the result of the collective synthesis of a learning process that has been organized and guided by the teacher.

Limitations and Directions for Future Research

When interpreting the findings of the current study, some limitations should be taken into account. First, the focus of this

study was on the teachers' perception of their own teaching approach, and to what extent these differences are accountable for variance in students' conceptual understanding of physics. However, prior studies have emphasized the existence of a gap between what teachers think constructivism is, and the way in which they actually teach (Mortimer and Scott, 2003). Thus, future research should complement the research design of this study by including classroom observation too, targeting all the components of the teaching approach (lecture, laboratory, classroom discussion, group work, field-trips, and the like).

Second, the conclusions that we are able to draw on the influence of the teaching approach on students' conceptual understanding of physics is limited to the topic of force and motion. The topic was chosen as students generally present several pre-instructional conceptions about it, and struggle to think in a Newtonian way even after being exposed to a Physics course. Other topics might impose different affordances to the learning context. Students might have fewer pre-instructional conceptions about phenomena that are rare in everyday life, or certain topic might be more difficult to observe and be connected in a clear way to concrete situations.

Finally, in the present study we aimed at controlling as many confounding variables as possible (i.e., teaching experience, grade taught, program content), which restricted the pool of eligible teachers. Future studies we aim at verifying whether the results of this study apply also when the controlled variables are manipulated.

CONCLUSION

Despite the limitations, the present study contributes to the literature on students' conceptual learning of physics in several ways. It contributed to create a semi-structured interview that includes several components, all associated to students' learning performance: beliefs about the nature of science (Lederman et al., 2002), beliefs about teaching and learning science (Tsai, 2002), and use of laboratory (not as a separate moment from the classroom lecture, but as a key moment of knowledge building when integrated with other components such as discussion and group work) in the teaching practices (Kang and Wallace, 2005).

In specific, in the interview we asked questions on their ideal teaching approach and their actual teaching approach, asking them to anchor their answers to the way they had taught force and motion during the school year (Mansour, 2009).

It also contributed to our understanding of which component of the teaching approach is associated with students' progress in physics and critical thinking skills. Several studies have investigated the influence of teachers' beliefs about the nature of science, but in this study all teachers held equivalent views, and differed on other crucial components. Having sophisticated epistemological beliefs is a necessary but not sufficient condition to create a learning environment that fosters conceptual learning. Finally, results of the study contributed to our understanding of the role that specific components of the teaching approach should have. Simply going to the laboratory does not foster a constructivist learning in students, unless it is matched with reflection. Specifically, our results suggest that the laboratory should be a guided experience that should be offered to students at the beginning of a teaching unit, with the purpose of making them aware of the difference between their pre-instructional conceptions and the manifestation of phenomena.

AUTHOR CONTRIBUTIONS

LB and CT designed and conducted the study, and wrote up the manuscript. CF and PF participated in the designing of the study and in the discussion of results. FS participated in the data analysis and in the discussion of results.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02474/full#supplementary-material

REFERENCES

Aerts, M., Molenberghs, G., Ryan, L. M., Geys, H., Molenberghs, G., Ryan, L. M., et al. (2002). *Topics in Modelling of Clustered Data*. New York, NY: Chapman and Hall/CRC. doi: 10.1201/9781420035889

Angell, C., Guttersrud, Ø., Henriksen, E. K., and Isnes, A. (2004). Physics: frightful, but fun. Pupils' and teachers' views of physics and physics teaching. *Sci. Educ.* 88, 683–706. doi: 10.1002/sce.10141

Angeli, C., and Valanides, N. (2009). Instructional effects on critical thinking: performance on ill-defined issues. *Learn. Instr.* 19, 322–334. doi: 10.1016/j.learninstruc.2008.06.010

Aretz, S., Borowski, A., and Schmeling, S. (2016). A fairytale creation or the beginning of everything: students' pre-instructional conceptions about the Big Bang theory. *Perspect. Sci.* 10, 46–58. doi: 10.1016/J.PISC.2016.08.003

Bereiter, C. (1994). Implications of postmodernism for science, or, science as progressive discourse. *Educ. Psychol.* 29, 3–12. doi: 10.1207/s15326985ep2901-1

Bevan, B. (2017). The promise and the promises of Making in science education. *Stud. Sci. Educ.* 53, 75–103. doi: 10.1080/03057267.2016.1275380

Bigozzi, L., Biggeri, A., Boschi, F., Conti, P., and Fiorentini, C. (2002). Children "scientists" know the reasons why and they are "poets" too. Non-randomized controlled trial to evaluate the effectiveness of a strategy aimed at improving the learning of scientific concepts. *Eur. J. Psychol. Educ.* 17, 343–362. doi: 10.1007/BF03173590

Bigozzi, L., Tarchi, C., Falsini, P., and Fiorentini, C. (2014). "Slow science": building scientific concepts in physics in high school. *Int. J. Sci. Educ.* 36, 2221–2242. doi: 10.1080/09500693.2014.919425

Bigozzi, L., Vezzani, C., Tarchi, C., and Fiorentini, C. (2011). The role of individual writing in fostering scientific conceptualization. *Eur. J. Psychol. Educ.* 26, 45–59. doi: 10.1007/s10212-010-0031-8

Boddy, N., Watson, K., and Aubusson, P. (2003). A trial of the five es: a referent model for constructivist teaching and learning. *Res. Sci. Educ.* 33, 27–42. doi: 10.1023/A:1023606425452

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Carey, S. (1985). *Conceptual Change in Childhood.* Cambridge, MA: Bradford Books.

Conley, A. M. M., Pintrich, P. R., Vekiri, I., and Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemp. Educ. Psychol.* 29, 186–204. doi: 10.1016/j.cedpsych.2004.01.004

Creswell, J. W., and Plano Clark, V. L. (2011). *Designing and Conducting Mized Method Research.* Thousand Oaks, CA: SAGE Publications.

Ennis, R. H. (1989). Critical thinking and subject specificity: clarification and needed research. *Educ. Res.* 18, 4–10. doi: 10.3102/0013189X018003004

Franco, G. M., Muis, K. R., Kendeou, P., Ranellucci, J., Sampasivam, L., and Wang, X. (2012). Examining the influences of epistemic beliefs and knowledge representations on cognitive processing and conceptual change when learning physics. *Learn. Instr.* 22, 62–77. doi: 10.1016/j.learninstruc.2011.06.003

Galbraith, S., Daniel, J. A., and Vissel, B. (2010). A study of clustered data and approaches to its analysis. *J. Neurosci.* 30, 10601–10608. doi: 10.1523/JNEUROSCI.0362-10.2010

Greene, J. A., Sandoval, W. A., and Bråten, I. (2016). *Handbook of Epistemic Cognition.* New York, NY: Routledge. doi: 10.4324/9781315795225

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *Phys. Teach.* 30, 141–158. doi: 10.1119/1.2343497

Hewson, P. W., and Hewson, M. G. A. (1987). Science teachers' conceptions of teaching: implications for teacher education. *Int. J. Sci. Educ.* 9, 425–440. doi: 10.1080/0950069870090401

Hofer, B. K., and Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Rev. Educ. Res.* 67, 88–140. doi: 10.2307/1170620

Hofstein, A., and Lunetta, V. N. (1982). The role of the laboratory in science teaching: neglected aspects of research. *Rev. Educ. Res.* 52, 201–217. doi: 10.3102/00346543052002201

Johnson, R. B., and Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educ. Res.* 33, 14–26. doi: 10.3102/0013189X033007014

Kang, N.-H., and Wallace, C. S. (2005). Secondary science teachers' use of laboratory activities: linking epistemological beliefs, goals, and practices. *Sci. Educ.* 89, 140–165. doi: 10.1002/sce.20013

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educ. Psychol.* 41, 87–98. doi: 10.1207/s15326985ep4102

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* 8, 355–362. doi: 10.1177/1948550617697177

Larkin, D. (2012). Misconceptions about "misconceptions": preservice secondary science teachers' views on the value and role of student ideas. *Sci. Educ.* 96, 927–959. doi: 10.1002/sce.21022

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., and Schwartz, R. S. (2002). Views of nature of science questionnaire: toward valid and meaningful assessment of learners' conceptions of nature of science. *J. Res. Sci. Teach.* 39, 497–521. doi: 10.1002/tea.10034

Linn, M. C. (2000). Designing the knowledge integration environment. *Int. J. Sci. Educ.* 22, 781–796. doi: 10.1080/095006900412275

Mansour, N. (2009). Science teachers' beliefs and practices: issues, implications and research agenda. *Int. J. Environ. Sci. Educ.* 4, 25–48.

Mason, L., Boldrin, A., and Ariasi, N. (2010). Epistemic metacognition in context: evaluating and learning online information. *Metacogn. Learn.* 5, 67–90. doi: 10.1007/s11409-009-9048-2

Mason, L., Boscolo, P., Tornatora, M. C., and Ronconi, L. (2013). Besides knowledge: a cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science. *Instr. Sci.* 41, 49–79. doi: 10.1007/s11251-012-9210-0

McCloskey, M. (1983). Intuitive physics. *Sci. Am.* 248, 122–131. doi: 10.1038/scientificamerican0483-122

Millman, J., Tomko, T. N., and Ennis, R. H. (2005). *Cornell Critical Thinking Tests Levels X & Z.* Seaside, CA: Critical Thinking Company.

Miri, B., David, B. C., and Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: a case of critical thinking. *Res. Sci. Educ.* 37, 353–369. doi: 10.1007/s11165-006-9029-2

Mortimer, E. F., and Scott, P. (2003). *Meaning Making in Secondary Science Classrooms.* Philadelphia, PA: Open University Press.

Niu, L., Behar-Horenstein, L. S., and Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educ. Res. Rev.* 9, 114–128. doi: 10.1016/J.EDUREV.2012.12.002

Organisation for Economic Co-operation Development [OECD] (2018). *PISA 2015 Results in Focus.* Available at: https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf [accessed May 24, 2018]

Ramlo, S. (2002). The force and motion conceptual evaluation. *Paper Presented at the Annual Meeting of the Mid-Western Educational Research Association*, Columbus, OH.

Ramlo, S. (2008). Validity and reliability of the force and motion conceptual evaluation. *Am. J. Phys.* 76, 882–886. doi: 10.1119/1.2952440

Schneider, R. M., Krajcik, J., and Blumenfeld, P. (2005). Enacting reform-based science materials: the range of teacher enactments in reform classrooms. *J. Res. Sci. Teach.* 42, 283–312. doi: 10.1002/tea.20055

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *J. Educ. Psychol.* 82, 498–504. doi: 10.1037/0022-0663.82.3.498

Stathopoulou, C., and Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemp. Educ. Psychol.* 32, 255–281. doi: 10.1016/j.cedpsych.2005.12.002

Stephens, M., Landeros, K., Perkins, R., and Tang, J. H. (2016). *Highlights from TIMSS and TIMSS Advanced 2015: Mathematics and Science Achievement of U.S. Students in Grades 4 and 8 and in Advanced Courses at the End of High School in an International Context. NCES 2017–2012.* Washington, DC: National Center for Education Statistics.

Thornton, R. K., Kuhl, D., Cummings, K., and Marx, J. (2009). Comparing the force and motion conceptual evaluation and the force concept inventory. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 5:010105. doi: 10.1103/PhysRevSTPER.5.010105

Thornton, R. K., and Sokoloff, D. R. (1998). Assessing student learning of newton's laws: the force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *Am. J. Phys.* 66, 338–352. doi: 10.1119/1.18863

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., and Janssen, R. (2017). Measuring critical thinking in Physics: development and validation of a critical thinking test in electricity and magnetism. *Int. J. Sci. Math. Educ.* 15, 663–682. doi: 10.1007/s10763-016-9723-0

Tiruneh, D. T., Verburgh, A., and Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: a systematic review of intervention studies. *High. Educ. Stud.* 4, 1–17. doi: 10.5539/hes.v4n1p1

Tobin, K. (1986). Secondary science laboratory activities. *Eur. J. Sci. Educ.* 8, 199–211. doi: 10.1080/0140528860080208

Tsai, C. C. (2002). Nested epistemologies: science teachers' beliefs of teaching, learning and science. *Int. J. Sci. Educ.* 24, 771–783. doi: 10.1080/09500690110049132

Tsai, C. C., Ho, H. N. J., Liang, J. C., and Lin, H. M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learn. Instr.* 21, 757–769. doi: 10.1016/j.learninstruc.2011.05.002

Urdan, T., and Schoenfelder, E. (2006). Classroom effects on student motivation: goal structures, social relationships, and competence beliefs. *J. Sch. Psychol.* 44, 331–349. doi: 10.1016/J.JSP.2006.04.003

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E. R., and Manlove, S. (2005). Co-Lab: research and development of an online learning environment for collaborative scientific discovery learning. *Comput. Hum. Behav.* 21, 671–688. doi: 10.1016/J.CHB.2004.10.039

Vosniadou, S. (2013). *International Handbook of Research on Conceptual Change*, 2nd Edn. New York, NY: Routledge. doi: 10.4324/9780203154472

Vygotsky, L. S. (1978). *Mind in Society.* Cambridge, MA: Harvard University Press.

Wei, B., and Li, X. (2017). Exploring science teachers' perceptions of experimentation: implications for restructuring school practical work. *Int. J. Sci. Educ.* 39, 1775–1794. doi: 10.1080/09500693.2017.1351650

World Medical Association (2013). Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 27, 2191–2194. doi: 10.1001/jama.2013.281053

Yoon, H. G., and Kim, B. S. (2016). Preservice elementary teachers' beliefs about nature of science and constructivist teaching in the content-specific context. *Eurasia J. Math. Sci. Technol. Educ.* 12, 457–475. doi: 10.12973/eurasia.2016.1210a

Zhang, L. (2018). Withholding answers during hands-on scientific investigations? Comparing effects on developing students' scientific knowledge, reasoning, and application. *Int. J. Sci. Educ.* 40, 459–469. doi: 10.1080/09500693.2018.1429692

"Turn Around and Forget": Assessment of the Cognitive Inhibitory Effect of Working Memory Information Using the List-Before-Last Paradigm

Xiaojun Zhao†, Changhao Liu† and Changxiu Shi*

School of Education, Hebei University, Baoding, China

This study mainly discusses whether the cognitive inhibitory effect of working memory information is affected by the nature of the signified information and the number of retrieval cues in the inhibitory information. Experiment 1 of our study examined the effect of concreteness on the information retrieval phase under different cognitive inhibition scenarios that were distinguished by the nature of the signified information and the number of retrieval cues in the inhibitory information. Experiment 2 of our study examined the effect of the number of retrieval cues in the inhibitory information on the cognitive inhibitory effect under different cognitive inhibition scenarios. The results of both experiments showed that information displaying more concrete characteristics exerted a greater the cognitive inhibitory effect during the working memory task, and a greater cognitive inhibitory effect was produced when all of inhibition retrieval information clues are provided than when none of the clues are provided in the working memory task. Based on these results, the concreteness effect on cognitive inhibition exists, and when all retrieval clues for inhibitory information are provided, the cognitive inhibitory effect might be greater.

Keywords: list-before-last paradigm, working memory, cognitive inhibitory effect, the nature of the signified information, the number of retrieval cues

INTRODUCTION

Imagine a situation in which we cannot find something but would like to try to remember concrete information to answer the question "Where did I last see it?" At this point, our memory system begins to function. According to the theory of information processing, our memory system initially encodes and stores concrete information and then helps us retrieve this information when needed (Howard and Kahana, 2002; Lehman and Malmberg, 2009). Fortunately, we are able to retrieve concrete information with the help of our memory system. However, we may fail to retrieve the information in some cases: "I just saw it, but I can't remember where it is." Therefore, even though we do not forget certain information for a long time, why do we forget? This question is the one we wished to solve in the present study; namely, what is the mechanism underlying the "forgetfulness" of working memory in the information retrieval stage, or what is the mechanism of "cognitive inhibition"?

The term "cognitive inhibition" in this study refers to the internal process in which an individual inhibits the retrieval of irrelevant information and maintains the information relevant to the task in working memory (Harnishfeger and Bjorklund, 1993).

The Theoretical Basis of the Cognitive Inhibitory Effect: Context Change Model and Temporal Context Model (TCM)

Context Change Model

First, the reason why we study working memory is to eliminate the interference of normal forgetting and better explain the abnormal forgetting phenomenon of "turn around and forget." Second, we describe the "forgetting mechanism" as "cognitive inhibition mechanism" because our experiments are based on the "context-change model," and this assumption is in fact a type of cognitive inhibition theory (Sahakyan and Kelley, 2002). Furthermore, the context change model assumes that different memory tasks lead to changes in people's internal situations, and the inconsistency between encoding and retrieval leads to cognitive inhibition (forgetting). Thus, the phenomenon of "turn around and forget" may be related to a change in our memory, and its nature may be "a cognitive inhibitory effect" on concrete working memory information.

Temporal Context Model (TCM)

According to a previous study (Sahakyan and Hendricks, 2012), the mechanism of individual context change situations may utilize a temporal context model (TCM). This theory postulates that each new unit of information processing will cause a change in the cognitive environment, namely, in our information processing ability. Therefore, our internal psychological status will also constantly change throughout the experiment. In addition, the successful retrieval of information may lead to a psychological context change. The TCM emphasizes that information processing occurs in a particular order and hypothesizes that information obtained later might affect information obtained earlier during memory and retrieval processes.

Although the TCM has not yet been supported by a large number of experiments, it offers a new concept while we research the cognitive inhibitory effect of working memory information. In the information retrieval phase, greater retrieval of irrelevant information about the task produces greater inhibitory effects and increases the difficulty in retrieving information relevant to the task.

Paradigm of Experimental Research on Cognitive Inhibition: The List-Before-Last Paradigm

The List-Before-Last Paradigm and Its Limitations

Our experiment uses the list-before-last paradigm to examine the context change during the cognitive inhibition process. This paradigm has been reported to be a reliable test of an individual's context change (Shiffrin, 1970; Jang and Huber, 2008; Sahakyan and Hendricks, 2012; Sahakyan and Smith, 2014). During the experiment, participants will first be distinguished according to the different types of memory tasks (e.g., restudy L1 task and retrieval L1 task). Then, they will sequentially memorize three lists (L1, L2, and L3), each of which contains a specific number of words. However, after memorizing L1 and L2, participants will memorize L1 again using different strategies, according to the different types of memory tasks we established. When participants have memorized all lists, they will be asked to provide free recall of L2 as the final test. Instigated by the different methods for retrieving L1, the contextual similarity/continuity between the adjacent lists (L2 and L3) was disrupted (what we called "context-change" in our study). Furthermore, reinstating the L2 context will become more difficult for participants because context drifted from one list to the next list in a somewhat gradual fashion and was disrupted by the retrieval of L1. Therefore, using this paradigm, we examined the degree of the interruption effect (caused by the retrieval of L1 using different strategies) by testing the effects of the L2 context during the final test. Greater free recall of the L2 context represents a smaller interruption effect of the retrieval of L1. This "interruption effect of the retrieval of L1" is defined as "the cognitive effect" in the present study. Meanwhile, in the list-before-last paradigm, we consider L1 as "inhibitory information" and L2 as "inhibited information."

However, previous studies have created cognitive inhibition situations by destroying the gradual context drifts through different L1 retrieval tasks, and an individual's context change also initially occurred during different cognitive inhibition tests (Sahakyan and Hendricks, 2012). However, we have not yet identified an experimental operational index to measure the extent of that context change. Therefore, we aim to solve this problem in the present study.

Localization of the List-Before-Last Paradigm

The experimental research on "Chinese words" as the experimental material in the list-before-last paradigm is insufficient. However, some research can help researchers localize this experimental paradigm. In the field of cognitive psychology, studies of "Chinese lexical information processing and storage methods" indicated that the method used to divide words into concrete words (such as 手机—mobile phone; 铅笔—pencil) and abstract words (such as 思想—ideology; 方面—division) is more mature than the other methods of classifying words because it will help the "Chinese processing context" become more pertinent to the "English processing context"; namely, when the memory and understanding of a same concrete word or an abstract word are employed, the meaning will be clearer and more precise for both Chinese and English speakers than the use of other word classification methods (Sui et al., 2016). This finding provides new insights into the suitability of using the list-before-last paradigm for Chinese words. And Chinese experimental processing situations may be more consistent with the original English experimental processing situations through this word classification method.

Additionally, previous studies have used three different lists of words containing 12 nouns per list (Sahakyan and Hendricks, 2012; Sahakyan and Smith, 2014). However, in a "directed forgetting" study, where participants memorized and recalled two different lists (A and B) of words containing 12 nouns per list in

order (A–B), the first four words of list B have the highest recall rate among all words in list B (Pastötter et al., 2012). Thus, during the memorization and recall of a list of words, the semantic information of the first four words is the most accurate for the participants. And depending on this conclusion, the disruption of contextual continuity between adjacent lists (context change) and the cognitive inhibitory effect caused by the L1 retrieval task may be more likely reflected in the first four words of L2 in the list-before-last paradigm. When the number of words in each list was reduced to four in the list-before-last paradigm, maybe most of the words could process equally by participants throughout the experiment and the serial position effect (SPE) with more memory load in each list could avoid partly. Accordingly, the opportunity of each item of L2 to be recalled first in the final test was approximately equal. And the disruption of contextual continuity between neighboring lists will indeed inhibit L2. If this hypothesis is true, according to context change model, L2 words will be more difficult to recall in the free recall task (Task 1); and according to the TCM, more L2 words will be recited out-of-order among all three lists of words in the memory sorting task (Task 2). And as more position units changed, a larger inhibitory effect caused by the later information (retrieval L1 task) was observed.

Factors Influencing the Cognitive Inhibitory Effect: Types of Memory Tasks and Types of Memory Materials

Different Memory Tasks Based on Different Numbers of Retrieval Cues for Inhibitory Information

In the presence of different numbers of retrieval cues for inhibitory information (L1), this inhibitory information produces varying degrees of context change (Sahakyan and Smith, 2014). Specifically, in the list-before-last paradigm, tests in which all retrieval cues for inhibitory information are provided (task of "restudy L1") cause greater free recall of L2 than tests in which a portion of the clues are provided (task of "retrieval L1"). Thus, a smaller context change may be observed in participants who are provided with all retrieval cues for inhibitory information than in participants who are provided a portion of the clues. Using this strategy, our research should consider the number of retrieval cues for inhibitory information as a factor that may affect the cognitive inhibition of working memory.

The Concreteness Effect Is Based on the Processing of Material Referring to Something of a Different Nature

The processing of words that refers to the different nature of things will be affected by the "concreteness effect." Processing of concrete words is more accurate and faster than processing of abstract words, particularly when words are presented separately (James, 1975; Schwanenflugel and Shoben, 1983; Kroll and Merves, 1986; Bleasdale, 1987; Schwanenflugel et al., 1988). An ERP study focused on explaining the mechanism of the concreteness effect showed that the processing of concrete words evoked greater N400 (N400 is associated with semantic processing) than the processing of abstract words, indicating

that the processing of concrete words in the information processing phase may activate more semantic information than the processing of abstract words (Kounios and Holcomb, 1994). Therefore, the degree of change in a mental situation will be greater when an individual is processing concrete words than when an individual is processing abstract words. Thus, in the list-before-last paradigm, the interruption of the context of L2 to L3 concrete words context (by L1 retrieval tasks) may activate more semantic information than abstract words, which also represents a high degree of context change. Under the same interruption condition, a greater amount of inhibition of the semantic information in L1 would occur in the concrete words group than in the abstract words group. According to our former hypothesis, a high degree of cognitive inhibitory effect will be observed with a high degree of context change, which will make the concrete words more difficult to retrieve. Does this "concreteness effect" exist in the information retrieval phase? We have not been able to determine a unified answer from the existing studies. We will explore this question in the present study.

Research Purpose and Hypothesis Examined in This Study

We aimed to explore the cognitive inhibition mechanism of working memory in the information retrieval phase. In this study, we conducted two experiments to address whether different memory tasks, which are based on different numbers of retrieval cues for inhibitory information, influence the retrieved results. We hypothesized that the task "restudy L1," which contains large amounts of retrieval cues for inhibitory information in the process of information processing, would produce a smaller cognitive inhibitory effect than the task "retrieval L1," which contains fewer retrieval cues for inhibitory information in the process of information processing. In the experiment, a greater number of free recall L2 words was presented to the restudy group than to the retrieval group.

We aimed to explore whether the "concreteness effect" existed in the information retrieval phase. We hypothesized that concrete words would cause larger cognitive inhibitory effects than abstract words. In the experiment, more free recall L2 words were presented to the abstract word group than to the concrete word group.

EXPERIMENT 1

In Experiment 1, the cognitive inhibition mechanism of concrete words and abstract words in different memory tasks was investigated under the condition of stimulus-alone-appear. The experiment consists of two tasks. Task 1, "free recall L2 (the inhibited information)", aims to test the presence of the concreteness effect on the working memory information in the retrieval phase and the different cognitive inhibition modes based on the different numbers of retrieval cues for inhibitory information that will exert different cognitive inhibitory effects. Task 2, "memory sorting L2 (the inhibited information)," aims to test information processing in working memory that will

lead to a psychological context change; this "context change" was statistically and simultaneously managed. Furthermore, the logical relationship between these two tasks is described below. After Task 1 confirms the concreteness effect and the cognitive inhibitory effect on the retrieval of working memory information, Task 2 will prove that the mechanism of cognitive inhibition is consistent with the context change model. The participants must complete Task 1 first (time is 60 s) and then complete Task 2 (time is 120 s). The entire experiment lasted approximately 260 s.

Before the experiment, the first four words of each separate list in the experiment were effectively and accurately equal to take advantage of the participants' memories, and to test the context change, we will control the working memory capacity of the participants using the Operation Span (OSPAN) task (Turner and Engle, 1989). This method has been reported to display high correlation and reliability in measuring the working memory capacity of individuals (Klein and Fiss, 1999; Conway et al., 2005). In this test, we used words to replace letter strings, enabling the test be more similar to the experiment. In the test, participants were required to first determine whether a math equation was correct and then memorize the word according to the math equation. The operation string, for example, might be "(9÷3)–2 = 2? Uncle." As the number of operation strings gradually increases, the number of words correctly recalled by the participants will comprise their working memory span. However, in our study, we directly established two conditions to improve efficiency: "2 operation strings" for practice and "4 operation strings" in the experiment. All participants were required to report all 4 words, and the correct rate was 100%. All strings were presented sequentially in a random order using the E-prime2.0 software. Each string (in white bold typeface, font size 48 points) was presented for 4 s, and the time interval between the strings was 1 s. The string was presented in the middle of the screen against a black background.

After the experiment, we used a Likert scale (from 1, "when you see the word, you feel very sad," to 7, "when you see the word, you feel very happy") to control for the emotional valence of all of the experimental materials and avoid the potential interference from differences in the participants' emotional valences of the experimental material on the experimental results. All participants were required to complete the scale after the experiment and evaluate those words they observed in the experiment. The results were concrete word, $M = 4.2733$, and abstract word, $M = 4.2656$. The difference was not significant.

Methods
Participants
Sixty volunteers from Hebei University participated in the study. Participants ranged in age from 18 to 24 years ($M = 21.08$, $SD = 2.04$), were not color blind and had normal vision or normal corrected vision. In Experiment 1, participants were randomly assigned to one of four experimental groups: abstract word retrieval, L1 group; abstract word restudy, L1 group; concrete word retrieval, L1 group; and concrete word restudy, L1 group. Each group consisted of 15 participants. After the experiment,

each person was rewarded with a gift (stationery, such as pens or notebooks).

Materials
First, we identified 18 concrete words (0.0072 average frequency, 17 average strokes) and 18 abstract words (0.0097 average frequency, 16.5 average strokes) from "the most frequently used 3000 words" in *Modern Chinese Frequency Dictionary*. All words were double syllable nouns. Second, all words were randomly assigned to a 6×6 format, and 10 non-psychology students (who did not participate in either of the two studies) assessed the concreteness of each word using a Likert scale (from 1, "the word is concrete," to 7, "the word is abstract"). Before assessing the words, these studies were told that "the word is concrete," which meant that "the word expresses a concrete image and can also be touched," and that 1 to 7 points indicated that the word's concreteness was able to be gradually enhanced. The final results were as follows: concrete word, $M = 6.2722$, $SD = 0.8641$; and abstract word, $M = 2.0444$, $SD = 0.7155$. The differences between the two categories was significant ($p < 0.01$). In the formal experiment, L1, L2, and L3 each contained 4 words, and the other words were used in the practice experiment, with 2 words per list. All words were presented randomly using the E-prime2.0 software. Each word (in white bold typeface, font size 48 points) appeared in the middle of a black background screen for 4 s. The time interval between the words was 1 s.

In Task 2, a "Memory sorting L2 test paper" (**Figure 1**) was used to avoid a SPE that may occur when information is retrieved. In this test paper, all words were randomly placed in a "circle," and when the participants finished Task 2, they were required to label the memorized sequence of each word on the test paper.

Design
The study used a 2 memory material (concrete words/abstract words) × 2 memory task (retrieval L1/restudy L1)
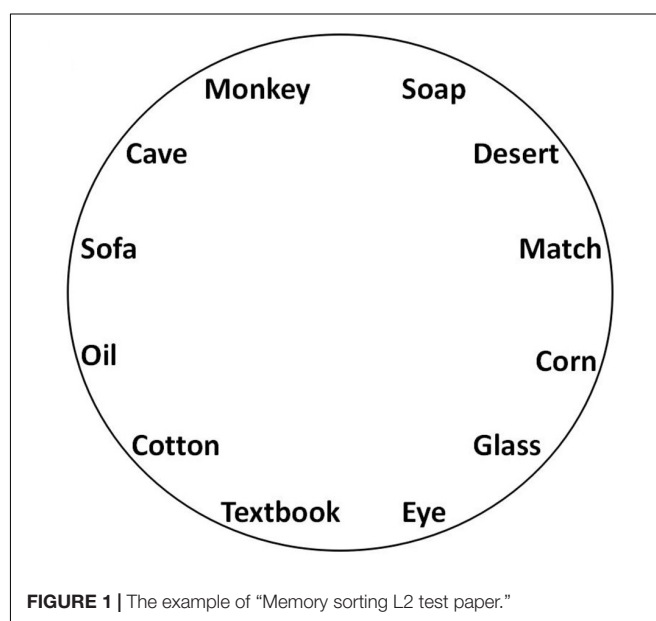

FIGURE 1 | The example of "Memory sorting L2 test paper."

between-subjects design. In Task 1, the dependent variable is the amount of retrieved inhibitory information (L2) and its statistical indicators represent the amount of free recall of L2. In Task 2, the dependent variable is the position change amount of inhibitory information (L2) in the information retrieval phase. Its statistical indicator is the number of position change units, which represents the degree of context change of inhibitory information (L2) in the information retrieval phase. In the present study, "a context-change unit" indicates that every difference in the recall order of a word compared with the original presentation order of each word will be recorded as a context change unit.

Procedure

For the retrieval groups, we first presented the following instructions on the screen: "This experiment aims to study our memory. It is divided into three lists. You must memorize all lists, and each list is separated by a plus sign '+'. A green plus sign means 'Continue the memory task,' and a red plus means 'Please restudy L1 based on the clue.' When you are ready, press the space bar to begin." Then, we presented L1, a green plus sign for 1 s, L2, a red plus sign for 1 s, L1 (each word had only the first word, such as "猴"), and L3. Finally, we presented the instruction, "The experiment is over." Then, the participants first completed Task 1 followed by Task 2.

For the restudy groups, we first presented the following instructions on the screen: "This experiment aims to study our memory. It is divided into three lists that you must memorize. Each list is separated by a plus sign '+'; a green plus means 'Continue the memory task,' and a red plus means 'Please restudy L1.' When you are ready, press the space bar to begin." Then, we presented L1, a green "+" for 1 s, L2, a red "+" for 1 s, L1, and L3. Finally, we presented the instruction: "The experiment is over." Then, the participants first completed Task 1 and then Task 2 (**Figure 2**).

Before the commencement of the formal experiment, all participants were required to perform the practice experiment

to become familiar with the experimental procedures and understand the instructions. In the practice experiment, each list comprised two words, and other procedures were consistent with the formal experiment. The practice experiments were not experimental tasks to ensure that the participants were blinded to the purpose of the experiment.

Results
Task 1

In the 2 memory material (concrete words/abstract words) × 2 memory task (retrieval L1/restudy L1) design, the two-factor complete random analysis of variance showed a significant main effect of memory materials [$F_{(1,56)} = 4.248$, $p < 0.05$, $\eta_p^2 = 0.071$]. The *post hoc* comparison revealed a significant greater number of correct answers in the free recall L2 task in the abstract words group ($M = 1.8667$, $SD = 0.730$) than in the concrete words group ($M = 1.4667$, $SD = 1.042$, $p < 0.05$). The main effect of the memory tasks was not significant [$F_{(1,56)} = 3.121$, $p > 0.05$]. The interaction effect was not significant [$F_{(1,56)} = 0.780$, $p < 0.05$, $\eta_p^2 = 0.014$] (**Table 1**).

Task 2

In the 2 memory material (concrete words/abstract words) × 2 memory task (retrieval L1/restudy L1) design, the two-factor complete random analysis of variance did not reveal significant main effects of memory materials [$F_{(1,56)} = 0.001$, $p = 0.976$] or memory tasks [$F_{(1,56)} = 0.669$, $p = 0.417$]; the interaction effect was not significant [$F_{(1,56)} = 0.669$, $p = 0.417$] (**Table 2**).

EXPERIMENT 2

According to Experiment 1, the explanation for the lack of a significant main effect of the "memory task" may be attributed to two points: (1) each list contained too few words,
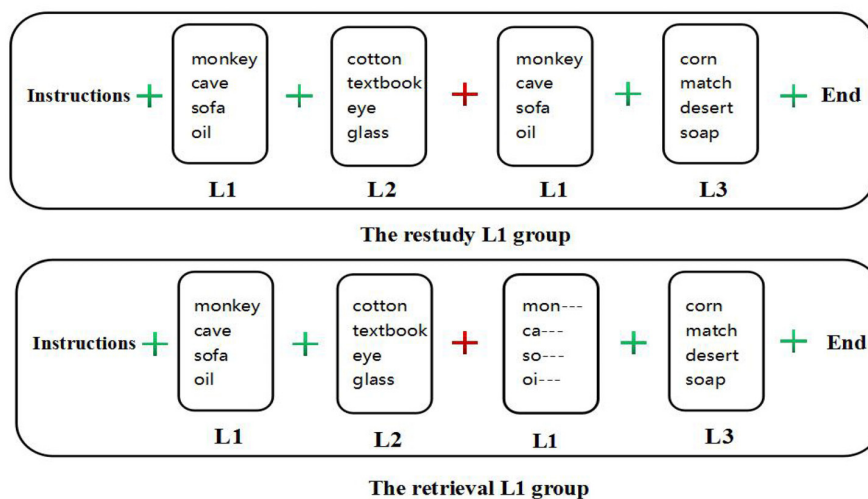

FIGURE 2 | The procedure of Experiment 1.

TABLE 1 | Analysis of variance in Task 1 (Experiment 1).

Source	SS	df	MS	F	p
Memory materials	3.267	1	3.267	4.248	0.044
Memory tasks	2.400	1	2.400	3.121	0.083
Memory materials × Memory tasks	0.600	1	0.600	0.780	0.381
Error	43.067	56	0.769		

TABLE 2 | Analysis of variance in Task 2 (Experiment 1).

Source	SS	df	MS	F	p
Memory materials	0.017	1	0.017	0.001	0.976
Memory tasks	12.150	1	12.150	0.669	0.417
Memory materials × Memory tasks	12.150	1	12.150	0.669	0.417
Error	1016.667	56	18.155		

and (2) the difference in the context change, which was caused by the "retrieval L1 task" and the "restudy L1 task," was not significant. However, our study focuses on working memory; therefore, changing the word items in each list is not appropriate. If each list contains three or fewer words, the memory items of all experiments will be equal to or less than nine words. At this time, our independent variable will be confused with the differences in the short-term memory abilities of the participants. If each list contains five or more words, the participants' memory of items will be exhausted after the memory of L2. Therefore, L3 will exist in name only, and the first effect and the recent effect will be more prominent.

Therefore, we should consider changing the "retrieval L1" task. In previous studies using the list-before-last paradigm, when the "memory tasks" variable contained the "retrieval L1" task and the "mathematical problem-solving task" for two levels, the context change caused by these tasks exhibited significant differences, and a significantly greater number of correct answers for free recall L2 was observed in the mathematical problem-solving group (Sahakyan and Hendricks, 2012). Perhaps by significantly reducing the number of retrieval cues for inhibitory information, we will observe a significant difference in context change between the different memory tasks. Therefore, we will replace the "retrieval L1 task" with "free recall L1 task" in Experiment 2, and we expect that different memory tasks, which are based on the presentation of all or no retrieval cues, will produce a significant context change between the different groups.

According to two previous studies (Sahakyan and Hendricks, 2012; Sahakyan and Smith, 2014), the "retrieval L1" group recorded the fewest number of free recall L2 words compared with the "restudy L1" group and the "mathematical problem-solving" group. We do not know which of the latter groups recorded the greatest number of free recall L2 words. However, a greater number of free recall L2 words was recorded in the no-retrieval-clues for inhibitory information condition than in the yes-retrieval-clues for the inhibitory information condition. Therefore, we hypothesize that in Task 1 of Experiment 2, a greater number of free recall L2 words

will be recorded by the "free recall L1" group than by the "restudy L1" group. In Task 2, more position change units will be observed for the "restudy L1" group than for the "free recall L1" group.

Additionally, the emotional valence values for all words in Experiment 2 were as follows: concrete word, $M = 4.2689$; and abstract word, $M = 4.2633$. The difference between the two types of words was not significant.

Methods
Participants
We recruited a group of 60 volunteers from Hebei University who ranged in age between 18 and 25 years ($M = 22.07$, $SD = 2.10$), were not color blind and had normal vision or normal corrected vision. In Experiment 2, participants were randomly assigned to the following four experimental groups: the abstract words free recall L1 group; the abstract words restudy L1 group; the concrete words free recall L1 group; and the concrete words restudy L1 group. Each group contained 15 participants. After the experiment, each person was rewarded with a gift (stationery, such as pens and notebooks).

Materials
All words were the same as those used in Experiment 1.

Design
The study employed a 2 memory material (concrete words/abstract words) × 2 memory task (free recall L1/restudy L1) between-subjects design. Other details were the same as Experiment 1.

Procedure
In the free recall L1 group, the instructions were as follows: "This experiment aims to study our memory. It is divided into three lists that you must memorize. Each list is separated by a plus sign '+'; a green plus means 'Continue the memory task,' and a red plus means 'Please keep looking at the red plus and memorize L1.' When you are ready, press the space bar to begin." Other details were the same as in Experiment 1 (**Figure 3**).

Results
Task 1
In the 2 memory material (concrete words/abstract words) × 2 memory tasks (free recall L1/restudy L1) design, the two-factor complete random analysis of variance showed a significant main effect of memory materials [$F(1,56) = 14.097$, $p < 0.05$, $\eta_p^2 = 0.201$]. The *post hoc* comparison showed a greater number of correct answers for free recall L2 in the abstract words group ($M = 2.467$, $SD = 1.074$) than in the concrete words group ($M = 1.567$, $SD = 1.357$, $p < 0.05$). The main effect of memory tasks was significant [$F(1,56) = 39.157$, $p < 0.05$, $\eta_p^2 = 0.412$]. The *post hoc* comparison showed a significantly greater number of correct answers for free recall L2 in the free recall group ($M = 2.767$, $SD = 0.935$) than in the restudy group ($M = 1.267$, $SD = 1.172$, $p < 0.05$). The
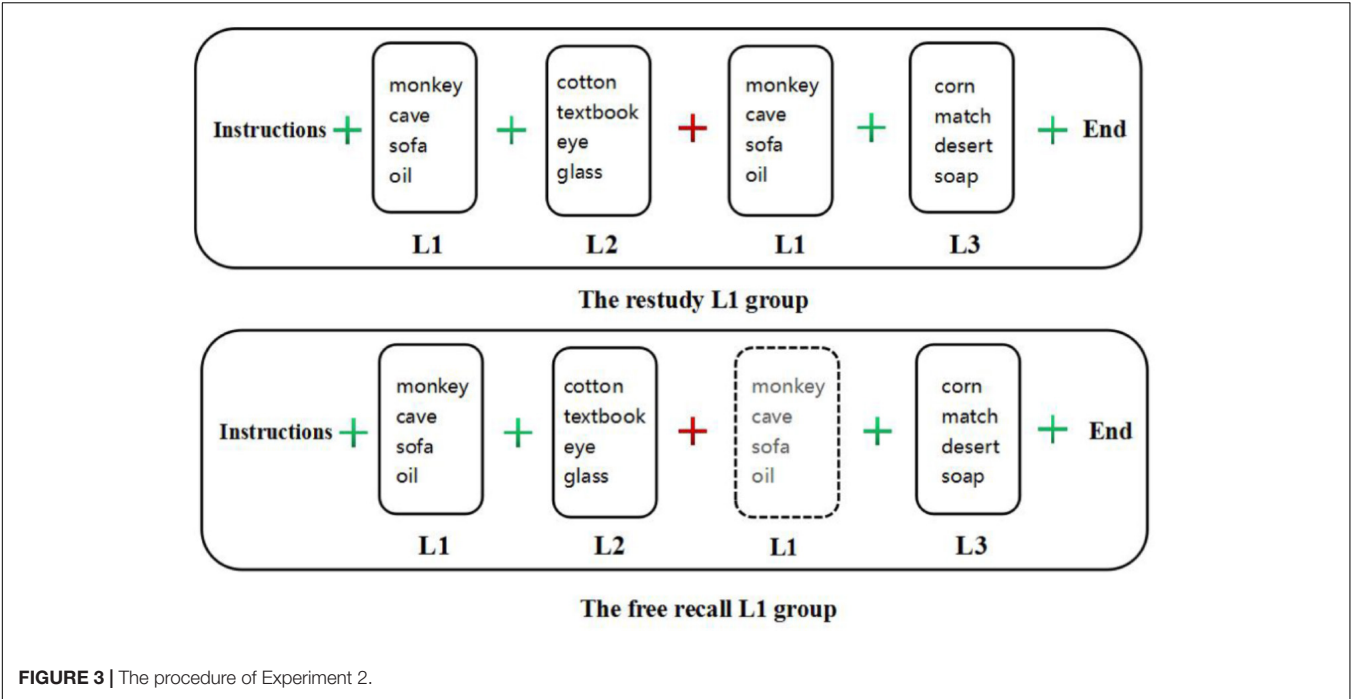
FIGURE 3 | The procedure of Experiment 2.

interaction effect was significant [$F(1,56) = 5.588$, $p < 0.05$, $\eta_p^2 = 0.091$] (**Table 3**). The simple effect analysis showed that the participants restudying L1 in the abstract words recorded a significantly greater number of correct responses for free recall L2 ($M = 2.000$, $SD = 1.069$) than the participants restudying L1 in the concrete words ($M = 0.533$, $SD = 0.743$, $p < 0.05$) (**Figure 4**).

Task 2

In the 2 memory material (concrete words/abstract words) × 2 memory task (free recall L1/restudy L1) design, the two-factor complete random analysis of variance revealed a significant main effect of memory materials [$F(1,56) = 11.677$, $p < 0.05$, $\eta_p^2 = 0.173$]. The *post hoc* comparison showed a significantly greater number of position change units of L2 in the concrete words group ($M = 7.300$, $SD = 4.893$) than in the abstract words group ($M = 5.033$, $SD = 4.056$, $p < 0.05$). The main effect of memory tasks was significant [$F(1,56) = 117.818$, $p < 0.05$, $\eta_p^2 = 0.678$]. The *post hoc* comparison showed a significantly greater number of position change units of L2 in the restudy L1 group ($M = 9.767$, $SD = 3.266$) than in the free recall L1 group ($M = 2.567$, $SD = 2.359$, $p < 0.05$). The
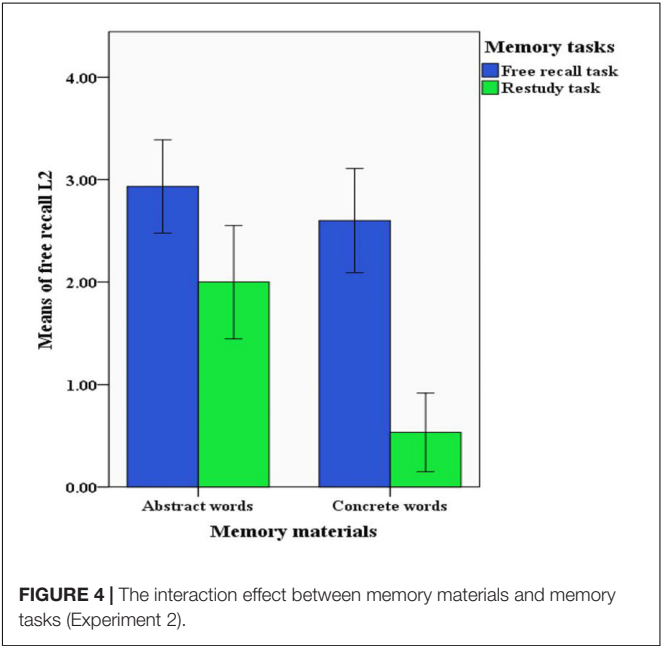
FIGURE 4 | The interaction effect between memory materials and memory tasks (Experiment 2).

interaction effect was not significant [$F(1,56) = 3.646$, $p = 0.061$] (**Table 4**).

GENERAL DISCUSSION

The present study found the cognitive inhibitory effect on working memory was influenced by the nature of the processed information and the number of inhibitory information retrieval cues.

TABLE 3 | Analysis of variance in Task 1 (Experiment 2).

Source	SS	df	MS	F	p
Memory materials	12.150	1	12.150	14.097	0.000
Memory tasks	33.750	1	33.750	39.157	0.000
Memory materials × Memory tasks	4.817	1	4.817	5.588	0.022
Error	48.267	56	0.862		

TABLE 4 | Analysis of variance in Task 2 (Experiment 2).

Source	SS	df	MS	F	p
Memory materials	77.067	1	77.067	11.677	0.001
Memory tasks	777.600	1	777.600	117.818	0.000
Memory materials × Memory tasks	24.067	1	24.067	3.646	0.061
Error	369.600	56	6.600		

In Experiment 1, both tasks, we didn't examine a significant difference between retrieval L1 task and restudy L1 task as the previous studies did. And we hypothesized that the number of fewer items in each list may had contributed to this result compared with previous studies. However, we examined a significant difference between concrete words and abstract words in Task 1. This finding was consist with the concreteness effect theory.

And in Experiment 2, both tasks observed the concreteness effect and cognitive inhibitory effect during the information retrieval phase. Especially in Task 2, the results indicated by our new statistical indicator "the amount of position change units in L2" was consist with Task 1 results, this finding further proved that the position change amount of inhibition information may be used as a statistical standard of individual's context change.

Concreteness Effects of Cognitive Inhibition on Working Memory

The concreteness effect may also exist in the information retrieval phase. In our study, a greater number of correct answers for the free recall of L2 abstract words was observed compared with concrete words. According to a previous study, the processing of concrete words in the information processing phase may activate more semantic information than the processing of abstract words (Kounios and Holcomb, 1994). In other words, a greater context change will occur when we process concrete words. Therefore, when we experimentally controlled the information retrieval time and mode, we observed a smaller cognitive inhibition effect during the concrete words processing procedure than in the abstract words processing procedure.

Clearly, the information we process in our daily life is more concrete, and the cognitive inhibitory effect on these types of information will thus be greater. For example, if we were asked at noon, "Did you eat in the morning?" we may easily answer this question, but if we were asked, "What did you eat in the morning? What did you eat first and what did you eat afterwards?" we may need to think for a while. Moreover, the concreteness effect not only exists in the information processing phase but also in the information retrieval phase; it can cause difficulty in retrieving information our daily life. Namely, if we specifically remember one thing, we may completely forget another concrete thing, and after a period of time, with the degree of the information's concreteness decreasing naturally, the inhibited information may be easier for us to retrieve as the degree of the concreteness of the information decreases naturally. However, the information will become more confusing,

and its summary may even be wrong at this later time point.

Relationship Between the Number of Inhibitory Information Retrieval Cues and the Cognitive Inhibitory Effect

The result in Experiment 1 differs from a previous study in which the number of correct answers for free recall L2 were greater in the restudy group than in the retrieval group (Sahakyan and Smith, 2014). This discrepancy may be caused by the difference in the number of items in each list. In the previous study, each list contained twelve words. Considering the short-term memory span (7 ± 2 units) and "the first four words effect" as we already have said (Pastötter et al., 2012), we used four words in each list in this study. That is to say, whether inhibitory information retrieval cues were processing partly or totally, the effect of cognitive inhibition on working memory may not be obvious with less memory load. And only when the working memory load reaches a certain amount (more than four), the differential amount of inhibitory information retrieval cues will cause a significant cognitive inhibition effect. Additionally, The lack of a significant main effect of memory tasks performed in Experiment 1 may also indicate that the cognitive inhibitory effect was not affected by the number of retrieval cues for inhibitory information; at least, the effect is not significant. But, in Experiment 2, under two extreme conditions in which the retrieval cues for inhibitory information were all provided or no cues were provided, a greater cognitive inhibitory effect was observed for the former condition. However, the relationship between the number of inhibitory information retrieval cues and the cognitive inhibitory effect may not be linear because previous studies have reported that the presentation of inhibitory information retrieval cues actually resulted in the greatest cognitive inhibitory effect (Sahakyan and Hendricks, 2012; Sahakyan and Smith, 2014). This finding is interesting.

The study also found that using an "all or none" classification method for inhibitory information retrieval cues. The number of retrieval cues for inhibitory information may lead to different levels of cognitive inhibitory effect, even with a less memory load. Accordingly, different levels of cognitive inhibitory effect can be reflected through the number of position change units of inhibition information. In Experiment 2, two tasks' results were consistent with each other and partially confirmed our earlier hypothesis that in the list-before-last paradigm when all retrieval cues for the inhibitory information are presented, a greater cognitive inhibitory effect will be observed than for the inhibitory information with no retrieval cues. Furthermore, participants recalled most of the L1 words in two experiments. However, when we ignored the number of L1 words participants retrieved during the final test, participants actually retrieved all 4 words in L1 under both conditions; therefore, the effect of L1 differed because of the different processing pathways. According to the results of Experiment 2, if participants retrieved L1 words through the use of all retrieval cues, then the inhibitory effect of L1 will be greater than another pathway in which L1 words were retrieved with no retrieval cues. Thus, with the exception of the effect of the amount of inhibitory information on the extent

of the inhibitory effect, the pathways used to retrieve inhibitory information may also be effective.

A high-level cognitive inhibitory effect based on all amount of inhibited information cues was more likely to occur in the information processing of low cognitive resource consumption. According to context change model, the greatest number of retrieval cues for concrete words condition produced a greater context change effect (inhibitory effect) that transferred to the other homogeneous concrete words than the fewest (none) number of retrieval cues condition. As shown in previous studies, the "concreteness effect" is present when individuals are processing information (James, 1975; Schwanenflugel and Shoben, 1983; Kroll and Merves, 1986; Bleasdale, 1987; Schwanenflugel et al., 1988). However, based on our findings, differences in degree of the activated semantic information between abstract words and concrete words may also have existed when individuals retrieved information. According to the TCM, the represented concrete information may activate more semantic changes and cause greater changes in the memory of the order of the other homogeneous concrete information. Unfortunately, we did not observe a significant interaction effect in Task 2. We must reconsider the rationality of the dependent variable and its statistical indicator in Task 2.

Moreover, in our experimental memory task, inhibitory information was not irrelevant information about the memory task goal. In fact, regardless of whether inhibitory information (L1) or inhibited information (L2) was related to the completion of the experimental memory task, the information belonged to the same information processing sequence and the same cognitive inhibitory situation in the experiments. Although we did not count the number of participants' free recall L1 words, participants were able to correctly recall most of the words in L1 when they performed Task 1 in each experiment. Therefore, in the list-before-last paradigm, the method in which we considered L1 as the inhibitory information was somewhat reasonable.

Mechanism of the Cognitive Inhibitory Effect on Working Memory

The mechanism of this "turn around and forget" phenomenon, which was considered a cognitive inhibition phenomenon in our study, may theoretically consist of the context change model and may practically operate under the TCM. Because we observed a significant retrieval position change in L2 words in the retrieval phase of Experiment 2, the information retrieval sequence had already been affected by the cognitive inhibitory effect, in which the retrieval of the latter information was limited by the retrieval of the former information.

Notably, the observation of a greater context change does not mean that the individual experienced difficulty in processing or retrieving information. Although a difference in the difficulty of memory between the two levels of independent variable (memory tasks), which are distinguished by the number of the retrieval cues for inhibitory information, this type of difference does not represent evidence of the ability to distinguish the two levels of independent variables. Furthermore, the essence of independent variable in the two levels of memory tasks is the

difference in the number of inhibitory information retrieval cues, and the essence of the difference in the number of inhibitory information retrieval cues is the difference in the cognitive inhibitory modes. The cognitive inhibition situations, which are caused by the different cognitive inhibitory modes, represent the crux of the list-before-last paradigm. This hypothesis enables the list-before-last paradigm to prove the context change model.

Applicability of New Statistical Indicators of Context Change and Limitations of the Experiment

Using the number of position change units as a statistical indicator of internal context change is a open question. In our two experiments, the results from Tasks 1 and 2 were not completely matched in each experiment; therefore, we were not able to definitively conclude that the position change of inhibited information (L2) in the retrieval phase represents a statistical indicator of context change. We should consider the limitations of our experiments to explore the reasons for the inconsistent results.

On one hand, although the entire experiment was performed in 5 min, we were not able to avoid individual differences in normal forgetting in our experiments; therefore, the retrieval sequence of each word may not be equivalent when the participants retrieved the individual words in Task 2. We asked the participants to complete Tasks 1 and 2 separately in our experiments to ensure efficiency. However, according to the TCM (Sahakyan and Hendricks, 2012), the retrieval of the former information can influence the retrieval of the latter information. In our experiments, the participants' memory of concrete information in Task 1 may have affected their retrieval procedure in Task 2, and this experimental error was not controlled for in the experiments.

On the other hand, we controlled for participants' working memory abilities regarding the individual differences before the experiments, and previous studies confirmed that our methodology was appropriate (Klein and Fiss, 1999; Conway et al., 2005). However, the control measure that uses the participants' working memory span to represent their working memory capacity might still cause experimental error. In particular, in Task 2, our participants found it challenging to memorize all twelve words. Moreover, in a previous study (Buczny et al., 2015), the participants' cognitive inhibition types and the loss of cognitive inhibition caused an implicit attitude change toward the same tasks. Specifically, the "directed cognitive inhibition type" (individuals who prefer to inhibit the irrelevant concrete information when completing a task) participants may have an advantage in completing the task than the "undifferentiated cognitive inhibition type" (individuals who prefer to inhibit all processed information when completing a task) participants. Simultaneously, the natural loss of the cognitive inhibition of the participants over time will have a certain effect on target completion. Therefore, these factors that were considered a source of systematic error in our experiments that we were not able to control, and thus may also be a source of experimental error in our studies.

Research Development and Prospects

The use of Chinese materials as experimental materials in our study might be feasible under the condition that distinguished Chinese words by their different nature, and we observed a significant concreteness effect in our experiments. However, we still do not know the applicability of other types of Chinese experimental materials, such as the Chinese adjectives (which may relate to an individual's emotion and motivation) or Chinese verbs (which may relate to an individual's embodied cognition), which may also affect information processing and cognitive inhibition in the working memory task using the list-before-last paradigm. On this issue, future research needs to be further refined.

Furthermore, considering the item numbers of each list when researcher transform the list-before-last paradigm seems to be necessary. The most obvious difference compared with original study was "the number of items in each list" and we suspect that it was the main reason for the inconsistency results. Additionally, when participants freely recalled L2 in the final test, recent research observed significant differences in L3 intrusions between the math group (using a distraction task between L2 and L3 in list-before-last paradigm) and the retrieval L1 group (Sahakyan and Hendricks, 2012). The math group had significantly more intrusions than the retrieval group. However, in our study, particularly in Experiment 2, we created a free recall L1 group that was similar to the math group, but we still did not observe significant L3 intrusions in the final free recall L2 test.

In addition to the number of position change units, time estimates have also been shown to represent a marker of internal context change (Sahakyan and Smith, 2014). In terms of verbal estimates, the retrieval group recorded significantly longer time estimates throughout the experiment than the restudy group in the list-before-last paradigm, although the duration of the experiment was equal in both groups. Although this finding has not been confirmed by a sufficient number of studies, it might still represent a reference marker of internal context change.

In addition to experimental materials and experimental paradigms, future research can also focus on the methods used to present experimental materials due to the current rapid development of augmented reality (AR) and virtual reality (VR) technology. The perception of different spatial scales (particularly the large-scale space) by individuals might also affect the mechanism of the cognitive inhibitory effect on working memory. The application of additional neural science technology might be useful in investigations of the mechanism underlying the cognitive inhibitory effect on working memory at the technical level.

ETHICS STATEMENT

The authors have read and approved the submission of the manuscript. It has not been published in this or a substantially similar form (in print or electronically, including on a web site), nor accepted for publication or consideration thereof elsewhere, in whole or in part, in any language. The study was approved by the academic and ethics committee of school of education in Hebei University. The academic and ethics committees approved this consent procedure. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

AUTHOR CONTRIBUTIONS

XZ, CL, and CS conceived and designed the study. XZ and CL performed the study. XZ and CS analyzed the data. XZ, CL, and CS wrote the paper.

FUNDING

REFERENCES

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: separate lexical organization for concrete and abstract words. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 582–594. doi: 10.1037/0278-7393.13. 4.582

Buczny, J., Layton, R. L., and Muraven, M. (2015). The role of implicit affective responses and trait self-control in ego resource management. *Motiv. Emot.* 39, 669–679. doi: 10.1007/s11031-015-9489-7

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychon. Bull. Rev.* 12, 769. doi: 10.3758/BF0319 6772

Harnishfeger, K. K., and Bjorklund, D. F. (1993). The ontogeny of inhibition mechanisms: a renewed approach to cognitive development. *Emerg. Themes Cogn. Dev.* 1, 28–49. doi: 10.1007/978-1-4613-9220-0_2

Howard, M. W., and Kahana, M. J. (2002). A distributed representation of temporal context. *J. Math. Psychol.* 46, 269–299. doi: 10.1006/jmps.2001. 1388

James, C. T. (1975). The role of semantic information in lexical decisions. *J. Exp. Psychol. Hum. Percept. Perform.* 1, 130–136. doi: 10.1037/0096-1523.1. 2.130

Jang, Y., and Huber, D. E. (2008). Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 112–127. doi: 10.1037/0278-7393.34. 1.112

Klein, K., and Fiss, W. H. (1999). The reliability and stability of the turner and engle working memory task. *Behav. Res. Methods Instrum. Comput. A J. Psychon. Soc. Inc.* 31, 429. doi: 10.3758/BF03200722

Kounios, J., and Holcomb, P. J. (1994). Concreteness effects in semantic processing: erp evidence supporting dual-coding theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 804–823. doi: 10.1037/0278-7393.20.4.804

Kroll, J. F., and Merves, J. S. (1986). Lexical access for concrete and abstract words. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 92–107. doi: 10.1037/0278-7393. 12.1.92

Lehman, M., and Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 970. doi: 10.1037/a0015728

Pastötter, B., Kliegl, O., and Bäuml, K. H. T. (2012). List-method directed forgetting: the forget cue improves both encoding and retrieval of postcue information. *Mem. Cogn.* 40, 861–873. doi: 10.3758/s13421-012-0206-4

Sahakyan, L., and Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Mem. Cogn.* 40:844. doi: 10.3758/s13421-012-0198-0

Sahakyan, L., and Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1064. doi: 10.1037/0278-7393.28.6.1064

Sahakyan, L., and Smith, J. R. (2014). "A long time ago, in a context far, far away": retrospective time estimates and internal context change. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 86–93. doi: 10.1037/a0034250

Schwanenflugel, P. J., Harnishfeger, K. K., and Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *J. Mem. Lang.* 27, 499–520. doi: 10.1016/j.bandl.2013.01.005

Schwanenflugel, P. J., and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology Learning Memory & Cognition* 9, 82–102. doi: 10.1037/0278-7393.9.1.82

Shiffrin, R. M. (1970). Forgetting: trace erosion or retrieval failure? *Science* 168, 1601–1603.

Sui, X., Gao, M., Meng, L., and Yingxue, M. A. (2016). A probing analysis of the effect of concreteness effect of word property and word frequency. *J. Liaon. Normal Univ. (Soc. Sci. Edition)* 39, 52–57. doi: 10.16216/j.cnki.lsxbwk.201601052

Turner, M. L., and Engle, R. W. (1989). Is working memory capacity task dependent? *J. Mem. Lang.* 28, 127–154. doi: 10.1016/0749-596X(89)90040-5

Training for Coherence Formation When Learning From Text and Picture and the Interplay With Learners' Prior Knowledge

Tina Seufert*

Department for Learning and Instruction, Institute of Psychology and Education, Ulm University, Ulm, Germany

Learning with text and pictures requires learners to integrate the given information into one coherent mental representation. Since learners often fail to integrate text and pictures, the study investigates the effects of a training for text processing strategies, picture processing strategies and strategies to map text and picture onto each other. It was assumed that learners' prior knowledge would affect the effects of such a training with more beneficial effects for learners with high prior knowledge. The training comprised an introduction on how to process, integrate and reflect on texts and pictures with an additional training phase of 3 weeks. The study ($N = 30$) analyzed the effects of the training with regard to recall and comprehension performance in contrast to the no training group, which received an alternative program that was not related to text-picture integration. A regression analysis showed that the integration training was not overall beneficial but only for learners with increased levels of prior knowledge. Hence, training for coherence formation is beneficial for learning only when adequate knowledge structures are available to conduct the recommended steps of understanding and integrating text and picture.

Keywords: coherence formation, multimedia learning, text-picture-integration, effects of prior knowledge, aptitude-treatment-interaction

INTRODUCTION AND THEORETICAL BACKGROUND

When taking a look at modern learning material in books, on websites, in learning apps or conveyed by teachers, the most prominent presentation formats are texts accompanied by pictures. While texts are used to provide facts and details, pictures are often used to illustrate the topic and to provide an overview. In biology textbooks, for example, texts describe the processes that take place within a human cellular system like movements or transactions. The accompanying picture complements the understanding of these processes by providing an overview of the cellular structures. Together, text and picture help to understand a complex learning topic and both forms of representations convey different but interdependent information that has to be linked in the learner's mind (Ainsworth, 2006).

There is large evidence on the so called multimedia effect that when learning from text and picture learners' recall and comprehension performance is in fact higher than when learning from text alone (for an overview see Butcher, 2014). This fostering effect can be explained with Paivio's Dual Coding Theory (Paivio, 1986), i.e., text and pictures are first processed in separate memory

systems and hence lead to two separate memory traces. This dual coding of information enhances the probability to retrieve information from long-term memory. The positive effect on conceptual understanding or even on transfer of knowledge nevertheless depends on the above-mentioned integration of text and picture into one coherent representation within the learner's mental system (Mayer, 2009; see also Seufert, 2003; Scheiter et al., 2017). Hence, it is worth taking a closer look at the cognitive processes of integrating text and picture and their specific challenges in detail.

Learning From Text and Pictures

The most prominent model that describes processing of text and picture is Mayer's Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009). Based on Paivio's dual coding assumption (1986), information from text and picture is first selected in two separate subsystems. The next process of organizing requires the association of information within the text and picture subsystem, and hence encompasses the construction of separate mental models of the two information sources. In a last step, the information of the two mental models is integrated into one coherent mental representation by using prior knowledge stored in long-term memory. Mayer (1997) specifies this process of integration as referential processing by one-to-one-mappings. Thus, corresponding elements and relations of the single representations are related to each other in order to extract the underlying structure of the two representations together.

The model of integrative text and picture comprehension (ITPC) from Schnotz and Bannert (2003) also states separate processing systems for textual and pictorial information. However, the model especially differentiates the affordances of text processing in different steps. Learners first have to syntactically process the information without necessarily understanding the meaning. Only in a second step a semantic analysis helps to extract the meaning and leads to the deduction of propositions. These are then connected in a propositional network representation, which is still verbally coded. In the last step learners construct an analog mental model. Symbolic information therefore has to be translated into analog information. The external picture is also an analog representation and therefore using the picture as a scaffold can ease the construction of the mental model. Based on ITPC, the integration of text and picture information into one mental model means that the analog structure of the picture can build the frame for the mental model, which is then enriched by propositions from the text and is connected to learners' prior knowledge. Overall, in both models the processes of coherence formation and integration are accompanied by top down processes, i.e., by using prior knowledge.

Integrating Text and Picture

Both models describe the process of learning from text and picture and highlight the necessity to mentally integrate both sources. However, there is still no explicit model that describes the process of integration. Based on Mayer's (1997) description

of one-to-one-mapping, the integration process can be seen as a process of structure mapping. Learners identify relevant concepts or statements in the text and picture, compare them and link them if possible. The idea of identifying and linking corresponding elements is also explained in models of understanding multiple documents (for an overview see Barzilai et al., 2018). With reference to Gentner's (1983) structure mapping theory, Seufert and Brünken (2006) describe the process of integrating text and pictures, or multiple representations in general, as a process of mapping elements or relations in order to construct a coherent mental representation. Thus, this process is called coherence formation. Learners have to find corresponding elements that can be mapped onto each other (element-to-element mapping). Moreover, more comprehensive structures of elements and their interrelations have to be mapped onto each other (relation-to-relation-mapping). With reference to ITPC, one essential aspect for connecting representational structures is to translate between different sign systems. For example on the one hand, a pictorial element has to be verbalized in order to be able to relate it to other verbal elements. And on the other hand, verbal items must be translated into graphical structures in order to integrate them into one overall mental picture (e.g., Schnotz and Bannert, 2003).

Moreover, according to Seufert (2003) the mapping and translation processes can be conducted on different levels; syntactically or semantically. Information can only be processed superficially in order to extract the relevant surface features, e.g., shape or color (in a picture) or nouns and verbs (in a text). Hence, when a learner identifies corresponding surface features (e.g., when important parts are marked in red in the picture as well as in the text) and uses them as a hint for integration, this kind of mapping is called *syntactic mapping*. However, this does not necessarily come along with a deeper understanding of the single representations and consequently it does not ensure a comprehension of the overall relations. Instead, it would be desirable to animate learners to *semantic mapping*: In this case, elements and relations of single representations are mapped onto each other because learners really understand the semantic correspondences. Consequently, the resulting integrated knowledge structure is coherent and a basis for deeper understanding, appliance and transfer processes (Seufert, 2003).

It is obvious that finding relevant elements and relations within text and picture in order to map them onto each other can be a complex and effortful process for learners. This is especially the case when learners are not even able to identify what is relevant within the text or picture due to a lack of prior knowledge or strategies to understand texts or pictures. Thus, integrating text and picture on a semantic level has the potential to cause difficulties. There is evidence from eye-tracking studies that learners could in fact gain better learning outcomes when they show intensive transitions between text and pictures (Hegarty and Just, 1993; Scheiter and Eitel, 2015; Schüler, 2017) but that only a part of the learners actually showed such integrative behavior (Mason et al., 2013). Very often, learners pay attention to texts while they only briefly regard

the pictures (Hannus and Hyönä, 1999). Hence, they fail to successfully integrate both sources. Renkl and Scheiter (2017) also underline the challenges of integrating visual displays with other representations.

Consequently, there have been a lot of studies during the last decade dealing with the possibilities to foster learning with text and picture in order to cope with the difficulties and to profit from potential positive effects of an integrated mental representation.

Fostering Text-Picture-Integration

In general, there are two possible strategies to foster text-picture-integration. The first one is to add additional information like signals or explanations that help learners to identify corresponding elements in text and picture. Corresponding colors, connecting lines or the spatial integration of texts within pictures have frequently been used as signals for correspondences. A meta-analysis of Richter et al. (2016) on signaling revealed a small to medium fostering effect of signals on learning performance. Nevertheless, signals can only give hints of what could be mapped onto each other. Therefore, they are only low-key prompts for integration on a surface level. To ensure that learners actually engage in structure mapping on a semantic level, one could provide explicit information about the correspondences between representations (e.g., Seufert and Brünken, 2006). Such explicit explanations of semantic references turned out to be helpful, especially when combined with signals on the surface feature level (Seufert and Brünken, 2006). Instead of providing explanations on references, learners could also be prompted to find correspondences themselves. Studies in which integration prompts were used revealed positive effects on learning (e.g., Bodemer et al., 2005; Leopold et al., 2015). There is nevertheless evidence that learners can only profit from prompts, if the references they draw are actually correct (Leopold et al., 2015). Bodemer and Faust (2006) also could confirm that a drag and drop-integration task did not foster learning due to erroneous connections made. This was especially the case for learners with low levels of prior knowledge. These results reveal the shortcomings of prompts as a mean to foster text-picture integration. They can only help to overcome a production deficiency, i.e., learners are prompted to conduct a procedure they already know and master (Bannert, 2009). The reported studies nevertheless indicate that learners are not necessarily able to successfully integrate.

This leads to the second principal approach to foster the integration of text and picture: training learners to implement a successful strategy of coherence formation. With such a training one could enable learners to deal with text and pictures in general. Hence, the training would be more enduring and easier to transfer (e.g., Dignath et al., 2008).

Training to Integrate Text and Picture

A training that helps learners to integrate text and pictures can be seen as a training of a strategy that can be used for every combination of text and picture. Based on a vast amount of training studies for learning strategies in general (for an overview see Dignath et al., 2008), one can determine

crucial issues for effective strategy trainings. The first crucial issue is based on studies on training of cognitive strategies which point out that single elements of the strategy have to be mastered before the separate parts can be combined into one complex skill (McNamara et al., 2004). The second one is that learners should be provided with the crucial steps of the strategy, i.e., the cognitive aspects of the strategy as well as metacognitive strategies to regulate their strategy use (Berthold et al., 2007). When conducting the separate steps of a complex strategy, learners are then able to monitor their progress and can readjust their behavior, thus improving their strategy skills.

While these recommendations are crucial for strategy trainings in general, the above mentioned models of text-picture integration (Schnotz and Bannert, 2003; Mayer, 2009) as well as the concept of coherence formation as structure mapping (Seufert, 2003) provide specific guidelines for developing a training for text-picture integration.

As text and picture are processed separately in the first place, the trained strategy should comprise specific steps for text processing and picture processing. According to Mayer (2009), *text processing* starts with a selection process which, according to the ITPC model, mainly refers to surface features of the text. Thus, learners could start with getting an overview by scanning the headlines, the structure with its columns or chapters and first sentences. This bottom-up-process should be accompanied by top-down-processes. Thus, learners should activate their prior knowledge by reflecting on what they already know about the content. This fosters the process of organizing the information and deducing relevant propositions out of the text. Thereby, the crucial part of text processing strategies is to identify the relevant elements and relations that are the basis for subsequent mapping processes (e.g., McNamara et al., 2004). The last step, the construction of a mental model then requires learners to integrate the identified separate aspects into one coherent mental representation of the text, again by linking the text content to their prior knowledge. With reference to the ITPC model this last step of a mental model construction requires learners to mentally translate the verbal content into an analog mental structure. This process can be eased particularly when learners are prompted to self-explain the overall meaning of the text in their own words (McNamara et al., 2006).

The same strategy could be used for *picture processing*. Learners have to scan the picture to grasp the overall spatial structure. There is evidence that a first glance at the picture – and even a very short one – turned out to be highly effective for later text-picture processing (Eitel et al., 2013). The authors argue that learners can use the external picture as a scaffold for mental model construction. Learners then again should activate their prior knowledge in order to foster the selection and organization process. Due to their analog nature, pictures are often only viewed as one unit in a superficial way. Hence, Stalbovs et al. (2015) stress the importance of decomposing pictures into meaningful parts. This step should therefore be assisted by the instruction to identify and mark relevant elements and relations. However, for organizing and constructing a mental model of the picture content, learners need to bring the elements and relations

together again into one meaningful unit and they have to draw inferences from the picture (Hegarty, 2005).

However, the picture processing strategy and especially the steps of identifying relevant elements and relations differ notably for either realistic or logical pictures (Schnotz and Bannert, 2003). Thus, learners have to be provided with additional meta-representational information about the features of realistic pictures and with reference to the study of Miller et al. (2016) especially of logical pictures. They found that providing information about the conventions of diagrams in short warming-up tasks could increase understanding of logical pictures.

When learners possess effective strategies to deeply process and understand the text as well as the picture, the prerequisites for *structure mapping*, and hence for *integration* are given. With reference to the differentiation into surface and semantically oriented mapping processes (Seufert and Brünken, 2006) the learners could again start to use the surface features they extracted for their mapping process, like headlines or salient features. CTML as well as the ITPC model point out that the integration process is facilitated by using prior knowledge. Thus, the mapping strategy should also start with the activation of prior knowledge with respect to the overall content. This also enables semantic mapping processes and the linking of corresponding elements as well as relations between text and picture. As mentioned above, meaningful mapping is the crucial step for successful learning with text and picture. This is underlined by the eye-tracking study of Mason et al. (2013) where learners who integrated text and picture by looking intensively back and forth outperformed low-integrators.

Given that a training for text-picture integration should equip learners to deal with every possible combination of text and pictures learners should also learn how to evaluate the representational functions of text and picture. Based on Ainsworth's DEFT model (2006), representations can serve different functions. They can, for example, complement or constrain each other or they can even be redundant. Thus, as a part of the mapping strategy learners should evaluate whether text and picture or parts of them are redundant or complementary or whether there are parts that do not refer to the other representation or that are maybe even irrelevant. This reflection helps learners to gain meta-representational knowledge about texts and pictures and their functions. Schwonke et al. (2009) could show that providing such information about representational characteristics can deepen learners understanding and intensify the mapping process.

There is evidence from two previous studies that such a comprehensive training or support for integrating text and pictures can foster learning. Schlag and Ploetzner (2011) conveyed a short presentation of a step-by-step plan for text-picture-processing in their study and let learners practice these steps afterwards. The training turned out to be effective compared to a control group without training for all levels of understanding (factual, conceptual and transfer). The second study from Stalbovs et al. (2015) used implementation intentions as a specific strategy to support the essential steps of text-picture-integration. Learners should internalize if-then-plans, so that whenever they

are in a specific situation (e.g., if I have opened a new page) they will conduct a specific operation (e.g., then I will carefully study the title first). Stalbovs et al. (2015) instructed learners to internalize different variations of such implementation intentions that either addressed deepened text-, picture- or text-and-picture-processing. Learners were best supported when all three aspects were covered by the implementation intentions, which will also be the case in our training.

Thus, overall there is evidence that learning from text and pictures can be improved when learners are provided with a strategy training that comprises the crucial steps of text-processing, picture-processing and text-picture integration. However, based on prior studies on the effectiveness of help for coherence formation when learning from multiple representations one can ask whether training effects might depend on learners' prior knowledge.

Aptitude-Treatment-Interaction Effect of Help for Coherence Formation

Considering the affordances of the above mentioned strategies to understand text and picture and to integrate them, learners should be able to identify and map relevant elements between text and picture. With insufficient prior knowledge in the domain, learners lack appropriate cognitive schemata to identify the relevant elements in the single representations. In scientific domains and mathematics, it has often been proven that learners only concentrate on surface features (for an overview see Ainsworth, 2006) and hence cannot map between the representations semantically. They also often face problems with translating between different representational codes (Baker et al., 2001). It is also plausible that novice learners experience increased intrinsic cognitive load because they cannot build meaningful chunks (Bannert, 2002). Thus, their cognitive resources can easily be overloaded when they try to meaningfully integrate text and pictures. To additionally handle an unknown strategy could be even more strenuous and thus a strategy training might not be effective (Bjorklund and Coyle, 1995). These theoretical assumptions are in line with empirical results on the effectiveness of help for coherence formation. Based on a study of Seufert (2003), one can assume that prior knowledge is actually relevant for the effectiveness of help for coherence formation. The study revealed that the hints for integrating different representations only turned out to be effective for learners with a medium level of prior knowledge, whereas learners with too low or too high levels of prior knowledge did not improve when help was provided. The paper argues that especially novices lack the abilities to use such help adequately even though they would need it (see also Bodemer and Faust, 2006). Learners with high levels of prior knowledge also do not profit from help because they do not need it any longer. Only learners with a medium level of prior knowledge will still need some assistance and have enough resources and conceptual background to use the help effectively. They will be met in their zone of proximal development (Vygotsky, 1978) where help can effectively be used to accomplish the next level of expertise. In the present study it will be analyzed whether these moderating effects

can also be revealed when learners are provided with coherence formation strategies in a pre-training.

RESEARCH QUESTIONS AND HYPOTHESES

The present study investigates, whether strategies for integrating text and picture can effectively be conveyed in a pre-training, and whether the effects of training depend on learners' prior domain specific knowledge.

Based on the different levels of processing when dealing with text and picture, the training was designed to be helpful for both, recall and comprehension. But as the strategy especially aims at integrating text and picture on a semantic level, the effects should be stronger on comprehension as a higher level of processing.

However, the training is not assumed to be effective in general. Instead, learner's prior knowledge should affect the effectiveness of the training. Only with a sufficient level of prior knowledge should learners be able to apply the strategy. Their existing schemata will help to identify the relevant elements and relations in text and picture and to map them onto each other on a semantic level. Learners with lower levels of prior knowledge should have difficulties in applying the strategy and even if they manage to extract and map relevant information, they might not be able to build semantically meaningful chunks. To handle the newly acquired strategy would pose additional load. Consequently, the training might even be harmful for them compared to a no-training condition. In terms of Mayer (1997), we hypothesize an enhancing effect of prior knowledge for the effectiveness of the training. If the sample would also include experts with high levels of prior knowledge one could expect no or even detrimental effects for them, as they should be able to accomplish the integration task without any further help. The additional information about implementing the strategies could thus lead to an additional mental effort, as the experts would have to actively ignore it. Thus, the so-called expertise reversal effect (Kalyuga, 2007) could be expected.

MATERIALS AND METHODS

Participants and Design

Thirty university students of psychology and teacher training programs participated in the experiment. 14 of them were female and the average age was 23.53 (SD = 3.25). The sample only comprised learners with low to medium prior knowledge. Thus, no expertise-reversal effect will be analyzed as there were no expert learners in the given sample.

Participants were randomly assigned to one of the two treatment groups [experimental group with training (EG; n = 15) and control group without training (CG; n = 15)]. As dependent variables learners' performance was measured, differentiated for recall and comprehension.

In a linear regression analysis the effects of the treatment as a categorical factor (with or without training), prior knowledge as a continuous factor, and the interaction of both by including the product of both variables were analyzed. Learners' spatial abilities and working memory capacity were correlated with the performance measures and thus were included in the model as covariates.

Materials and Procedure

The experiment was part of an advanced seminar in educational science. Students, nevertheless, could decide for themselves whether they wanted to participate in the study or not. The experiment was conducted in three sessions: one pretest session, the training session (or the alternative session for the control group) with a 3 weeks practicing phase afterwards and the posttest in a separate session to prevent exhaustion.

Pretest Session

In the pretest session, we analyzed prior knowledge of the content domain, which will be used in the posttest (the function of an Otto engine) with 4 open questions and one picture-labeling task. The prior knowledge test comprised questions on a recall level, like "Name the 4 strokes of the Otto engine cycle" and on a comprehension level, like "explain the two processes that causes the warming of the air-fuel-mixture." Maximally 11.5 points could be reached and Cronbach's α = 0.84 was sufficient. Additionally, a test for spatial abilities (% correct) was conducted (Paper Folding and Card Rotation test; Ekstrom et al., 1976). At last, working memory capacity was assessed (memory updating numerical, Oberauer et al., 2000). The score in this test reflects the number of related elements learners can process simultaneously and it usually ranges from 1 to 6 (with a theoretical maximum of nine). Overall, all pretests took about 1 h.

Training Program for the Experimental Group

The training consisted of a training session and a 3 weeks practice phase afterwards. The training session took place 1 week after the pretest and lasted 90 min. The students of the experimental group worked individually with a workbook to train the coherence formation strategy. The individual learning phase allowed individual pacing. The workbook comprised three strategy parts: (1) a text reading strategy, (2) a picture reading strategy for realistic and logical pictures and (3) a strategy for integrating texts and pictures. For each of these strategies the workbook provided a step-by step explanation of how to apply the strategy. The different steps are outlined in **Table 1**.

Each step is formulated as a task or a question that has to be answered, e.g., what are the relevant data points in the diagram or what does the text explain that cannot be seen in the picture. After having read the introduction of each strategy the workbook provided a worked example for either a text, a picture or a text-picture combination, where the different steps of the strategy were implemented and annotated (like in the study of Berthold et al., 2007). Only then learners were asked to apply the strategy on their own with a new text, picture or text-picture combination. The training materials were all in the domain of natural science but in different areas, like geography, biology or ecology (for an

TABLE 1 | Overview of the trained strategies and their crucial steps.

Type of strategy	Strategy steps	Rationale for strategy steps
Text reading	Get a first overview	Scanning/bottom-up
	Think of what you already know	Activating prior knowledge/top-down
	Read the text and mark relevant words	Identifying relevant elements
	Re-read the text, mark relevant sections and annotate them with short summaries	Identifying relevant relations
	Summarize the main statements of the text with reference to your prior knowledge either verbally or by sketching a picture	Structuring and elaborating to ensure understanding and long term recall
Picture reading	Identify the type of picture	Sensitize for different types of pictures
	Get a first overview	Scanning/bottom-up
	Think of what you already know	Activating prior knowledge/top-down
	Search the picture for relevant picture parts (e.g., figure-ground and highlighting in realistic pictures/a legend, labels of axis, salient features in logical pictures)	Identifying relevant elements
	Look for overall structures in the picture (e.g., interlinked picture parts in realistic pictures/slopes or contrasts of trends for logical pictures)	Identifying relevant relations
	Summarize the main statement of the picture with reference to your prior knowledge and write it down	Structuring and elaborating to ensure understanding and long term recall
Text-picture integration	Get a first overview by reading the title of the text and scanning the picture	Scanning/bottom-up
	Think of what you already know	Activating prior knowledge/top-down
	Read and understand the text by using the text reading strategy	Text processing
	Examine and understand the picture by using the picture reading strategy	Picture processing
	Identify corresponding elements in text and picture	Element- to-element-mapping
	Identify corresponding statements/structures in text and picture	Relation-to-relation-mapping
	Identify complementary elements or statements/structures in text and picture	Evaluating the function/relation of text and picture
	Summarize the main statement of the text-picture-combination with reference to your prior knowledge and write it down	Structuring and elaborating to ensure understanding and long term recall
Overall	Using the strategies checklist as a guideline and reflection tool	Metacognitive monitoring

example see **Figure 1**). The worked examples were always in a different scientific area than the practicing examples.

After this session, participants of the experimental group exercised the strategies by using the workbook for 3 weeks. In the 1st week they practiced the text reading strategy, in the second the picture reading strategy and in the 3rd week the mapping strategy with texts and pictures. Participants were reminded via email. They had to apply the strategies by using representations from their daily live or from current lessons of their study program. With this, we intended to provide a more meaningful setting and thus to enhance compliance and strategy transfer. We collected participants practicing materials every week at the beginning of the seminar course and checked for traces of strategy use. We found clear evidence for strategy use in all texts, pictures and text-picture-combinations.

Alternative Program for the Control Group

The treatment of the control group also comprised a 90-min session on-site and a 3-week elaboration phase outside the classroom. During the seminar session, students had to work on the pros and cons of the use of new media in school. First, students had to read introducing texts and discussed them afterwards in a teamwork discussion during the seminar. Subsequently, they further discussed this issue in a 3-week

lasting online discussion forum. Thus, the topic was not related in any way to strategies for reading or integrating texts and pictures. Both groups had a 3 weeks period to work on their tasks. The material that was handed in by the training group as well as the statements in the discussion forum indicates that they all spent a reasonable amount of time. Due to ethical reasons the control group also received the strategy training material after the last session. The learning material for the experimental as well as for the control group can be seen in the **Supplementary Datasheets S1, S2**.

Posttest Session

In the posttest session (at the end of the 3-week training or online-discussion session) both groups received learning material about the function of a four-stroke Otto engine. The experimental group was instructed to use the acquired learning strategies whereas the control group had no further instruction besides the task description. The material comprised a brief introduction to the function of the Otto engine and a labeled picture of the engine's structure. The processes of the 4 strokes were nevertheless only described verbally while the four pictures of each of the four strokes were given unordered at the end of the learning material. While studying the material, learners had to relate the appropriate picture to the corresponding description of
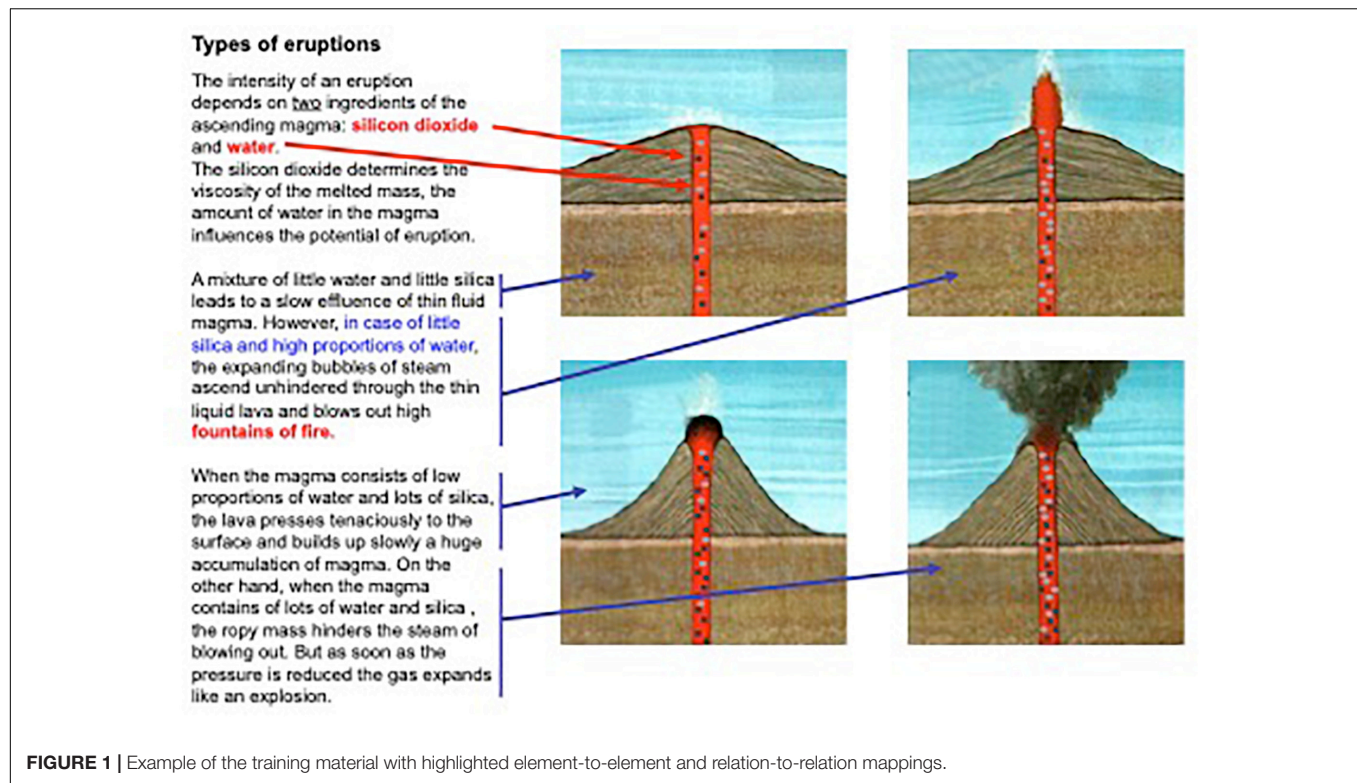
FIGURE 1 | Example of the training material with highlighted element-to-element and relation-to-relation mappings.

each stroke. The number of the correct relations indicated global coherence formation and was included in the comprehension measure of the post-test that was conducted after learning. The test comprised 3 open-ended recall tasks, asking for the most important propositions of the text. In addition, learners had to sketch the picture with its labels. Comprehension was measured with 3 open questions where learners had to draw inferences from what they learned from the text and picture. In addition, the score of the text-picture relation task was integrated. The recall test had a maximum score of 16.5 points, 11 points could be reached in the comprehension test. Cronbach's Alpha was sufficient for the recall measure ($\alpha = 0.74$) but lower for the comprehension test ($\alpha = 0.63$) due to its various inference tasks.

RESULTS

Descriptive Results

As can be seen in **Table 2**, Learners' prior knowledge was overall on a very low level and their spatial abilities as well as their working memory capacity were on a medium level. A MANOVA with the treatment (training versus no training) as independent variable and prior knowledge, spatial ability and working memory capacity as dependent variables revealed that the groups did not differ concerning their prior knowledge, and their spatial abilities ($Fs < 1$), ns and also not significantly concerning their working memory capacity [$F(1,28) = 2.22$, $p = 0.15$]. However, the Kolmogorov–Smirnov-Test revealed that prior knowledge was not normally distributed, [$D(28) = 0.22$,

$p < 0.01$], but the Levene-test showed homogeneous variances as well ($F < 1$, ns). Thus, the results have to be interpreted carefully, mainly based on a descriptive level.
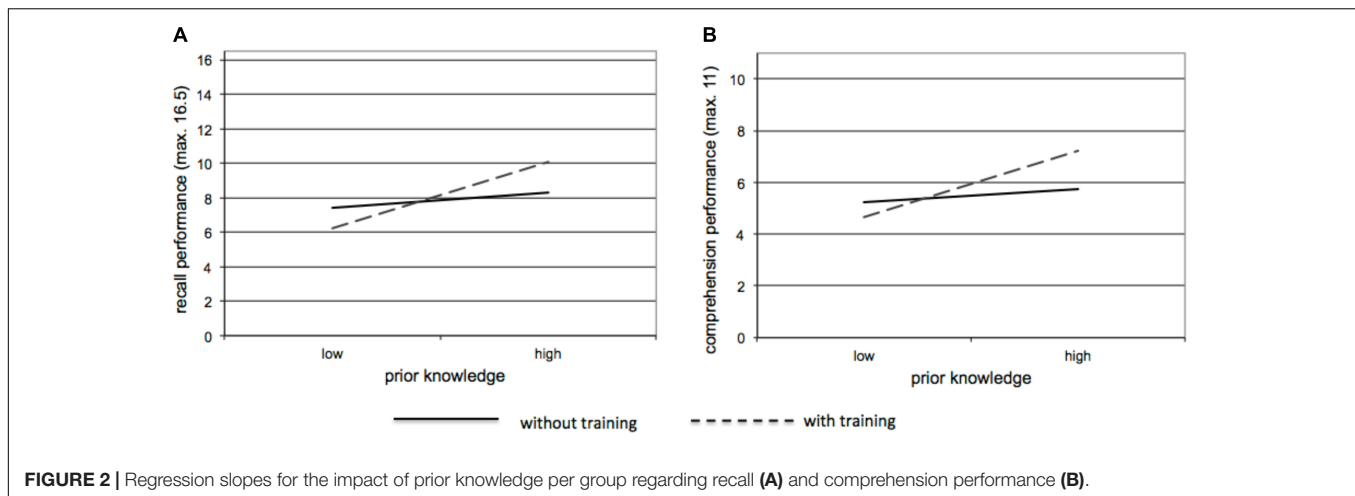
As both control variables, i.e., spatial ability and working memory capacity were positively correlated with recall ($r_{spatial} = 0.52$, $p < 0.01$; $r_{wmc} = 0.31$, $p < 0.10$) and comprehension measures ($r_{spatial} = 0.50$, $p < 0.01$; $r_{wmc} = 0.44$, $p < 0.05$) we entered them as covariates in the subsequent analyses.

Treatment Effects in Interaction With Learners' Prior Knowledge

In order to test the hypotheses whether the training is effective compared to no training and whether these effects depend on learners' prior knowledge a regression model was analyzed for recall and comprehension performance with the following predictors: treatment, prior knowledge, the product term treatment * prior knowledge, working memory capacity

TABLE 2 | Means (standard deviations) for control variables and dependent variables in both groups.

	No Training ($n = 15$)	Training ($n = 15$)
Prior knowledge (maximum 11.5)	2.87 (3.55)	2.17 (2.96)
Spatial abilities (%)	68.76 (17.84)	67.88 (13.31)
Working memory capacity	4.00 (1.41)	3.27 (1.28)
Recall (maximum 16.5)	8.70 (3.87)	9.20 (3.90)
Comprehension (maximum 11)	6.13 (2.80)	6.47 (2.97)

FIGURE 2 | Regression slopes for the impact of prior knowledge per group regarding recall **(A)** and comprehension performance **(B)**.

and spatial abilities. At first, treatment was coded with 0 for the control group and 1 for the training group. In a second step the treatment factor was recoded (control = 1, training = 0) and the regression analysis was conducted again. With this method of "recentering", proposed by Aiken and West (1991), it is possible to analyze the specific impact of prior knowledge for the respective group which is coded with 0 as reference group. First, it has to be noted that the Kolmogorov–Smirnov-Test revealed that the recall data were normally distributed [$D(28) = 0.11$, $p > 0.05$], but that the comprehension scores were not [$D(28) = 0.173$, $p < 0.05$]. In addition the Levene-test revealed that the variances for both outcomes measures were homogeneous [recall: $F(1,28) = 4.06$, $p > 0.05$; comprehension: $F(1,28) = 1.142$, $p > 0.05$].

For *recall* performance the regression model was significant, [$F(5,29) = 6.25$, $p = 0.001$, R^2adj = 0.48]. The treatment factor (training versus no training) was not significant [beta = 0.43, $t(29) = 0.26$, $p = 0.79$]. Learners in the two groups showed almost the same performance (for all outcome measures see **Table 2**). However, the aptitude-treatment-interaction was significant [beta = 0.42, $t(29) = 2.17$, $p = 0.04$]. Thus, the influence of learners' prior knowledge differed significantly between the groups, as **Figure 2A** depicts: while prior knowledge had no influence in the CG [beta = 0.19, $t(29) = 1.04$, $p = 0.31$] it had a significant influence in the EG [beta = 0.81, $t(29) = 3.72$, $p = 0.001$]. With increasing prior knowledge learners showed increased learning performance in the training group. Spatial abilities also turned out to be significantly predictive [beta = 0.37, $t(29) = 2.46$, $p = 0.02$]. Working memory capacity had a positive but non-significant influence [beta = 0.30, $t(29) = 1.89$, $p = 0.07$].

For *comprehension* performance we found almost similar results (see **Figure 2B**). The overall model was significant [$F(5,29) = 6.18$, $p = 0.001$, R^2adj = 0.47]. The training had no overall effect [beta = 0.08, $t(29) = 0.51$, $p = 0.61$]. The interaction pattern is also significant [beta = −0.48, $t(29) = −2.04$, $p = 0.05$]. Again, the differentiated analyses revealed that prior knowledge had no influence in the CG [beta = 0.15, $t(29) = 0.81$, $p = 0.43$] but significantly predicted comprehension

performance in the EG [beta = 0.73, $t(29) = 3.35$, $p = 0.003$]. Again learners with increasing prior knowledge revealed higher comprehension scores in the training group. Comprehension was not significantly influenced by spatial abilities [beta = 0.30, $t(29) = 1.99$, $p = 0.06$] but by working memory capacity [beta = 0.46, $t(29) = 2.90$, $p = 0.008$].

SUMMARY AND DISCUSSION

Texts are often enriched with pictures and based on the well-known multimedia principle learners can profit from such a combination (Butcher, 2014). However, the beneficial effects of an additional picture only pay off when learners actually integrate text and picture information into one coherent mental representation (Ainsworth, 2006). In this study a training was developed and analyzed that provides learners with the crucial steps of understanding and integrating text and picture combinations. Overall, it was assumed that the training could be helpful for learning but that these effects will be moderated particularly by learner's prior knowledge.

In fact, we found no overall positive effect of the training, neither for recall performance nor for comprehension performance. Thus, the training is not effective in general.

However, we could confirm the expected moderating effect of prior knowledge for *recall performance*. The first and main result is that prior knowledge especially affected the results in the training condition. As assumed, learners could only profit from the training with sufficient prior knowledge, i.e., we found an enhancing effect of prior knowledge. With insufficient prior knowledge the training was not effective or even hindered learning. These findings are in line with previous studies on situational help for coherence formation (Seufert, 2003; Seufert et al., 2007). They also found that help is only effective for learners with sufficient but not too high levels of prior knowledge. These learners still are in need of help and are capable of using it. In our sample only 15% of the learners reached at least half of the possible scores in the pre-test, thus we only have very few learners with higher expertise. So we can ask how experts would

have performed with the training. Based on the expertise reversal effect (Kalyuga, 2007) where expert learners are actually hindered by unnecessary help, one could assume that our training may also produce such reversal. Experts would not need the strategy, because they can extract the semantic structure of text and picture based on their knowledge. Moreover, the proposed strategy might even interfere with their existing strategies and has to be ignored actively causing unnecessary burden on learners' resources.

The second interesting aspect of the interaction pattern we found is that prior knowledge has no significant influence on recall in the control condition. Without any further help even higher knowledgeable learners show only medium performance scores. This is further evidence for the argument that many learners have substantial difficulties in integrating text and picture and that assistance is needed (see Ainsworth, 2006; Renkl and Scheiter, 2017). Nevertheless, there is further evidence needed with a greater sample with normally distributed scores of learners' prior knowledge. Until then, the results should be interpreted carefully mainly based on a descriptive level, that shows the different slopes of the two groups.

Concerning the effects of the training on *comprehension performance* we also found an influence of learners' prior knowledge, but with a smaller effect. We again found the same pattern that with increasing prior knowledge learners profited from the training and once more prior knowledge showed no effect in the no training group. Again, one has to consider the effects with care as the prior knowledge scores as well as the comprehension scores were not normally distributed. Nevertheless, the slopes show different increases but we would have expected even stronger effects on comprehension performance as the training explicitly aimed at semantic mapping processes. And especially when it comes to comprehension, learners should profit from their prior knowledge as this could help to link new and existing knowledge and to build meaningful schemata. One could speculate why learner's prior knowledge does not have the expected stronger enhancing function while using the trained strategy for comprehension. Maybe learners do not make the link between their existing knowledge and the new information or they do not aim at understanding the material even with prior knowledge. Instead they might integrate on a surface level by syntactic mappings. When taking a closer look at the structure of the strategy training, it is also plausible that learners tend to follow the strategy instructions stepwise in a successive order. Thus, they first primarily focus on elements and relations in the text, then on the picture and only afterwards they link both structures. With these fine-grained analyses of the two sources, the overall picture might get lost and learners do not strive at building an overall network of all the information where they could effectively use their existing network of knowledge.

But all these possibilities remain speculative since we have no further indicators for the processes learners actually execute. Process data like thinking aloud protocols or eye-tracking data could provide further information about if and how the strategy is applied, whether it needs to be refined or whether additional help is needed. One could also learn more about the interplay with learners' prior knowledge. Despite the processes, it would also be valuable to analyze not only cognitive but also motivational effects of the training. The effects of the training will surely depend on the commitment the learners have toward the strategy and this in turn surely depends on whether they actually evaluate it as useful. Seufert (2018) suggests that the amount of regulation, in our case the intensity of using the trained strategy, depends on the necessity of this strategy to accomplish the goal, the available resources to accomplish the strategy and the resulting load imposed by the strategy use. While we analyzed learner's prior knowledge as one crucial resource, we did not take into account the necessity or the appraisal of usefulness as suggested above. Additionally, we did not investigate the experienced load when using the strategy. As argued above the use of a newly trained strategy could impose additional load in terms of extraneous load as it is not yet automated and requires resources for conducting and monitoring the proceeding steps. In contrast, one could also assume that strategy assistance could also relieve learners as they are guided step by step. The study of de Bruin et al. (2005) provides evidence for such a relieving effect of a strategy instruction. However, as the strategy in our study was very complex and surely cannot be automatized after only 1 h or even 3 weeks of occasional exercises, we would assume an increase in cognitive load. Moreover, we additionally asked learners to reflect on their strategy what could impose an additional metacognitive load (Bannert, 2002). In terms of germane processing one could also assume that learners who are able and willing to follow the strategy would also invest germane resources. Thus, a differentiated measurement of learners' perceived extraneous and germane load could further enlighten the actual effects of the training (Klepsch et al., 2017).

Another important point is that the training should be compared to a stronger control condition, which also provides a strategy training, like e.g., on metacognition. In both groups learners would then have to handle an additional strategy while learning and hence the cognitive affordances would be comparable. Only then one could qualify the effects on learning outcomes as effects of a training on coherence formation in contrast to an alternative training.

Overall, we developed a training that can be helpful for coherence formation if learners have sufficient prior knowledge and are thus able to deal with the possible additional burden and therefore can accomplish the strategy successfully. With this constraint one cannot actually recommend to implement the strategy training in instructional settings. To ensure that also learners with low prior knowledge can benefit from strategy instruction the affordances have to be further decreased. This could be accomplished by either providing pre-training where the most relevant concepts of the learning domain are conveyed or by segmenting the elements of the coherence formation strategy (Ayres, 2013). Learners could first be provided with strategies for single representations like text or reading strategies. Only when these strategies are automated the next level of coherence formation should be addressed. Segmenting could lessen the intrinsic cognitive load – in this case of the strategy – and therefore even novices could learn successfully. Moreover, with an extended research program with variations of the training, it would also be possible to analyze differential effects of the training components. Are all training components necessary, which of

them produce the strongest effects, for which processes and for whom? Hence, it would be interesting and necessary to take a deeper insight into the learner's mind by asking them to think aloud or to evaluate their load repeatedly and differentially. In spite of a successive implementation of the strategy parts one could also think of a fading strategy to ensure that learners are able to conduct the strategy autonomously. Studies on fixed versus faded prompts show promising effects on strategy transfer in the long run (Davis, 2003). Generally, it could be interesting to introduce a follow-up measure to see whether the positive effects for high prior knowledge learners persist or whether there are any sleeper effects for low prior knowledge learners: it would be possible that the trained strategies are practiced in the meantime, so that they can be carried out in the follow-up test with less mental effort, resulting in improved learning outcomes. If this were the case, we would have a strong argument for enlarged training programs, which could be implemented in classroom teaching over a longer period of time.

However, even if the study provides some first insights in how and for whom strategies for text-picture integration can be trained it also has some major shortcomings. The major problem is the very small sample size that hampers a broad generalizability and restricts the statistical power. In addition, the sample also mainly consist of low prior knowledge learners. Thus, we could not ensure normal distribution of the data. In the naturalistic setting (with the training being part of a whole course with repeated training or testing phases), which was chosen to ensure the external validity of the study, it was not possible to obtain a greater number of participants with less skewed data. The complex procedure with high affordances for the students' commitment can be seen as an additional flaw. Whether participants' commitment was actually high cannot be ensured, but at least it should have been assessed in an appropriate way. This could have helped to qualify the intensity of strategy use. Also the students' products when using the strategy with their own study materials in the 3 weeks after the strategy

training session could have been analyzed. However, as they are further needed in their courses they could not hand them over. For replicating the effects of the training one should ensure a larger sample in a classroom setting over a longer period of time where one could implement the strategy as inherent part of the curriculum. This could allow a deeper insight in the processes and products and more complex analyses of mediating or moderating effects. Based on these, one could refine the training and might even find an adaptation mechanism to ensure effective trainings for learners based on their individual learner characteristics and on their individual progress.

ETHICS STATEMENT

This study was exempt from an ethic committee approval due to the recommendations of the German Research Association: All subjects were in no risk out of physical or emotional pressure, we fully informed all subjects about the goals and process of this study and none of the subjects were patients, minors or persons with disabilities. Participation was voluntary and all subjects signed a written informed consent and were aware that they had the chance to withdraw their data at any point of the study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00193/full#supplementary-material

REFERENCES

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage Publications.

Ainsworth, S. (2006). DeFT: a conceptual framework for learning with multiple representations. *Learn. Instr.* 16, 183–198. doi: 10.1016/j.learninstruc.2006.03.001

Ayres, P. (2013). Can the isolated-elements strategy be improved by targeting points of high cognitive load for additional practice? *Learn. Instr.* 23, 115–124. doi: 10.1016/j.learninstruc.2012.08.002

Baker, R. S., Corbett, A. T., and Koerdinger, K. R. (2001). "Toward a Model of Learning Data Representations," in *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, eds J. Moore and K. Stenning (New Jersey, NJ: Erlbaum), 44–50.

Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. *Learn. Instr.* 12, 139–146. doi: 10.1016/S0959-4752(01)00021-4

Bannert, M. (2009). Promoting self-regulated learning through prompts. *German J. Educ. Psychol.* 23, 139–145. doi: 10.1024/1010-0652.23.2.139

Barzilai, S., Zohar, A. R., and Mor-Hagani, S. (2018). Promoting integration of multiple texts: a review of instructional approaches and practices. *Educ. Psychol. Rev.* 30, 973–999. doi: 10.1007/s10648-018-9436-8

Berthold, K., Nückles, M., and Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learn. Instr.* 17, 564–577. doi: 10.1016/j.learninstruc.2007.09.007

Bjorklund, D. F., and Coyle, T. R. (1995). "Utilization Deficiencies in the Development of Memory Strategies," in *Memory Performance and Competencies*, eds F. E. Weinert and W. Schneider (Mahwah, NJ: Lawrence Erlbaum Associates), 161–180.

Bodemer, D., and Faust, U. (2006). External and mental referencing of multiple representations. *Comput. Hum. Behav.* 22, 27–42. doi: 10.1016/j.chb.2005.01.005

Bodemer, D., Plötzner, R., Bruchmüller, K., and Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instr. Sci.* 33, 73–95. doi: 10.1007/s11251-004-7685-z

Butcher, K. R. (2014). "The multimedia principle," in *The Cambridge Handbook of Multimedia Learning*, ed. R. E. Mayer (Cambridge: Cambridge University Press), 174–206. doi: 10.1017/CBO9781139547369.010

Davis, E. A. (2003). Prompting middle school science students for productive reflection: generic and directed prompts. *J. Learn. Sci.* 12, 91–142. doi: 10.1207/S15327809JLS1201_4

de Bruin, A. B., Schmidt, H. G., and Rikers, R. (2005). The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural

equation modeling approach. *Acad. Med.* 80, 765–773. doi: 10.1097/00001888-200508000-00014

Dignath, C., Buettner, G., and Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educ. Res. Rev.* 3, 101–129. doi: 10.1016/j.edurev.2008.02.003

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., and Holmqvist, K. (2013). How a picture facilitates the process of learning from text: evidence for scaffolding. *Learn. Instr.* 28, 48–63. doi: 10.1016/j.learninstruc.2013.05.002

Ekstrom, R. B., French, J. W., and Harmann, H. H. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests.* Princeton, NJ: Educational Testing Services.

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cogn. Sci.* 7, 155–170. doi: 10.1207/s15516709cog0702_3

Hannus, M., and Hyönä, J. (1999). Utilization of Illustrations during learning of science textbook passages among low- and high-ability children. *Contemp. Educ. Psychol.* 24, 95–123. doi: 10.1006/ceps.1998.0987

Hegarty, M. (2005). "Multimedia learning about physical systems," in *The Cambridge Handbook of Multimedia Learning*, ed. R. E. Mayer (New York, NY: Cambridge University Press), 447–465. doi: 10.1017/CBO9780511816819.029

Hegarty, M., and Just, M. A. (1993). Constructing mental models of machines from texts and diagrams. *J. Mem. Lang.* 32, 717–742. doi: 10.1006/jmla.1993.1036

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educ. Psychol. Rev.* 19, 509–539. doi: 10.1007/s10648-007-9054-3

Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.3389/fpsyg.2017.01997

Leopold, C., Doerner, M., Leutner, D., and Dutke, S. (2015). Effects of strategy instructions on learning from text and pictures. *Instr. Sci.* 43, 345–364. doi: 10.1007/s11251-014-9336-3

Mason, L., Pluchino, P., and Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: evidence from eye movements. *Read. Res. Q.* 48, 199–214. doi: 10.1002/rrq.41

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educ. Psychol.* 32, 1–19. doi: 10.1207/s15326985ep3201_1

Mayer, R. E. (2009). *Multimedia Learning.* Cambridge: Cambridge University. doi: 10.1017/CBO9780511811678

McNamara, D. S., Levinstein, I. B., and Boonthum, C. (2004). iSTART: interactive strategy training for active reading and thinking. *Behav. Res. Methods* 36, 222–233. doi: 10.3758/BF03195567

McNamara, D. S., O'Reilly, T. P., Best, R. M., and Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *J. Educ. Comput. Res.* 34, 147–171. doi: 10.2190/1RU5-HDTJ-A5C8-JVWE

Miller, B. W., Cromley, J. G., and Newcombe, N. S. (2016). Improving diagrammatic reasoning in middle school science using conventions of diagrams instruction. *J. Comp. Assisted Learn.* 32, 374–390. doi: 10.1111/jcal.12143

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., and Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Pers. Individ.* 29, 1017–1045. doi: 10.1016/S0191-8869(99)00251-2

Paivio, A. (1986). *Mental Representations: A dual Coding Approach.* Oxford: Oxford University Press.

Renkl, A., and Scheiter, K. (2017). Studying visual displays: how to instructionally support learning. *Educ. Psychol. Rev.* 29, 599–621. doi: 10.1007/s10648-015-9340-4

Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–36. doi: 10.1016/j.edurev.2015.12.003

Scheiter, K., and Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learn. Instr.* 36, 11–26. doi: 10.1016/j.learninstruc.2014.11.002

Scheiter, K., Schüler, A., and Eitel, A. (2017). "Learning from multimedia: Cognitive processes and Instructional Support," in *The Psychology of Digital Learning*, eds S. Schwan and U. Cress (Cham: Springer International Publishing), 1–19.

Schlag, S., and Ploetzner, R. (2011). Supporting learning from illustrated texts: conceptualizing and evaluating a learning strategy. *Instr. Sci.* 39, 921–937. doi: 10.1007/s11251-010-9160-3

Schnotz, W., and Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learn. Instr.* 13, 141–156. doi: 10.1016/S0959-4752(02)00017-8

Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: evidence for text-picture integration. *Learn. Instr.* 49, 218–231. doi: 10.1016/j.learninstruc.2017.03.001

Schwonke, R., Berthold, K., and Renkl, A. (2009). How multiple external representations and how they can be made more useful. *Appl. Cogn. Psychol.* 23, 1227–1243. doi: 10.1002/acp.1526

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learn. Instr.* 13, 227–237. doi: 10.1016/S0959-4752(02)00022-1

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educ. Res. Rev.* 24, 116–129. doi: 10.1016/j.edurev.2018.03.004

Seufert, T., and Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Appl. Cogn. Psychol.* 20, 321–331. doi: 10.1002/acp.1248

Seufert, T., Jänen, I., and Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Comput. Hum. Behav.* 23, 1055–1071. doi: 10.1016/j.chb.2006.10.002

Stalbovs, K., Scheiter, K., and Gerjets, P. (2015). Implementation intentions during multimedia learning: using if-then plans to facilitate cognitive processing. *Learn. Instr.* 35, 1–15. doi: 10.1016/j.learninstruc.2014.09.002

Vygotsky, L. S. (1978). *Mind and Society: The Development Of Higher Mental Processes.* Cambridge, MA: Harvard University Press.

The Development of Students' Understanding of Science

Stella Vosniadou*

College of Education, Psychology and Social Work, Flinders University, Adelaide, SA, Australia

Children construct intuitive understandings of the physical world based on their everyday experiences. These intuitive understandings are organized in skeletal conceptual structures known as framework theories. Framework theories are different from currently accepted science and impose constraints on how students understand the scientific explanations of phenomena causing the creation of fragmented or synthetic conceptions. It is argued that in order to understand science students need to make important changes in the way they represent and explain the physical world as well as in their ways of reasoning. During the development of science knowledge students must also create new concepts and new belief systems which do not necessarily supplant their framework theories but co-exist with them. These developments are gradual and slow and follow a learning progression. In order to be effective science education needs to make students aware of their intuitive understandings, provide scientific information gradually and in agreement with students' learning progressions and develop students' reasoning abilities and executive function skills.

Keywords: science education, misconceptions, intuitive theories, conceptual change, cognitive conflict

In the last 50 or so years, research in science education has provided a great deal of information about how students develop an understanding of science concepts. In the pages that follow I will focus on three aspects of this development: the creation of intuitive understandings, the process of science learning, and the presence of conceptual co-existence. I will then discuss their implications for science education.

INTUITIVE UNDERSTANDINGS

Students are not blank slates when they are first exposed to the learning of science. On the contrary they bring to the science learning task intuitive understandings of the physical world, which can be very different from the scientific concepts and theories presented in the science classroom (Driver and Easley, 1978; Clement, 1982; McCloskey, 1983; Novak, 1987). Researchers agree on the presence of these intuitive understandings, but disagree when they try to describe their nature. There are three main points of view on this matter. The first, known as the classical approach, claims that students conceptions have the status of unitary intuitive theories, often resembling earlier theories in the history of science. The second approach, known as "knowledge-in-pieces, claims that students" conceptions consist of a multiplicity of phenomenological principles or p-prims, which are abstracted from experiential knowledge. According to the third approach, known

as framework theory, students' conceptions consist of a collection of beliefs and presuppositions, which are organized in loose but relatively coherent framework theories.

There is some evidence to support the claim that students' conceptions represent relatively stable and deeply held intuitive theories. For example, McCloskey (1983) showed that there are systematic beliefs about the motion of objects that influence people's interactions with objects in the real world. These systematic beliefs are at variance from Newtonian mechanics and resemble a medieval theory of motion known as impetus theory. According to the impetus theory the motion of an object is maintained by a force internal to the object (impetus) which was acquired when the object was originally set in motion (McCloskey, 1983).

However, not all of students' conceptions can be characterized as unitary and systematic intuitive theories. According to Chi (2013), in addition to false intuitive theories, people also have false beliefs and false mental models. There are also constraints on students' ways of reasoning, such as constraints on the nature of causal explanations, which can give rise to misinterpretations of scientific information. For example, people often rely on a generalized version of a *Direct-Causal schema* to produce misconceived causal explanations for emergent processes, such as diffusion, natural selection, and heat transfer for which a direct-causal schema does not apply (Chi et al., 2012). Emergent processes do not have a single identifiable causal agent or an identifiable sequence of stages. On the contrary, they result from the simultaneous interactions of all agents.

At the opposite end of the intuitive theory position is the claim that students' initial understandings consist of knowledge-in-pieces (diSessa, 1993). diSessa has provided evidence from extensive interviews with students to support the position that students do not hold systematic and unitary intuitive theories but are internally inconsistent and fragmented and that their knowledge fragments can best be characterized in terms of p-prims. The "knowledge-in-pieces" position can account for the inconsistencies often observed in students' explanations, especially when the students are asked to explain the same physical phenomena in different situational contexts. It is problematic, however, when it comes to interpreting students' more complex, theory-like constructions which have been found to be resistant to instruction such as the intuitive theories discussed earlier (Clement, 1982). It also cannot explain constraints on students' causal explanations such as the ones described by Chi (2013), which can give rise to erroneous interpretation of scientific information.

Both the "intuitive theory" and the "knowledge-in-pieces" positions are based on empirical evidence coming from interviews with secondary school or University students and lay adults. In contrast, Vosniadou and her colleagues (Vosniadou and Brewer, 1992, 1994; Vosniadou, 2013; Vosniadou and Skopeliti, 2017) have argued that it is important to make a distinction between students' conceptions formed before exposure to science instruction and after being exposed to science. They have used empirical evidence from interviews with young children before they were exposed to science instruction to argue that children interpret their everyday experiences in

the context of lay culture to form beliefs, which are organized in loose but relatively coherent framework theories (Vosniadou, 2013; Vosniadou and Skopeliti, 2014).

A framework theory is different from an intuitive theory. An intuitive theory is a cohesive, unitary theory, which might contain misconceptions of scientific information. On the contrary, a framework theory is considered to be a skeletal conceptual system that grounds our most fundamental ontological categorizations and causal devices in terms of which we understand the world and on the basis of which new information is built, before any exposure to science (Wellman and Gelman, 1998). A framework theory lacks the systematicity, consistency, and explanatory power of scientific theories and it is not explicit and socially shared. It is however a principle-based system with learning mechanisms, such as categorization and causal attribution, capable of giving rise to explanation of phenomena and prediction (Gopnik et al., 2001; Slousky, 2003). For example, infants make an ontological distinction between objects with or without self-initiated movement (animate vs. inanimate). This distinction can then be used productively to categorize new, previously unseen, objects and attribute to them characteristics of animate or inanimate objects, such as solidity, need for support and the presence or absence of intentionality (Vosniadou and Brewer, 1992, 1994).

The framework theory approach (Vosniadou, 2013) does not exclude the possibility that knowledge elements such as p-prims might be present in our knowledge system. However, they are considered to be organized in loose conceptual structures from early on in childhood. Take for example the well-known Ohm's p-prim—that more effort leads to more effect and more resistance leads to less effect (diSessa, 1993). Although the Ohm's p-prim might serve to schematize a phenomenological experience, it can only be formulated in a conceptual system in which a distinction has already been made between animate and inanimate objects and in which it is already known that effort is usually exerted by the pull or push of animate agents, that forces are implicated, and that the size and weight of the agents and of the objects in question are important (Ioannides and Vosniadou, 2002). In other words, the very generation of an explanatory principle such as a p-prim already presupposes the presence of a skeletal conceptual system, such as a framework theory. Indeed, for researchers who employ a complex systems approach to science learning (e.g., Brown and Hammer, 2008, 2013), also advocated by diSessa (1993), the creation of integrative conceptual structures such as framework theories is not inconsistent with the knowledge-in-pieces approach.

THE PROCESS OF SCIENCE LEARNING

The position one takes regarding the nature of students' intuitive understandings can have important implications about how one interprets the process of science learning. If students' conceptions have the form of intuitive theories then the process of science learning cannot be seen as one of accretion or enrichment of prior knowledge. What is needed is instead theory change, or otherwise known, conceptual change. Posner et al. (1982) argued

that conceptual change requires the replacement of intuitive theories with the correct scientific ones. This replacement was described as the result of a rational process during which students need to become aware of the fundamental assumptions and epistemological commitments that characterize their intuitive theories and to realize their limitations and inadequacies vis. a vis. the scientific theory.

In the years that followed, the so called "classical approach" became subject to a number of criticisms. One issue of contention was the proposal that intuitive understandings are replaced by scientific theories. Arguments regarding the co-existence of intuitive understandings and scientific concepts were put forward early on (e.g., Caravita and Halldén, 1994) but became supported by empirical evidence in recent years and will be discussed in greater detail later.

Contrary to the sudden theory replacement via cognitive conflict view of science learning, the knowledge-in-pieces approach promoted the idea that the process of science learning should be seen as one of conceptual integration, during which the multiplicity of p-prims become organized into coherent scientific theories under the influence of instruction (diSessa, 1993, 2008). Smith et al. (1993) argued that cognitive conflict is not a good instructional strategy because it is inconsistent with a constructivist approach to learning; namely that learning is a process of building new knowledge on what we already know. They proposed instead that intuitive understandings are productive ideas that can serve as resources for science learning, and which evolve and become integrated in cohesive conceptual structures such as scientific theories through appropriate instruction. The emphasis on integration and discrimination rather than on confrontation and cognitive conflict is the hallmark of the knowledge-in-pieces approach to instruction (see also Clark and Linn, 2008).

I will support a different view of science learning, one consistent with the framework theory approach. According to this view students organize their intuitive understandings in loose and narrow but nevertheless relatively cohesive framework theories before they are exposed to science instruction. Framework theories are fundamentally different from scientific theories in their explanations, in their concepts, and in their ontological and epistemological presuppositions. When students who operate with an understanding of the physical world such as the one described as a framework theory of physics are first exposed to an incompatible and counter-intuitive scientific theory, they are not capable of understanding it. Assuming that these students use constructive learning mechanisms they will interpret the new scientific information in light of their prior knowledge. This constructive process will almost necessarily result in the creation of misconceptions which are hybrids—i.e., conceptions that have elements both of intuitive understandings and of scientific information. In a text comprehension study that tested the above proposition directly, Vosniadou and Skopeliti (2017) showed that many elementary school students who gave intuitive explanations of the day/night cycle at pretest either ignored the scientific information altogether, or created misconceptions when exposed to the counter-intuitive scientific explanation. These misconceptions were hybrids that could be

distinguished into fragmented and/or synthetic conceptions. A fragmented conception is one that combines intuitive understandings with scientific information without concern for internal consistency or explanatory power (e.g., day/night happens because the sun goes behind the mountains and also because the earth "moves"). A synthetic conception also combines intuitive understandings with scientific information but does so in ways that show some concern for internal consistency and explanatory power. Vosniadou and Skopeliti (2017) concluded that science learning is not produced through sudden insights but it is a slow and gradual process and that the generation of misconceptions is a natural outcome of this process. In other words, many misconceptions are not accidental errors but fragmented or synthetic conceptions produced when students use constructive learning mechanisms that connect incompatible scientific information with their prior knowledge.

CO-EXISTENCE OF INTUITIVE UNDERSTANDINGS AND SCIENTIFIC CONCEPTS

Recent research has shown that intuitive understandings are not completely replaced by scientific theories, not even in expert scientists. Rather, intuitive understandings co-exist with scientific concepts and may interfere with their access in scientific reasoning tasks. For example, Kelemen et al. (2013) showed that when tested under the pressure of time, with information processing capacity taxed, even expert scientists were likely to endorse non-scientific, teleological explanations of phenomena. In another study, Shtulman and Valcarel (2012) showed that college-educated adults were less accurate and slower to verify scientific concepts that were inconsistent compared to those that were consistent with naïve theories, suggesting that naïve theories continue to exist and interfere in the processing of scientific theories (see also Babai et al., 2010; Potvin et al., 2015).

Masson et al. (2014) used functional magnetic resonance imaging (fMRI) to compare brain activation in experts and novices when evaluating the correctness of simple electric circuits. Their results showed that experts, more than novices, activated brain areas involved in inhibition when evaluating non-scientific circuits, presumably because they were suppressing misconceptions encoded in their brain's neural networks.

The phenomenon of the co-existence of intuitive understandings and scientific concepts and theories raises important problems for theories of science learning and instruction as well as for theories of knowledge organization and representation. If earlier belief systems are not supplanted by information acquired later, how consistent is our knowledge base? How is it possible for the inconsistent old and new belief systems to co-exist, and for the inconsistencies not to be detected?

One way to explain the puzzle of the co-existence of intuitive understandings and scientific concepts is to see them not as incompatible representations organized within the same belief system, but as different belief systems encapsulated in overlapping but partly distinct neural networks within particular

domains of knowledge (Vosniadou, in press). This view is more consistent with the results of cognitive neuroscience research, which show that conceptual knowledge is represented in distributed networks located in different parts of the adult brain (Allan et al., 2014; Fugelsang and Mareschal, 2014). In such a system, coherence is not an attribute of the organization of information in the knowledge base but the outcome of an effective executive function system capable of selecting, integrating, or inhibiting information from different belief systems in ways that are appropriate for the task at hand.

The role of executive function and its relation to academic learning and conceptual change has become an important area of research in recent years. Executive function is a set of neurocognitive skills, such as working memory, cognitive flexibility, and inhibitory control. These skills are fundamental for engaging in goal-directed thought and action and for learning, particularly the learning of counter-intuitive concepts in science and mathematics. Research has shown that executive function skills are significantly related to academic achievement and to conceptual change learning, even when intelligence and prior knowledge are controlled for (Allan et al., 2014; Fugelsang and Mareschal, 2014; Vosniadou et al., 2018). The learning of science and mathematics concepts that are inconsistent with intuitive understandings has been associated specifically with the executive function skill of inhibitory control (see also Zaitchick et al., 2014; Carey et al., 2015).

IMPLICATIONS FOR TEACHER EDUCATION AND PROFESSIONAL DEVELOPMENT

Different theoretical approaches to science learning have proposed different recommendations for science instruction. The classical approach (Posner et al., 1982) considered cognitive conflict as the main instructional strategy for science learning. Cognitive conflict works by presenting the learner with conflicting evidence. This conflicting evidence is designed to produce dissatisfaction with the learners' intuitive theory and the recognition that it needs to be replaced by the scientific theory. One of the problems with instructional uses of cognitive conflict is that it does not guarantee that learners will experience the intended external conflict as internal cognitive dissonance. Chinn and Brewer (1993) have presented persuasive arguments that indicate that learners can respond to conflicting evidence in different ways. Indeed, many students and teachers hold inconsistent beliefs without being seemingly aware of the inconsistencies.

Contrary to what is known as the classical approach, the "knowledge-in-pieces" (diSessa, 1993) approach emphasizes the integration of students' p-prims into coherent scientific theories. This approach is based on the assumption that p-prims are productive and that what is needed is to find a way to integrate them into internally-consistent scientific theories. It does not, however, tell us what to do with intuitive understandings that might not be productive when it comes to learning a scientific theory.

From the perspective of the framework theory there are three main points that need to be emphasized regarding instruction. First, science learning is a constructive process that gradually builds on and modifies prior knowledge. Depending on the learners' prior knowledge, learning the correct scientific explanation is not something that happens immediately and suddenly; rather, it may take some time to be accomplished—there is a learning progression involved (Vosniadou and Brewer, 1992, 1994; Wiser and Smith, 2008; Vosniadou and Skopeliti, 2017, 2018). Indeed the whole idea of building learning progressions is to capture the intermediate steps in the learning of science concepts and theories (Corcoran et al., 2009; Duschl et al., 2011). When science educators are aware of the students' learning progression in a given subject matter area, they can provide scientific information that is less likely to be misunderstood.

Second, cognitive conflict can be used in the process of learning science but mainly in order to increase students' metacognitive awareness and understanding of the gap between their existing beliefs and the new scientific information rather than to prove that intuitive understandings are wrong and need to be replaced. Intuitive understandings are resistant to instruction because they are immediate and common-sense interpretations of everyday experience and because they are constantly reinforced by this experience. On the contrary, scientific concepts are usually not supported by everyday experience and require the construction of new, abstract, and complex representations that do not have a one-to-one correspondence to the things they represent. Students need to be facilitated to create these new, counter-intuitive representations, understand that they are based on different, non-egocentric perspectives and that they have much greater explanatory power.

Last but not least science instruction needs to develop students' reasoning abilities, their epistemological beliefs and their executive function skills. Science learning requires complex spatial reasoning, the ability to take different perspectives, construct complex and abstract models and representations and inhibit prior knowledge so that new, conflicting information can be entertained. The cultivation of these skills and ways of reasoning should be an integral part of science instruction.

CONCLUSIONS

It has been argued that children start the knowledge acquisition process by forming beliefs based on their everyday experiences and lay culture. These beliefs are not isolated but organized in loose and narrow but relatively coherent framework theories. Although framework theories are implicit, not socially shared and lack the systematicity and explanatory power of scientific theories, they are principle-based systems with learning mechanisms such as categorization and causal attribution that can give rise to explanation and prediction. Scientific concepts and theories are very different in their concepts, organization, ontological and epistemological presuppositions and in their representations from framework theories. They require major conceptual changes to take place in order to be

fully understood. These conceptual changes take time to be accomplished. The development of science knowledge is a long and gradual process during which students use constructive learning mechanisms to assimilate new, scientific, information into their prior knowledge causing hybrid conceptions—or misconceptions. Science instruction needs to help students become aware of their experience-based beliefs that might constrain science learning causing misconceptions, provide

information gradually based on students' learning progressions and develop students' scientific reasoning and executive function skills.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., and Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: a meta-analysis. *Dev. Psychol.* 50, 2368–2379. doi: 10.1037/a0037493

Babai, R., Sekal, R., and Stavy, R. (2010). Persistance of the intuitive conception of living things in adolescence. *J. Sci. Educ. Technol.* 19, 20–26. doi: 10.1007/s10956-009-9174-2

Brown, D., and Hammer, D. (2013). "Conceptual change in physics," in *International Handbook of Research on Conceptual Change,* 2nd Edn, ed S. Vosniadou (New York, NY: Routledge), 121–137.

Brown, D. E., and Hammer, D. (2008). "Conceptual change in physics," in *International Handbook of Research on Conceptual Change*, ed S. Vosniadou (New York, NY: Routledge), 121–137.

Caravita, S., and Halldén, O. (1994). Re-framing the problem of conceptual change. *Learn. Instr.* 4, 89–111. doi: 10.1016/0959-4752(94)90020-5

Carey, S., Zaitchik, D., and Bascandziev, I. (2015). Theories of development: in dialog with Jean Piaget. *Dev. Rev.* 38, 36–54. doi: 10.1016/j.dr.2015.07.003

Chi, M. (2013). "Two kinds and four sub types of misconceived knowledge, ways to change it and the learning outcomes," in *International Handbook of Research on Conceptual Change*, 2nd Edn, ed S. Vosniadou (New York, NY: Routledge), 49–71.

Chi, M. T. H., Roscoe, R., Slotta, J., Roy, M., and Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cogn. Sci.* 36, 1–61. doi: 10.1111/j.1551-6709.2011.01207.x

Chinn, C. A., and Brewer, W. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Rev. Educ. Res.* 63, 1–49. doi: 10.3102/00346543063001001

Clark, D. B., and Linn, M. C. (2008). "The knowledge integration perspective: connections across research and education," in *International Handbook of Research on Conceptual Change*, ed S. Vosniadou (New York, NY: Routledge), 520–559.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *Am. J. Phys.* 50, 66–70.

Corcoran, T., Mosher, F. A., and Rogat, A. (2009). *Learning Progressions in Science. An Evidence-Based Approach to Reform. Consortium for Policy Research in Education Report #RR-63.* Philadelphia, PA: Consortium for Policy Research.

diSessa, A. (1993). Towards an epistemology of physics. *Cogn. Instr.* 10, 105–225. doi: 10.1080/07370008.1985.9649008

diSessa, A. A. (2008). "A bird's-eye view of the 'pieces' vs 'coherence' controversy (from the pieces side of the fence)," in *International Handbook of Research on Conceptual Change*, ed S. Vosniadou (New York, NY: Routledge), 35–60.

Driver, R., and Easley, J. (1978). Pupils and paradigms: a review of literature related to concept development in adolescent science students. *Stud. Sci. Educ.* 5, 61–84. doi: 10.1080/03057267808559857

Duschl, R., Maeng, S., and Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Stud. Sci. Educ.* 47, 123–182. doi: 10.1080/03057267.2011.604476

Fugelsang, J., and Mareschal, D. (2014). "The development and application of scientific reasoning," in *Educational Neuroscience,* eds D. Mareschal, B. Butterworth, and A. Tolmie (Chichester: Wiley-Blackwell), 237–267.

Gopnik, A., Sobel, D. M., Schulz, L., and Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Dev. Psychol.* 37, 620–629. doi: 10.1037//0012-1649.375.620

Ioannides, C., and Vosniadou, S. (2002). The changing meanings of force. *Cogn. Sci. Quart.* 2, 5–62.

Kelemen, D., Rottman, J., and Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: purpose-based reasoning as a cognitive default. *J. Exp. Psychol. Gen.* 142, 1074–1083. doi: 10.1037/a0030399

Masson, S., Potvin, P., Riopel, M., and Brault-Foisy, L.-M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind Brain Educ.* 8, 44–55. doi: 10.1111/mbe.12043

McCloskey, M. (1983). Intuitive physics. *Sci. Am.* 248, 122–130. doi: 10.1038/scientificamerican0483-122

Novak, J. D. (1987). "Introduction," in *Proceedings of the Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics* (Ithaca, NY; New York, NY: Cornell University).

Posner, G. J., Strike, K. A., Hewson, P. W., and Gertzog, W. A. (1982). Accommodation of a scientific conception: towards a theory of conceptual change. *Sci. Educ.* 66, 211–227. doi: 10.1002/sce.3730660207

Potvin, P., Masson, S., Lafortune, S., and Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: a reaction time study with different levels of interference. *Int. J. Sci. Mathem. Educ.* 13, 21–34. doi: 10.1007/s10763-014-9520-6

Shtulman, A., and Valcarel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition* 124, 209–215. doi: 10.1016/j.cognition.2012.04.005

Slousky, V. M. (2003). The role of similarity in the development of categorization. *Trends Cogn. Sci.* 7, 246–251. doi: 10.1016/S1364-6613(03)00109-8

Smith, J. P., diSessa, A. A., and Rochelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *J. Learn. Sci.* 3, 115–163. doi: 10.1207/s15327809jls0302_1

Vosniadou, S. (2013). "Conceptual change in learning and instruction: the framework theory approach," in *The International Handbook of Conceptual Change*, 2nd Edn, ed S. Vosniadou (New York, NY: Routledge), 11–30.

Vosniadou, S. (in press). "Teachers' beliefs and knowledge," in *Problem Solving in Learning and Teaching: A Festschrift for Emeritus Professor Mike Lawson*, eds J. Orrell and H. Askell-Williams.

Vosniadou, S., and Brewer, W. F. (1992). Mental models of the earth. *Cogn. Psychol.* 24, 535–585. doi: 10.1016/0010-0285(92)90018-W

Vosniadou, S., and Brewer, W. F. (1994). Mental models of the day/night cycle. *Cogn. Sci.* 18, 123–183. doi: 10.1207/s15516709cog1801_4

Vosniadou, S., Pnevmatikos, D., and Makris, N. (2018). The role of executive function in the construction and employment of science and mathematics concepts that require conceptual change learning. *Neuroeducation* 5, 58–68. doi: 10.240.46/neuroed.20180502.58

Vosniadou, S., and Skopeliti, A. (2017). Is it the Earth that turns or the Sun that goes behind the mountains? Students' misconceptions about the day/night cycle after reading a science text. *Int. J. Sci. Educ.* 39, 2027–2051. doi: 10.1080/09500693.2017.1361557

Vosniadou, S., and Skopeliti, E. (2018). Evaluating the effects of analogy enriched text on the learning of science. The importance of learning indexes. *J. Res. Sci. Teach.* doi: 10.1002/tea.21523. [Epub ahead of print].

Vosniadou, S., and Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Sci. Educ.* 23, 1427–1445. doi: 10.1007/s11191-013-9640-3

Wellman, H. M., and Gelman, S. A. (1998). "Knowledge acquisition in foundational domains," in *Cognition, Perception and Language. Volume 2 of the Handbook of Child Psychology,* 5th Edn, eds D. Kuhn and R. Siegler (New York, NY: Wiley), 523–573.

Wiser, M., and Smith, C. L. (2008). "Learning and teaching about matter in grades K-8: when should the atomic-molecular theory be introduced?," in *International Handbook of Research on Conceptual Change*, ed S.Vosniadou (New York, NY: Routledge), 205–239.

Zaitchick, D., Iqbal, Y., and Carey, S. (2014). The effect of executive function on biological reasoning in young children: an individual differences study. *Child Dev.* 85, 160–175. doi: 10.1111/cdev.12145

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Check for updates

Engaging Students in Science: The Potential Role of "Narrative Thinking" and "Romantic Understanding"

Yannis Hadzigeorgiou [1,2]* and Roland M. Schulz [1]*

[1] Imaginative Education Research Group (IERG), Simon Fraser University, Vancouver, BC, Canada, [2] University of the Aegean, Rhodes, Greece

Engaging students in science and helping them develop an understanding of its ideas has been a consistent challenge for both science teachers and science educators alike. Such a challenge is even greater in the context of the "Science for All" curriculum initiative. However, Bruner's notion of "narrative thinking" and Egan's "romantic understanding" offer an alternative approach to teaching and learning science, in a way that engagement and understanding become a possibility. This chapter focuses on students' "narrative mode of thought," as a bridge to understanding science—which has traditionally been based more upon the use of logico-mathematical thinking in the upper grades—and on a distinctive form of understanding the world, characteristic of students of the age range from 8 to 15 years. This latter form of understanding, that the educational theorist Kieran Egan calls "romantic understanding," has features that can be readily associated with the natural world and its phenomena. Therefore its development could be fostered in the context of school science learning, and in a way that the narrative mode would also be taken into consideration.

Keywords: science, engagement, narrative thinking, romantic understanding, story, language

INTRODUCTION

Science as a school subject to be taught and learned, has always presented a challenge to both teachers and students. On the one hand, understanding science (as content, inquiry and process skills) is a challenging task for students, as it involves a construction process, which is complex and iterative—not a linear one—and which also takes time and effort. An important implication of this construction process, as constructivist-oriented research in the 1980s and 1990s showed, is the possibility for students to *construct* not only a conceptual framework that lacks the coherence of true scientific knowledge, but to equally construct alternative ideas that are different from the canonical scientific ones. Other implications that were discovered is that the construction process is influenced by several interrelated factors, such as students' prior conceptions and views on the nature of science (their epistemologies—Kalman, 2008/2017; Matthews, 2015), their interest and motivation, the classroom culture, the opportunities they have for social interaction, dialogue, and argumentation, the generation of representations (for the use of modeling and analogies), and also their opportunities for cognitive dissonance and conceptual change, as well as for applying new knowledge to new contexts (Resnick, 1983; Hadzigeorgiou, 1997, 1999, 2015; Stefanich and Hadzigeorgiou, 2001; Tytler et al., 2013).

On the other hand, teaching science is a challenging task for teachers, because, in addition to providing students with opportunities for constructing scientific understanding, they have to

primarily engage and motivate students with science, its content and techniques (e.g., concepts, equations, laws, and laboratory skills). For it is obvious that without some degree of engagement, understanding cannot truly take place. Even though some degree of understanding may very well motivate students to learn, the initial engagement with science seems to be a prerequisite for understanding and long-term learning. And needless to say, motivation on the part of students to learn does not guarantee an understanding of science, especially science content (Hadzigeorgiou, 2005a, 2015).

Thus, at least as far as school science education is concerned, one can very well talk about a two-fold challenge: how can students be engaged with science content—but in a way that true understanding of science could also become a possibility? This paper will discuss the possibility of engagement with science content learning by focusing on the potential of two ideas, namely, "narrative thinking" and "romantic understanding." But first a look at the problem of engagement itself, which is central to the teaching/learning process, and, as such, central to the process of understanding science.

THE PROBLEM OF STUDENTS' ENGAGEMENT WITH SCIENCE CONTENT KNOWLEDGE

The problem of how to engage students in science, as mentioned, has always been challenging and pressing. Even though engagement does not necessarily entail, or result in, understanding, especially when it comes to the case of learning science, engaging students in science is a prerequisite for understanding. However, what may not be obvious is that the process of engagement itself is a complex one. Even though engagement may very well be encouraged by students' *interest,* there are other key factors which are also involved, such as personal identity, maturity, purpose for learning science, and students' awareness of the significance of the object or topic of study. Such factors can influence to a large extent, or may even determine, students' engagement with science (Hadzigeorgiou, 2005a; Hadzigeorgiou and Stivaktakis, 2008). Furthermore, the variety of ways in which the term "engagement" has been interpreted in the literature poses an additional problem in regards to what the findings of the various studies on student engagement really mean. As Godec et al. (2018) point out, engagement has been construed as enjoyment and interest, but also as motivation toward science, as well as future orientations toward science. Moreover, it has been taken to mean the degree (frequency) of students' participation in science related activities, as well as intensity of such participation.

Although a conceptual clarification of the notion of engagement is beyond the scope of this paper, it is important nevertheless to accentuate here that "engagement" should not be conflated with the *motivation* to learn. Even though the two terms could be used interchangeably—and in fact they often are—there is a subtle and nonetheless important difference between them. For there has always been a question about whether students' motivation for learning resides mainly in students' object of study

per se, that is, the content and/or the processes of science, or if in fact other factors are involved (Hadzigeorgiou and Stivaktakis, 2008; Hadzigeorgiou, 2012; Hadzigeorgiou and Schulz, 2014, 2017). Indeed, it is quite evident that students can be motivated to participate in learning activities but for reasons that may vary and where their motives primarily reside in things other than the immediate topic of study (e.g., involving such factors as teaching style and teacher personality, humor, peer social interactions in group activities, flashy demonstrations, etc.). Furthermore, the notion of engagement should not be conflated specifically with the notion of student "interest" either. Apart from the conceptual problems inherent in the notion of interest itself, there is empirical evidence that what students think is interesting (e.g., a topic, an issue, an idea) does not necessarily motivate them to study it, let alone to study it further—that is, to try to learn more about it and move beyond the class situation (Hadzigeorgiou and Schulz, 2017).

Thus, there is a *distinction* to be made between peripheral things involved in pedagogy (albeit linked to content knowledge), that are supposedly interesting and motivating, and the actual or intimate engagement with the students' personal scientific object of study, namely content and processes. And such a distinction is a crucially important one: the reason being that this engagement *with* the actual science content has the potential to discourage what the American philosopher of education Dewey (1934, 1966) had previously called the "spectator theory of knowledge." What he meant was the dualistic learning framework that created an emotional-cognitive gap between the *subject* (the student) and the *object* (content) that could very well be fostered by, and inherent to, common instructional sequences and curricula. And this despite the reform initiated "constructivist" and "guided inquiry" intentions of both teachers and curriculum designers (see Dahlin, 2001; Hadzigeorgiou, 2005c, 2016). For example, a dualistic learning framework can be unknowingly encouraged by science teachers when (as one of their main instructional strategies) they try to figure out how to "sugar-coat" difficult science ideas and topics (like the mole concept in chemistry, or dynamics equations in physics) using flashy demonstrations, hence by focusing on peripheral things and not, as Pugh (2004) pointed out, on the science content itself.

The crucial importance of true engagement with science content can be seen in its potential to encourage the application of classroom learning in "free-choice" contexts, also the expansion of perception (that is, the ability to see objects, events, and issues *through* the lens of the science content), as well as an appreciation of the value of this content for its role in enriching everyday experience (see Pugh, 2011; Pugh et al., 2017). Certainly, such a learning experience with such characteristics may be considered ideal, and, to a certain extent it is. However, it deserves to be recognized that it is indeed a pedagogical possibility (see Hadzigeorgiou, 2016; section The Problem of Students' Engagement with Science Content Knowledge). In this paper though, as was previously said, the focus will be on two ideas, namely, the narrative mode of thinking and the romantic mode of understanding, with the focus on their potential to encourage engagement *with* science content. Moreover, if it is indeed true that personal engagement with a school subject, like

science, has the potential to take the science knowledge (i.e., what students learn at school) beyond the walls of the classroom, and equally if it has the ability to transform one's outlook on the world—which some philosophers of education, physicists and cognitive scientists identify with significant learning (Feynman, 1968; Hirst, 1972; Hadzigeorgiou, 2016)—then the problem of how to engage and motivate students with actual science content should become a central concern for school science education.

NARRATIVE THINKING AS A BRIDGE TO UNDERSTANDING SCIENCE

No doubt science can be an exciting subject, yet a difficult one to teach, simply because science, as both a body of knowledge and a way of reasoning or thinking, is different from everyday knowledge and thinking (leaving aside the linkage to science as a mode of experimental inquiry). Even though the viewpoint that scientific thinking is a refinement of everyday thinking contains an element of truth, this refinement process nonetheless, in the case of students, as research evidence suggests, takes time and requires specific strategies (Stefanich and Hadzigeorgiou, 2001; Hadzigeorgiou and Fotinos, 2007; Schulz, 2009). Central to this process of understanding has been the use of what can be called logico-mathematical reasoning, that is, "logico-scientific" thinking (Bruner, 1986, p. 12), which is responsible for the formation of hypotheses, the development of arguments, creative modeling, the solutions of problems, the descriptions and construction of systems and their inter-relationships (Piaget, 1970; Bruner, 1986; Giere, 1991). Secondary and tertiary science education is known to make use of inductive-empirical and hypothetico-deductive variations of scientific reasoning (Cawthron and Rowell, 1978; Duschl, 1994), though it tends to become overly simplified and known to degenerate into talk of a "step-wise scientific method" supposedly used by all scientists, which is a myth (Bauer, 1992).

However, not all thinking is like this, when humans seek to understand and interpret the world around them. It was this observation which led, in the mid-1980s, the psychologist Jerome Bruner to propose another kind of thinking that is not predominately logical, mathematical, abstract, and seeks to construct and model ideal systems. Bruner's observations (1985; 1986), as a forerunner of the "cognitive revolution," were based on common experience and empirical evidence.

> There are two irreducible modes of cognitive functioning—or more simply, two mode of thought—each meriting the status of a "natural kind". Each provides a way of ordering experience, of constructing reality and the two (though amenable to complementary use) are irreducible to one another (p. 97).

According to Bruner (1985), the status "natural kind" refers to the fact that each mode of thinking comes spontaneously into being, and always under minimal contextual constraint. These two modes Bruner called *paradigmatic* (or logico-mathematical) and *narrative*. The former is concerned with the formation of hypotheses, the development of arguments, solutions to problems, finding proofs, and with rational thinking in general.

According to Bruner (1986, p. 12), it fulfills "the ideal of a formal, mathematical system of description and explanation" by employing "categorization or conceptualization and the operations by which categories are established, instantiated, idealized, and related to one another to form a system." The latter, on the other hand, is concerned with what Bruner calls "verisimilitude," that is, life-likeness, and the creation of meaning. It seeks explications that are context sensitive and particular (not context-free and universal). It is entirely divergent—in sharp contrast to paradigmatic mode, which is convergent—and employs literary devices, such as stories, metaphors, similes, even hyperboles, in order to create meaning. In looking at those two modes of thinking, it is quite evident that while the paradigmatic mode is about "logico-mathematical thinking" *per se*, the narrative mode is about people (i.e., human emotions, ambitions, intentions, successes and failures, human actions, and experiences). In other words, while the paradigmatic mode presupposes distancing oneself from emotions and the human element in general, the narrative mode presupposes personal involvement with the object of thought. However, according to Bruner (1985, 1986), the two modes are complementary.

But what does the narrative mode of thinking have to do with science, that is, a field of study characterized by logical analysis, and which (field) has been developed as a result of logical arguments and scientific explanations (e.g., in the form of hypotheses, mathematical models and theories)? To answer this question one has to consider the fact that many scientific (and mathematical for that matter) hypotheses did indeed start their lives as stories and metaphors (Hadzigeorgiou, 2016). This view is in line with the one held by the philosopher of science Popper (1972), who argued that today's science is built upon the science of yesterday and that the older scientific theories were built upon "prescientific myths" (p. 346). Thus, the narrative mode of thinking can be considered equally important to science.

One can, of course, very well argue that the narrative mode of thinking (as the source of the creation of a myth or a story) can result in the construction of unreal or even impossible worlds. However, as Bruner (1985) points out, "the narrative mode is not as unconstrainedly imaginative as it might seem to the romantic" (p. 100). In science, therefore, the constructions that result from the use of the narrative mode cannot just refer to any kind of world (or reality), or even to all kinds of impossible worlds. The reason is that the paradigmatic (or logico-mathematical) mode of thinking, as a mode of thinking that is inextricably tied to the real world of things, does test concepts and ideas (i.e., the constructions of the narrative mode) through the use of evidence, experimentation, argumentation, and so on. Nonetheless, Bruner's hypothesis about the existence of the two modes of thinking, although a bold one, does shed light on the development of scientific language and knowledge, which cannot be explained solely in terms of paradigmatic (logico-mathematical) thinking. Sutton (1996) in fact has illustrated how the language of a scientific concept changes from its initial formulation to how it later becomes rephrased, codified, and depersonalized through the different stages of publication—from original discovery to research paper, handbook and finally textbook—and uncovers the often neglected aspect of

the development of scientific concepts themselves. In other words, the *historicity* of scientific language and theories. The original creative, speculative and often very personal narrative occurs when discoveries are made—"where wonder and curiosity abound"—and where the language can be figurative and even metaphorical (e.g., discovery of electron, DNA and quarks) during the early stages of research or "frontier science." However, by the time the much later stage of "textbook science" has been reached, the concepts and discoveries have been codified, often abstracted out of the historical matrix, while the language has shifted from narrative or lived-story to depersonalized transmission and exposition. Too many textbooks create the false impression that science does not start as an exciting, arduous exploratory process but rather arrives as a "finished product" whose ideas, facts and equations are to be memorized and manipulated (Stinner, 1995; Kalman, 2008/2017; Schulz, 2014b).

The interplay between the two modes of thinking, that has been central to the historical and philosophical development of science, has been also empirically documented in the context of school science education (Kurth et al., 2002). It is indeed this interplay between the two modes of thinking, that is, the narrative and the paradigmatic, which helps children to make sense of the natural world. But whether the two modes of thinking are really as mutually exclusive as Bruner (1985, 1986, 1990) hypothesized, is debatable, and not our concern here. And yet the very nature of "final form science" as codified in language of increasing technicality in textbooks at the upper grades, reinforces the problem of engagement, as they further distance the student from the object of study.

It should be noted at this point that the importance, in fact the centrality, of the narrative mode of thinking is captured in the notion of "mind as a narrative concern" (Sutton-Smith, 1988). Such a notion can help explain not only the "irrational character" of some kinds of scientific thinking (Kuhn, 1970; Feyerabend, 1993; Di Trocchio, 1997), but also the creation of scientific ideas that necessitated mental leaps, even "jumps of the imagination," also famous thought experiments, which could not have become possible only through strictly logical causal-type thinking (Hadzigeorgiou, 2016). In addition, such a notion can help explain certain facts, which are important to consider when approaching the general problem of student engagement with science. One such fact, has been pointed out by White (1981, p. 1): despite the fact that people are not capable of understanding "the specific thought patterns of another culture" they have "less difficulty understanding a story coming from another culture, however exotic that culture may appear." Another fact—and this is crucially important when it comes to the problem of engaging students in science—is that the narrative mode of thinking is used by people in everyday life. Indeed people of all ages use their narrative mode not only to make sense of their experiences, but also to communicate and to plan their future actions (Bruner, 1990). And this is why Bruner (1991), called the narrative mode, the "default mode" of thinking. If this is true, then the argument that people are, or become more, competent at thinking in the narrative mode, in comparison with thinking with the logical mode (Hadzigeorgiou, 2016), can provide food for thought when it comes to planning for curriculum and instructional sequences,

which consider the students' own inclination toward a narrative mode of thinking.

In his *The Storytelling Animal. How Stories Make Us Human*, Jonathan Gottschall provides a compelling argument that we are storytelling animals because of evolutionary reasons (Gottschall, 2012). His argument is based on research in psychology, neuroscience, and evolutionary biology. Even if one remains skeptical about what specific scientific studies Gottschall has drawn upon in order to advance his argument (e.g., does reading fiction cause people to modify or change their attitudes and behavior?), there is still plenty of evidence from a variety of experiments that seems to support Bruner's (1985; 1986) hypothesis about the narrative mode of thinking. In addition, Gottschall's work also supports Egan's (1997, 2005) work on the development of the "educated mind". It is of note that Egan transcended some dilemmas regarding the development of the mind by focusing neither on knowledge *per se* nor on child psychology, but instead on the notion of "cognitive tool," that is, a tool that facilitates thinking and understanding. Cognitive tools are picked up by children as they grow up and become socially enculturated through a language community. One such tool is "story," and the educational process could be conceived, according to Egan (1997), as a process that provides students with an array of cognitive tools, which (tools) are also associated with particular *kinds of understanding*—more broader and general socio-cultural tools (see also next section in this chapter).

The implications of narrative thinking is that narratives and especially stories become indispensable teaching/learning tools. Indeed narratives and stories can be used for communicating important ideas *of* and *about* science. This mode of introducing students to science is engaging for a number of reasons. First, "narratives and stories are more appropriate in describing what we learn about the world" (Hadzigeorgiou, 2016, p. 90), according to research based on the constructive nature of human sense- and meaning-making (see also Egan, 1986, 1988, 1999). Second, narratives, particularly those produced by the students themselves, can foster science learning, by bridging the gap between students' everyday knowledge (and quite frequently naïve conceptions) and scientific conceptions (Zabel and Gropengiesser, 2015). Only through dialogue and the opportunity to partake of using science language in specific class settings can the so-called "three language problem" (i.e., specialist science language, everyday language, science education language), be gradually overcome, according to recent socio-linguistic-based research (Wellington and Osborne, 2001; Yore and Treagust, 2006; Schulz, 2014b). Third, narratives and stories can be considered the means of translating "knowing into telling" (Avraamidou and Osborne, 2009, p. 1,012), an idea that is crucially important in science education, where abstract scientific knowledge must be presented in a meaningful way to the student.

Fourth, stories provide the context for a "silent" dialogue between the teller and the listener, which, by its very nature, is engaging. According to Solomon (2002), a story can be considered as a dialogue. Indeed, despite the fact that the student/listener does not actively participate in the telling of the story, she/he tries to create meaning by listening attentively to the story. Finally fifth, narratives and stories have the potential

to break barriers and dichotomies between epistemic subject and epistemic object, something that has been stressed from both a post-modernist perspective on teaching and learning, and a hermeneutic approach (Kalman, 2011; Schulz, 2014b; Although a strong caution should be brought to bear when some post-modernist perspectives are employed in science education; Nola and Irzik, 2005; Schulz, 2007). Indeed, from such a post-modern perspective, understanding the world involves a rejection of traditional stark dichotomies, like those between fact and fiction, reality and epistemic subject. Gough (1993) has convincingly argued for a pedagogy, which "tacitly embraces [...] the relatedness of the observer and the observed and the personal participation of the knower in all acts of understanding" (p. 607). Likewise from a hermeneutic perspective, meaning-making through language and interpretation is seen to be prerequisite for any understanding to take place at all, which involves the learners' very *being* involved in an interpretive act (a form of *intersubjectivity*) between knower and object, in contrast to knowledge "possession" by isolated individual cognition, according to the standard (epistemological) spectator theory of knowing (Eger, 1992). Borda (2007) has even suggested how some Hermeneutic *dispositions* (doubt, humility, strength) could be fostered in science learners to increase their engagement, to help overcome the textbook content-based and classroom-based language barriers, and approach science as a hermeneutic endeavor[1] [see also Kalman (2008/2017, 2011), on the advantage of "reflective writing" when using the "hermeneutic circle method" in tertiary physics and engineering classes].

It should be noted that narratives and stories can be very engaging (compared with other teaching methods), not only because students become emotionally involved with content knowledge on a deeper level, but also because they have the benefits of experiential learning due to high levels of the listeners' active engagement. Moreover, narratives and stories can appeal to a wide range of intelligences as well as a variety of learning styles (see Hadzigeorgiou, 2016). It should also be noted that storytelling, in particular, satisfies all three elements of effective learning, based on brain-based research (Caine et al., 2005, p. 233): (a) *Relaxed Alertness* (i.e., a state of mind created in a low-threat atmosphere, which also creates a sense of community), (b) *Planned Immersion* (i.e., the creation of an environment in which students become involved with the objectives of the lesson) and (c) *Active Processing* (i.e., utilization of learning methods, which encourage reflection and integration of the information in a meaningful way)[2].

---

[1]The physicist and philosopher Martin Eger in a series of papers 1992; 1993a; 1993b has skillfully shown how the field of philosophical hermeneutics (the study and interpretation of texts), can be applied to science education when learners seek to find personal meaning and understanding when reading and interpreting textbooks, and participating in classroom dialogue (see also Schulz, 2014b).

[2]Even though "active processing" may be considered something that cannot be encouraged through storytelling, one should bear in mind that storytelling does encourage "active processing," in the sense that the listener is not a passive recipient of information, but one who tries to create meaning by relating new information to prior knowledge. In addition, the listener, in his/her attempt to understand also employ higher order thinking skills, like analysis and synthesis. Who indeed, can doubt the fact that those who listen attentively to a story and try to create meaning do not put the past, the present, and the future in a relationship? This is

In light of the above, narrative thinking becomes indispensable if engagement with science content is itself a main goal of pedagogy. This, in turn, means that narratives and stories can play the role of bridges to the world of science, between the learner, and the science content. Narratives and stories can introduce students to science content ideas and to ideas about the *history and nature of science* (NoS), if these ideas are embedded in the narratives and the plot of the stories, and especially if the actual historical background is respected (Allchin's warnings signs about using pseudohistory and pseudoscience is to be heeded, 2013). The empirical evidence thus far, although limited, is quite encouraging (see Hadzigeorgiou, 2016, for a review of studies on the use of narratives and storytelling in science education). Certainly there are some limitations to be considered, according to Hadzigeorgiou (2017)—e.g., narrative explanations are more suitable for the historical sciences, like geology and cosmology, and for unique events, like the disappearance of dinosaurs, whereas it is difficult to create narratives for all phenomena and for all science concepts because of the need to use deductive and descriptive explanations. While these can be presented in a narrative form, possibly also through the use of anthropomorphism, but these are more suitable for younger children. But it is their potential to engage students emotionally and cognitively that we should keep them in mind, and the instructional sequences that we design should take this potential into consideration, too. In particular, special attention must be paid so that the narratives and stories we create (fictitious or based on the history of science), should have specific features (i.e., narrative elements), according to the literature on narratives (see Klassen and Froese-Klassen, 2014a). Such caution is more readily understood in the case in which one seeks to create a narrative or a story with "romantic features," with the aim of fostering in students a *romantic understanding* of science (Hadzigeorgiou et al., 2012). This we discuss in the next section.

"ROMANTIC UNDERSTANDING" AS A WAY TO BE ENGAGED WITH THE CONTENT OF SCIENCE

"Romantic Understanding" is a term coined by the educational theorist Kieran Egan, who used it to describe a kind or form of understanding that children develop approximately between the ages of 8 and 15 years. It is one of five *forms* of understanding that students can develop throughout their participation in the educational process of schooling. According to Egan's socio-linguistic theory of "imaginative education" (*The Educated Mind*, 1997), educational development can be conceived as a process or *recapitulation*, during which students' minds are socio-culturally shaped to recapitulate, that is, repeat, the forms of understandings, as these have appeared in our extended cultural history. These forms, also termed socio-cultural cognitive tools of mind, Egan called "Somatic," "Mythic,"

---

the power of the story, that many teachers and educators have not really grasped. It is not just about interesting stories that can be used in order to convey important information. It is also about creating meaning through various relationships and associations that the listener constructs (Hadzigeorgiou, 2016).

"Romantic," "Philosophic," and "Ironic," and postulated most cultures moved through these stages as civilizations progressed, although at a diverse pace and performance. Egan's grand theory (or "metatheory") is grounded on the fact of the *historicity* of language in human anthropology and cultural development and how this has managed to shape—albeit in ways not yet entirely understood—both the brain and the mind. "Without the historicity of language, human nature and the human mind remain essentially unchanged in history" (Polito, 2005, p. 486). (See Schulz, 2009, 2014b, for a more indepth discussion).

One can certainly maintain, with little controversy when examining the anthropological record, that there has come to be a general cultural progression of the human race from plain mimicry and artifact construction (common to our primal *homo sapiens* ancestors—"somatic"), to oral language use and society ("mythic"), to creating literacy with the written word ("romantic"), and finally to more complex forms of language symbolism and use, including a shift to theoretical ("philosophic") and even ironic thinking, as noted by others (Donald, 1991). "The exceedingly long historico-cultural development since our early hominid prehistory, which appears to be neither inevitable nor 'progressive' (in the older 19th century evolutionary sense), has nonetheless brought with it the discovery and invention of both *physical* and especially *cognitive* tools, which, according to their own sequence and time, have wrought technological advance as well as expanded the human capacity to reason and make sense of themselves and the world" (Schulz, 2009, p. 262).

Here, however, the focus is strictly on "Romantic Understanding" which is itself a *transitional* kind of understanding, between "Mythic" (i.e., a kind of understanding associated with orality and developed by children in the age range 2–7, who rely on oral language to interact and understand the world), and "Philosophic" (i.e., conceptual or "theoretic" understanding, for those learners in the age range of about 15–20 years). It is important though to point out that Egan's notion of "Romantic Understanding," as a transitional kind of understanding is quite unique (Egan, 1990). The reason is that neither Donald's (1991) distinction between mythic and rational thinking, nor Bruner's (1986) distinction between narrative and paradigmatic (or logico-mathematical) thinking, can explain or account for a transitional stage of understanding (i.e., from mythic *to* narrative understanding *to* more advanced conceptual understanding at the upper grade levels). In other words, Egan's "Romantic Understanding" is a quite distinctive mode of understanding, which is not to be confused or conflated with narrative understanding in general.

Although one could argue that both mythic and romantic understandings are narrative in their nature (i.e., both very young children and teenagers rely on the narrative mode of thinking to make sense of the world and their experiences), these two kinds of understanding represent two distinct ways of making sense of the world. This becomes easily understood if one looks at the specific *characteristics* for these kinds of understanding. For "Romantic Understanding" in particular these characteristics are the following: (a) the humanization of meaning (i.e., students' awareness of the human context of the

knowledge and content to be learned); (b) an association with heroes and heroic qualities (i.e., students' association with things or people with heroic qualities, so they gain confidence that they, too, can face and deal with the real world); (c) an attraction to the limits of reality and extremes of experience (i.e., the limits of any new environment and human experience enables students to gain security and confidence in dealing with reality); (d) the experience of a sense of wonder (i.e., astonishment mingled with bewildered curiosity, admiration, and the awareness that one's knowledge is incomplete or erroneous or that some extraordinary phenomenon-exists), and finally, (e) revolt and idealism (i.e., contesting of conventional ideas and all kinds of conventions).

From this general conception of a "Romantic Understanding," an *operational definition* of romantic understanding in the context of school science education can be construed as follows: "A narrative kind of understanding which enables students to become aware of the human context of the subject content that they are supposed to learn, by associating, at the same time, such content with heroic human qualities, with the extremes of reality and experience, with a contesting of conventional ideas, and also by experiencing a sense of wonder" (Hadzigeorgiou et al., 2012, p. 1,112). This definition of "Romantic Understanding," while different from that of conceptual or "theoretic" understanding, is very relevant to school science education in the sense that it relates to the content of many different science subjects. Indeed, the content of science is full of extremes, it can evoke a sense of wonder, and can provide opportunities for associating the subject concepts with people and even things that have "heroic" qualities. It can also provide opportunities for associating the content with the contesting of convention, as in the case of scientists who struggled against conventional and prevailing ideas and beliefs, and dealt with in proper historical context (i.e., Copernicus, Kepler, Galileo, Lavoisier, Priestly, Joule, Young, Darwin, Hutton, Wegener, Tesla, etc.).

It deserves to be pointed out that the humanistic element/context, the heroic element, and the sense of wonder, are similar to the characteristics of "romantic science," which had its roots in the movement of "Romanticism" (as a revolt against many Enlightenment era doctrines), that took place in Europe between 1780 and 1840 (see Poggi and Bossi, 1994; Hadzigeorgiou and Schulz, 2014). Watson (2010) sees the movement as a major contribution to the "second scientific revolution." And even though the term "romantic science" may sound like an oxymoron, even a paradox, given that the prevalent view of science sees its development primarily due to an emphasis on rationalism, deductive thinking, experimentation, reductionism, and the mathematization (modeling) of nature, "there is now widespread recognition of the importance of particular romantic contributions to the natural sciences" (Cunningham and Jardine, 1990, p. 19). This revised historical assessment of "romantic science" can make science teachers and science educators more attentive to Egan's (1997) recapitulation theory, and specifically, to the potential of "Romantic Understanding."

What should be pointed out though is that the development of "Romantic Understanding" of science presupposes that students are given the opportunity to relate the science content *with* the romantic features. Even though students of the age range 7 or 8 to 15 generally understand and relate to the world romantically, that is, by associating reality (e.g., a mountain, a neighborhood, a building, a friendship, a human relationship) with the romantic features, it does not follow that they will understand science romantically. (Quite the contrary, they often find themselves at this age alienated from the content and language as presented in textbooks and classroom dialogue, what Lemke (1990) refers to as the "mystique" of science). Hence, it is vital, if the development of "Romantic Understanding" and ultimately engagement with content knowledge is to be an instructional goal, that students be given opportunities to experience a sense of wonder, to explore the extremes and the limits of reality and human experience, and to associate with heroic qualities, and also to become aware of the human context in which scientific knowledge is discovered and developed. Whether the instructional topic is forces and motion, photosynthesis, electric current, biodiversity, or radioactivity, it should be associated with all the above named features of the mind-set. Perhaps, the best way to associate all the aforementioned features of "Romantic Understanding" with science content is to create a narrative or, better, a story, whose plot incorporates all of them. Such an approach gives students the opportunity to use their narrative mode of thinking and to understand science content "romantically." Egan (1992) had also pointed out that a narrative context for the romantic characteristics "can enhance their power to stimulate and develop the imagination" (p. 72).

One could, of course, very well raise the issue of what empirical evidence exists, as regards the development of "Romantic Understanding." The anecdotal evidence about the educational benefits and perhaps about the effectiveness of romantic understanding are insufficient when it comes to informing instructional and curriculum planning, let alone educational policy. It is true that no empirical study can be found in the literature except the study conducted by Hadzigeorgiou et al. (2012). This study investigated "The Effect of the Nikola Tesla Story" on grade 9 students' understanding of the concept of alternating current. This story, based exclusively on historical events, included all the romantic elements [i.e., the characteristic features of "Romantic Understanding" according to Egan's theory (Egan's, 1997)] and the researchers used a quasi-experimental design (i.e., a two-group, pre-test/intervention/post-test design). This means that the students (ninth graders) who participated in the study were not randomly assigned to two groups. Thus, two classrooms from each of 19 schools (from the wider metropolitan area of a European capital) that participated in the intervention formed the control and the experimental group respectively, with a total of 197 students. More specifically, the intervention was conducted over a period of 10 weeks, with the first 4 weeks spent on the teaching of prerequisite knowledge (i.e., fundamentals of current electricity), the fifth and sixth weeks spent on assessment, while the next 3 weeks were devoted to

teaching both groups about alternating current and the idea of the wireless transmission of electrical power. The final assessment of both groups took place during the tenth week. However, 8 weeks later, that is, on the eighteenth week a delayed post-test was also administered to the two groups. The students of the control group were taught about alternating current and the wireless transmission of electricity through direct instruction, and more specifically through the mastery model (see Stefanich and Hadzigeorgiou, 2001), while the experimental group were taught exactly the same content through storytelling (i.e., the Nikola Tesla story)[3].

However, it is important to point out that that study did provide evidence for a significant difference between the control group and the experimental group, in terms of engagement with science content knowledge, retention and understanding. Regardless of the interpretation of significant differences between the two groups (e.g., novelty of the instructional sequence through storytelling, the specific curricular content that was covered, such as current electricity, the Hawthorn effect), the fact that the story helped foster in the students of the experimental group a "Romantic Understanding" of science content knowledge (as all the characteristics of "Romantic Understanding" were identified through content analysis of students' optional journal entries), cannot be disputed.

Certainly more empirical evidence is imperative, but it quite evident that a "Romantic Understanding" of science relates to what the philosopher of science Yehuda Elkana had called "personal science" (as opposed to "public science"). He had argued that the methods of logic are insufficient for describing science as a human endeavor: "logical tools are of limited use in understanding the development of science or, what is even more important, in the teaching of science" (Elkana, 2000, p. 473). Private science, as Hadzigeorgiou and Schulz (2017) argued, is inevitably phenomenological but the prevailing insistence on the "logic" of science, when formulated in public language of "final form science" as found in textbooks, does not give students the picture of science as a human activity, or even a proper historical activity (though the presented history is too often mythical—Allchin, 2013), as pointed out by several previous researchers (see Matthews, 1994, 2015; Hodson, 1998; Donnelly, 2004). A "Romantic Understanding" of science, if it takes place in a narrative learning context, in addition to

---

[3]There are, no doubt, certain limitations regarding the intervention. As with any quasi-experimental design, the two groups in each school were not similar, even though the students' characteristics, like academic achievement, socio-cultural and economic background, and even their general interests, were considered similar. In addition, the novelty of the intervention for the experimental group students, and not the intervention per se, should also be considered a factor that played a role in the results of the intervention. Moreover, the story itself was quite powerful, not only because it included all the elements. However, the limitations of the intervention should not downplay its effectiveness with regard to student engagement and understanding. The interest, in particular, that was generated by the Tesla story, that is, a story with all the characteristic features that encourage the development of romantic understanding, needs to be seriously considered. Indeed, as Klassen and Froese-Klassen (2014b) have pointed out, "The insights provided by romantic understanding and its success in achieving improved student learning could add an enriching new dimension to the research on interest" (p. 140).

encouraging engagement with science content, gives students the opportunity to understand science as an arduous and exciting, but also error prone, human activity, embedded in a socio/cultural context (Hadzigeorgiou et al., 2012). Moreover, taking a wider view and considering the vocational aspect of school science, a romantic understanding can present science as "a grand adventure," something of vital importance, according to the late Nobel prize physicist Feynman (1964), if we want to attract young students to the world of science and, hence, educate future scientists.

CONCLUDING COMMENTS

This paper discussed the potential of "narrative thinking" and "romantic understanding" to engage students in science, particularly science content ideas and ideas about the nature of science. Even though engagement does not guarantee understanding, the latter always presupposes some degree of emotional and/or cognitive engagement. In light of the fact that both "narrative thinking" and "romantic understanding" are about students' making sense of the world and meaning making, they can both "help us answer two fundamental questions in educational theory: What is significant for students? What is meaningful to them"? (see also Hadzigeorgiou, 1997, 2005b, p. 31; Schulz, 2014a,b; Krevetzakis, 2019). It may very well be argued that the degree to which students become engaged with science, through the opportunities they have to use their narrative mode of thinking and also to understand science "romantically," namely, by being helped to associate science ideas with the characteristic features of "romantic understanding," can show teachers the degree to which students perceive science as something significant and meaningful (Hadzigeorgiou, 2016).

Certainly, the complexity of the processes of both engagement with science and understanding science, one the one hand, and the multiplicity of factors involved in both of them, on the other, makes one cautious about the effectiveness of the use of narratives and stories, and of the "romantic" approach, as was discussed in this paper, to encourage engagement that will, in turn, result in understanding. Putting aside the empirical evidence that exists to date, what should be noted is that what teachers and curriculum designers wish to achieve is to increase the possibilities for students to understand science. Apparently, because of their *inherent* nature, narrative and romantic understanding increase such possibilities (for understanding science). The message from a recent study by Godec et al. (2018) should be a reminder of that. Even though their study—which they approached from a sociological/Bourdieusian perspective—showed that student engagement with science became possible only when students' "habitus" (i.e., set of deeply embedded and internalized dispositions) aligned with the "field" (i.e., the social environment of the classroom with certain sets of rules, relationships, and expectations), they did acknowledge the possibility of broadening the notion of "field," so that more opportunities for more diverse students could be provided.

Thus, in recognizing and valuing the individual capacities of students, teachers could offer more opportunities to more

students. Narrative thinking is indeed an individual capacity as is a "romantic understanding" of the world, at least in the case of students approximately in the age range 8–15. In actual fact, such individual capacities are students' "capital," which ought to be considered by teachers and curriculum designers. For it should be noted that it is the "field" that determines whether something (e.g., an individual capacity) can be considered as "capital" (see Godec et al., 2018). Future research is certainly needed to more clearly articulate how such "capital" can be tapped from the various perspectives on teaching and learning science (e.g., sociocultural, conceptual change) so that we better understand and appreciate its potential. But we should be reminded, nonetheless, that this potential has been indirectly hinted at by the educational philosopher Maxine Greene: "the problem in education is how to help students discover the imaginative mode of awareness" (Greene, 1978, p. 186). Both narrative thinking and romantic understanding can facilitate such discovery.

Hence, it is important, in closing this chapter, to point out that more attention should be paid by the science education community to the development of students' *imagination*, by seriously considering the role of the narrative mode of thinking and the development of romantic understanding in the context of school science education. Regardless of the fact that the history of science has provided ample evidence that scientific discovery and scientific understanding are indeed imaginative endeavors (Hadzigeorgiou and Stefanich, 2001; Hadzigeorgiou and Garganourakis, 2010; Hadzigeorgiou, 2016; Lindholm, 2018), in the context of education in general, especially early childhood education, the value of imagination needs to be reclaimed. What the educational theorist Kieran Egan has pointed out should be seriously and carefully considered:

> *A feature of young children's mental life that is commonly asserted as an implication 'of research on their logico-mathematical thinking is that their thought is perception-dominated. If we focus instead on their imaginative lives we can see rather an enormously energetic realm of intellectual activity that is conception-driven.* (Egan, 1999, p. 9).

DATA AVAILABILITY

No datasets were generated or analyzed for this study.

AUTHOR CONTRIBUTIONS

YH is the first author as he had run the original research study in Greece as mentioned in the paper, and whose major research has been concerned with the process of engagement with science content knowledge using imaginative approaches in teaching and learning. However, both have co-authored on several papers together involving engagement and history of science. YH wrote the first draft and RS added substantial aspects (concepts, clarifications, additions, research literature) to the manuscript as it developed. Both authors have read, revised, and contributed to the final version of the submitted paper, and approved the final submission.

REFERENCES

Allchin, D. (2013). *Teaching the Nature of Science: Perspectives and Resources*. St. Paul, MN: Ships Press.

Avraamidou, L., and Osborne, J. (2009). The role of narrative in communicating science. *Int. J. Sci. Educ.* 31, 1683–1017. doi: 10.1080/09500690802380695

Bauer, H. H. (1992). *Scientific Literacy and the Myth of the Scientific Method*. Chicago, IL: University of Illinois Press.

Borda, E. J. (2007). Applying Gadamer's concept of dispositions to science and science education. *Sci. Educ.* 16, 1027–1041. doi: 10.1007/s11191-007-9079-5

Bruner, J. (1985). "Narrative and paradigmatic modes of thought," in *84th Yearbook of NSSE, Learning and Teaching the Ways of Knowing* (Chicago, IL: University of Chicago Press).

Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. (1991). The narrative construction of reality. *Critic. Inquiry* 18, 1–21. doi: 10.1086/448619

Caine, R., Caine, G., McClintic, C., and Klimic, K. (2005). *Brain/Mind Learning Principles in Action*. Thousand Oaks, CA: Morgan Press.

Cawthron, E. R., and Rowell, J. A. (1978). Epistemology and science education. *Stud. Sci. Educ.* 5, 31–59. doi: 10.1080/03057267808559856

Cunningham, A., and Jardine, N. (Eds.). (1990). *Romanticism and the Sciences*. Cambridge; New York, NY: Cambridge University Press.

Dahlin, B. (2001). The primacy of cognition – or of perception? A phenomenological critique of the theoretical bases of science education. *Sci. Educ.* 10, 453–475. doi: 10.1023/A:1011252913699

Dewey, J. (1934). *Art as Experience*. New York, NY: Perigee Books.

Dewey, J. (1966). *Democracy and Education*. New York, NY: MacMillan.

Di Trocchio, F. (1997). *Il Genio Incompreso*. Milan: Mondadori.

Donald, M. (1991). *Origins of the Modern Mind*. Cambridge, MA: Cambridge University Press.

Donnelly, J. (2004). Humanizing science education. *Sci. Educ.* 88, 762–784. doi: 10.1002/sce.20004

Duschl, R. (1994). "Research on history and philosophy of science," in *Handbook of Research on Science Teaching and Learning*, ed D. Gabel (New York, NY: Macmillan), 443–465.

Egan, K. (1986). *Teaching as Story-Telling*. Chicago, IL: University of Chicago Press.

Egan, K. (1988). *Primary Understanding*. Chicago, IL: University of Chicago Press.

Egan, K. (1990). *Romantic Understanding*. Chicago, IL: University of Chicago Press.

Egan, K. (1992). *Imagination in Teaching and Learning*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226244136.001.0001

Egan, K. (1997). *The Educated Mind: How Cognitive Tools Shape Our Understanding*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226190402.001.0001

Egan, K. (1999). *Children's Minds, Talking Rabbits and Clockwork Oranges*. New York, NY: Teachers College Press.

Egan, K. (2005). *An Imaginative Approach to Teaching*. San Francisco, CA: Jossey-Bass.

Eger, M. (1992). Hermeneutics and science education: an introduction. *Sci. Educ.* 1, 337–348. doi: 10.1007/BF00430961

Eger, M. (1993a). Hermeneutics as an approach to science: part I. *Sci. Educ.* 2, 1–29. doi: 10.1007/BF00486659

Eger, M. (1993b). Hermeneutics as an approach to science: part II. *Sci. Educ.* 2, 303–328. doi: 10.1007/BF00488169

Elkana, Y. (2000). Science, philosophy of science and science teaching. *Sci. Educ.* 9, 463–485. doi: 10.1023/A:1008652109868

Feyerabend, P. (1993). *Against Method*. London: Verso.

Feynman, R. (1964). "The value of science," in *Science and Ideas*, eds A. Arons and A. Bork (Englewood Cliffs, NJ: Prentice Hall), 3–12.

Feynman, R. (1968). What is science? *Phys. Teach.* 6, 313–320. doi: 10.1119/1.2351388

Giere, R. N. (1991). *Understanding Scientific Reasoning, 3rd Edn*. Orlando, FL: Harcourt Brace Jovanovich.

Godec, S., King, H., Archer, L., Dawson, E., and Seakins, A. (2018). Examining student engagement with science through a Bourdieusian notion of field. *Sci. Educ.* 27, 501–521. doi: 10.1007/s11191-018-9988-5

Gottschall, J. (2012). *The Storytelling Animal: How Stories Make Us Human*. Boston, MA; New York, NY: Houghton Mifflin Harcourt.

Gough, N. (1993). Environmental education, narrative complexity and postmodern science/fiction. *Int. J. Sci. Educ.* 5, 607–625. doi: 10.1080/0950069930150512

Greene, M. (1978). *Landscapes of Learning*. New York, NY: Teachers College Press.

Hadzigeorgiou, Y. (1997). Relationships, meaning and the science curriculum. *Curric. Teach.* 12, 83–89. doi: 10.7459/ct/12.2.08

Hadzigeorgiou, Y. (1999). On problem situations and science learning. *Sch. Sci. Rev.* 81, 43–49.

Hadzigeorgiou, Y. (2005a). *On Humanistic Science Education*. Eric Document (ED506504) Washington, DC.

Hadzigeorgiou, Y. (2005b). Romantic understanding and science education. *Teach. Educ.* 16, 23–32. doi: 10.1080/1047621052000341590

Hadzigeorgiou, Y. (2005c). Science, personal relevance and social responsibility: integrating the liberal and the humanistic traditions of science education. *Educ. Pract. Theor.* 27, 87–103. doi: 10.7459/ept/27.2.07

Hadzigeorgiou, Y. (2012). Fostering a sense of wonder in the science classroom. *Res. Sci. Educ.* 42, 985–1005. doi: 10.1007/s11165-011-9225-6

Hadzigeorgiou, Y. (2015). "Young children's ideas about physical science concepts," in *Research in Early Childhood Science Education*, eds K. Trundle and M. Sackes (Dordrecht; Heidelberg; New York, NY; London: Springer), 67–97. doi: 10.1007/978-94-017-9505-0_4

Hadzigeorgiou, Y. (2016). *Imaginative Science Education. The Central Role of Imagination in Science Education*. Cham: Switzerland: Springer International. doi: 10.1007/978-3-319-29526-8_8

Hadzigeorgiou, Y. (2017). Teaching the nature of science through storytelling: some empirical evidence from a grade 9 classroom. *SFU Educ. Rev.* 10, 1–11. doi: 10.21810/sfuer.v10i2.318

Hadzigeorgiou, Y., and Fotinos, N. (2007). Imaginative thinking and the learning of science. *Sci. Educ. Rev.* 6, 15–22.

Hadzigeorgiou, Y., and Garganourakis, V. (2010). Using Nikola Tesla's story and experiments, as presented in the film "The Prestige", to promote scientific inquiry. *Interchange* 41, 363–378. doi: 10.1007/s10780-010-9136-x

Hadzigeorgiou, Y., Klassen, S., and Froese-Klassen, C. (2012). Encouraging a "Romantic Understanding" of school science: the effect of the Nikola Tesla story. *Sci. Educ.* 21, 1111–1138. doi: 10.1007/s11191-011-9417-5

Hadzigeorgiou, Y., and Schulz, R. M. (2014). Romanticism and romantic science: their contribution to science education. *Sci. Educ.* 23, 1963–2006. doi: 10.1007/s11191-014-9711-0

Hadzigeorgiou, Y., and Schulz, R. M. (2017). What really makes secondary school students "want" to study physics? *Educ. Sci.* 7:84. doi: 10.3390/educsci7040084

Hadzigeorgiou, Y., and Stefanich, G. (2001). Imagination in science education. *Contemp. Educ.* 71, 23–29.

Hadzigeorgiou, Y., and Stivaktakis, S. (2008). Encouraging involvement with school science. *J. Curric. Pedag.* 5, 138–162. doi: 10.1080/15505170.2008.10411692

Hirst, P. (1972). "Liberal education and the nature of knowledge," in *Education and the Development of Reason*, eds R. Dearden, P. Hirst, and R. Peters (London: Routledge), 391–414.

Hodson, D. (1998). Science fiction: the continuing misrepresentation of science in the school curriculum. *Curric. Stud.* 6, 191–216. doi: 10.1080/14681369800200033

Kalman, C. (2008/2017). *Successful Science and Engineering Teaching at Colleges and Universities, 2nd Edn*. Charlotte, NC: IAP. doi: 10.1007/978-1-4020-6910-9

Kalman, C. (2011). Enhancing student's conceptual understanding by engaging science text with reflective writing as a hermeneutical circle. *Sci. Educ.* 20, 159–172. doi: 10.1007/s11191-010-9298-z

Klassen, S., and Froese-Klassen, C. (2014a). "Science teaching with historically based stories: theoretical and practical perspectives," in *International Handbook of Research in History, Philosophy and Science Teaching*, ed M. Matthews (Berlin: Springer), 1503–1529. doi: 10.1007/978-94-007-7654-8_47

Klassen, S., and Froese-Klassen, C. (2014b). The role of interest in learning science through stories. *Interchange* 45, 133–151. doi: 10.1007/s10780-014-9224-4

Krevetzakis, E. (2019). On the centrality of physical/motor activities in primary education. *J. Adv. Educ. Res.* 4, 23–33. doi: 10.22606/jaer.2019.41003

Kuhn, T. (1970). *The Structure of Scientific Revolutions, 2nd Edn.* Chicago, IL: University of Chicago Press.

Kurth, L. A., Kidd, R., Gardner, R., and Smith, E. L. (2002). Student use of narrative and paradigmatic forms of talk in elementary science conversations. *J. Res. Sci. Teach.* 39, 793–781. doi: 10.1002/tea.10046

Lemke, J. (1990). *Talking Science: Language, Learning, Values.* Norwood, NJ: Ablex.

Lindholm, M. (2018). Promoting curiosity? Possibilities and pitfalls in science education. *Sci. Educ.* 27, 987–1002. doi: 10.1007/s11191-018-0015-7

Matthews, M. R. (1994). *Science Teaching: The Role of History and Philosophy of Science.* New York, Ny: Routledge.

Matthews, M. R. (2015). *Science Teaching: The Contribution of History and Philosophy of Science.* New York, NY: Routledge. doi: 10.4324/9781315811642

Nola, R., and Irzik, G. (2005). *Philosophy, Science, Education and Culture.* Dordrecht: Springer.

Piaget, J. (1970). *Genetic Epistemology.* New Yok, Ny: Norton. doi: 10.7312/piag91272

Poggi, S., and Bossi, M. (Eds.). (1994). *Romanticism in Science: Science in Europe, 1790-1840.* Dordrecht; Boston, MA; London: Kluwer Academic Publishers. doi: 10.1007/978-94-017-2921-5

Polito, T. (2005). Educational theory as theory of culture: a vichian perspective on the educational theories of John Dewey and Kieran Egan. *Educ. Philos. Theor.* 37, 475–494. doi: 10.1111/j.1469-5812.2005.00136.x

Popper, K. (1972). *Objective Knowledge. An Evolutionary Approach.* Oxford: Clarendon Press.

Pugh, K. (2004). Newton's laws beyond the classroom walls. *Sci. Educ.* 88, 182–196. doi: 10.1002/sce.10109

Pugh, K., Bergstrom, C., and Spencer, B. (2017). Profiles of transformative engagement: Identification, description, and relation to learning and instruction. *Sci. Educ.* 101, 369–398. doi: 10.1002/sce.21270

Pugh, K. J. (2011). Transformative experience: an integrative construct in the spirit of Deweyan pragmatism. *Educ. Psychol.* 46, 107–121. doi: 10.1080/00461520.2011.558817

Resnick, L. (1983). Mathematics and science learning: A new conception. *Science,* 220, 477–488.

Schulz, R. M. (2007). Lyotard, postmodernism and science education. A rejoinder to Zembylas. *Educ. Philos. Theor.* 39, 633–656. doi: 10.1111/j.1469-5812.2007.00301.x

Schulz, R. M. (2009). Reforming science education: part II. Utilizing Kieran Egan's educational metatheory. *Sci. Educ.* 18, 251–273. doi: 10.1007/s11191-008-9168-0

Schulz, R. M. (2014a). "Philosophy of education and science education: a vital but underdeveloped relationship," in *International Handbook of Research in History, Philosophy and Science Teaching,* ed M. R. Matthews (Berlin: Springer), 1259–1315. doi: 10.1007/978-94-007-7654-8_39

Schulz, R. M. (2014b). *Rethinking Science Education: Philosophical Perspectives.* Charlotte, NC: IAP Publishing.

Solomon, J. (2002). Science stories and science texts: what can they do for our students? *Stud. Sci. Educ.* 37, 85–106. doi: 10.1080/03057260208560178

Stefanich, G., and Hadzigeorgiou, Y. (2001). Models and applications," in *Science Teaching in Inclusive Classrooms,* ed G. Stefanich (Cedar Falls, IA: Woolverton), 61–90.

Stinner, A. (1995). "Science textbooks: their present role and future form," in *Learning Science in Schools: Research Reforming Practice,* eds S. M. Glynn and R. Duit (Mahwah, NJ: Lawrence Erlbaum Associates), 275–296.

Sutton, C. (1996). Beliefs about science and beliefs about language. *Intl. J. of Sci Ed.,* 18, 1–18.

Sutton-Smith, B. (1988). "In search of the imagination," in *Imagination and Education,* eds K. Egan and D. Nadaner (New York, NY: Teachers College Press), 3–29.

Tytler, R., Prain, V., Hubber, P., and Waldrip, B. *(Eds.).* (2013). *Constructing Representations to Learn Science.* Rotterdam: Sense Publishers. doi: 10.1007/978-94-6209-203-7

Watson, P. (2010). *The German Genius: Europe's Third Renaissance, the Second Scientific Revolution and the Twentieth Century.* New York, NY: Harper Perennial.

Wellington, J., and Osborne, J. (2001). *Language and Literacy in Science Education.* Buckingham: Open University Press.

White, H. (1981). "The value of narrativity in the representation of reality," in *On Narrative,* ed W. Mitchell (Chicago, IL: University of Chicago Press), 1–23.

Yore, L. D., and Treagust, D. F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *Int. J. Sci. Educ.* 28, 291–314. doi: 10.1080/09500690500336973

Zabel, J., and Gropengiesser, H. (2015). What can narrative contribute to students' understanding of scientific concepts, e.g. evolution theory? *J. Eur. Teach. Educ. Netw.* 10, 136–146.

Check for updates

Using Phenomenography to Tackle Key Challenges in Science Education

Feifei Han* and Robert A. Ellis

Office of Pro-Vice-Chancellor (Arts, Education and Law), Griffith University, Brisbane, QLD, Australia

This article describes how phenomenography, as a qualitative research method, can be used to tackle key challenges in science education. It begins with an overview of the development of phenomenography. It then describes the philosophical underpinnings of phenomenographic inquiry, including ontological and epistemological roots, and its unique second-order perspective. From theoretical background to practicality, the paper uses rich examples to describe in detail the procedures of conducting a phenomenographic study, including sampling and data collection, analyzing phenomenographic data, and communicating key findings. The paper concludes by showing how the phenomenographic method can be used to develop students' conceptual understanding of scientific concepts, to inform effective instructional design in science teaching, and to identify and improve evidence-based factors in student learning to enhance learning outcomes in science.

Keywords: phenomenography, qualitative research method, theoretical underpinnings, second-order perspective, key challenges in science education

INTRODUCTION

How to assist students in achieving better quality of learning in science subjects is an ongoing agenda in science education. With a purpose to impact on real-world educational practice in science education, researchers from different methodological camps bring their own ontological (why things exist the way they do) and epistemological (how learning occurs) perspectives to the advancement of theories in science teaching. However, no matter what background they come from, there are some common challenges faced by science educators today. This article draws on national reports into challenges for science education and describes a research method for addressing them known as phenomenography.

Current educational dilemmas facing science education are highlighted in national reports in the United Kingdom (Hoyle, 2016), United States (National Research Council, 2012), and in many other countries (Alberts, 2013). One of the grand challenges for science education is to improve students' conceptual development of scientific concepts, including helping students modify their prior mistaken concepts, and/or moving novice concepts toward professional ones (Osborne et al., 2016). To achieve this goal, it is important to begin with identifying what concepts students already have, whether the concepts are aligned with scientific explanations, and if not, what aspect(s) make it variant from what is commonly understood (National Research Council, 2012). The phenomenographic method is illuminating, because the content-rich phenomenographic data can be used evaluate students' initial understanding and the evolvement of that understanding of scientific concepts (Minasian-Batmanian et al., 2006).

Moreover, to facilitate students' understanding of scientific concepts and guide them away from pathways that lead to misunderstandings, especially for abstract and difficult concepts, science educators should develop innovative instructional strategies from various angles in order to help students understand scientific concepts more holistically (National Research Council, 2012). Phenomenography is useful to achieve this aim because it serves as a basis for using the variation theory of learning to improve pedagogical design for presenting scientific concepts (Lo and Chik, 2016; Pang and Ki, 2016).

Another key challenge faced by science educators is to identify key aspects of student learning experience which are able to explain learning outcomes so as to take targeted actions to improve the learning experience. Using the phenomenographic method, researchers in science education have identified variations in conceptions of and approaches to learning science subjects, and perceptions of the teaching quality and learning environment, all of which account for qualitatively different learning outcomes (Hardy et al., 2014; Kapucu, 2014). Once these variations have been identified, educators can implement corresponding strategies to change the less desirable variation(s) of these elements (e.g., fragmented conceptions, surface approaches, and negative perceptions) to the more desirable ones (e.g., coherent conceptions, deep approaches, and positive perceptions) to enhance quality of science learning.

Before we unpack how to apply phenomenography in tackling these issues, we first introduce the philosophical background of the method and explains practical issues in conducting phenomenographic studies using representative examples in published studies. The following provides a brief historical account of phenomenography and how and where it has been used. It highlights theoretical underpinnings of the method and explains key procedures of conducting a phenomenographic study, including data collection, sampling methods, principles and procedures of phenomenographic analysis, and ways of communicating findings. The last section discusses how the research method can be meaningfully used to tackle the three key challenges in science education.

RESEARCH FOCI AND HOW PHENOMENOGRAPHY HAS BEEN USED

Phenomenography was initially developed by a body of educational researchers in Sweden in the late 1970s to study variations of how students learn and understand concepts (Marton and Säljö, 1976a,b; Marton and Svensson, 1979; Säljö, 1979). In its subsequent development, the research foci have been expanded. The method examines "qualitatively different ways in which people experience, conceptualize, perceive, and understand various aspects of, and various phenomena in the world around them" (Marton, 1986, p. 31).

Phenomenography is now known as a well-established qualitative research method and has been widely adopted to research education in multiple disciplines, such as technology (Englund et al., 2017; Hsieh and Tsai, 2017), engineering (Case and Light, 2011; Magana et al., 2012), mathematics

(Kapucu, 2014; Gordon and Nicholas, 2015); and terrains beyond education, like management, computer programming, organizational studies, library and information research, nursing, medical and health care research (Yates et al., 2012; Stenfors-Hayes et al., 2013; Teeter and Sandberg, 2016). In the last couple of decades, the method has been especially appealing to science educators (Brown et al., 2006; Olympiou and Zacharia, 2012; Lee et al., 2013; Chiu et al., 2016; Howitt and Wilson, 2018).

## THEORETICAL UNDERPINNINGS OF PHENOMENOGRAPHY

Ontologically speaking, phenomenography believes that "an individual cannot experience without something being experienced" (Marton and Pang, 2008, p. 535). This means that phenomenographic researchers do not treat a phenomenon separately from people who experience it (Sin, 2010). Marton (2000, p. 105) further elaborated the ontology of phenomenography:

"There are not two worlds: a real, object world, on the one hand, and a subjective world of mental representations, on the other. There is only one world, a really existing world, which is experienced and understood in different ways by human beings. It is simultaneously objective and subjective."

Using an example of approaches to learning as a research object to illustrate, phenomenographic researchers consider that the approaches adopted by students are not an inherent trait, but may vary from one learning context to another, depending on factors, such as students' understanding of the disciplinary contents, their perceptions of the course design, and their views of the learning environment. This means that the same student may adopt a deep approach (e.g., being proactive, taking initiatives, and seeking in-depth meaning of the subject matter, Prosser and Trigwell, 1999; Vermunt and Donche, 2017) to learning *biology*, but he/she may adopt a surface approach (e.g., following formulas, rote memorization, reproducing the contents in the textbooks, and completing the learning tasks with little reflections, Prosser and Trigwell, 1999; Vermunt and Donche, 2017) to studying *chemistry*, because the student may find difficult to understand the learning goals in chemistry.

Turning toward the epistemological stance, which reflects a person's view on the nature of knowledge, phenomenography is grounded in the "intentionality" of human behaviors, which is characterized by purposefulness and consciousness, involving different foci of an awareness of a phenomenon. Such intentionality can generate two sources responsible for the qualitative variations in an experience. For one thing, people may experience different parts of a phenomenon. For another, even if they experience the same parts, these parts may not in the foreground of their awareness (Yates et al., 2012). This is why some people can share the same experience but come away with different meanings from it.

The phenomenographic method present sources of variations in an unique analytical framework known as "the anatomy of experience," which describes the two components of the

FIGURE 1 | The anatomy of experience (adapted from Marton and Booth, 1997, p. 88).
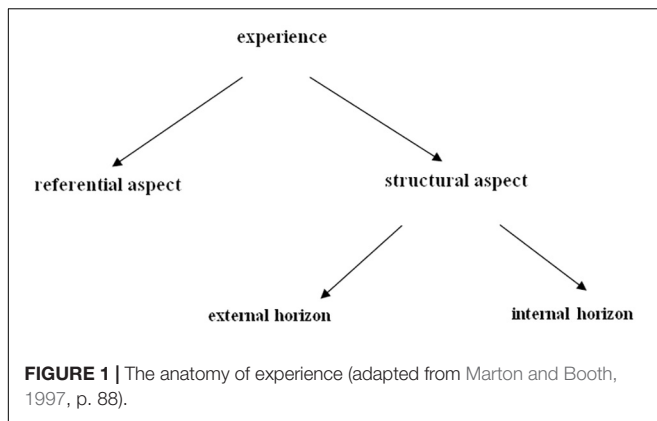
conscious awareness of an experience, namely a referential aspect and a structural aspect. While the former refers to the meaning of an experience, the latter is related to the structure of that experience (Marton and Pong, 2005). The two aspects simultaneously occur and are intertwined (Marton and Booth, 1997). The structural aspect can be further distinguished between an external and an internal horizon. The external horizon, the "discernment of the whole from the context," enables the experience to be differentiated from its context and background (Marton and Booth, 1997, p. 87); whereas the internal horizon, the "discernment of the parts and their relationships within the whole," denotes the internal relationship of various parts in an experience, how the parts are distinctive from each other, and how the parts jointly form a cohesive entity (Marton and Booth, 1997, p. 87) (see **Figure 1** for an visual representation "the anatomy of experience"). To be cognizant of all aspects of a phenomenon is to be consciously aware of its referential and structural components.

We use 'conceptions of learning science' as a research object to illustrate different aspects of "the anatomy of experience." A student describes his/her conceptions of learning science: "When learning science, I need to memorize many concepts, facts, symbols, and equations. Sometimes, I feel that I am learning social studies such as history and language while learning science. . ." (Tsai, 2004, p. 1739). The learner assigns "memorizing many things" as the meaning of learning science, which is the referential aspect. The learner distinguishes "learning science" from the background of learning other subjects (i.e., external horizon of the structural aspect), even though his/her experience finds learning these subjects share similarities. The learner describes that the experience of memorizing includes a number of parts, such as concepts, facts, symbols, and equations; and recognizes that these parts together constitute the things needs to be memorized (i.e., the internal horizon of the structural aspect) in order for learning to occur. These aspects can be visually represented in the anatomy of experience of "learning science" in **Figure 2**.

Another important theoretical underpinning of phenomenography is its unique second-order perspective, which emphasizes the collective meaning and variations in a phenomenon *as experienced by people* (Marton and Pang,

2008). This contrasts sharply with the first-order perspective, which focuses on explicating the general and invariant essence of a phenomenon *through people* (Richardson, 1999; Marton and Pang, 2008). The detailed explanations of the first- and second-order perspectives are given by Åkerlind (2018) in the following:

"From a second-order perspective, human experience and variation in experience is the core of the investigation; from a first-order perspective, human experience is but the medium for collecting data, and variation in human experience (within the same experimental conditions) is white noise, to be filtered by statistical tests of significance to better determine the reality underlying the noise." (p. 6)

Such fundamental difference between the first- and second-order perspectives is also reflected in the research questions addressed by phenomenography and methods adopting first-order perspective. For instance, "What are the different approaches college students adopt to learn physics?" is more suitable to be answered using the phenomenographic method, because the research purpose is to gain an understanding of *various ways* of learning physics *in the lenses of college students*. On the other hand, the research question "How do college students learn physics?" is more appropriate to be investigated using the first-order perspective, as the focus is on describing *the common features* which characterize tertiary physics learning.

Having described the theoretical underpinnings of phenomenography, the next section explains practical issues of conducting a phenomenographic study by using accessible examples.

KEY PROCEDURES OF CONDUCTING A PHENOMENOGRAPHIC STUDY

Three key procedures for conducting a phenomenographic study are described in the following: (1) data collection and sampling, (2) principles of phenomenographic data analysis, and (3) effective communication of the phenomenographic results.

Data Collection and Sampling Methods

There are multiple ways to collect phenomenographic data, such as using semi-structured interviews, open-ended questionnaires, think-aloud methods, and observation, each of which offers different strengths and limitations to the research process. When there are a relatively large number of participants, using an open-ended questionnaire is advantageous as it is easy to administer and allows a wider range of experiences of a phenomenon to be captured. Think-aloud methods, which require participants to verbalize their thoughts while performing a task, are more suitable to uncover a process-oriented phenomenon, like carrying out a scientific experiment. While think-aloud methods are able to reflect detailed concurrent thinking, an obvious drawback is that data collection is time-consuming and the essential training of participants adds an extra burden. Used to a much lesser extent, observation is used to reflect how people perceive a phenomenon through what they act upon (Dall'Alba, 1994; Marton, 2015). Observation has a merit to collect the information

FIGURE 2 | The anatomy of experience of science learning.
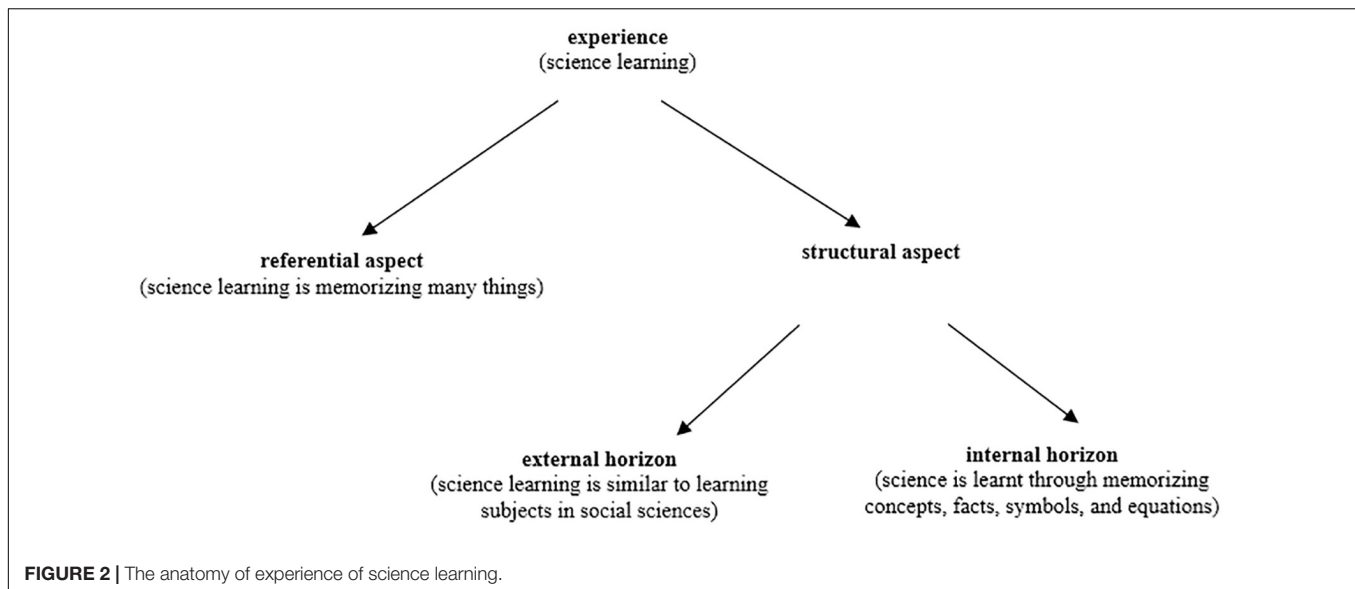
of both the process (e.g., video clips and field notes of dissecting specimens in a laboratory) and the product of an activity (e.g., the dissected organs), providing triangulation from multiple data sources (Lam, 2017).

The most popular phenomenographic data collection method is semi-structured interviews, which are often conducted using a set of pre-defined interview questions as well as the information emerging from participants' responses (Stenfors-Hayes et al., 2013). While other qualitative interviews either focus on the participant or the phenomenon itself, the phenomenographic interviews emphasize the relation between the participant and the phenomenon (Bruce, 1997). Hence, the interview questions should be carefully constructed to allow participants to reflect on their experience (Yates et al., 2012). For instance, to find out conceptions of "learning science," a question like: "What do you understand by 'learning science'?" is more appropriate than "What is 'learning science'?", because the former is on the interplay between the interviewee and science learning, whereas the latter is on science learning itself, which does not necessarily involve the interviewee's personal experience.

To secure a rich understanding of the students' perspectives in interviews, researchers should give them freedom to expand their understandings, and researchers should ask follow-up questions to explore interesting themes from the responses. When constructing follow-up questions, neither should researchers ask leading questions nor should they introduce ideas that has not been expressed by the interviewees to avoid collecting biased data (Åkerlind et al., 2005). A question like: "What are the differences between learning science and learning social sciences subjects?" would be leading because it presumes that learning science differs from learning social sciences subjects. A more appropriate question would be: "Do you consider learning science and learning social sciences to be the same thing? Why or why not?" However, if the interviewee has responded: "To me, science learning is quite different from learning social sciences, such as history and language," then asking "What are the differences?"

is not leading. In this scenario, researchers should explore the differences between "learning science" and "learning history" or "learning language" experienced by the interviewee rather than introducing another social sciences subject.

With regard to sampling method, phenomenographic inquiry adopts purposeful sampling, which resembles most of other qualitative methods (Marton, 1986; Booth, 1997). To select participants, researchers should consider whether the potential participants have experienced the phenomenon under investigation; and whether the number of the participants are sufficiently large for variations to be revealed. However, purposeful sampling by no means just targets a particular type of individuals, as this will result in danger of undermining variations and violating the validity of the study (Ashworth and Lucas, 2000). For instance, when a researcher intends to explore first year undergraduates' approaches to learning science, he/she should not only target those with good academic performance in science subjects. Otherwise, opportunities to capture approaches to learning science from students with poor academic performance will be lost.

In phenomenographic research practice, using both semi-structured interviews and open-ended questionnaires to collect data is often favored as such combination allows both breadth and depth of variations to be covered in the data. Because the semi-structured interviews are able to provide rich and in-depth descriptions, whereas the open-ended questionnaires are suitable for collecting data from relatively large number of participants to cover a wider range of experience for variations to be revealed (see Kapucu, 2014; Chiu et al., 2016 as examples).

Principles and Processes of Analyzing Phenomenographic Data

The main aim of phenomenographic data analysis is to identify a set of qualitatively different categories representing variations of individuals' experience of a phenomenon. There

are a set of special principles to follow in phenomenographic data analysis to achieve this. The most important principle is that data analysis is iterative rather than sequential (Yates et al., 2012). This principle alerts researchers to not to make quick decisions on the number of categories arising from the data. Another principle is that analyses should focus on searching for collective meaning of responses rather than describing each individual's response (Åkerlind, 2005). Thirdly, researchers should avoid merely presenting participants' responses without identifying variations and relations amongst them (Bruce, 1997). Interestingly, there is no singular agreed upon analytical procedure about how to analyze phenomenographic data (Ashworth and Lucas, 2000). For this reason, **Table 1** summarizes the main stages proposed by different researchers. Although the number and the name of the stages vary, there are some similarities in terms of key stages.

TABLE 1 | Phenomenographic data analysis processes.

Marton et al. (1992)	Dahlgren and Fallsberg (1991), McCosker et al. (2004)	Säljö (1997)
	1. Familiarization: The data are viewed for researchers to be familiar with the details of the data.	**1. Familiarization:** Similar to stage 1 in Dahlgren and Fallsberg (1991), McCosker et al. (2004)
1. Identification: Data which is related to the phenomenon being described are identified.	**2. Condensation:** The most representative statements are selected to uncover the patterns of the data.	**3. Identification:** Similar to stage 1 in Marton et al. (1992)
2. Sorting: The identified data are sorted into 'pools of meaning' according to similarities.	**3. Comparison:** Unpack similarities and differences to identify sources of variations.	**4. Sorting:** Similar to stage 2 in Marton et al. (1992)
3. Contrasting and categorizing: The 'pools of meaning' are contrasted, and categories are generated with descriptions.	**4. Grouping:** The statements are sorted by similarities. **5. Articulating:** The essence of similarities are extracted, categorized, and described. **6. Labeling:** The categories are represented linguistically. **7. Contrasting:** The categories are contrasted.	**5. Contrasting and categorizing:** Similar to stage 3 in Marton et al. (1992)
4. Reliability checking: The reliability is checked by having a portion of the data coded by independent researchers and the inter-coded reliability is calculated.		**6. Reliability checking:** Similar to stage 4 in Marton et al. (1992)

Table 1 reveals that researchers seem to agree that phenomenographic data analysis commences with a stage of familiarization, which is normally realized by viewing and reading through the transcripts of the interviews or the responses in the open-ended questionnaires. The purpose of familiarization is for researchers to develop a good sense of the breadth and depth of the participants' responses. Following familiarization stage is data reduction and condensation stage, which is given different names by different researchers [e.g., "identification" in Marton et al. (1992) and Säljö (1997); "condensation" in Dahlgren and Fallsberg (1991) and McCosker et al. (2004)]. Reduction and condensation is achieved through identifying the most relevant and important parts in the responses, allowing patterns of the responses to be revealed more easily. The third main stage is classification of responses, which is achieved through comparing and contrasting similarities and differences in order to generate an initial set of the categories. Each category should stand distinctly to reflect *the variation* of the experience rather than singular experience (Bowden, 2000). The next stage is labeling categories using appropriate descriptors which best represent the theme of each category. Due to the iterative nature of phenomenographic data analysis, classifying and labeling stages often take place multiple times, during which the initially formed categories and their descriptions are refined and modified to reach a final set of categories, which should best represent the qualitative variations of the phenomenon from the participants' responses.

When deriving categories, it is important to remember three points in order to provide the most meaningful and transferable outcomes. First, each category should reveal some distinctness from other categories. The distinctness can be either from the referential aspect focusing on differences in the meaning or from the structural aspect focusing on different parts or combinations of parts (Marton and Booth, 1997). Second, the number of categories should be parsimonious. Third, the type of the logical relations amongst the categories should be clearly specified (Marton and Booth, 1997). The process of specifying logical relations amongst the categories helps pinpoint whether the variation is caused by: (1) failure to distinguish the phenomenon from its context; (2) unawareness of some parts of the phenomenon; (3) having different perceptions of the structural relations between the parts; or (4) a combination of these.

In **Table 1**, it should be noted that only Marton et al.'s (1992) and Säljö's (1997) procedure has a stage of reliability checking. Unlike the quantitative research methods whose reliability is on replicability in other research contexts, the term of "reliability" in phenomenographic research places emphasis on consistency of assigning data using the generated categories by other researchers (Marton, 1988). Marton et al. (1992) advises that two or more researchers should apply the categories and analyze the data independently. Disagreement can be discussed and resolved to minimize researcher bias (Tight, 2016). The inter-judge reliability (also called as inter-judge communicability) can be computed based on the disagreement after the discussion (Cope, 2004).

To illustrate the key stages in phenomenographic data analysis, we use students' responses about "conceptions of learning science" as an example (adapted from Tsai, 2004).

Extract A: *"I just have an impression that in science classes, the teachers often state manyspecial terms and formula in which I am supposed **to memorize**."*

Extract B: *"The major purpose of learning science is **to pass the exams** and have high exam scores, and then get into good colleges."*

Extract C: *"Learning science indicates the **acquisition of scientific knowledge**. I have more knowledge derived from science instruction."*

Extract D: *"Learning science is **preparing for tests**. Science, for us, is a major subject for the College Entrance Examination."*

Extract E: *"The purpose of learning science is **to acquire** more **knowledge** about natural phenomena and living things."*

Extract F: *"Learning science is to acquire some knowledge and skills to solve real-life problems. Science needs **to be applied to** solve practical problems."*

Extract G: *"When learning science, I need **to memorize** many concepts, facts, symbols, and equations. Sometimes, I feel that I am learning social studies such as history and language while learning science…"*

Extract H: *"Learning science helps us obtain knowledge. The knowledge can **be applied to** invent more products to improve the quality of our life."*

Using the stages in **Table 1** as a guide, in the familiarization and identification stages, researchers may mark or take notes of the key words (bolded in extracts A to H), such as: "to memorize" (in A), "to pass the exams" (in B), "acquisition of…knowledge" (in C), which reveal some distinct features of the conceptions. In the next stage, by comparing and contrasting these features, the responses sharing similar features, such as A and G (memorizing), B and D (preparing for tests), C and E (acquiring knowledge), and F and H (applying) can be grouped to form an initial set of categories. Then researchers can start to describe each category by paying attention to the marked key words in the responses. For instance, in the category made up by A and G, "to memorize" appears to be the main theme, which conceives learning science as memorizing different things, including "special terms," "formula," "concepts," "facts," "symbols," and "equations." Therefore, possible labels for this category could be "learning science is to memorize," "learning science is a process of memorization," or "learning science involves memorizing many things."

Communicating Results in Phenomenographic Research

The phenomenographic results are presented as an outcome space, which is defined as a "logically structured complex" (Marton, 2000, p. 105), "a diagrammatic representation" (Bruce, 1997, p. 87), and "a map of a territory" (Säljö, 1988, p. 44). The outcome space has two essential elements: descriptions of each category and selections of illustrative statements accompanying each category (Marton, 1994; Bowden, 2000). The outcome space can be represented in various formats, such as in tables, in diagrams, or in figures (Yates et al., 2012). Corresponding to the structural relationship between the categories, three types of outcome space are recommended in phenomenographic data presentation. The most common type is a hierarchically inclusive outcome space, in which the categories are arranged from lower-order to higher-order categories, and the lowest level represents

the most simplistic way, whereas the highest level indicates the most sophisticated and developed way of experiencing the phenomenon (Tight, 2016). The outcome space can also be arranged chronologically (temporal ordering), which denotes the evolution of the participants' experience of a phenomenon (Englund et al., 2017). The outcome space presented in a climatic order is adopted when the categories are arranged according to the level of the explanatory power (Laurillard, 1993).

In the following, we present a sample outcome space of "conceptions of learning science" (adapted from Tsai, 2004) (**Table 2**) and discuss the structural and referential aspects of the categories (**Table 3**).

As shown in **Table 2**, there are seven qualitatively different ways of learning science conceived by high school students. Structurally, these categories are hierarchically related, with "memorizing" as the most simplistic conception and "seeing in a new way" as the most sophisticated one in the hierarchy. The level of sophistication increases as the categories move from 1 to 7. Referentially, the categories offer qualitatively different meaning in three dimensions. In terms of forms of knowledge acquisition and standards for evaluation of outcomes, there is a marked shift between categories 1–4 and categories 5–7. While categories 1–4 consider the value for learning science is knowledge reproducing and use the quantity to evaluate learning outcomes; categories 5–7 conceive learning science as applying theories to solve real life problems and providing new perspectives to understand the nature, and these categories are more concerned with the quality of learning.

USING THE PHENOMENOGRAPHIC METHOD TO TACKLE CHALLENGES IN SCIENCE EDUCATION

As outlined in the introduction that the phenomenographic method are suitable to tackle some current international challenges in science education, this section will explain these in detail. First, the phenomenographic method is a good way to evaluate students' understanding of scientific concepts and identify sources of misunderstanding because the phenomenographic data not only offer rich and contextual descriptions of students' understanding but are able to unpack a holistic understanding into "different patterns of awareness and non-awareness of component parts" (Åkerlind, 2018, p. 3), which allows the sources of misconceptions to be revealed more easily (Newton and Martin, 2013; Svensson, 2016). Educators can ask students to talk about a scientific concept and audio-record the answers, or they can ask students to write down their understanding. Then the educators can pinpoint the source of misunderstanding following the procedure we have described in "analyzing phenomenographic data." Once these sources are found, teachers may group students according to categories of misunderstandings, and present different information to different groups of students by highlighting the parts which they are unaware of or directly explain the structural relations between the parts, depending on the sources

TABLE 2 | An outcome space of conceptions of learning science.

Categories	Descriptions	Representative statements
1. Memorizing	Memorizing definitions, formulae, laws, and special terms	*When learning science, I need to memorize many concepts, facts, symbols, and equations. Sometimes, I feel that I am learning social studies such as history and language while learning science. There are often a lot to be remembered. I often need to rehearse these concepts and equations again and again to keep them strictly in my mind.*
2. Testing	Preparing for tests, passing the examinations or achieving high scores	*The major purpose of learning science is to pass the exams and have high exam scores, and then get into good colleges.*
3. Calculating	Calculating, practicing tutorial problems, and manipulating formulae and numbers	*Learning science involves the application or rearrangement of certain formulae to compute a right answer.*
4. Increase	Knowledge acquisition and accumulation of scientific knowledge	*Learning science indicates the acquisition of scientific knowledge. I have more knowledge derived from science instruction.*
5. Applying	Applying scientific knowledge in practical situations	*Learning science is to acquire some knowledge and skills to solve real-life problems. Science needs to be applied to solve practical problems.*
6. Understanding	Obtaining deep understanding, and constructing integrated and theoretically consistent knowledge	*Learning science needs a deep understanding of scientific knowledge. If you do not really understand, you will encounter a lot of conflict. And, you will not make sense of its concepts.*
7. Seeing in a new way	A process to get a new perspective and a new way to interpret natural phenomena	*Learning science brings new ways to see natural phenomena for me. Often, the scientific knowledge challenges my intuitions, and I finally know that I was incorrect in seeing something.*

TABLE 3 | Structural relations amongst categories of conceptions of learning science.

Categories	Forms of knowledge acquisition		Motivational orientation		Evaluation of learning outcomes	
	Reproducing	Extending and developing	External	Internal	How much is learnt	How well it is learnt
1. Memorizing	x		x		x	
2. Testing	x		x		x	
3. Calculating	x		x		x	
4. Increase	x			x	x	
5. Applying		x		x		x
6. Understanding		x		x		x
7. Seeing in a new way		x		x		x

of misunderstandings. Using an example from Lo and Chik (2016) to illustrate, in assessing students' understanding of an astronomical occurrence – solar eclipses, students were asked if it is possible that solar eclipses occurred 12 times in a year. Student A responded that she thought that it is possible because the Moon travels around the Earth once a month, that is 12 times in a year, therefore, the Moon should block the Sun from the Earth 12 times in a year, producing 12 solar eclipses. This response reveals that the source of misunderstanding of the solar eclipses formation is her unawareness of the critical feature that "the Moon has an orbit that is tilted at an angle to the plane of the Earth's orbit" (p. 301). Once this is identified, the teacher can highlight this critical feature that the orbit of the Moon is tilted to the Earth's orbit in the instruction or in the learning activities. This will clarify students' misconceptions that when the Moon is in between the Earth and the Sun, they are always on the same straight line.

Second, the phenomenographic method can be applied in science teaching to create facilitative conditions for learning difficult and abstract scientific concepts. The application of the phenomenographic method in instructional design is known as the variation theory of learning (Pang and Marton, 2013). It recognizes the qualitative variations of people's experience and interpretation of phenomena. In applying variation theory to instruction, the general principle is to introduce the variation of a critical aspect(s) of an object of learning (e.g., a scientific concept) to enable learners to discern and focus on this aspect while keeping the other aspects (the unfocused aspects) invariant (Pang and Ki, 2016). In this process, the phenomenographic data analysis can be used to identify "the critical features and aspects, relevance structure, and patterns of variation" for the object of learning (Lo and Chik, 2016, p. 296). Phenomenographic research has identified four patterns of variation: namely *separation*, *contrast*, *generalization*, and *fusion* (Marton, 2015). Using these patterns, science teachers can

manipulate the conditions of how information is presented to students in different ways to draw students' attention to the critical aspect(s) that the students need to discern in order to learn a scientific concept. The instructors can separate the critical aspects and the non-critical aspects of a concept (*separation*); keep some critical aspects of a concept invariant (*generalization*) while another varies (*contrast*); clarify the interrelationships amongst the critical aspects and the part-whole relationships within a concept (*fusion in the internal horizon*); and delineate the relationship between a concept and its background (*fusion in the external horizon*) (Lo and Chik, 2016). Pang and Ling (2012), for example, described an instructional design, which aimed to help secondary school students understand an important chemical concept – "whether the volume of the reactant or the concentration level of the reactant affects the rate of a chemical reaction." It showcased how chemistry teachers kept some critical aspects invariant while another varied in two sets of experiments. In the first set, the mass of $CaCO_3$ was kept invariant, the concentration of acid was kept invariant, but the volume of acid varied. These experiments helped students discern that "the volume of the reactant does not affect the rate of the chemical reaction when the concentration level of the reactant remains the same." In the second set of experiments, the mass of $CaCO_3$ was kept invariant, the volume of acid was kept invariant, but the concentration of acid varied. The second set of experiments enabled students to discern that "the concentration level of the reactant affects the rate of the chemical reaction even though the volume of the reactant remains the same."

Third, to improve the quality of science learning, another issue for science educators to deal with is continuous identification of factors (e.g., students' approaches to learning science, and how students perceive science learning environment) which contribute to the learning outcomes (Hardy et al., 2014). Past research in science education has consistently demonstrated that qualitatively different conceptions of learning science are logically related to how students go about learning it, and levels of learning outcomes (Minasian-Batmanian et al., 2006). These studies reported that students who hold fragmented conceptions of learning science tend to adopt more surface approaches, and achieve relatively poorly; whereas those with cohesive conceptions are more likely to adopt deep approaches to learning science, and have relatively better academic performance. The phenomenographic method can be used to identify variations of other factors in students' learning experience, such as students' perceptions of the course design, students' understanding of laboratory experiments, and students' approaches to teamwork and collaborations. This may help science educators decide which factor(s) they should act upon to move students from less undesirable to more desirable variation of learning experience in order to enhance their learning outcomes in science subjects.

For example, once teachers find that some students believe that learning science does not have any practical applications in everyday life at all, and that science learning is merely rote memorization of scientific formula without needing to

understand the principles behind them, teachers may try to help students change such fragmented conceptions and relate science learning to solving real life issues. Teachers may design learning activities for students to conduct scientific investigation of practical problems in their lives and local communities related to a class theme, such as "Does the weather affect your pulse?", "Which soil is the best growing medium?", "Does exercise improve your memory?" (Forbes and McCloughan, 2010; Forbes and Skamp, 2019). Through participation in authentic scientific activities, students will become more engaged in every process of scientific inquiry, including observing phenomena related to personal and societal contexts, questioning, predicting, testing, collecting, analyzing, reasoning and arguing, so that they may start to value scientific investigation in finding real life solutions and appreciate the beauty of scientific reasoning. It is through the phenomenographic method, which is concerned with the interplay between a phenomenon and people who experience it, science educators are able to continuously locate and modify undesirable variations of learning experience to help students learn science better.

CONCLUSION

The purpose of this article has been to introduce readers to the phenomenographic research method, which can be usefully designed to tackle contemporary challenges in science education. A fundamental purpose of the method is to describe people's collective experience of the world and variations in that collective experience. This is particularly useful for educators interested in understanding why some students learn more deeply and successfully than others, even though they all experience the same course assessment and activities. In order to provide science educators a theoretical appreciation of the method and capacity to implement it in practice, we have described the origin and development of the phenomenographic framework, including its ontological and epistemological assumptions, and its unique second-order perspective. We have then illustrated the key procedures of conducting phenomenographic research using examples. The article continues with an account of how the method can be applied to: (1) identify sources of students' misunderstanding of scientific concepts; (2) implement effective instructional design for teaching difficult and abstract scientific concepts; and (3) locate actionable elements in student experience of learning science which are likely to impact on quality of learning outcomes. The number of research studies adopting the phenomenographic method has been growing rapidly in science education (Tight, 2016), hence, we hope this paper can serve as a primer to implement phenomenography in educational practice to improve science learning of students.

DATA AVAILABILITY

No datasets were generated or analyzed for this study.

AUTHOR CONTRIBUTIONS

Both authors made substantial contribution to the conception of the work, drafting the work and revising it critically for important intellectual content, approving the final version of the manuscript to be published, and agreeing to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

Åkerlind, G. (2005). Variation and commonality in phenomenographic research methods. *High. Educ. Res. Dev.* 24, 321–334. doi: 10.1016/j.ijnurstu.2015.10.004

Åkerlind, G. (2018). What future for phenomenographic research? On continuity and development in the phenomenography and variation theory research tradition. *Scand. J. Educ. Res.* 62, 949–958. doi: 10.1080/00313831.2017.1324899

Åkerlind, G., Bowden, J., and Green, P. (2005). "Learning to do phenomenography: a reflective discussion," in *Doing Developmental Phenomenography*, eds J. Bowden and P. Green (Melbourne: RMIT University Press), 74–102.

Alberts, B. (2013). Prioritizing science education. *Science* 340, 249–249. doi: 10.1126/science.1239041

Ashworth, P., and Lucas, U. (2000). Achieving empathy and engagement: a practical approach to the design, conduct and reporting of phenomenographic research. *Stud. High. Educ.* 25, 295–308. doi: 10.1080/713696153

Booth, S. (1997). On phenomenography, learning and teaching. *High. Educ. Res. Dev.* 16, 135–158. doi: 10.1080/0729436970160203

Bowden, J. (2000). "The nature of phenomenographic research," in *Phenomenography*, eds J. A. Bowden and E. Walsh (Melbourne: RMIT University), 1–18.

Brown, P. L., Abell, S. K., Demir, A., and Schmidt, F. J. (2006). College science teachers' views of classroom inquiry. *Sci. Educ.* 90, 784–802. doi: 10.1002/sce.20151

Bruce, C. (1997). *The Seven Faces of Information Literacy*. Adelaide: Auslib Press.

Case, J., and Light, G. (2011). Emerging methodologies in engineering education research. *J. Eng. Educ.* 100, 186–210. doi: 10.1002/j.2168-9830.2011.tb00008.x

Chiu, Y., Lin, T., and Tsai, C. C. (2016). The conceptions of learning science by laboratory among university science-major students: qualitative and quantitative analyses. *Res. Sci. Technol. Educ.* 34, 359–377. doi: 10.1080/02635143.2016.1222518

Cope, C. (2004). Ensuring validity and reliability in phenomenographic research using the analytical framework of a structure of awareness. *Qual. Res. J.* 4, 5–18.

Dahlgren, L., and Fallsberg, M. (1991). Phenomenography as a qualitative approach in social pharmacy research. *J. Soc. Adm. Pharm.* 8, 150–156.

Dall'Alba, G. (1994). "Reflections on some faces of phenomenography," in *Phenomenographic Research: Variations in Method: The Warburton Symposium*, eds J. Bowden and E. Walsh (Melbourne: RMIT University Press), 73–88.

Englund, C., Olofsson, A., and Price, L. (2017). Teaching with technology in higher education: understanding conceptual change and development in practice. *High. Educ. Res. Dev.* 36, 73–87. doi: 10.11124/jbisrir-2015-1919

Forbes, A., and McCloughan, G. (2010). Increasing student participation in science investigations in primary schools: the MyScience initiative. *Teach. Sci.* 56, 24–30.

Forbes, A., and Skamp, K. (2019). 'You actually feel like you're actually doing some science': primary students' perspectives of their involvement in the myscience initiative. *Res. Sci. Educ.* 49, 465–498. doi: 10.1007/s11165-017-9633-3

Gordon, S., and Nicholas, J. (2015). What do bridging students understand by 'assumed knowledge' in mathematics? *Int. J. Innov. Sci. Math. Educ.* 23, 10–20.

Hardy, J., Hancock, S., Johnson, O., Galloway, R., Wallace, M., Draper, S., et al. (2014). *Fostering Learning Improvements in Physics*. Available at: http://eprints.whiterose.ac.uk/117465/1/FLIP_final_report_jan_final.pdf (accessed May 5, 2019)

Howitt, S. M., and Wilson, A. N. (2018). Reflecting on the use and abuse of scientific data facilitates students' ethical and epistemological development. *Sci. Educ.* 102, 571–592. doi: 10.1002/sce.21333

Hoyle, P. (2016). Must try harder: an evaluation of the UK government's policy directions in STEM education. *Paper presented at ACER Research Conference: Improving STEM learning: What will it take?* (Melbourne: Australian Council for Educational Research).

Hsieh, W., and Tsai, C. C. (2017). Taiwanese high school teachers' conceptions of mobile learning. *Comput. Educ.* 115, 82–95. doi: 10.1016/j.compedu.2017.07.013

Kapucu, S. (2014). University students' conceptions of the relationship between mathematics and physics and the relationship between mathematics and physics learning. *J. Baltic Sci. Educ.* 13, 625–636.

Lam, H. (2017). Using phenomenography to investigate the enacted object of learning in teaching activities: the case of teaching chinese characters in hong kong preschools. *Scand. J. Educ. Res.* 61, 169–186. doi: 10.1080/00313831.2015.1119728

Laurillard, D. (1993). *Rethinking University Teaching: A Framework for the Effective Use of Educational Technology*. London: Routledge.

Lee, M. H., Lin, T. J., and Tsai, C. C. (2013). Proving or improving science learning? Understanding high school students' conceptions of science assessment in Taiwan. *Sci. Educ.* 97, 244–270. doi: 10.1002/sce.21046

Lo, M., and Chik, P. (2016). Two horizons of fusion. *Scand. J. Educ. Res.* 60, 296–308.

Magana, A., Brophy, S., and Bodner, G. (2012). Instructors' intended learning outcomes for using computational simulations as learning tools. *J. Eng. Educ.* 101, 220–243. doi: 10.1002/j.2168-9830.2012.tb00049.x

Marton, F. (1986). Phenomenography – A research approach to investigating different understandings of reality. *J. Thought* 21, 28–49.

Marton, F. (1988). "Phenomenography: exploring different conceptions of reality," in *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*, ed. D. M. Fetterman (New York, NY: Praeger), 176–205.

Marton, F. (1994). "On the structure of awareness," in *Phenomenographic Research: Variations in Method*, eds J. Bowden and E. Walsh (Melbourne: RMIT University), 89–100.

Marton, F. (2000). "The structure of awareness," in *Phenomenography*, eds J. Bowden and E. Walsh (Melbourne: RMIT University), 102–116.

Marton, F. (2015). *Necessary Conditions of Learning*. New York, NY: Routledge.

Marton, F., and Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Erlbaum Associates.

Marton, F., Carlsson, M., and Halasz, L. (1992). Differences in understanding and the use of reflective learning in reading. *Br. J. Educ. Psychol.* 62, 1–16. doi: 10.1111/j.2044-8279.1992.tb00995.x

Marton, F., and Pang, M. (2008). "The idea of phenomenography and the pedagogy of conceptual change," in *International Handbook on Research of Conceptual Change*, ed. S. Vosniadou (New York, NY: Routledge), 533–559.

Marton, F., and Pong, W. (2005). On the unit of description in phenomenography. *High. Educ. Res. Dev.* 24, 335–348. doi: 10.1080/07294360500284706

Marton, F., and Säljö, R. (1976a). On qualitative differences in learning I – Outcome and process. *Br. J. Educ. Psychol.* 46, 4–11. doi: 10.1111/j.2044-8279.1976.tb02980.x

Marton, F., and Säljö, R. (1976b). On qualitative differences in learning II - Outcome as a function of the learner's conception of the task. *Br. J. Educ. Psychol.* 46, 115–127. doi: 10.1111/j.2044-8279.1976.tb02304.x

Marton, F., and Svensson, L. (1979). Conceptions of research in student learning. *High. Educ.* 8, 471–486. doi: 10.1007/bf01680537

McCosker, H., Barnard, A., and Gerber, R. (2004). A phenomenographic study of women's experiences of domestic violence during the childbearing years. *Online J. Issues Nurs.* 9, 76–88.

Minasian-Batmanian, L., Lingard, J., and Prosser, M. (2006). Variation in student reflections on their conceptions of and approaches to learning biochemistry in a first-year health sciences' service subject. *Int. J. Sci. Educ.* 28, 1887–1904. doi: 10.1080/09500690600621274

National Research Council (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. Washington, DC: National Academies Press.

Newton, G., and Martin, E. (2013). Blooming, SOLO taxonomy, and phenomenography as assessment strategies in undergraduate science education. *J. Coll. Sci. Teach.* 43, 78–90.

Olympiou, G., and Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: an effort to improve students' conceptual understanding through science laboratory experimentation. *Sci. Educ.* 96, 21–47. doi: 10.1002/sce.20463

Osborne, J., Henderson, J., MacPherson, A., Szu, E., Wild, A., and Yao, S. (2016). The development and validation of a learning progression for argumentation in science. *J. Res. Sci. Teach.* 53, 821–846. doi: 10.1002/tea.21316

Pang, M. F., and Ki, W. W. (2016). Revisiting the idea of "critical aspects". *Scand. J. Educ. Res.* 60, 323–336. doi: 10.1080/00313831.2015.1119724

Pang, M. F., and Ling, L. M. (2012). Learning study: helping teachers to use theory, develop professionally, and produce new knowledge to be shared. *Instr. Sci.* 40, 589–606. doi: 10.1007/s11251-011-9191-4

Pang, M. F., and Marton, F. (2013). Interaction between the learners' initial grasp of the object of learning and the learning resource afforded. *Instr. Sci.* 41, 1065–1082. doi: 10.1007/s11251-013-9272-7

Prosser, M., and Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Buckingham: SRHE and Open University Press.

Richardson, J. (1999). The concepts and methods of phenomenographic research. *Rev. Educ. Res.* 69, 53–82. doi: 10.3102/00346543069001053

Säljö, R. (1979). Learning about learning. *High. Educ.* 8, 443–451.

Säljö, R. (1988). "Learning in educational settings: Methods of enquiry," in *Improving Learning: New Perspectives*, ed. P. Ramsden (London: Kogan), 32–48.

Säljö, R. (1997). Talk as data and practice – A critical look at phenomenographic inquiry and the appeal to experience. *High. Educ. Res. Dev.* 16, 173–190. doi: 10.1080/0729436970160205

Sin, S. (2010). Considerations of quality in phenomenographic research. *Int. J. Qual. Methods* 9, 305–319. doi: 10.1186/1471-2296-10-4

Stenfors-Hayes, T., Hult, H., and Dahlgren, M. (2013). A phenomenographic approach to research in medical education. *Med. Educ.* 47, 261–270. doi: 10.1111/medu.12101

Svensson, L. (2016). Towards an integration of research on teaching and learning. *Scand. J. Educ. Res.* 60, 272–285. doi: 10.1080/00313831.2015.1120233

Teeter, P., and Sandberg, J. (2016). Constraining or enabling green capability development? How policy uncertainty affects organizational responses to flexible environmental regulations. *Br. J. Manage.* 28, 649–665. doi: 10.1111/1467-8551.12188

Tight, M. (2016). Phenomenography: the development and application of an innovative research design in higher education research. *Int. J. Soc. Res. Methodol.* 19, 319–338. doi: 10.1080/13645579.2015.1010284

Tsai, C. C. (2004). Conceptions of learning science among high school students in Taiwan: a phenomenographic analysis. *Int. J. Sci. Educ.* 26, 1733–1750. doi: 10.1080/0950069042000230776

Vermunt, J. D., and Donche, V. (2017). A learning patterns perspective on student learning in higher education: state of the art and moving forward. *Educ. Psychol. Rev.* 29, 269–299. doi: 10.1007/s10648-017-9414-6

Yates, C., Partridge, H., and Bruce, C. (2012). Exploring information experiences through phenomenography. *Libr. Inform. Res.* 36, 96–119. doi: 10.29173/lirg496

Large Scale Scientific Modeling Practices That Can Organize Science Instruction at the Unit and Lesson Levels

Maria Cecilia Nunez-Oviedo[1]* and John J. Clement[2]

[1] *Departamento de Curriculum e Instrucción, Universidad de Concepción, Concepción, Chile,* [2] *College of Education, University of Massachusetts Amherst, Amherst, MA, United States*

Science educators today still struggle with finding better ways to help students develop strong conceptual understandings as opposed to memorizing isolated facts. Recently there has been increased attention on learning explanatory models as a key to conceptual understanding. Science educators also struggle with how to teach students scientific thinking practices, and sometimes this goal is seen as being in competition with content goals for conceptual understanding. In this study we ask whether whole class discussion can contribute to both of these goals at the same time and whether there are ways that a teacher can support this. We describe the results of a case study of an experienced teacher leading modeling discussions in a series of three middle school life sciences classes. A qualitative microanalysis of the videotaped whole class discussions led to the identification of a variety of modeling processes operating across the lessons at two different time scale levels. These include model competition, in which students compare and evaluate their models, and model evolution, in which the models go through stepwise evaluation and improvement. The latter process involves a smaller time scale pattern of model generation, evaluation, and modification cycles. All of these processes are similar to those found in recent studies of practices of expert scientists. Implications from the case study suggest that: (1) A teacher need not be limited to the two opposing interaction styles of Open Discussion vs. Authoritative lecture. Rather, there are there intermediate discussion styles between these that involve co-construction and cognitive scaffolding; (2) It is possible to start from student-generated models that conflict with the target model in a number of ways, and still arrive at the target model for the lesson through discussion. Processes of model competition and disconfirmation, as well as model evolution, both supported by the teacher's cognitive scaffolding, were central in this accomplishment; (3) In doing so, it is possible for a teacher to foster student modeling practices, as a type of scientific thinking, at the same time that they are teaching science content, by scaffolding the two levels of model construction processes identified.

Keywords: science learning, science teaching, classroom discussion, scaffolding, scientific thinking, scientific practices, modeling

INTRODUCTION

We are encouraged by current educational reform ideas such as emphases on teaching for scientific thinking and teaching for conceptual understanding, as well as the use of small group and whole class discussions that draw out student ideas and thinking. Our particular focus in this article is on how whole class discussions can contribute to the learning of conceptual models in science, as well as fostering modeling practices as a central form of scientific thinking.

However, in our experience in educating preservice and in-service teachers, we have become sensitive to several tensions or dilemmas experienced by them in engaging such reforms, even for those who have accepted the desirability of teaching both scientific thinking practices and models, as a form of disciplinary content goals, in their syllabus. We will distinguish three types of tension (see **Figure 1**):

Conceptual Dissonance Tension: In trying to utilize open whole class discussion to tap into and start from students' ideas for explaining scientific phenomena, they uncover some useful ideas but also uncover some ideas that are in conflict with their target model for the lesson (Scott et al., 2006, 210).
Lesson Objective Tension: Since discussions for scientific thinking can be time consuming and many are under pressure to cover content, there is a perceived tension or competition between science content goals and scientific thinking goals.
Tension Between Opposing Teaching Approaches: When they try to pursue both of the latter goals, they tend to associate inquiry methods such as open discussion with thinking goals, and there is a tension about not knowing when to use open discussion and when to use a more authoritative approach (e.g., lecture; Scott et al., 2006, p. 606).

In this paper we attempt to identify practices occurring in productive whole class discussions by conducting a case study of an experienced science teacher who appears to have found a method that resolves or reduces these tensions. She appears to use several different modes of discussion and to have ways of cognitively scaffolding student thinking during critical pieces of the discussion. Our goal is to identify new ways of describing the most important classroom interactions and processes that take place in this class sequence.

Doing a qualitative microanalysis of discussions over three consecutive lessons has the potential to develop such a set of descriptive concepts along with developing a multi-level framework of model construction processes, in order to provide a foundation for broader studies. The challenge is to identify the set of model construction processes that allow students to participate in building a complex scientific model (In this article we will use "scientific (thinking) processes" and "scientific practices" as synonyms).

PREVIOUS LITERATURE AND THEORETICAL FRAMEWORK
Important Previous Work on Tensions

In an important precedent for our study, Scott et al. (2006, p. 606) point to a growing interest in studying the role of discussion or discourse in the science classroom. They cite curriculum initiatives based on student inquiry and argumentation, where discussion appears to be central in drawing out student ideas. They conclude that discussion is key for relating new concepts to the student's everyday prior knowledge. On the other hand, they conclude that the teacher is key in guiding and supporting classroom discussion as s/he takes into account students' perspectives but also has the goal of arriving at targeted models for a unit. But they also acknowledge that discussion is still notably absent from science classrooms around the world. We will take these as assumed starting points for the present study.

They also set out to study classroom interactions in small group and whole-class discussions. They conducted a classroom case study of an experienced teacher leading four lessons designed to develop 14–15 years old students' models of heat transfer and temperature change. They observed classes alternating between what they called dialogical (open discussion) and authoritative (teacher centered) passages of interaction. In other words, there is a "turning point" in the flow of discourse as the teacher brings together everyday and scientific views and makes an authoritative case for the scientific view by making a direct juxtaposition of everyday and scientific views. They stated: "We see a tension between authoritative and dialogic approaches as being an inevitable characteristic of meaning making interactions in the science classroom" (Scott et al., 2006, p. 606). We represent this idea in the top half of **Figure 1** as a tension between two opposite approaches or teaching styles. An implication is that it may be difficult for a teacher to decide when to use each approach (For reasons of space, we do not present all aspects of their more complex framework, but focus on those relevant here). They see skillful transitions between dialogic and authoritative interactions as being fundamental to support meaningful learning of disciplinary knowledge. However, they point out that in many typical classrooms, dialogic discussion can tend "to fade out altogether" and teachers need strategies in order to prevent this.

Remaining Gaps
However, which teaching approach to use at different times often cannot be mapped out in advance by the teacher, since it depends on the interests and concerns of the students. Finding principles for doing this is one of the objectives of the present paper. Scott et al. (2006, p. 607) indicate that the skillful shifts in teaching approaches resonate with the principles of "productive disciplinary engagement" (Engle and Conant, 2002, p. 400-401), "accountable talk" (Resnick, 1999, p. 40), and "reflective discourse" (van Zee and Minstrell, 1997a, p. 209–210).

We value Scott et al.'s description of the need for both discussion and authoritative input as contrasting approaches to teaching. However, the gap between open discussion and authoritative approaches seems very wide, as represented in **Figure 1**. In this paper we ask whether there are other productive approaches in between these that may reduce that tension.

The lower half of **Figure 1** also represents the other two tensions we identified above. Studies such as Clement (1993) and Minstrell and Kraus (2005) have documented useful ideas that students bring into the science classroom. But it is well known that opening the classroom to students' naive ideas in science can

FIGURE 1 | Tensions in classroom interaction approaches and in products.
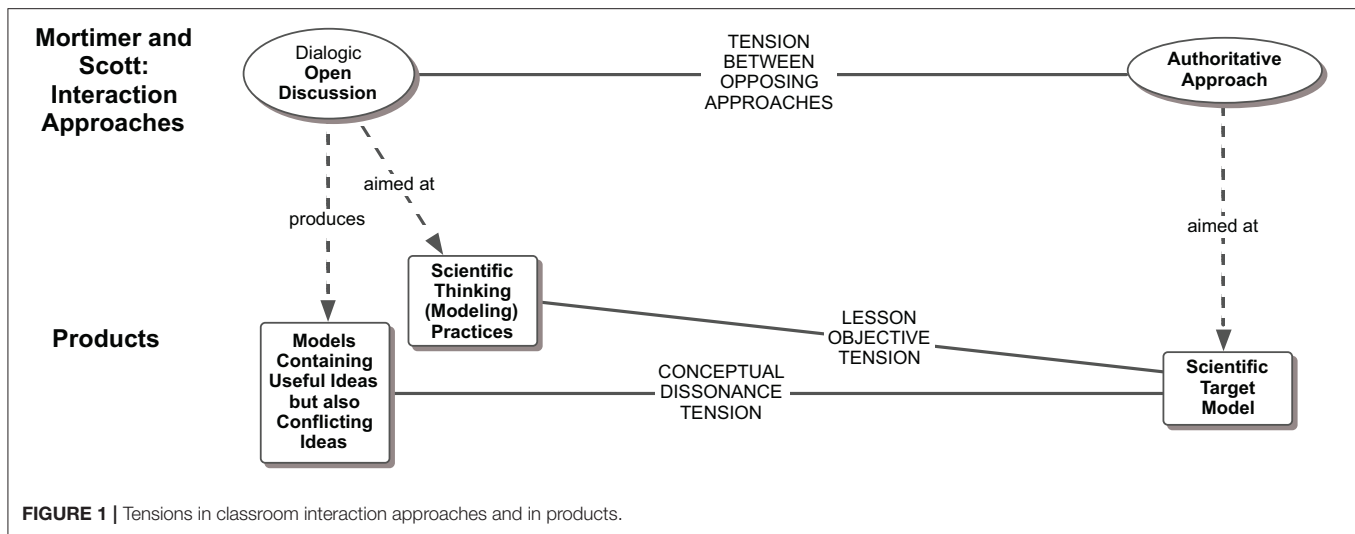
also lead to faulty alternative models. If not dealt with, these can conflict with the target models of a unit. We call this a Conceptual Dissonance Tension.

In addition, since discussions for scientific thinking can be time consuming, and with teachers under pressure to cover content, there can be a tension between pursuing science content goals and scientific thinking goals. We call this a Lesson Objective Tension. We will also ask how these other two tensions in **Figure 1** can be dealt with.

Models in the History of Science

In order to discuss prior work on modeling in science classrooms, we will first say how we are using the term "model." Campbell (1920) and Harré (1961) argued that scientists often think using theoretical *explanatory models*, such as neutrons, electro-magnetic waves, and fields that are a separate kind of hypothesis from empirical laws, and we will focus on using the term model in this way. Harré proposed that there are four types of knowledge used in science and placed them in a continuum from the more empirical toward the more theoretical: (1) Observations; (2) Empirical law hypotheses (mathematical and verbal descriptions of patterns in the observations); (3) Explanatory models; and (4) Formal principles. Explanatory models then are conjectured theoretical hypotheses providing a picture of a hidden description or process that explains why a pattern in observations occurred. Thus, an explanatory model is not simply a condensed summary of empirical observations but an invention of new theoretical terms and images that are not "implied by" the data. The scientist is free to generate such models via conjecture or analogies or other means, but to survive, the model also needs to be evaluated with respect to a number of criteria, such as empirical testing, simplicity, aesthetic appeal, and consistency with other accepted models.
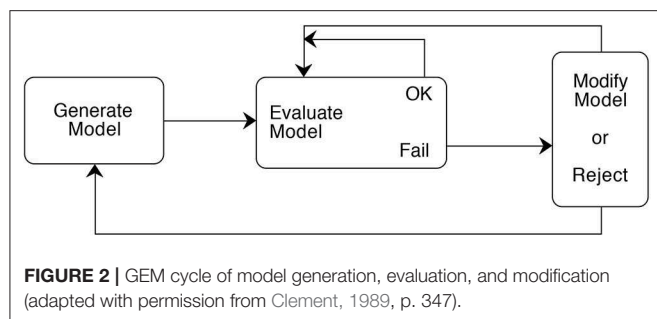
Machamer et al. (2000) used the term "mechanisms" with a meaning similar to "explanatory models" and described how scientific disciplines are multi-layered (e.g., subatomic, atomic, molecular layers). This can apply to how educators organize a

subject to be taught via hierarchical substructures. For example, the concept of blood can be modeled as blood cells and plasma. But one can go further to unpack the structure of blood cells into parts of the cell and their functions, which can be thought of as starting a new cycle of modeling at a lower level, producing nested layers of models.

Kuhn (1970) indicated that the advance of scientific knowledge is the result of a revolutionary process by which an older theory or paradigm is rejected and replaced by an incompatible new one. The process begins when the old paradigm cannot solve new problems and anomalies accumulate producing a crisis in the scientific community that turns to search for new paradigms. Kuhn argues that these paradigms compete until one survives because either most of the researchers are converted to the newer paradigm or are removed by attrition.

On the other hand, Toulmin (1972, p. 202–204) indicated that the advance of scientific knowledge is often the result of an evolutionary process that is comparable to Darwin's natural selection theory. The process involves two steps: selection and innovation. Starting from an existing model, the selection part involves a process of critical evaluation and debate about problems with the model, while the innovation part involves originating new conceptual modifications of the model while until leaving in place the best parts of the model. In this way science can evolve better and better models. This contrast between the views of Kuhn and Toulmin on learning in scientist raises the interesting question of whether scientific practices students engage in classrooms are better thought of as revolutionary or evolutionary.

Nersessian (1995, 2008) and Darden (1991) have conducted important historical studies of the modeling practices of experts in the domains of electro-magnetic theory and genetics, respectively. Clement (1989, 2008a) conducted think aloud studies of problem solving and modeling practices of modern scientists. All three found that experts use a variety of reasoning processes such as analogies, discrepant events, imagistic simulation, and thought experiments. They also found

FIGURE 2 | GEM cycle of model generation, evaluation, and modification (adapted with permission from Clement, 1989, p. 347).

a larger pattern in which expert models can go through a series of successive refinements to produce a chain of better and better models. These are generated by cycles of model Generation, Evaluation and Modification (GEM) processes. Clement (1989, p. 347) used **Figure 2** to describe this cyclical process of creative generation of a hidden structure or process to account for the phenomenon, evaluation that can be rationalistic and/or empirical, and modification or rejection (termed a GEM Cycle). As shown in the diagram when the evaluation of a hypothesis is strongly negative, it may be completely disconfirmed. But if it is evaluated only somewhat negatively, it can be improved through modification. Such a cycle can generate a series of successively improved models. We will attempt to use this concise summary of basic central practices in science as one starting point hinting at how one might interpret modeling discussions in classrooms.

Recent Emphases on Models in Education

Authors early on such as Hestenes (1987), White and Frederiksen (2000), and Clement and Steinberg (2002) have emphasized the importance of learning models and modeling as central to the learning of science in schools. In addition there is recent work urging teachers to develop qualitative models during discussions, e.g., Krajcik et al. (2014, p. 163); Louca et al. (2012, p. 1845–1847); Reiser et al. (2012, p. 6–7); Schwarz et al. (2009, p. 640–643), and Windschitl et al. (2012, p. 884–885). They have emphasized that if we want students to learn modeling as a practice, we need to find ways to involve them in generating models. Indeed in the USA, the NGSS (2013) standards now emphasize modeling as a key scientific practice objective for teaching students to *do* science as well as learning content. That is, it not only calls for students to learn models, but dozens of its performance expectations call on students to *develop* models.

We will use the term "model based curriculum" to refer to a curriculum that was designed to focus on students making contributions to the construction of explanatory models as a foundation for conceptual understanding. Although they are not yet widely adopted at a national level, there have now been a number of model-based curricula developed at elementary, secondary levels for different subjects, such as: electricity (Capacitor-Aided System for Teaching and Learning Electricity (CASTLE) Curriculum, Steinberg and Wainwright, 1993 and Clement and Steinberg, 2002); mechanics (Preconceptions in Mechanics, Clement, 1993 and Camp et al., 1994); particle theory (Children's Learning in Science (CLIS) Research Group at the

University of Leeds in the UK, Driver and Scott, 1996); and life sciences (Energy in the Human Body Curriculum, EHBC, Rea-Ramirez, 1998 and Rea-Ramirez et al., 2004).
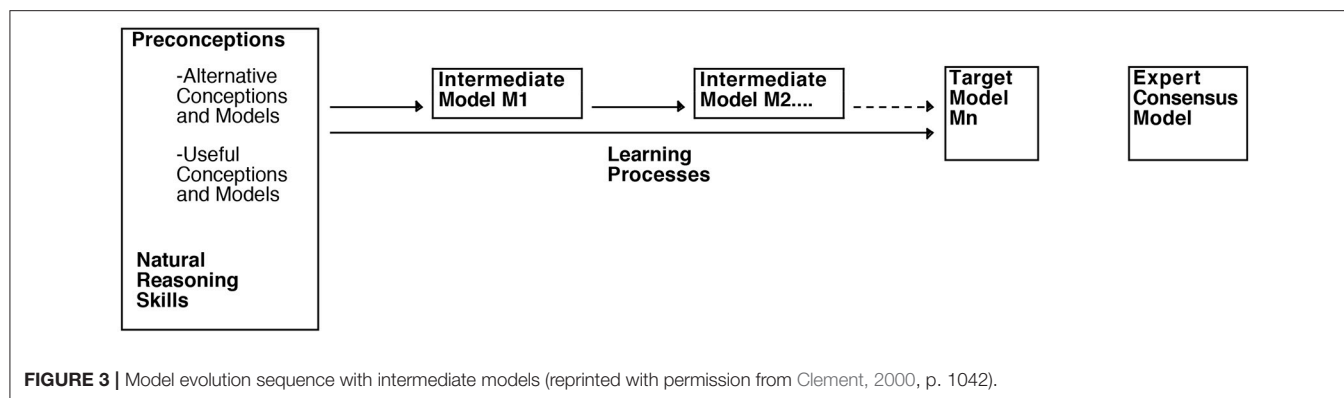
These curricula emphasize content learning via model construction and revision processes. They emphasize understanding the dynamic causal mechanisms, not just static structure, of qualitative explanatory models. However, from the teachers' and student's point of view these models are quite complex, and teaching and learning them can involve many conceptual steps. But all these curricula have the following common characteristics: (1) they begin from students' ideas and move toward a conceptual target; (2) instruction was conducted via both small and large group discussions; (3) large significant gains in students' understanding were measured from pre and post-tests; (5) the curricula were field tested in classrooms and revised over multiple years; (6) teachers contributed to the development and testing of the curricula.

Each of these curricula use a teaching sequence that contains up to six steps to support the modeling process but they each use different names for the steps. Clement (2008b) summarized broadly how these curricula use four common steps in supporting students' modeling, namely: Introducing Problems, Building Model Parts, Synthesis (Consolidation), and Application. These are candidates for organizing a high level sequence that may foster modeling processes through classroom interactions. However, these broad steps are still too rough for guiding teachers in moment-to-moment scaffolding of modeling practices. In particular, these four steps do not provide a detailed description of the modeling processes that are taking place within each section particularly within the Building Model Parts section which we consider to be the most challenging.

With respect to discussion leading strategies, van Zee and Minstrell (1997b), Hogan and Pressley (1997), and Chin (2007) have identified important moment-to-moment questioning strategies that teachers use to guide discussions. A few of these are cognitive strategies aimed at specific conceptual processes, but most are broader strategies designed to sustain dialog in general. One important role of discussions is to provide formative assessment feedback to the teacher Minstrell et al. (2011, p. 2–3). Other research groups have described modeling practices that appear to have a large time scale of 2–6 lessons (e.g., generating vs. consolidating ideas) (e.g., Driver and Scott, 1996, p. 99; Windschitl et al., 2012, p. 891 and Brewe, 2008, p. 1158) as well as smaller patterns that appear to occur in smaller 5–20 min segments in classrooms (e.g., using analogies, written records of discussion, and argumentation) (e.g., Hammer, 1995, p. 423–427 and Schwarz et al., 2009, p. 640–643).

Remaining Gaps, Purposes, and Plan for This Paper

While papers in the sections above have highlighted the importance of modeling and of drawing out and having students debate ideas in sustained discussions, they still have not provided a clear description of how students can engage in model construction as a teacher guides a discussion toward a target model, while navigating the tensions described in **Figure 1**. Some

FIGURE 3 | Model evolution sequence with intermediate models (reprinted with permission from Clement, 2000, p. 1042).

processes have been proposed, but we still lack ways of organizing them into a coherent "big picture" collection of (at least partially) ordered processes and subprocesses.

In the next section we will review some recent work on cognition in expert scientists as an important resource. We will be asking whether those descriptions apply in the classroom or not, and review some of our own group's previous work on this problem. We then formulate more specific research questions and attempt to apply these pieces along with ideas from other researchers to our case study of a series of classroom discussions.

Applying Expert Modeling Practices to Instruction

Clement (1989) and Nersessian (1995, 2008) argued that generation, evaluation and modification cycles are processes that need to take place in students who are learning to comprehend scientific models with conceptual understanding. But neither of these early studies provided guidelines for teachers and curriculum developers about how to guide instruction. In addition, they did not provide descriptions to explain how the processes of model construction and revision take place when multiple subjects are participating in the construction of the same mental model.

Clement (2000) proposed that for complex models too large or too unintuitive to learn all at once that it made sense for students to learn via GEM cycles in a sequence of steps from model M_n to model M_{n+1} (see **Figure 3**). The emergence of successive intermediate mental models is also called a "learning pathway" (Scott, 1992, p. 221). A learning pathway can be envisioned as a chain starting on the left from common misconceptions and possible positive conceptions to build on, and progressing to the right toward a target model for the unit that is usually a simplified version of the expert consensus model. In between are intermediate models that may be model elements or partial approximations. These have the potential to provide a more fine-grained guide to the teacher concerning what pathways of learning can make sense to the student and lead to deeper conceptual understanding.

Findings From Our Own Group on Modeling Practices in the Classroom

GEM cycles leading to model evolution sequences like those in **Figure 3** have been documented by our own group in classroom

teaching in chemistry Khan (2008), middle school life sciences Nunez-Oviedo et al. (2002, 2008) and Clement and Steinberg (2002) did so for high school electricity tutoring. They called this "teaching via Model Evolution," which meant fostering a series of "model criticisms and revisions" to parts of the students' models, often by using dissonance producing techniques and analogies. More recently, Williams and Clement (2015) described GEM cycles in high school physics discussions, and showed how they were supported by other smaller strategies like analogy and discrepant events.

Rea-Ramirez (1998) described the teacher's role as being constantly aware of the students' mental models so as to foster criticism and revision cycles. Thus, an important role of both small and whole class discussion is to allow the teacher to listen in deeply to students' points of view so that s/he is aware of student models. Co-construction is the process that occurs during the cooperative construction of a mental model through which the teacher and the students both contribute ideas to build and evaluate a model. Nunez-Oviedo (2004) developed models to explain the processes within co-construction.
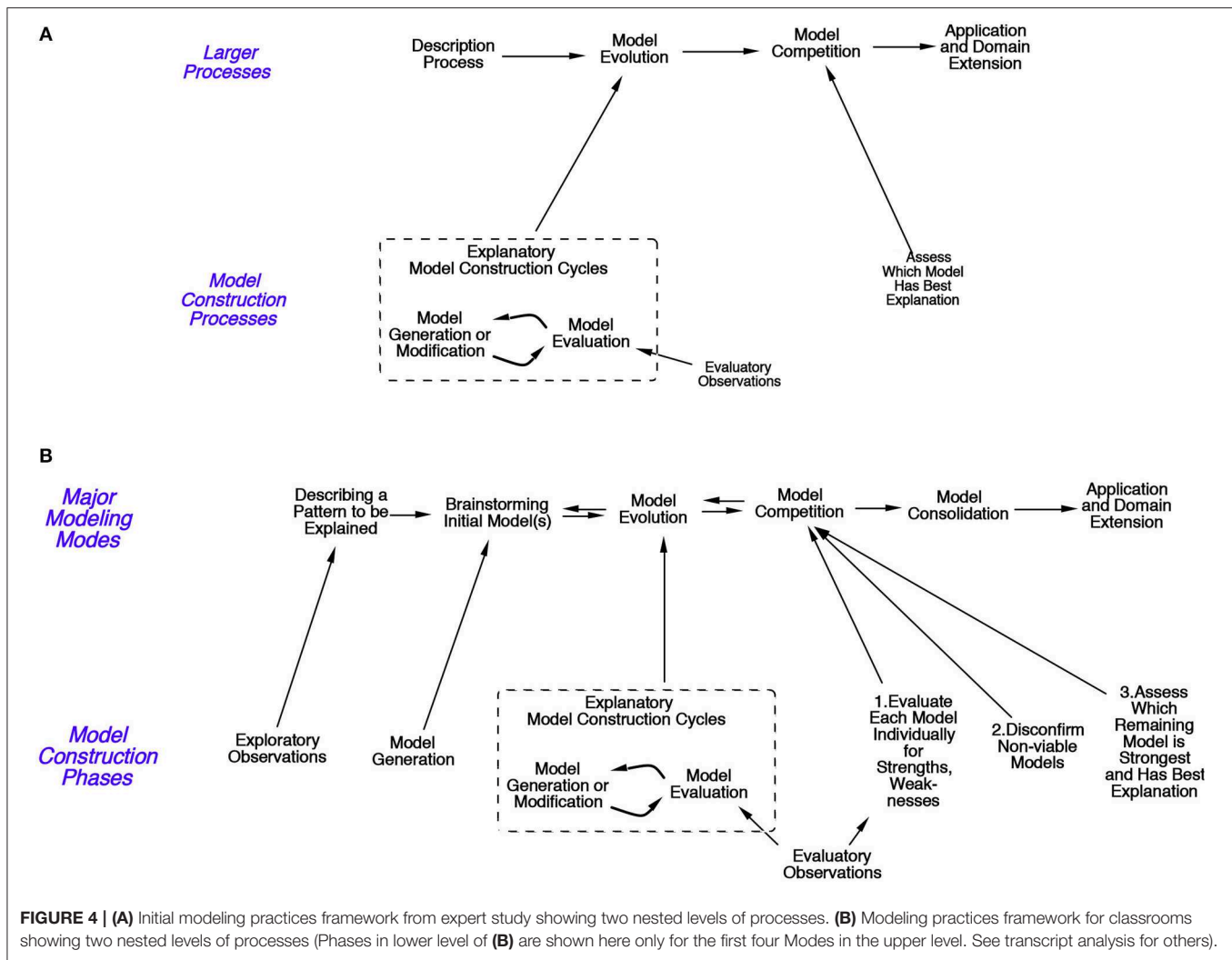
Work on Developing a Multi-Level Modeling Framework

Based on the previous findings (Clement, 2008b) proposed the generation of a larger organizing framework called "Multiple Time Scale Levels of Organization" that includes modeling strategies used by experts as well as the common strategies found in curricula. The framework includes six levels that reflect different time scales, ranging from those strategies operating over months to those operating over seconds. Lower-level strategies are then nested within higher-level strategies as follows:

Level F Curriculum unit integration strategies,
Level E Unit-sized modeling strategies,
Level D Lesson strategies,
Level C Single model element strategies,
Level B Individual cognitive strategies,
Level A Dialogical strategies.

The present study maps most closely to levels D and E in his scheme and seeks to identify detailed structuring and substrategies for scaffolding processes within those.

In addition to GEM cycles producing model evolution, Nunez-Oviedo and Clement (2008) identified different long

FIGURE 4 | (A) Initial modeling practices framework from expert study showing two nested levels of processes. **(B)** Modeling practices framework for classrooms showing two nested levels of processes (Phases in lower level of **(B)** are shown here only for the first four Modes in the upper level. See transcript analysis for others).

time scale Macro Processes called "Model Competition." In Model *Competition* if different models generated by the students are compared, such as a single tube vs. double tube model of tubes leading down from the throat, the teacher or the students can make positive and/or critical or negative evaluations of each model, encouraging students to eventually confirm or disconfirm a model such as the single tube model.

Clement (2008a, 2017) described modeling processes in experts at different grain size levels. By analyzing data from videotaped protocols of experts thinking aloud about unfamiliar explanation problems he attempted a synthesis in including several levels of modeling practices, the most encompassing and highest of which are shown in **Figure 4A**. The upper level includes four, large time scale processes. These are supported by processes at the lower level that include several medium scale practices—GEM cycles and assessing competing models.

In this study, we will use this Expert Framework as a departure point for thinking about cognitive processes involved in producing classroom model construction of the kind shown in **Figure 3**, by asking whether any of its elements can be seen in classroom whole class discussions. It may suggest initial hypotheses or concepts for what large scale and medium size scientific knowledge construction processes are taking place in the classroom. We will then reject, modify expand or add to elements of the initial framework where needed. Our final framework is shown in **Figure 4B** as an advanced organizer and will be discussed in the results section.

Questions Motivating This Study
Thus, many researchers have worked on pieces of this problem and the challenge for the present case study is to build from these studies to describe a coherent framework of multi-level processes involved in the teaching and learning of explanatory models along with modeling practices. If a viable framework is found, it should be able to describe repeated patterns of processes occurring over multiple lessons.

General Long Term Background Questions

The following three items are not in themselves specific research questions but they describe some general long-term motivations for our work (derived from the three tensions in **Figure 1**).

Opposing Teaching Approaches Question

Are Teachers limited to two opposing choices for discussion: Open Discussion vs. Authoritative? Or are there intermediate modes between these?

Conceptual Dissonance Question

Given the topic of this study, we are assuming the teacher has a goal of conceptual understanding for a concept in the form of a target model, and assuming student discussions for active learning are desirable. Can a class start from student-generated models that may conflict with the target model in a number of ways, and arrive at the target model through discussion? How? Won't students' faulty models interfere?

Lesson Objective Tension Question

Can a teacher guide or scaffold discussions to foster model construction (reasoning) practices, as a type of scientific thinking, at the same time that they are learning science content? How? (see **Figure 1**).

Specific Research Questions

Assuming we find a class with some or all of the characteristics listed above, we can ask more specific research questions for a case study of how such discussions evolve. Scott et al. (2006) focused on finding a sequence of *sociocultural* discourse patterns. To complement this, we ask, for a teacher experienced in fostering the learning of conceptual models in biology, what are the major *cognitive* processes involved in constructing a scientific target model within the social context of classroom discussions?

RQ1. Is there a pattern of large model construction Modes that occurs over a large time scale of 1–5 lessons?
RQ2. Is there a pattern of smaller model construction phases or processes that occurs over a medium sized time scale of 5–20 min cycles within lessons?
RQ3. If present, how are these patterns connected?

These questions are about: (1) *generating new descriptions and hypotheses* for teaching and learning processes; (2) *providing existence demonstrations* for several newly described types of discussion modes and phases. For this reason a descriptive case study is the method of choice.

METHODS

Context

We conducted a video case study microanalysis of whole class discussions in a series of three classes taught by an experienced science teacher. The broader topic of the unit was "how the glucose goes to the cells through the blood stream." During the lesson sequence, the 24 7th grade students examined the processes and structures that allow glucose absorption in the small intestine. The teacher had nearly 20 years of teaching experience, very good content domain and classroom management skills, and conducted her teaching by taking into account students' contributions. Students were organized into six groups or tables of four students. The teaching episodes took place almost at the end of the school year and consisted of three 45 min lessons that were videotaped and transcribed verbatim. We chose to look at a teacher who had strong content goals for the students learning particular models of the digestive system and also had goals for their doing scientific thinking.

The lessons were part of a model-based curriculum whose goal is to teach middle school students the function of major body systems by starting from their own conceptions. Examples of topics they had studied in previous chapters were: cellular respiration (Chapter V); the circulatory system (Chapter VI); and the respiratory system (Chapter VII). The teacher sometimes asked students to work individually or in pairs and write down or draw their ideas in the curriculum workbook or on their small group's shared whiteboard. At other times, students were asked to share, compare, draw, and discuss their ideas within their small group until reaching consensus. In whole class discussions the students shared and evaluated their ideas, often by displaying drawings they had made on their white boards, and these discussions are what we focus on here.

Data Analysis

Transcripts were analyzed to find large scale and medium scale reasoning patterns that were occurring in this three-lesson sequence by conducting a micro analysis of the teacher and students exchanges, including questions, answers, and drawings. To develop viable constructs for the processes taking place, we employed a construct development cycle (Miles and Huberman, 1994, p. 308) leading to the progressive refinement (Engle et al., 2007, p. 240) of hypotheses about modeling processes. This consisted of: (a) segmenting the transcript into meaningful teacher and student statements or turns, (b) making observations from a cluster of statements, (c) utilizing the framework in **Figure 4A**, where possible, to form a tentative classification of the process behind the statement (or if that failed, generating a tentative new construct), (d) examining the data to look for more confirming or disconfirming instances of the process, (e) criticizing and modifying or extending the hypothesized category to be consistent with, or differentiated from, other instances, and adding it to the framework; (f) returning to the data in (c) again to apply the modified construct, and so on. Because of the difficulty involved in studying a relatively unexplored area of large and medium sized complex processes with high inference coding, coding was done jointly by the two authors. Triangulation from (1) both analysts having to reach agreement and (2) from checks on the ability to use the same constructs across all three lessons and subtopics served to improve and support viability and validity.

RESULTS: CASE STUDY FINDINGS

In Modeling Processes Identified we will first summarize the process modes and phases found at two levels that provide our main answers speaking to research questions 1 and 2. The

TABLE 1 | Major modeling modes.

Major modeling modes	Description of the mode
Describing a pattern to be explained mode	Process of describing a pattern in the form of a question that the students may generate or receive that calls for an explanation. The pattern may be: a. An exploratory observation or a set of observations leading to a pattern that could also be presented through a demonstration or film. b. A description or model of a system that is already accepted, leading to a deeper question that the students answer by generating an underlying explanation or mechanism at a different level. In the present study the teacher asks: "*How* does the glucose get out of the intestine and into the blood?" Here, students already have the model that glucose goes from the intestine into the blood; but they are asked to explain *how* that happens by generating a model of the intestinal wall at a smaller, more detailed grain size.
Brainstorming initial model(s) mode	Process of having students brainstorm initial explanatory models that will provide the explanation or prediction for the above pattern. The teacher supports the students in generating models via individual writing/drawing, and small group and whole class discussion, without evaluating them. Students' ideas will include useful conceptions, but may also include gaps and alternative conceptions in conflict with the target model for the lesson.
***Model evolution mode**	Process of having students improve their models, sometimes several times, via model Generation, Evaluation and Modification cycles. Evaluatory Observations may also be involved in these processes.
***Model competition mode**	Process of identifying and evaluating two or more alternative models that compete for providing an explanation. Models are: (1) Each evaluated for strengths and weaknesses (evaluatory observations may also be involved here); (2) Non-viable models are disconfirmed; and (3) Remaining models are assessed to determine which model is strongest and has the best explanation.
Model consolidation mode	Process of summarizing and making any final repairs to the new model by disclosing the scientific model and encouraging students to review and compare the new model to their original models. Can include student articulation of support for the model.
Model application and/or Domain extension mode	Process of applying the new model to a new case for explanation or prediction. If the case is outside the initially perceived domain of application of the model, it may stretch or extend that domain.

*Preliminary versions of these Major Modeling Modes were identified in Nunez-Oviedo (2004).

TABLE 2 | Model construction phases.

Model construction phases	Description of the phase
Exploratory observations	The teacher or student statement either asks for or provides an empirical observation; they are not sought for purposes of evaluating any particular theory or model (in which case they would be called evaluatory observations). They may be new observations or ones gathered from memories of previous occurrences. Multiple observations can form a pattern to be explained.
Model generation*	In the *Generation* **(G)** phase the teacher or student statement either asks for or provides a new theory, explanatory model, partial model, or conception. This can be done with varying degrees of speaker confidence in the correctness of the statement and can be done in either a declarative or interrogative manner by the teacher or student. Examples of key phrases that help identify student Model Generation processes or the teacher's attempt to scaffold them: What ideas do you have about…, I think what is happening…, What explanation can you think of for…, I think that maybe what's going on is…., etc.
Model evaluation*	In the *Evaluation* **(E)** phase the teacher or student statement refers to a theory, model, conception or explanation that has previously been or is currently under discussion. The statement either asks for or provides an evaluation, judgment, refutation, criticism, support, or endorsement of a particular explanatory model. These can refer to either evaluatory observations or rationalistic evaluations, e.g., for coherence with thought experiments or previously accepted theory. Examples of phrases that help identify Model Evaluation processes or the teacher's attempt to scaffold them: Do you (I) agree or disagree with …, That makes sense…, I don't (Do you) believe that…, Are you sure you can have…, Do you think that is the way…, etc.
Model modification*#	In the *Modification* **(M)** phase the teacher or student statement either asks for or provides a suggested change, adjustment, or modification to a theory, explanation or addition or could introduce a revised model with multiple changes from the original. Sometimes the modification statement comes with little verbal evidence that a prior evaluation process has been underway as students often engage in this process internally. If the statement appears to make little or no reference to the previous model, it is instead considered to be in the Generation category above. Examples of phrases that help identify Model Modification processes or the teacher's attempt to scaffold them: Could one see it a different way… Could one change…, Could it be explained another way…, etc.
Model confirmation#	Models are confirmed when a model that had been evaluated positively or at least more positively than other candidate models is accepted by the class and confirmed by the teacher.
Model disconfirmation#	Models are disconfirmed when a serious problem with the model is found and no one could think of a way to modify and repair the model.

*[Identified in Williams and Clement (2015, 88-89), Clement and Steinberg (2002)].
#[Identified in Nunez-Oviedo (2004, 115)].

TABLE 3 | Pattern to be explained mode, and brainstorming initial models mode (Part 1).

Lesson	Topic	Description	Major modes of modeling	Model construction phases
1	1. T: "How does the glucose go into the blood at the small intestine?" (Main question introduced)	The teacher asks the class: "How does the glucose go into the blood at the small intestine... and then to the cell?"	Pattern to be Explained	
1	2. "Could you draw your ideas?" (Drawing ideas)	The teacher asks the students to draw their ideas individually.	Brainstorm Initial Model(s)	Model Generation
1	3. "Can you come up with a team model?" (Sharing ideas)	She asks then to share their ideas in their small groups and to create a team model.	Brainstorm Initial Model(s)	Model Generation
1–2	4. "Would you share your team drawing in whole class?" (Sharing ideas)	The teacher asks small groups to present their ideas in whole class. Students' ideas are not evaluated.	Brainstorm Initial Model(s)	Model Generation

section Transcript Analysis below presents detailed findings in the form of a transcript analysis according to the modes and phases identified. The section Findings by Research Question collects these together to give more general findings that speak to each of our three research questions.

Modeling Processes Identified

Our video transcript analysis yielded evidence for two levels of large and medium sized modeling processes, with the large processes operating over a longer time scale and the medium sized processes nested within the larger ones. The two levels are termed Major Modeling Modes and medium sized Model Construction Phases. Each level can be cyclical.

Table 1 shows six Major Modeling Modes at larger time scales that were identified. Four of these modes are similar to those from studies in expert reasoning shown in Figure 4A. Two other modes were found that were not observed in expert protocols yet but appear to be important steps in classroom modeling (see Figure 4B).

Table 2 shows six smaller Model Construction Phases or processes identified in the transcript analysis that can take place within some of the larger Modes above (especially within Model Evolution Mode).

For example within the Model Evolution Mode in Table 1, we will see the teacher and students engaging in the model Generation, Evaluation, and Modification phases in Table 2, as they try to improve or repair a model they have generated, and as it evolves toward the teacher's target model. While we have presented the two levels of processes in the tables above as an advanced organizer, each table is the culmination of a long process of transcript analysis, evaluation, and revision to arrive at stable categories that fit the protocol.

Transcript Analysis

In this section we use the above categories in presenting our case study microanalysis of a three-lesson sequence. We display the fit between the processes and the transcript episodes, showing how a small number of processes can be seen to underlie a relatively long and complex transcript sequence of statements by the teacher and students. Readers who wish to see a summary of the analysis as another advanced organizer can preview Figure 6. We are not displaying the teaching here as either a perfect example or a bad example; rather our purpose is to develop

constructs to describe the processes she attempts to foster. Later we will discuss what we consider to be positive and negative elements of the discussion.

We first divided the transcript into 21 "topics," numbered in the second column of Table 3. The third column contains a narrative of the dialogue with teacher and student quotations; the rather long transcript over three lessons is condensed for reasons of space. The fourth column and fifth columns show the Major Modeling Modes and the Model Construction Phases, respectively. The topics of the three lessons are grouped into six major parts.

Transcript Part 1—Pattern to Be Explained, and Brainstorming Initial Models Modes

In the first lesson, the teacher introduced the topic to the class and then asked the students, "How does the glucose go into the blood at the small intestine and then to the cell?" (Topic 1). We consider this question to be an example of fulfilling the Identifying a Pattern To Be Explained Mode because the students do not have a model of how the glucose is able to get into the blood through the intestinal wall (In other cases a series of exploratory observations might lead to a pattern of observations to be explained).

The teacher then asked the students to draw out their ideas individually for about 7 min and then asked the students to share their drawings in small group and come up with a team model (Topic 2). The teacher then called on the groups to share their ideas in whole class (Topic 3). During the students' presentations in this period the teacher did not evaluate drawings or encourage the students to do so. Through their presentations, students explained processes and used concepts such as villi and absorption but none of the six groups had a working model of the transfer of glucose to the blood at the small intestine. In other words, the students' ideas were still far from the target model of the lesson.

In Topics 2–3, we classified the activities as belonging to a mode called Brainstorming Initial Models that primarily involves the process of Model Generation. It is worth noting, that the students did not build the initial model in one step. Instead, the teacher asked the students to thinking individually, then to share their ideas at their small group, and finally each team presented their ideas in whole class.

Lesson	Topic	Description	Major modes of modeling	Model construction phases
2	5. T: "Villi are what?"	A student generated the idea that villi "are like hands to grab nutrients.	Model Evolution	Model Generation
		Teacher fostered evaluation of the student's idea by saying "Is it like a hand?"		Model Evaluation
		Another student modified the initial idea by saying that villi "look like fingers."		Model Modification
2	6. "What is the purpose of villi?"	A student generated the idea that villi "just grab little bits of glucose like fingers.	Model Evolution	Model Modification
		The teacher fostered evaluation, saying: "are they scooping the stuff in like a mouth?"		Model Evaluation
		A student modified the initial idea by saying that villi "are absorbing this glucose."		Model Modification
2	7. "What is the advantage of having these villi?" Episode 1	A student generated the idea that, "maybe there is a kind of filter and it is just the semipermeable membrane for the cell that let out some stuff that they did not want to, the villi get the glucose and the vitamins and other stuff in.	Model Evolution	Model Generation
		The teacher evaluated the answer positively by saying "if villi are made of cells and each cell has a cell membrane then I am not sure how villi are going to act like a filter, and that is a good idea, isn't it?... It is very creative… villi have cells and…what is the inherent filter in the cells...?"		Model Evaluation
		Students answered, "membrane." They concluded that villi "cells have semipermeable membrane that might act as sort of a filter."		Model Confirmation
2	8. (Repeats) "What is the advantage of having these villi?" Episode 2	A student generated the idea that "when it is bumpy there is like more space to absorb glucose?	Model Evolution	Model Generation
		The teacher elaborated on and lent support to the student's answer by reminding to the students that earlier in the curriculum they had discussed that "if the small intestine is stretched out it will reach the size of a tennis court."		Model Evaluation
		The teacher and the students then modified the initial idea by saying that villi "increase the surface area" to increase "the amount of nutrients that is absorbed."	Model Consolidation	Model Modification; Model Confirmation

Part 2—Evolution and Consolidation Modes

The next section is one of the most important but also challenging part of the lessons because it is where the teacher scaffolds the students' modeling processes as she moves toward the target model. The interaction pattern of the discussion shifts markedly here from an open discussion (brainstorming) pattern wherein ideas are not evaluated, to a pattern of Model Evolution wherein the teacher fosters model evaluation or modification, usually implicitly by hinting that models could be improved and then modified by the students.

The teacher focused on a small segment of the small intestine to foster student models and their evolution toward the target model (see **Table 4**; Students had also studied the idea previously that cells had semi-permeable membranes, and one student brings this up).

Beginning in **Table 4** the teacher conducted a different large scale modeling process that we call Model Evolution Mode" (see **Figures 4A,B**) through which students' ideas evolve from "villi are like hands to grab nutrients" to villi "look like fingers"; villi "grab little bits of glucose like fingers"; to villi "are absorbing

this glucose"; "maybe it is a kind of filter…"; to "villi's cells have semipermeable membrane that might act sort of a filter"; to "when it is bumpy there is like more space to absorb glucose?" to "villi increase the surface area to increase the amount of nutrients to be absorbed." The evolution of these ideas was the result of four cycles of medium sized (medium time scale) processes shown on the column located on the far right side of the table: Model Generation, Model Evaluation, and Model Modification (We use "Medium Sized" because in future publications, we plan to also discuss even smaller reasoning processes). Referring back to **Tables 1**, **2**, the reader can see that we so far have encountered the first three modes in **Table 1** and the first four phases in **Table 2**.

Part 3—New Pattern to be Explained, Brainstorming, and Competition Modes

The teacher then went further and asked the students to conjecture and discuss in their teams about "where to locate the capillaries to make villi an efficient absorbing machine" (Topic 9 in **Table 5**) and then to draw their ideas on an overhead (Topic

TABLE 5 | A new round of modeling (part 3).

Lesson	Topic	Description	Major modes of modeling	Model construction phases
2	9. T: Where to locate capillaries? (Main Question, see Topic 1)	Teacher asked where to locate capillaries to make villi an efficient absorbing machine and asks them to work in their groups.	Pattern to be Explained Mode	
2	10. "Where did you place capillaries with respect to villi?" (Drawing and sharing ideas, see Topic 2, 3 and 4)	She asked the groups to report their ideas in whole class and to draw them in the overhead. A student from group 4 drew three vertical lines outside of the small intestine (Model A in Fig. 5A). A student from group 3 said "they are on the tip of the villi" and the teacher drew a capillary on top of the villi on the overhead (Model B). A student from group 1 "drew a capillary with dead ends inside the villi" (Model C). No evaluations occurred.	Brainstorm Initial Model(s)	Model Generation
2	11. "What is the least efficient, Model A, Model B, or Model C?"	The teacher asked whether, "Model A, Model B or Model C was the least efficient to absorb nutrients?" The students said "Model A." The teacher asked "why Model A is the least efficient?" and a student said, "it is not even close."	Model Competition	Model Evaluation (by students)
		The teacher then said, in Model A the "nutrients would have to travel [far] and it probably is not our model."		Model Disconfirmation
2	12. "What is the most efficient, Model B or Model C?"	The teacher asked, "What might be the most efficient B or C?" Votes? Students voted that Model B was the most efficient. The teacher acknowledged this idea and said, "but there is one problem with this model ("B"). "Think in the environment in which you are finding these capillaries. It is friendly or hostile?" The students answered "Hostile" adding that there are "acids" in the intestine and that the little capillary could be damaged. The teacher and students also discussed that the walls of the blood vessels are not as strong as the walls of the small intestines and that they lack mucus to coat and protect them.	Model Competition	Model Evaluation
		She concluded that "even though Model B is the most efficient model, it might damage these hair-like, hair-size blood vessels."		Model Disconfirmation
2	13. "Is Model C the most efficient to place capillaries at the villi?"	The teacher and the students then concluded that "…probably this (Model C) is the most efficient and safest model."	Model Competition	Model Evaluation
		She asked the students whether "this make sense" and they answered "Uh huh…"		Model Confirmation

10). We consider this a new modeling question at a different micro level of detail or grain size, and it starts a new pass through the Major Modeling Modes. The students disagreed about where to locate the capillaries (see models A, B, C in **Figure 5A**).

Here the teacher challenged the students to generate ideas by conducting new episodes of the Pattern to be Explained Mode and the Brainstorm Initial Model Mode at a new level of detail. However, none of the three models were close to her target model shown in **Figure 5B**. As result, the teacher scaffolded the students in evaluating their ideas. With three alternative models in play, she enters a Model Competition Mode.

The teacher asked the students "what is the least efficient, Model A, B, or C?" and the discussion proceeded as shown in **Table 5**, Topic 11. The teacher and the students collaborated in evaluating ideas, with the teacher guiding the discussion with questions. The class concluded that Model C was the best model (Topic 13). In **Table 5** we show the class entering a Model "Competition" Mode where *several alternative models are evaluated* until one emerges as preferable (When working with teachers we call this a Model "Comparison" Mode instead to emphasize that the goal is a joint decision rather than to find a "winner"). This contrasts with the earlier Model Evaluation

Mode that worked with a *single model and tried to improve it*. As shown in the table, the Competition Mode was fostered by smaller processes of model Evaluation, Disconfirmation, and Confirmation.

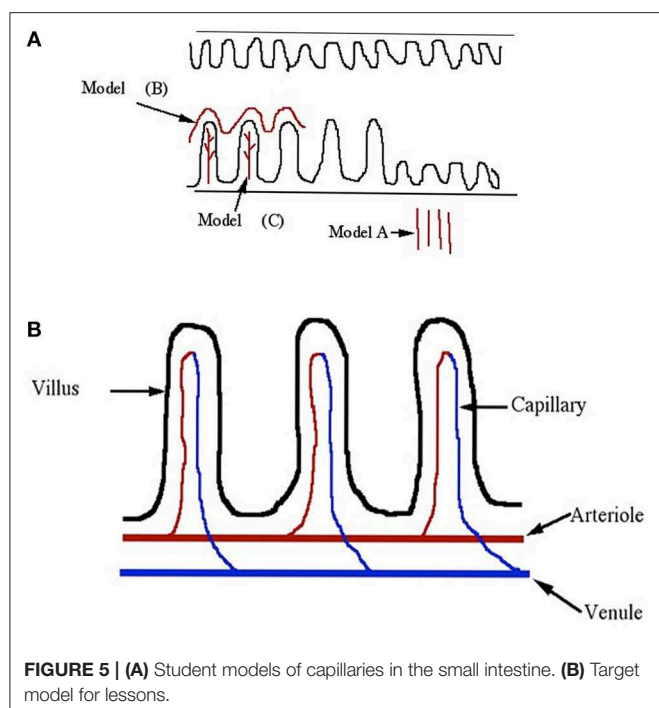### Phases involved in the model competition mode
Figure 4B shows a picture of the modeling practices framework we are assembling. It shows two nested levels of processes allowing us to summarize steps in the Model Competition Mode as:

1. Evaluate each model individually for strengths, weaknesses
2. Disconfirm non-viable model
3. Assess which remaining model is strongest and has best explanation

In this case the last step 3 was not needed because there was only one model remaining.

How the Interaction Is Different From Recitation
Since there are many interactions in the transcript with turns of the form TST or TSST… (where T, teacher; S, student), one might ask whether the teacher is just doing *recitation* (called IRE by

FIGURE 5 | (A) Student models of capillaries in the small intestine. **(B)** Target model for lessons.

some, but see Nassaji and Wells (2000, p. 393) for different views) in these classes. We can define recitation as the teacher:

- asking for *recall* of prior "School Knowledge," including "School Science knowledge,"
- getting a student response, and then,
- providing an *immediate judgment on the correctness* of the response.

There are some instances of recitation in the full transcript, especially where the teacher is trying to review what happened in the previous class, but most of the interactions are not recitation. The most obvious example of that is during Brainstorming Mode when the teacher is eliciting student models without evaluation. In addition, in many other sections, the teacher engages in TSTTST... exchanges where:

- instead of *recall*, the teacher is most often asking for students to *reason* in order to *generate, evaluate, or modify a model*;
- instead of an *immediate judgment on correctness*, she often asks them to *participate in evaluating* a proposed model. Even when she does indicate whether a model is correct, she is usually not just doing that, but *giving a reason* for her evaluation. For a positive evaluation she may state a reason or question that could add to or enhance the model. For a negative evaluation she gives a reason that explains important constraints on the modeling.

This has the atmosphere of the teacher helping the students *reason to construct a visual model together* rather than a "quiz game" of *reciting memorized words*. The repeated evaluations and modifications of student drawings add to this atmosphere of construction that appears to be very different from recitation.

So in the transcript we appear to have a whole spectrum of interaction styles between the poles of brainstorming on one end and recitation and lecture on the other end. The teacher was able to adaptively change her interaction style as she moved through the different Modeling Modes.

We should also caution that the processes we describe here are not all necessarily conscious or articulated strategies for the teacher. With her years of experience, she is somehow able to intuitively scaffold student thinking in ways that led to the process patterns identified here, without her knowing or using terms like "Evolution Mode" or "Model Modification Phase." Rather, those are constructs we have formulated to describe the patterns of reasoning that emerged from the discussions.

Part 4—Model Evolution Mode

In the next segment in **Table 6**, the teacher attempts to take the remaining Model C in **Figure 5A** and have students modify it to be closer to the target model by returning to Model Evolution Mode.

We infer that the teacher scaffolded the students in improving their shared Model C in **Figure 5A** toward the target model by returning to Model Evolution Mode. The student models evolved from "villi having dead ends" to "villi having loops" (Topic 15) and to "villi having loops and red and blue colors" (Topic 16). The red color indicated blood moving away from the heart. We also view the Evolution Mode as utilizing subprocesses—i.e., as being the result of three cycles of smaller scale Model Evaluation and Model Modification processes. This is shown in **Figure 4B** by the upward arrow pointing to Model Evolution.

Classroom dialogue diagram

The classroom dialogue diagram in **Figure 6** was created to depict a summary picture of events in Topics 9 to 16 above. It shows the student and teacher contributions to co-constructing an explanatory model by starting from the students' ideas.

Williams and Clement (2015) describe such classroom dialogue diagrams as follows:

The diagrams are (abridged) horizontal versions of the transcript with student statements [above] and teacher statements [below], with time running from left to right. The horizontal strip across the middle of the diagram contains short written phrases to describe the evolving explanatory model. These phrases represent our hypotheses for teacher's conception of what a student's addition to the model was at a given point in the discussion... It was assumed that the teacher was aiming to foster model construction based on their view of the students' model at that time, and how it differed from the target model... arrows that point from both teacher and student statements toward the explanatory model descriptions in the center strip indicate their shared contributions to the changes or additions in the models (p. 13).

The transcript in this diagram is highly abbreviated. The symbol A_X indicates Model A was disconfirmed. The figure highlights three aspects of our analysis of model construction processes:

Lesson	Topic	Description	Major modes of modeling	Model construction phases
2	14. T: "Could you draw the best model for absorbing nutrients?" Episode 1	The teacher asked the students to review individually the initial ideas they drew and change them if necessary.	Model Evolution	Model Evaluation
		Students worked on the task until the end of the lesson.		Model Modification
3	15. "Where are the loops and the colors in the capillaries?"	The teacher 1. asked the students to review what they have done the day before. 2. asked small groups, "to draw in your whiteboards the capillary system…able to absorb… glucose, ok?" 3. described Group 1's drawing (Model C in **Figure 5**) to the rest of the students. 4. polled the class to know how many groups had drawn something similar. 5. Many students raised their hands and she said that she has a problem with that model… The teacher 6. asked the students to think of the capillary system and to remember what they had learned about capillaries and their function earlier. They said that capillaries are a site of "exchange." 7. She commented "If they (capillaries) are exchanging something your models look to me as if they are dead ends.	Model Evolution	Model Evaluation
		8. She asked them to remember how they had drawn capillaries before, that they deliver and absorb substances, that they need to complete the loop. She agreed with the students, that they **should** assume villi each have only 'one' capillary inside (a simplification **for drawing purposes**). 9. The teacher asked the small group students to add capillary loops in their drawings but did not say how. In addition, she asked the students to find a colorful way to draw the capillary loop.		Model Modification
3	16. "Where are the colors in the capillaries?"	Soon the teacher realized that only half of the students had found a way to draw color coded capillaries. She asked the students what colors they have used before for capillaries and they said "red" and "blue." The teacher and the students discussed that we color capillaries differently when they are arteries or veins.	Model Evolution	Model Evaluation
		The teacher asked the students to modify the colors of the capillary loops by using red and blue colors but did not tell them where each color goes.		Model Modification

(1) The arrows in the diagram provide a picture of how statements from both the teacher and the students contribute to the model co-construction process taking place from left to right in row 3.

(2) It illustrates discussion modes of Model Competition and Model Evolution that are intermediate between purely open discussion and pure lecture.

(3) The bottom two rows depict processes that we hypothesize are simultaneously occuring in parallel during instruction. It shows the large scale modes (5) operating over longer time scales than the smaller scale phases (6), and each smaller scale phase contributing as a subprocess to the large scale mode above it.

The overall pattern that set up the conditions for these intermediate discussion modes was:

● After a Pattern to be Explained has been identified, the teacher elicits model ideas proposed by the students. These will naturally include flaws and gaps from the scientific target point of view;

● The teacher assesses the proposed students' ideas on the fly to determine their distance from the scientific model. In other words, the teacher needs to determine what the students' ideas

are that need to be discarded and what ideas they can build on and modify to reach the target;

● The teacher scaffolds the students in examining and discarding those students' ideas that are farthest from the scientific model. We described this as a Competition Mode that was fostered by evaluation processes—that disconfirmed Models A and B;

● The teacher scaffolds the students in improving or repairing the most promising ideas until they are close to the scientific model. The transcript provides evidence that the teacher worked in a Model Evolution Mode that was comprised of two GEM modeling cycles to repair Model C until it was very close to the target model.

● We believe that the sequence in Topics 9–16 above is not atypical: the Model Competition mode ended with one surviving model, but that model still needed improvement via the Model Evolution mode.

Part 5—Consolidation Mode

Due to space limitations, we do not include later phases of the instruction in the classroom dialogue diagram. But we will describe them here. In what we call Model Consolidation Mode, the teacher then showed the students four transparencies (see

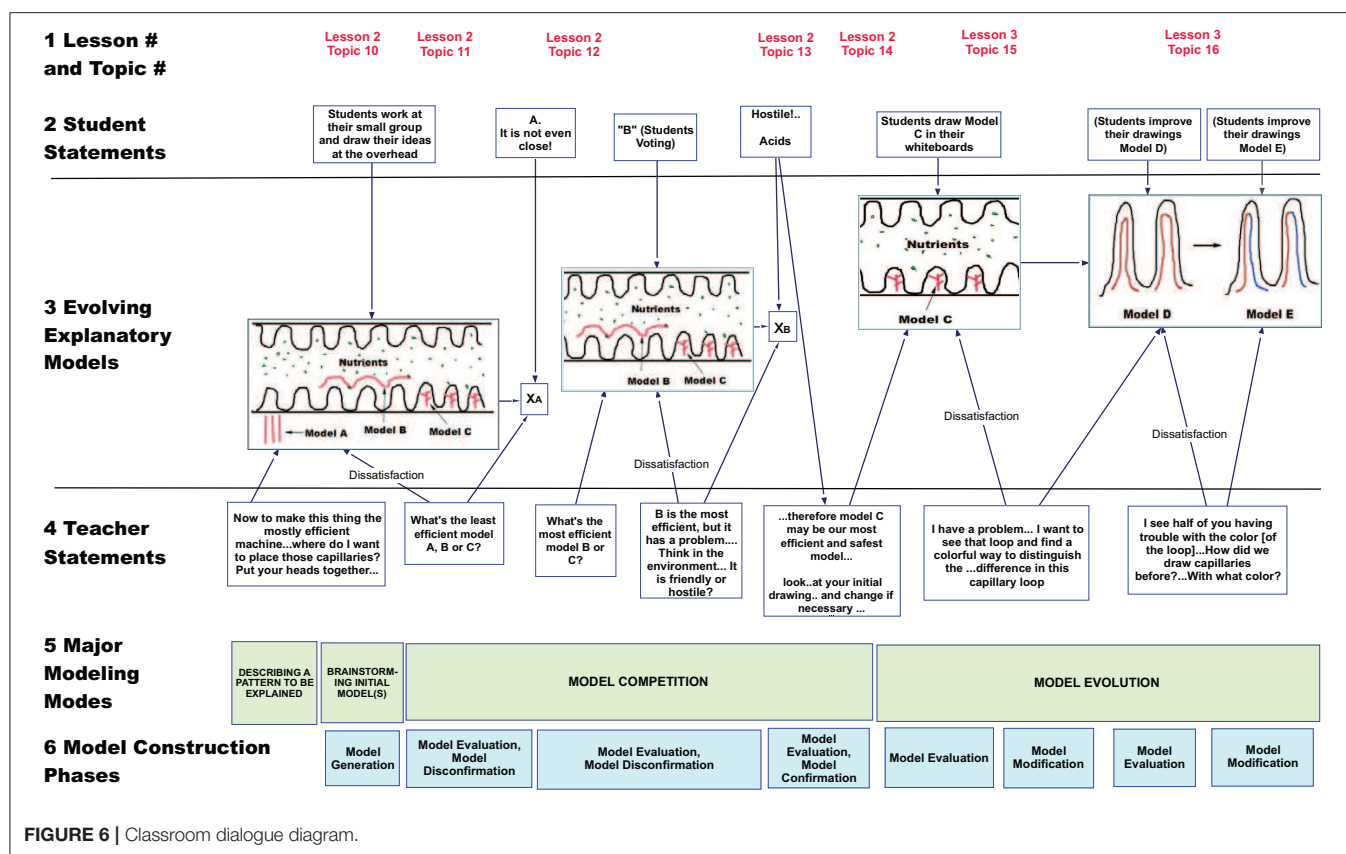**FIGURE 6 |** Classroom dialogue diagram.

Table 7) that contained the scientific version of the target model, with connections from the vessels in each villus to a common artery and vein. The teacher also asked the students to go back to their initial drawing and asked them to compare it with what they had learned and change in case it was necessary (Topic 18).

In the first segment, the teacher showed the students the scientific model as an authority by using transparencies and then asked the students to compare what she was showing them with their ideas. We call this process "Consolidating the Scientific Model" (Topic 17). In the second segment, the teacher asked the students to compare the scientific model with their initial drawings and asked them to modify them if necessary to explain exactly how sucrose transfer occurs. We call this process "Explaining the Original Pattern" (Topic 18). During this time we infer that there were also smaller, Model Construction Phases occurring, in particular, Model Evaluation and Model Modification. We note that as the students get nearer to the target model, the teacher has become more proactive in evaluating their models, and suggesting directions for their modification. But she does not tell them exactly how to draw the modifications.

Part 6—Application Mode

The teacher then asked the students to apply their new understanding to an entirely different subject, the respiratory system in what we call Model Application and/or Domain Extension Mode. She asked them "if lungs are another site of exchange, what they might have?" (see **Table 8**).

In this part of the lesson, the teacher asked the students to transfer part of their understanding of villi as a site of exchange to alveoli at the lungs. We hypothesize that a Model Application activity may not only be a shortcut as a starting point for building the new alveoli model, but also may be another way of exercising and consolidating the model of the villi that they have just learned. However, the teacher was running out of time at this point and was actually only able to introduce the topic of gas transfer to the students in this brief segment, but it did serve to exercise the model they had just learned. In other cases where the application is not as big a leap as going from intestines to lungs (e.g., application to transfer of other substances besides glucose via the villi), the Application Mode may serve to simply extend the domain of application of the model they have just learned.

Findings by Research Question

Figure 4B summarizes the modeling processes identified at two major levels in our case study. It is somewhat surprising that processes derived from studies of sophisticated experts working on physics problems (in **Figure 4A**) could have parallels with some of the learning processes in a 7th grade life sciences classroom (in **Figure 4B**). This adds authenticity to the idea that the students were contributing to some real scientific reasoning practices.

TABLE 7 | Model consolidation mode (part 5).

Lesson	Topic	Description	Major modes of modeling	Model construction Phases
3	17. T: "Could you compare your model with the scientific model?"	The teacher showed the students several transparencies about villi including photos of their blood vessels. She also asked the students if their models were similar or not to the transparencies.	Model Consolidation Mode: Consolidating the scientific model	Evaluative Observations Model Evaluation
		Students viewed the transparencies attentively and reacted to teachers' questions and comments.		Model Modification
3	18. "Could you improve your villi model?" Episode 2 (See Topic 14 Episode 1)	The teacher then asked the students to go back to page 204 again and to compare what they have learned with their original drawing.	Model Consolidation Mode: Explaining the Original Pattern	Model Evaluation
		She also asked the students to modify original drawing if necessary or to create a new drawing on page 205 of the glucose absorption that occurs at the small intestine.		Model Modification

TABLE 8 | Model application and domain extension mode (part 6).

Lesson	Topic	Description	Major modes of modeling	Model construction Phases
3	19. T: "What is the general structure of the respiratory system?" (Main question and drawing)	The teacher asked them to draw on page 240 the respiratory system and the gas exchange in the lungs.	Model Application and/or Domain Extension	Model Generation
		The teacher said that most of the students did pretty well in drawing the pulmonary system.		Model Evaluation
		The teacher showed the students a transparency of the respiratory system that displayed its general structure and made comments about it.		Model Confirmation
3	20. "Are alveoli sort of villi?"	But the students did not include gas exchange in their drawings. The teacher asked the students, "if lungs are another site of exchange, what they might have?" A student answered that lungs might have something "sort of villi." The teacher accepted the student's answer and called it "alveoli" –a site of exchange.	Model Application and/or Domain Extension	Model Generation
		The teacher showed the students a transparency depicting a cluster of grapes wrapped with a string. The teacher and the students talked about the grapes representing air sacks and string representing blood vessels.		Model Modification
		The teacher showed the students a transparency of the alveoli and their blood vessels. They commented the overall structure, function, and flow of the gas exchange at the lungs.		Model Modification
3	21. "What do the colors of the blood vessels that surround alveoli represent?"	The teacher called the students attention to the capillaries' red and blue colors located around each alveolus. She asked the students to discuss in their small group the significance of the blood vessels' colors.	Model Application and/or Domain Extension	Model Modification

Research Question (1): Is There a Pattern of Large Model Construction Strategies That Occurs Over a Large Time Scale of 1–5 Lessons?

We found six large modes of model-based teaching within these lessons, called Major Modeling Modes, described in **Table 1** of the Results section which is our main answer to this question. The top level of **Figure 4B** also summarizes the Major Modeling Modes identified in the case study. Describing a Pattern to be Explained Mode is the starting point for modeling there.

It was of interest to us to see the different styles of interaction occurring in each of the subsequent modes. The Brainstorming Initial Models Mode was characterized by divergent open discussion with the teacher preventing evaluation of the models, in contrast to ensuing modes. Model Competition Mode served

to evaluate and disconfirm the least viable student models and the Model Evolution Mode served to repair the most promising models in a sequence of progressive refinements. Model Competition and Model Evolution Modes were characterized by the teacher's efforts to foster student contributions to those evaluation and repair processes as well as adding some herself. We call this interaction style *teacher-student co-construction*. As illustrated in **Figure 6**, it yielded many ideas and inferences that were student-generated as well as some that were teacher-generated. These two modes were the most lengthily ones in these discussions.

Near the right side of **Figure 4B**, Model Consolidation Mode was characterized by a very different mini-lecture style that was the most convergent style, followed by final model modifications

in student drawings. By convergent (as opposed to divergent) we mean that the number of models and the distance of the models under discussion from the target model was getting smaller. The Application and/or Domain Extension Modes were a mixture of co-construction and mini-lecture styles. Thus, a spectrum of styles was seen within these large scale modes, from student-generated ideas in open discussion, to teacher-generated ideas in mini-lectures, with a co-construction style of shared idea generation in between. These styles appeared to fall on a spectrum running from divergent to convergent. This resonates with others who have identified the need for both divergent and convergent discussions (see Scott et al., 2006 and Windschitl et al., 2012). The balancing of divergent and convergent thinking is also a hallmark of model construction work for scientists Clement (2008a). (See Lehesvuori et al., 2013 for contrasting case studies of two teachers who each moved in opposite directions between convergent and divergent styles).

The Mode sequence was cyclical, with the first round passing through the first three modes in the upper level of **Figure 4B** plus Model Consolidation, and the second round restarting with a new Pattern to be Explained in Topic 9 in **Table 5** and passing through all six modes. This reflects the view that that scientific models are nested (Machamer et al., 2000). The double arrows in the upper level of **Figure 4B** indicate that Model Evolution and Model Competition might occur in a different order, or even alternate, depending on when new models occur to students. So the upper level is intended to portray a loosely ordered sequence.

Research Question (2) Is There a Pattern of Smaller Model Construction Phases or Processes That Occurs Over a Medium Sized Time Scale of 5–20 min Cycles Within Lessons?

We found such a pattern that occurred repeatedly in this case study shown as Model Construction Phases in the lower level of **Figure 4B** and defined in **Table 2**. The three most frequent processes were Model Generation, Model Evaluation, and Model Modification (GEM), and these participated in another smaller GEM cycle within Model Evolution, although they could also occur as individual processes. We refer to repeated GEM cycles as a model evolution process capable of producing a sequence of more and more adequate models (**Figure 3**).

In addition, as shown in **Figure 4B**, Evaluatory Observation was proposed as a subprocess that could implement an Evaluation Phase, and Exploratory Observation was identified as a subprocess that could lead initially to a Pattern to be Explained, although the latter process did not appear in this particular case study.

Research Question (3) If Present How Are the Above Patterns Connected?

Another pattern in the transcript analysis is that the smaller time scale Model Construction Phases are nested within the larger Major Modeling Mode processes; the small phases are subprocesses that contribute to the purpose of the larger process, as shown in **Figures 4B**, **6**. For example we have found that individual Generation, Evaluation, and Modification processes are utilized within both the Model Evolution Mode and the Model Competition Mode. The Evolution mode utilizes all three GEM processes, while the Competition Mode was seen to utilize mostly the Model Evaluation process applied to *several different models*, leading to some models being disconfirmed while others being confirmed (see **Table 2**). In contrast, the Model Evolution mode is the focus on a *single partially correct model* that is modified by the GEM cycle pattern (hence the name Model Evolution mode).

Rows 5 and 6 of **Figure 6** show the two levels of processes *operating in parallel*, with the large scale modes (row 5) operating over longer time scales than the smaller scale phases (row 6), and each smaller scale phase contributing as a subprocess to the large scale mode above it. For example, if a teacher had evolution toward the target model as a goal during Model Evolution Mode, they could keep that goal in mind as they fostered repeated subprocesses of model Evaluation and Modification.

DISCUSSION

Connections to Previous Literature

Our general objective in this study was to describe a coherent framework of multi-level processes involved in the teaching and learning of explanatory models. The sequence of six Major Modes in the top row of **Figure 4B** shares several individual processes with those described by other researchers such as: Driver and Scott (1996, p. 99); Minstrell et al. (2011, p. 4); Driver and Scott (1996, p. 613); Windschitl et al. (2012, p. 887) and Campbell et al. (2012). The article by Driver and Scott (1996, p. 99) was particularly pioneering in anticipating several of the major modes described here at an early date. We have attempted to add clarity to options within the Identifying a Pattern to be Explained Mode, as well as to separating two levels of processes and especially adding new modes that we call Evolution and Competition Modes. Based on studies of the history of science reviewed earlier, we consider Model Evolution and Competition to be central and essential to active engagement in modeling. We will describe these additions to theory in more detail, moving from left to right in **Figure 4B** in what follows.

Describing a Pattern to be Explained is the first Mode there. Although identifying a major question for modeling in the unit is a mode that others such as Windschitl et al. (2012) have identified, a new feature to us was to include the possibility that its departure point was sometimes an already learned model rather than a pattern in observations. In this case students who had a model of glucose somehow going to the bloodstream from the small intestine were asked to open up the deeper level question of how that transfer takes place– as a pattern to be explained. The explanation was provided by generating a new model at that deeper level.

In this study we focused on the "big picture" of all six modes, but the interested reader can find other case studies focused on individual modes of Model Competition in Nunez-Oviedo and Clement (2008) and Model Evolution in Nunez-Oviedo et al. (2008). These two modes have a very rough but interesting analogy to, respectively, the revolutionary (Kuhnian) and evolutionary views of science discussed earlier. The analogy is weak here because none of the models in the present case

study are persistent enough to act like a resilient mini-paradigm, but the analogy would be closer for a persistent student model such as impetus-like ideas in physics (Even there, the analogy is controversial, but we think still interesting; see Smith et al., 1994, Clement, 2013, and Lattery, 2016).

The bottom row of **Figure 4B** shows the structure of the smaller model construction processes being used within several of the modes above them as the teacher scaffolds student thinking. These findings complement those of authors such as Hestenes (1987, p. 443); Minstrell and Kraus (2005, p. 480; 484; p. 489-491); Schwarz et al. (2009, p. 635); and Windschitl et al. (2012, p. 887) who have described students generating models and discussing them, accompanied by multiple revisions and teacher scaffolding. Here we have attempted to dissect the concept of "scaffolding" further to describe in some detail the nature of the cognitive processes being scaffolded for model development. And a distinctive feature that we have not seen discussed by other groups is the idea of different time scale *levels* of connected strategies for scaffolding these processes Clement (2008a,b); and Williams and Clement (2015).

We also observed that the Model Generation phase occurred less often than the model Evaluation and model Modification phases in the case study. We explain this by saying that once the teacher supports the students in generating an explanatory model that contains several elements, the teacher repeatedly guides the students in evaluating and modifying each one of the elements of the explanatory model until it gets close to the target model.

Regarding Our General Long Term Background Questions

On the first page of this article, we listed three tensions that we believe teachers face when drawing out students' ideas in whole class discussions, summarized in **Figure 1**. These were related to three long-term questions that challenge us.

1. **Opposing Approaches Tension** Are Teachers limited to two opposing choices for discussion represented in **Figure 1**: Open Discussion vs. Authoritative Lecture [described by Scott et al. (2006) as the tension between dialogic and authoritative discourse]. Or are there intermediate modes between these?

We did find that the teacher used open discussions (in Generation of Initial Models Mode) near the beginning of the sequence and some authoritative discussions (in Model Consolidation Mode) near the end. However, we observed that this teacher conducted other two discussions modes with interaction styles that do not fit this dichotomy. In particular, as shown in **Figure 7**, Model Competition and Model Evolution modes involve scaffolding on the part of the teacher that is somewhere in between these open and authoritative approaches. We have described these two newly identified modes as the core of a teaching approach that we call "guided co-construction." As a result, we believe that the present approach may involve a longer delay of closure than in Scott et al.'s case. But the intermediate approach there is not just a simple blend mixing open student discussion and teacher lecturing.

FIGURE 7 | Cognitive scaffolding of student modeling helps resolve tensions shown in **Figure 1**.
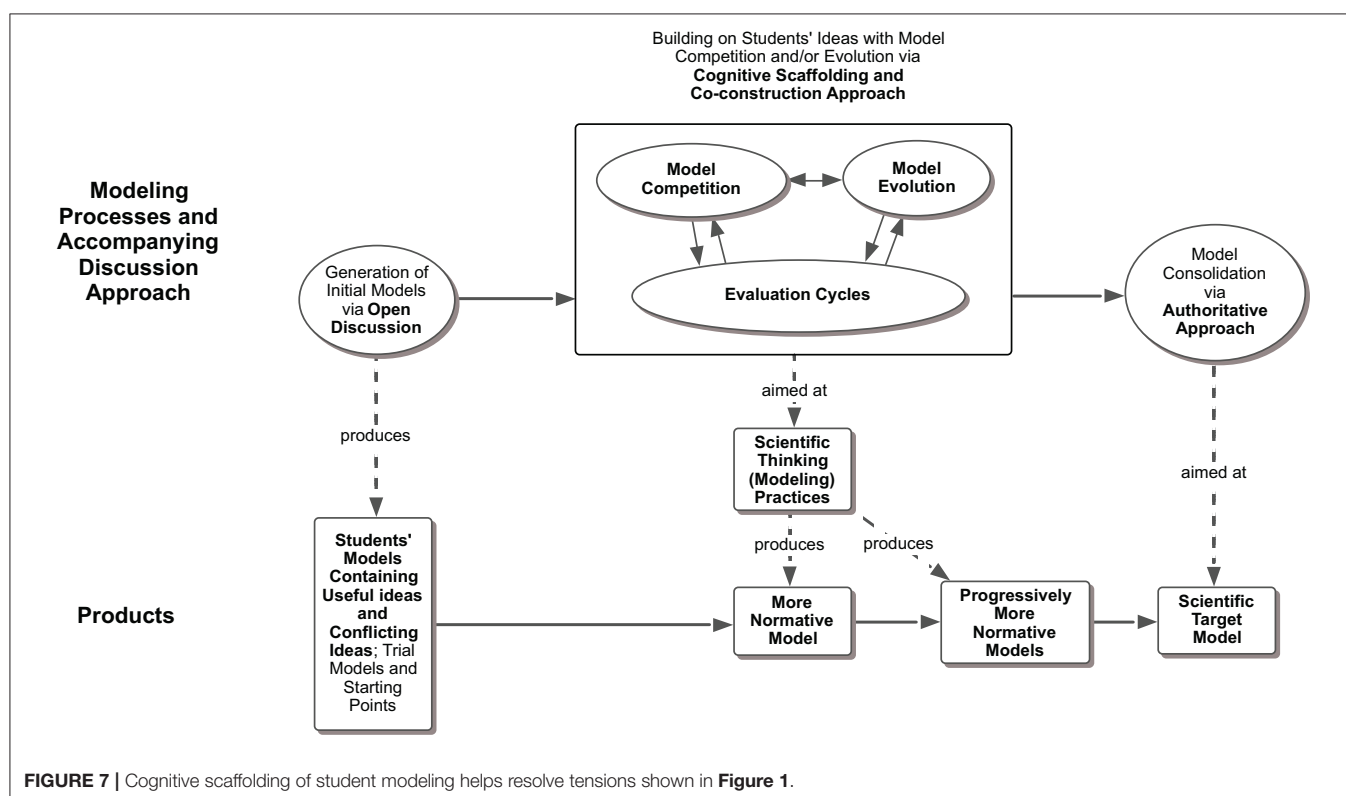
As shown in the box between these in **Figure 7**, and under Competition and Evolution in **Figure 4B**, we have tried to unpack an impressive set of nested processes that the teacher is supporting through a particular kind of scaffolding we call *cognitive scaffolding.*

In order to first say what we do *not* mean by cognitive scaffolding, researchers such as van Zee and Minstrell (1997b) and Williams and Clement (2015) have described less content specific and less cognitive forms of teacher support for participation in discussions such as paraphrasing student statements for clarity, using wait time, probing for clarification, and providing norms for respectful discussion. These are important general strategies for keeping any discussion going and fostering participation. In contrast to those moves we can define cognitive scaffolding as including moves that foster, guide, or support students' content-specific reasoning or idea formation about the topic of the lesson. This kind of scaffolded reasoning can take place for example through specific teacher questioning that involves students in doing model evaluation and modification. The goal is to provide just enough support to keep students in a zone where they are able to participate in the reasoning processes (here model construction). This is a broad definition of "scaffolding," because it does not include the idea of withdrawing support gradually, which is sometimes included, but not as relevant in this study over such a short time period. Scott (1998, 68–72) theorized that effective scaffolding involves a feedback loop that contains three steps (1) analyze the learner's situation; (2) assist the learner by using pedagogical means; (3) monitor the learner's progress. The present paper focuses on unpacking step (2).

Even within the Evolution and Competition Modes, cognitive scaffolding can vary from strong to weak. Some teacher statements classified as scaffolding model evaluation in **Table 4** for example are quite subtle, sometimes merely repeating a student's words in what van Zee and Minstrell (1997b) called a "reflective toss," and leaving most of the evaluation and modification process to be done by the student (e.g., Topics 5 to 8). Whereas in topic 15 in **Table 6**, much more, but not all, of the reasoning in the evaluation and modification phases is done by or strongly hinted at by the teacher. Williams and Clement (2019) found that successful teachers can differ strongly as to the ratio of teacher initiated to student initiated modeling moves (e.g., model generation, modification, or evaluation). Thus we can identify a spectrum of approaches from open inquiry, to scaffolded inquiry with various degrees of cognitive scaffolding, to recitation, to lecture. This spectrum of approaches along with the analysis of purposes and substrategies for each may help mitigate the dilemma of having only two opposite approaches and not knowing when to switch, as depicted in **Figures 1**, 7.

2. **Conceptual Dissonance Tension** Can a class elicit student-generated models that may conflict with the target model in a number of ways, and arrive at the target model through discussion? How? Won't students' faulty models interfere? (see **Figure 1**).

Once the teacher opens up the classroom to student generated models and ideas, it is true that a divergent variety of models can emerge. This is uncomfortable for many teachers. In this case study the students generated at least three models that conflicted with the scientific target model. However, two of them were disconfirmed with reasoned plausibility arguments, some drawn from the student and some from the teacher. Then evaluation and revision cycles evolved the third model until it was close to the scientific target concept. As shown at the bottom of **Figure 7**, such a sequence of progressively more normative models can bridge the gap between initial ideas and a scientific target model. Importantly this should allow the student to use some elements of their prior knowledge as meaningful building blocks, while disconfirming other elements as not relevant. We hypothesize that this grounding in prior knowledge may have advantages for imageability, meaningfulness, and memorability (Ausubel, 1968). In summary, the teacher did not consider these students' incorrect models as interfering with their learning. Instead, she engaged students in reasoning about why some were less viable and used others as a stepping stone to build toward the target. This is quite different than just juxtaposing them with the scientific target model in a lecture.

3. **Lesson Objective Tension** Can a class foster model construction (reasoning) practices, as a type of scientific thinking, at the same time that they are learning science content? How? (see **Figure 1**).

In the USA, despite calls from NGSS to integrate the teaching of disciplinary core ideas, practices, and cross-cutting concepts, many teachers still think about content goals separately from scientific thinking goals. The teacher in this study, like many teachers, also had a goal of engaging students in scientific thinking practices. However, given time pressures, such a goal is often seen as in conflict with content goals and may be neglected. But as shown in **Figure 7**, in these lessons the teacher appeared to foster scientific modeling practices as a *means* to arrive at the target model, a content goal. By starting from expert modeling practices in this study we have seen how this teacher scaffolded basic science practices while simultaneously guiding students toward the target model. This gave the students experience with the ideas that models can be invented, can compete, can be disconfirmed, can be evaluated and modified by asking challenging questions, can be confirmed, and can be transferred to new contexts. Here, the method of learning content *was* scientific thinking.

Limitations

Figure 4B is a simplified representation of sequences (horizontal arrows) and subprocesses (diagonal arrows) that we have found. However, there are certainly variations and exceptions to the sequence, as is partly indicated by the double horizontal arrows there. As we saw in Topics 1–8, the teacher may not complete every mode in a sequence before starting a new sequence.

We would describe many of the teacher's actions during the competition and evolution modes as "moderately strong

scaffolding." There are certainly other approaches that could have been taken here, some with stronger directness and some with less. Which produce more content learning and which produce more learning of scientific thinking are important issues that need further research.

This paper's focus is on the qualitative objective of unpacking and sorting out connections between two nested levels of processes occurring in classroom discussions. It does not consider the measurement of gains in comprehension. One study of very similar types of scaffolded discussions in high school physics, which fostered GEM cycles, did measure significant gain differences over controls in comprehension with a large effect size (Williams and Clement, 2015). But it did not identify connections to higher level Mode processes as we have done in this study.

We are also interested in finer-grained levels of processes below those shown in **Figure 4B**, but that is a large topic on its own for another paper. For example, the figure shows "Evaluatory Observations" as a process contributing to model evaluation, but there are other subsidiary processes for model evaluation such as thought experiments and coherence criteria (see Williams and Clement, 2015).

We should note that we did not find evidence of deep seated, persistent alternative conceptions in this case study. Units dealing with such conceptions will need to use multiple methods and revisit them over a longer time period, sometimes much longer (see Clement and Steinberg, 2002; Kalman and Lattery, 2018, 1; Lattery, 2016; Minstrell and Kraus, 2005).

INSTRUCTIONAL IMPLICATIONS

Positive Features of the Classroom Discussions

There were several positive features of the classroom dialog diagram analyzed:

(1) This was an advanced topic for 12 year olds, but students participated in a *co-construction* process, where they contributed ideas along with the teacher. Perhaps the most radical feature was that almost all of the initial model generation ideas came from the students before receiving instruction on them. Both the teacher and the students contributed to modifying ideas (see also Williams and Clement, 2019). Most model evaluations were initiated by teacher questions, with students contributing to the reasoning especially in the first two thirds of the transcript. The student contributions are key for their experiencing scientific modeling practices while learning about a scientific subject, and a source of any benefits from active learning.

(2) Evidence for student *engagement* in these lessons was provided by the large number of on topic contributions, the variety of models generated by the students and the inventiveness some of them display. Students' ideas such as placing the capillaries with respect to villi into three different positions (outside of villi, on top of villi, and inside of villi), illustrate the variety. In addition, some were quite inventive, e.g., that "villi looked like

hands that grab things," that "villi were like filters," and transferring the concept of semi-permeable membrane from cellular exchange in other parts of the body to the villi. There is wide agreement on the importance of engagement (see Resnick, 1999; Engle and Conant, 2002; van Zee and Minstrell, 1997a).

(3) We also see this teacher as doing *"responsive teaching."* The teacher skillfully navigated the class through all of the Modes shown in **Figure 4B**, as scientific practices, by appropriately using individual, small group, and whole class work. The teacher appeared to order the large scale modes of teaching depending on the topic, and the spontaneous models generated by students. For example, when students generated a wider variety of models, she fostered model comparisons (Competition). In addition she appeared to scaffold the lower level model construction phases in **Figure 4B** repeatedly. Since these are shorter processes, they require faster teacher decisions depending on what the students said.

Negative Features of the Classroom Discussions

There were also some negative features in our view: (1) many of the student answers to the teacher's questions were short. This meant that their opportunities for expression were not as great as they could have been. No doubt the teacher had a tradeoff with time on her mind. (2) In the last class the teacher was definitely running out of time. This meant that near the end she did not foster as many student contributions as she might have done; (3) the last section in Model Application and/or Domain Extension Mode was consequently quite short, and although students made a connection to the new topic of gas transfer in the lungs, we assume that this was too brief a segment for most to develop deep conceptual understanding of that area.

Time is unfortunately scarce in today's classrooms. Model based learning can take longer than lecture-based approaches. Resulting increases in conceptual understanding should save time later, but teachers under real institutional pressure to cover wide content may need to "pick their fights" in choosing which content areas they think are most valuable for significant student modeling practices. Driver and Scott (1996, P. 624) suggested prioritizing interactive ways of teaching when detecting strong differences or gaps between students' initial ideas and the scientific model, in which case stepwise model construction should be even more important for understanding. One tactic we have observed at the college level is to do modeling activities in small and large group in class but then assign readings and problem solving for the Consolidation and Application Modes. More innovation and research work is needed here.

General Instructional Implications for Teacher Education

In addition to being a lens for lesson microanalysis, we can consider whether the process patterns we have identified suggest

a set of model development strategies for teachers in teacher education courses. There are many teachers who do not conduct the kind and level of responsive teaching we saw in the present protocol. For those who want to learn to incorporate students' ideas, once student generated models have been admitted into classroom discussions, the teacher can be unsure of how to deal with the divergent variety of student ideas, and there is a need to have some strategies/guidelines for thinking about how to scaffold further modeling. We hypothesize that these teachers might learn to use the six Major Modeling Modes and their nested Model Construction (GEM) Processes in **Figure 4B** as large and medium scale strategies for scaffolding modeling. Although the items in that figure were described as student processes that were fostered in the classroom, each process can also be seen to identify a corresponding teaching strategy of scaffolding that particular process.

The larger time scale Major Modeling Modes there could provide an organization for design at a unit level as a "modeling sequence pattern." As we saw, this modes sequence can be repeated in a cycle for each major piece of a model in a unit. That would allow students to gain experience with modeling practices across different contexts. At the lesson design level, projecting from the framework in **Figure 4B** and the examples in the case study we can speculate that the different modeling modes may benefit from different styles of discussion leading. The Generating Initial Model(s) Mode would appear to benefit from an open style. The teacher used individual, small group, and whole class discussion formats for this (with the teacher mainly restricted to drawing out, rather than evaluating, ideas). The Competition Mode required more scaffolding, with the teacher clarifying the differences between models and prompting students to evaluate the different models. The teacher used both small group and whole class discussion for this. Model Evolution Mode required perhaps the most scaffolding and the most skill on the part of the teacher because the teacher will need to creatively figure out how to evolve certain models toward the target model through questioning. This may best be done in whole class discussion. Thus, it should be possible to use the framework for unit and lesson design, and as a guide to using different teaching styles at different times. The distinction between the Brainstorming Mode and subsequent Model Evolution or Model Competition Modes is important, because teachers using this approach could then learn first to withhold judgements and hints and the providing of correct answers in Brainstorming Mode. Only after practicing that would they build on it to add the more difficult skills of scaffolding Model Evolution or Competition.

However, the strategy sequence in **Figure 4B** is nothing like a full algorithm; a teacher who bravely opens the floor to discussion can receive many student ideas with varying degrees of distance from the target model, and they will have to make decisions about which to take up and in what order. This is part of the art of responsive discussion leading and it is a skill that takes a long time to learn. We believe it would need to be learned slowly, ideally with a support group in an in-service course. Scaffolding strategies would need to be simplified and introduced one major piece or level at a time (see Price et al., 2017; Krajcik and Merrit

(2012, 11–12); Williams and Clement, 2015; and Stephens and Clement, 2010).

Near the end of the sequence, in Consolidation Mode, the teacher consolidated and confirmed the target model for the class using mostly a mini-lecture style. But why, one may ask, make the effort to elicit and work with students' ideas if the teacher is going to present and confirm the target model at the end anyway? Isn't this inefficient? Certainly one important reason to do so is to pursue scientific thinking goals in addition to content goals. But even for content goals, the potential of the present method for fostering deeper conceptual understanding would seem to lie in eliciting students' ideas and engaging them in thinking, allowing them to build on their connections to prior knowledge, to talk about and evaluate how various models function dynamically and experience cognitive dissonance with some of them, to build difficult models more slowly with understanding, and to see *why* certain models are better than others. On the other hand, lectures can sometimes be very inefficient, either if the concepts are presented too quickly because the teacher does not have feedback from student discussion, or if the students do not discuss and make sense of the given information, or if they do not engage in active learning with the ideas.

We can speculate that the strategies in **Figure 4B** may apply to fields outside of science education. For example, historians generate and revise models. And the general strategies for designing a scientific model should not be far removed from those for designing systems in engineering, although some of the criteria for evaluation may be different, and the mode sequence would start from a problem to be solved, rather than a pattern to be explained.

CONCLUSION

We began this article by reviewing previous work identifying many individual processes involved in scientific modeling in classrooms, with a focus on whole class discussions. However, this work still lacked an overall coherent framework for how these processes fit together. Imposing the constraints of accounting for each episode in the microanalysis of a case study in a real classroom allowed us to identify a coherent set of modeling practices at two nested levels, summarized in abbreviated form by the scheme shown in **Figure 4B**. Most of those processes are similar to those found in recent studies of the modeling practices of expert scientists. Each process in the Framework can also be viewed as designating a strategy for scaffolding modeling.

Figure 6 shows the two levels of processes operating in parallel in a classroom interaction style we call teacher-student co-construction. **Figure 7** indicates how the framework strategies have the potential to remove or reduce the tensions described in **Figure 1**- tensions felt, we believe, by any teacher beginning to open up their classrooms to real modeling discussions:

- A teacher need not be limited to the two opposing interaction styles of Open Discussion vs. Authoritative lecture. Rather, there are there intermediate discussion

styles between these that involve co-construction and cognitive scaffolding.

- It is possible to start from student-generated models that conflict with the target model in a number of ways, and still arrive at a target model in a discussion. Processes of model competition and disconfirmation, as well as model evolution, both supported by the teacher's cognitive scaffolding, were central in this accomplishment in the case study.
- In doing so, it is possible for a teacher to foster student modeling practices, as a type of scientific thinking, at the same time that they are teaching science content.

This study is intended as a starting point for developing a more adequate picture of the modeling practices and scaffolding strategies involved in discussions for learning science. We look forward to evaluating and modifying elements of the theory as more studies are completed by ourselves and others.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

REFERENCES

Ausubel, D. (1968). *Educational Psychology: A Cognitive View*. New York, NY: Holt, Rinehart and Winston.

Brewe, E. (2008). Modeling theory applied: modeling Instruction in introductory physics. *Am. J. Phys.* 76, 1155–1160. doi: 10.1119/1.2983148

Camp, C. W., Clement, J., Brown, D., Gonzalez, K., Kudukey, J., Minstrell, J., et al. (1994). *Preconceptions in Mechanics: Lessons Dealing With Conceptual Difficulties, 2nd Edn*. College Park, MD: American Association of Physics Teachers.

Campbell, N. R. (1920). *Physics. The Elements*. Cambridge: Cambridge University Press; Republished in 1957 as The foundation of science. New York, NY: Dover.

Campbell, T., Oh, P. S., and Neilson, D. (2012). Discursive modes and their pedagogical functions in Model-Based Inquiry (MBI) classrooms. *Int. J. Sci. Educ.* 34, 2393–2419. doi: 10.1080/09500693.2012.704552

Chin, C. (2007). Teacher questioning in science classrooms: approaches that stimulate productive thinking. *J. Res. Sci. Teach.* 44, 815–843. doi: 10.1002/tea.20171

Clement, J. (1989). "Learning via model construction and criticism. Protocol evidence on sources of creativity in science," in *Handbook of Creativity: Assessment, Theory And Research*, eds G. Glover, R. Ronning, and C. Reynolds (New York, NY: Plenum), 341–381. doi: 10.1007/978-1-4757-5356-1_20

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *J. Res. Sci. Teach.* 30, 1241–1257. doi: 10.1002/tea.3660301007

Clement, J. (2000). Model based learning as a key research area for science education. *Int. J. Sci. Educ.* 22, 1041–1053. doi: 10.1080/095006900416901

Clement, J. (2008a). *Creative Model Construction in Scientists and Students. The Role of Imagery, Analogy, and Mental Simulation*. (Netherlands: Springer). doi: 10.1007/978-1-4020-6712-9

Clement, J. (2008b). "Six levels of organization for curriculum design and teaching," in *Model Based Learning and Instruction in Science*, eds J. Clement and M. A. Rea-Ramirez (Dordrecht: Springer), 260–277. doi: 10.1007/978-1-4020-6494-4_14

Clement, J. (2013). "Roles for explanatory models and analogies in conceptual change," in *International Handbook of Research on Conceptual Change*, ed S. Vosniadou (New York, NY: Routledge), 412–446. doi: 10.4324/9780203154472.ch22

Clement, J. (2017). *Four levels of scientific modeling practices in expert learning. Paper presented at the National Association for Research in Science Teaching (NARST)*, San Antonio, TX. Available online at: https://cadrek12.org/resources/four-levels-scientific-modeling-practices-expert-learning (accessesd June 27, 2017).

Clement, J., and Steinberg, M. S. (2002). Step-wise evolution of mental models of electric c ircuits: a learning-aloud case study. *J. Learn. Sci.* 11, 389–452. doi: 10.1207/S15327809JLS1104_1

Darden, L. (1991). *Theory Change in Science. Strategies From Mendelian Genetics*. Oxford: Oxford University Press.

Driver, R., and Scott, P. H. (1996). "Curriculum development as research: a constructivist approach to science curriculum development and teaching," In *Improving Teaching and Learning in Science and Mathematics*, eds F. David Treagust, R. Duit and J. Barry Fraser (New York, NY: Teachers College Press), 94–108.

Engle, R. A., and Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: explaining an emergent argument in a community of learners classroom. *Cogn. Instr.* 20, 399–483. doi: 10.1207/S1532690XCI2004_1

Engle, R. A., Conant, F. R., and Green, D. W. (2007). "Progessive refinement of hypotheses in video-supported research," in *Video Research in the Learning Sciences*, eds R. Goldman, R. Pea, B. Barron and Sharon J. Derry (Mahwah, NJ: Erlbaum), 240–243.

Hammer, D. (1995). Students inquiry in a physics class discussion. *Cogn. Instr.* 13, 401–430. doi: 10.1207/s1532690xci1303_3

Harré, R. (1961). *Theories and Things, History and Philosphy of Science*. London: Newman.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *Am. J. Phys.* 55, 440–454. doi: 10.1119/1.15129

Hogan, K., and Pressley, M. (eds.). (1997). *Scaffolding Student Learning: Instructional Approaches and Issues*. Cambridge, MA: Brookline Books.

Kalman, C. S., and Lattery, M. J. (2018). Three active learning strategies to address mixed student epistemologies and promote conceptual change. *Front. ICT* 5:19. doi: 10.3389/fict.2018.00019

Khan, S. (2008). "Co-construction and model evolution in chemistry," in *Model Based Learning and Instruction in Science*, eds J. Clement and M. A. Rea-Ramirez (Springer), 68–87. doi: 10.1007/978-1-4020-6494-4_4

Krajcik, J., Codere, S., Dahsah, C., Bayer, R., and Mun, K. (2014). Planning instruction to meet the intent of the next generation science standards. *J. Sci. Teacher Educ.* 25:157–175. doi: 10.1007/s10972-014-9383-2

Krajcik, J., and Merrit, J. (2012). Engaging students in scientific practices: what does constructing and revising models look like in the science classroom. *Sci. Child.* 49, 10–13. Available online at: https://www.jstor.org/stable/43747325

Kuhn, T. S. (1970). "Foundations of the Unity of Science," in *The Structure of Scientific Revolutions*, *2nd Edn*, eds O. Neurath, R. Carnap and C. Morris, Vol. I and II (Chicago, IL: The University of Chicago Press), 18–20; 49–50; 101–102.

Lattery, M. J. (2016). "Exploratory studies of Model-Based Reasoning," in *Deep Learning in Introductory Physics; Science & Engineering Education Sources,* ed S. Calvin Kalman (Charlotte, NC: Information AGE Publishing, Inc.).

Lehesvuori, S., Viiri, J., Rasku-Puttonen, H., Moate, J., and Helaakoski, J. (2013). Visualizing communication structures in science classrooms: tracing cumulativity in teacher-led whole class discussions. *J. Res. Sci. Teach.* 50, 912–939. doi: 10.1002/tea.21100

Louca, L. T., Zacharia, Z. C., and Tzialli, D. (2012). Identification, Interpretation—evaluation, response: an alternative framework for analyzing teacher discourse in science. *Int. J. Sci. Educ.* 34, 1823–1856. doi: 10.1080/09500693.2012.671971

Machamer, P., Darden, L., and Craver, C. F. (2000). Thinking about mechanisms. *Philos. Sci.* 67, 1–25. doi: 10.1086/392759

Miles, M. B., and Huberman, M. A. (1994). *Qualitative Data Analysis. An Expanded Sourcebook*. Thousand Oaks, CA: Sage.

Minstrell, J., Anderson, R., and Li, M. (2011). "Building on learner thinking: A framework for assessment in instruction," in *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*, ed Committee on Highly Successful Schools or Programs in K-12 StEM Education; National Research Council (Washington, DC: National Academies Press), 17.

Minstrell, J., and Kraus, P. (2005). "Guided inquiry in the science classroom," in *How Students Learn: History, Mathematics, and Science in the Classroom*, edited by M. Suzanne Donovan and D. John Bransford (Washington, DC: The National Academies Press), 475–513.

Nassaji, H., and Wells, G. (2000). What's the use of 'Triadic Dialogue'?: An investigation of teacher-student interaction. *Appl. Linguist.* 21, 376–406. doi: 10.1093/applin/21.3.376

Nersessian, N. J. (1995). Should physicists preach what they practice? Constructive modeling in doing and learning physics. *Sci. Educ.* 4, 203–226.

Nersessian, N. J. (2008). *Creating Scientific Concepts*. Cambridge MA: MIT Press. doi: 10.7551/mitpress/7967.001.0001

NGSS, Lead States (2013). *Next Generation Science Standards: For States, by States*. Washington, DC: The National Academies Press.

Nunez-Oviedo, M. C. (2004). *Teacher-Student Co-construction Processes in Biology: Strategies for Developing Mental Models in Large Group Discussions*. Amherst: Ed.D., TECS, Massachusetts (Doctoral Dissertations Available online at: https://scholarworks.umass.edu/dissertations/AAI3118319).

Nunez-Oviedo, M. C., and Clement, J. (2008). "A competition strategy and other modes for developing mental models in large group discussion," in *Model Based Learning and Instruction in Science*, edited by J. Clement and M. A. Rea-Ramirez (Dordrecht: Springer), 127–148. doi: 10.1007/978-1-4020-6494-4_7

Nunez-Oviedo, M. C., Clement, J., and Rea-Ramirez, M. A. (2008). "Developing complex mental models in Biology through model evolution," in *Model Based Learning and Instruction in Science,* edited by J. Clement and M. A. Rea-Ramirez (Dordrecht: Springer), 182–202. doi: 10.1007/978-1-4020-6494-4_10

Nunez-Oviedo, M. C., Rea-Ramirez, M. A., Clement, J., and ane Else, M. (2002). "Teacher-student co-construction in middle school life science," in *Proceedings of the AETS Conference*. (Charlotte, NC). Available online at: https://files.eric.ed.gov/fulltext/ED465602.pdf

Price, N., Lynn Stephens, A., Clement, J., and Nunez-Oviedo, M. C. (2017). Using imagery support strategies to develop powerful imagistic models. *Sci. Scope* 41, 40–49. doi: 10.2505/4/ss17_041_04_40

Rea-Ramirez, M. A. (1998). *Models of conceptual understanding in human respiration and strategies for instruction* (Doctoral Dissertations). Ed. D., School of Education, University of Massachusetts, Amherst, MA, United States. Available online at: https://scholarworks.umass.edu/dissertations/AAI9909208

Rea-Ramirez, M. A., Nunez-Oviedo, M. C., Clement, J., and Jane Else, M. (2004). *Energy in the Human Body Curriculum*. Amherst: National Science Foundation; University of Massachusetts. Available online at: http://www.cesd.umass.edu/energyinthehumanbody/

Reiser, B. J., Berland, L. K., and Kenyon, L. (2012). Engaging students in the scientific practices of explanation and argumentation. Understanding a framework for K-12 science education. *Sci. Child.* 35, 8–13. Available online at: https://www.jstor.org/stable/43747348

Resnick, L. (1999). Making america smarter. *Educ. Week* 18, 38–40.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *J. Res. Sci. Teach.* 46, 632–654. doi: 10.1002/tea.20311

Scott, P. H. (1992). "Pathways in learning science: a case study of the development of one student's ideas relating to the structure of matter," in *Research in Physics Learning: Theoretical Issues and Empirical Studies. Proceedings of an International Workshop held at the University of Bremen*, eds R. Duit, F. Goldberg, and H. Niederer (Kiel: IPN/Institute for Science Education), 203–224.

Scott, P. H. (1998). Teacher talk and meaning making in science classrooms: a Vygotskian analysis and review. *Stud. Sci. Educ.* 32, 45–80. doi: 10.1080/03057269808560127

Scott, P. H., Mortimer, E. F., and Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: a fundamental characteristic of meaning making interactions in high school science lessons. *Sci. Educ.* 90, 605–631. doi: 10.1002/sce.20131

Smith, J. P. I. I. I., diSessa, A. A., and Roschelle, J. (1994). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *J. Learn. Sci.* 3, 115–163. doi: 10.1207/s15327809jls0302_1

Steinberg, M. S., and Wainwright, C. L. (1993). Using models to teach electricity -the CASTLE project. *Phys. Teach.* 31, 353–357. doi: 10.1119/1.2343798

Stephens, A. L., and Clement, J. (2010). Documenting the use of expert scientific reasoning processes by high school physic students. *Phys. Rev. Spec. Phys. Educ. Res.* 6, 1–15. doi: 10.1103/PhysRevSTPER.6.020122

Toulmin, S. (1972). *Human Understanding. The Collective Use and Evolution of Concepts*, Vol. I. London: Oxford University Press.

van Zee, E. H., and Minstrell, J. (1997a). Reflective discourse: developing shared understanding in a physics classroom. *Int. J. Sci. Educ.* 19, 209–228. doi: 10.1080/0950069970190206

van Zee, E. H., and Minstrell, J. (1997b). Using questioning to guide student thinking. *J. Learn. Sci.* 6, 227–269. doi: 10.1207/s15327809jls0602_3

White, B. Y., and Frederiksen, J. R. (2000). "Metacognitive facilitation: An approach to making scientific inquiry accessible to all," in *Teaching in the Inquiry-Based Science Classroom*, eds J. Minstrell and E. van Zee (Washington, DC: American Association for the Advancement of Science), 20–21.

Williams, G., and Clement, J. (2015). Identifying multiple levels of discussion-based teaching strategies for constructing scientific models. *Int. J. Sci. Educ.* 37, 82–107. doi: 10.1080/09500693.2014.966257

Williams, G., and Clement, J. (2019). "Co-constructing models through whole class discussions in high school physics," in *Physics Teaching and Learning: Challenging the Paradigm,* eds D. Sunal, J. Shemwell, J. Harrell, and C. Sunal (Charlotte, NC: Information Age Publishing), 102–103.

Windschitl, M., Thompson, J., Braaten, M., and Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Sci. Educ.* 96, 878–903. doi: 10.1002/sce.21027

Learning Math: Two Principles to Avoid Headaches

Felipe Munoz-Rubke*, Daniela Vera-Bachmann and Alejandro Alvarez-Espinoza

Instituto de Psicología, Universidad Austral de Chile, Puerto Montt, Chile

INTRODUCTION

For the last 10 years South American nations have finished in mid to bottom positions in the Programme for International Student Assessment (PISA) math test, significantly behind dozens of countries around the globe. Regrettably, the lack of improvement over the past decade does not depict an optimistic future for this region (OECD, 2017). To reverse this trend, we believe that the recognition and adoption of two key principles could lead to substantial improvements in early math education: first, valuing each student's intuitive math knowledge; and second, focusing on the role that spatial skills play in learning math. We also suggest that both principles could be simultaneously put into practice by utilizing diagrams for teaching early mathematics.

Research shows that multiple interrelated factors explain the poor performance of South American students in mathematics (Cerda et al., 2017). Poverty remains one of the most notable obstacles (Hanushek and Luque, 2003; Kainz, 2019), though other variables at the school level are also relevant. Among these are each school's social climate and educational perspectives (Macneil et al., 2009; Gálvez-Nieto et al., 2015) and each country's public policies in education (Vegas and Petrow, 2007), just to mention a few. Despite this, research shows that the effectiveness of each school is mostly determined by their teachers; teachers' training, knowledge, and beliefs about how to teach mathematics seem to be more relevant than any other factor (Ball et al., 2008; Mapolelo and Akinsola, 2015).

At this level, two principles could be incorporated into early math teaching. Both are supported by considerable evidence and could reduce the sometimes painful experience of learning math. The first principle states that a strong understanding of early mathematics can be built using children's intuitive mathematical ideas as a foundation. This principle mirrors Vygotsky's ideas concerning the bridge that should exist between formal and spontaneous concepts, as the former operates as a zone of proximal development (ZDP) for the latter (Vygotski, 2001). The ZDP corresponds to the distance between current performance under no guidance and potential performance with guidance, and it highlights the linkage between what is currently known and what could be known provided enough support. The second principle indicates that students' spatial skills can influence how much they will get to enjoy and succeed in mathematics. Although there is evidence highlighting the importance of spatial skills in math performance, South American schools have yet to include spatial training in their academic curricula.

Improving math education is important because it could promote the development of South American countries by strengthening their human capital. It is imperative to have more and better professionals in Science, Technology, Engineering, and Mathematics (STEM), who can tackle the challenges that countries face in an increasingly complex and fast-changing economy (Schwab, 2017).

ROLE OF CHILDREN'S INTUITIVE IDEAS ON EARLY MATH LEARNING

Everyday mathematics refers to the use of intuitive mathematical notions in real-life contexts. In these situations, people are not directly concerned by specific mathematical principles, but instead use raw intuition to solve applied problems.

There are opposing perspectives on the role that everyday mathematics plays in formal learning. While some researchers see it as a foundation on which students can build meaningful understandings of concepts, others regard it as a source of interference (Carraher and Schliemann, 2002). These opposing views reflect differences in the social valuation of everyday experiences and academic practices (Civil, 2016).

Since the 1980s, researchers have highlighted the role that intuitive mathematical knowledge can play in improving school mathematics, especially in generating more meaningful learning experiences for students (Carraher et al., 1985; Wager, 2012). A seminal study by Carraher et al. (1985) illustrated the use of everyday mathematical knowledge by Brazilian children and adolescents working as street vendors. These participants demonstrated advanced proficiency in solving arithmetic problems when dealing with complex economic transactions, despite their lack of formal mathematical training. Interestingly, these participants made significant mistakes when attempting to solve similar mathematical problems through the traditional algorithmic procedures taught in schools. This disparity in performance made the investigators wonder how it was possible for participants to show high proficiency in one context, and a lack of it in another. In a follow-up study, the investigators showed that meaningful contexts, like those experienced by the street vendors, tend to evoke alternative problem solving strategies based on simple yet powerful heuristics (Carraher et al., 1987).

Previous research has also shown the benefits of Cognitively Guided Instruction (CGI), a professional development program for teachers that underscores the role of children's intuitive ideas in early math education (Carpenter and Fennema, 1992; Carpenter and Franke, 2004). This program does not encourage the application of specific instructional methodologies, but instead stimulates appreciation for the diverse problem solving strategies and distinct understandings that students have of mathematical principles. Upon acknowledging that students are active creators of their own knowledge (Cobb, 1988), CGI teachers ask children to explain their problem solving strategies, familiarize themselves with each children's preferred problem solving approaches, and promote the use of various problem solving methods (Carpenter et al., 1989; Peterson et al., 1989a,b). These behaviors positively correlate with students' problem solving performance.

ROLE OF VISUOSPATIAL THINKING ON EARLY MATH LEARNING

Spatial skills play an important role in STEM disciplines. Longitudinal studies have shown that people with higher spatial skills tend to enjoy, choose, and succeed in STEM areas (Shea et al., 2001; Wai et al., 2009; Lubinski, 2010).

For a long time, spatial abilities were seen as a stable and unmodifiable human trait (Newcombe, 2014). However, multiple investigations contradict this assumption. A recent meta-analysis summarizing the results of more than 200 studies showed that spatial abilities are malleable, that spatial training can promote long-lasting effects, and that training one specific ability can result in the enhancement of other untrained spatial skills (Uttal et al., 2013).

Some studies have focused on the positive direct effects that spatial training can have on mathematical learning. For instance, Cheng and Mix (2012) showed that mental rotation practice can lead to an increase in numerical calculation among 6 and 8 year-olds. In a more natural setting, Lowrie et al. (2017) implemented a 10 week spatial training program in the classrooms of 10-to-12 year-old students. The interventions were implemented by teachers and encompassed the direct training of different spatial skills like mental rotation, spatial orientation, and spatial visualization. Students who underwent this spatial intervention program increased both their spatial and mathematical skills more than the students who were part of the control group. A study by Hawes et al. (2017) used a somewhat different strategy. Instead of training spatial abilities directly, they created spatial games and dynamics to teach mathematical concepts. Their results suggested significant increases in spatial language, spatial reasoning, and numerical comparison following the intervention.

Although we do not yet have a complete understanding of the mechanisms linking spatial abilities and mathematical performance, some studies have already provided hints. A study by Hegarty and Kozhevnikov (1999) suggested that not all types visuospatial representations promote math problem solving. In their study, two visuospatial strategies were contrasted: one based on spatial-schematic imagery and another based on visual-pictorial imagery. Spatial-schematic imagery was defined as the creation of representations that included information about the parts of objects, their spatial relation to other objects, and their respective locations in space. Visual-pictorial imagery was defined as the creation of representations centered on the visual appearance of objects, including properties such as color and shape. Results showed that the use of spatial-schematic strategies, but not of visual-pictorial strategies, was associated with a higher rate of success in mathematical problem solving.

THE BRIDGE BETWEEN BOTH PRINCIPLES: USING DIAGRAMS FOR MATH PROBLEM SOLVING

The two aforementioned principles come together into a single pedagogical practice when diagrams are used to support math problem solving. This is by no means a new idea, as this methodology has been implemented in the educational systems of both Singapore (Ng and Lee, 2009; Kaur, 2018) and Japan (Murata, 2008), countries with outstanding international performances in mathematics.

Diagrams are visuospatial representations that depict significant information in a spatial display. Because diagrams are more abstract than objects/manipulatives but more concrete than mathematical symbols, they can provide a valuable bridge between initial and advanced learning stages. In their role as intermediate-level representations, they highlight relationships that could be difficult to spot in higher-level symbolic equations, particularly for novices (for an example, see **Figure 1**). This is of importance for early math students that are just becoming familiar with the discipline and who often struggle with abstract conceptualizations.

Previous research shows that diagrams encourage the use of alternative, intuitive problem solving strategies. For instance, they can facilitate the application of children's intuitive mathematical ideas during early arithmetic lessons and more advanced algebra lessons (Edens and Potter, 2008; Murata, 2008; Chu et al., 2017). For instance, in a study that included a brief intervention targeted at teaching seventh-grade American students to use diagrams to solve algebra problems, Chu et al. (2017) found that diagrams favored the utilization of informal problem solving strategies and led to significant gains in solving accuracy.

The role of diagrams as visual-spatial representations that favor the use of intuitive problem solving strategies is stressed by concreteness fading, a theory of instruction based on the ideas of Bruner (1966) and subsequently developed by Fyfe and Nathan (2018). This theory suggests that the best way to achieve a deeper understanding of a concept is to first ground it at a concrete level, and to then progressively expose the learner to more abstract instances of it. In the first representational level, interactions with objects and places represent the relevant concept (e.g.,

learning subtraction by counting apples). During the second representational level, students deal with representations that are more abstract but that still resemble concrete objects, places, and their relationships (e.g., learning subtraction by using diagrams). The third representational level corresponds to the symbolic stage, in which the representations have no obvious relation with objects and spaces (e.g., learning subtraction by using numbers).

Similar ideas applied to learning geometry have been endorsed by Battista (2007), who suggests that students should move from visualization, to abstraction, to formal deduction, until reaching higher mathematical rigor.

CONCLUSION

South American early math education could be improved through the adoption of these two central principles. The first principle indicates that learning formal concepts becomes more meaningful when teachers integrate what children already know. The second principle indicates that spatial abilities have a strong and positive effect on both the motivation to learn math and math performance itself. The evidence points out that spatial training at an early age can lead to improvements in the mathematical performance of students. While most early education programs consider the development of language and math skills, the development of spatial thinking has not received systematic attention.

Both principles are integrated into math problem solving through the use of diagrams. Diagrams, as intermediate representations between the concrete and the abstract, are highly effective in the development of mathematical learning. The
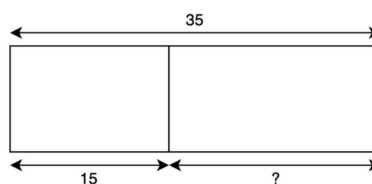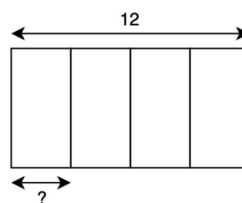

FIGURE 1 | Diagrams showing part-whole models. Both figures are based on Kaur's ideas (Kaur, 2018).

attractiveness and simplicity of diagrams can make it easier for children to build meaning around mathematical activity. That is, students can link abstract concepts with elements of their own experience in a way that allows the appropriation of concepts.

Although the success of Singapore and Japan in mathematics is certainly the result of multiple features, evidence suggests that incorporating the methodical use of diagrams during math lessons could have played a role. These initiatives were possible due to the existence of public policies in education that encouraged new practices guided by scientific evidence. South American countries, in contrast, have a notable gap between public policies, scientific evidence, and educational practices. This is important because public education has a strong impact on a country's social and economic development, and there is no doubt that well-formed human capital tends to generate innovation, a crucial factor for competing in a globalized world.

AUTHOR CONTRIBUTIONS

All team members contributed to this project. FM-R, DV-B and AA-E wrote and reviewed the final manuscript.

ACKNOWLEDGMENTS

REFERENCES

Ball, D. L., Thames, M. H., and Phelps, G. (2008). Content knowledge for teaching. What makes it special? *J. Teacher Educ.* 59, 389–407. doi: 10.1177/0022487108324554

Battista, M. T. (2007). "The development of geometric and spatial thinking," in *Second Handbook of Research on Mathematics Teaching and Learning,* ed F. Lester (Reston, VA: National Council of Teachers of Mathematics), 843–908.

Bruner, J. S. (1966). *Toward a Theory of Instruction.* Cambridge, MA: Harvard University Press.

Carpenter, T. P., and Fennema, E. (1992). Cognitively guided instruction: building on the knowledge of students and teachers. *Int. J. Educ. Res.* 17, 457–470. doi: 10.1016/S0883-0355(05)80005-9

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., and Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *Am. Educ. Res. J.* 26, 499–531. doi: 10.3102/00028312026004499

Carpenter, T. P., and Franke, M. L. (2004). "Cognitively guided instruction: challenging the core of educational practice," in *Expanding the Reach of Education Reforms,* eds T. Glennan, S. Bodilly, J. Galegher, and K. Kerr (Santa Monica, CA: RAND Corporation), 41–80.

Carraher, D. W., and Schliemann, A. D. (2002). Chapter 8: is everyday mathematics truly relevant to mathematics education? *J. Res. Math. Educ. Monogr.* 11:131. doi: 10.2307/749968

Carraher, T. N., Carraher, D. W., and Schliemann, A. D. (1985). Mathematics in the streets and in schools. *Br. J. Dev. Psychol.* 3, 21–29. doi: 10.1111/j.2044-835X.1985.tb00951.x

Carraher, T. N., Carraher, D. W., and Schliemann, A. D. (1987). Written and oral mathematics. *J.Res. Math. Educ.* 18:83. doi: 10.2307/749244

Cerda, G., Pérez, C., Casas, J., and Ortega-Ruiz, R. (2017). Teaching and learning of mathematics: the need for a multidisciplinary analysis. *Psychol. Soc. Educ.* 9, 1–10. doi: 10.25115/psye.v9i1.428

Cheng, Y.-L., and Mix, K. S. (2012). Spatial training improves children's mathematics ability. *J. Cogn. Dev.* 15, 2–11. doi: 10.1080/15248372.2012.725186

Chu, J., Rittle-Johnson, B., and Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving, 87, 273–287. doi: 10.1111/bjep.12149

Civil, M. (2016). STEM learning research through a funds of knowledge lens. *Cult. Stud.Sci. Educ.* 11, 41–59. doi: 10.1007/s11422-014-9648-2

Cobb, P. (1988). The tension between theories of learning and instruction in mathematics education. *Educ. Psychol.* 23, 87–103. doi: 10.1207/s15326985ep2302_2

Edens, K., and Potter, E. (2008). How students "Unpack" the structure of a word problem: graphic representations and problem solving. *Sch. Sci. Math.* 108, 184–196. doi: 10.1111/j.1949-8594.2008.tb17827.x

Fyfe, E. R., and Nathan, M. J. (2018). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educ. Rev.* 20, 1–20. doi: 10.1080/00131911.2018.1424116

Gálvez-Nieto, J. L., Vera-Bachmann, D., and and, Trizano, I. (2015). Estudio confirmatorio del Cuestionario para evaluar Clima Social del Centro Escolar en Chile. *Rev. Mexicana Psicol.* 32, 160−168.

Hanushek, E., and Luque, J. (2003). Efficiency and equity in schools around the world. *Econ. Educ. Rev.* 22, 481–502. doi: 10.1016/S0272-7757(03)00038-4

Hawes, Z., Moss, J., Caswell, B., Naqvi, S., and MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: effects of a 32-week intervention. *Cogn. Instr.* 35, 236–264. doi: 10.1080/07370008.2017.1323902

Hegarty, M., and Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *J. Educ. Psychol.* 91, 684–689. doi: 10.1037/0022-0663.91.4.684

Kainz, K. (2019). Early academic gaps and Title I programming in high poverty, high minority schools. *Early Childhood Res. Q.* 47, 159–168. doi: 10.1016/j.ecresq.2018.08.012

Kaur, B. (2018). The why, what and how of the "Model" method: a tool for representing and visualising relationships when solving whole number arithmetic word problems. *ZDM* 51, 151–168. doi: 10.1007/s11858-018-1000-y

Lowrie, T., Logan, T., and Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance 87, 170–186. doi: 10.1111/bjep.12142

Lubinski, D. (2010). Spatial ability and STEM: a sleeping giant for talent identification and development. *Pers. Individ. Dif.* 49, 344–351. doi: 10.1016/j.paid.2010.03.022

Macneil, A. J., Prater, D. L., and Busch, S. (2009). The effects of school culture and climate on student achievement. *Int. J. Leadersh. Educ.* 12, 73–84. doi: 10.1080/13603120701576241

Mapolelo, D., and Akinsola, M. (2015). Preparation of mathematics teachers: lessons from review of literature on teachers' knowledge, beliefs, and teacher education. *Am. J. Educ. Res.* 3, 505–513.

Murata, A. (2008). Mathematics teaching and learning as a mediating process: the case of tape diagrams. *Math. Think. Learn.* 10, 374–406. doi: 10.1080/10986060802291642

Newcombe, N. S. (2014). "Teaching space: what, how, and when," in *Space in Mind. Concepts for Spatial Learning and Education,* eds D. Montello, K. Grossner, and D. G. Janelle (Cambridge, MA: The MIT Press), 323–334.

Ng, S. F., and Lee, K. (2009). The model method: Singapore children's tool for representing and solving algebraic word problems. *J. Res. Math. Educ.* 40, 282–313.

OECD (2017). *PISA 2015 Results,* Vol. V.

Peterson, P. L., Carpenter, T., and Fennema, E. (1989a). Teachers "knowledge of students" knowledge in mathematics problem solving: correlational and case analyses. *J. Educ. Psychol.* 81, 558–569. doi: 10.1037/0022-0663.81.4.558

Peterson, P. L., Fennema, E., Carpenter, T. P., and Loef, M. (1989b). Teacher's pedagogical content beliefs in mathematics. *Cogn. Instr.* 6, 1–40. doi: 10.1207/s1532690xci0601_1

Schwab, K. (2017). *The Fourth Industrial Revolution*. New York, NY: Penguin Random House LLC.

Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *J. Educ. Psychol.* 93, 604–614. doi: 10.1037/0022-0663.93.3.604

Uttal, D. H., Miller, D. I., and Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: links to achievement in science, technology, engineering, and mathematics? *Curr. Dir. Psychol. Sci.* 22, 367–373. doi: 10.1177/0963721413484756

Vegas, E., and Petrow, J. (2007). *Raising Student Achievement in Latin America: The Challenge for the 21st Century*. Washington, DC: Latin American Development Forum; The World Bank.

Vygotski, L. (2001). "Chapter 6: Pensamiento y lenguaje," in *Obras Escogidas*, *Tomo II*, ed L. S. Vygotski (Madrid: Visor/MEC), 181–285.

Wager, A. A. (2012). Incorporating out-of-school mathematics: from cultural context to embedded practice. *J. Math. Teacher Educ.* 15, 9–23. doi: 10.1007/s10857-011-9199-3

Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127

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

The Long-Term Benefit of Video Modeling Examples for Guided Inquiry

Irina Kaiser* and Jürgen Mayer

Department of Biology Education, University of Kassel, Kassel, Germany

Inquiry-based learning can be considered a critical component of science education in which students can assess their understanding of scientific concepts and scientific reasoning skills while actively constructing new knowledge through different types of activity levels (Klahr and Dunbar, 1988; Bell et al., 2005; Hmelo-Silver et al., 2007; Mayer, 2007). However, engaging in inquiry activities can be cognitively demanding for students, especially those with low prior knowledge of scientific reasoning skills (reasoning ability). Learning new information when preexisting schemata are absent entails more interacting elements and thus imposes a high working memory load, resulting in lower long-term learning effects (Paas and van Merriënboer, 1994; Kirschner et al., 2006). Borrowing knowledge from others via video modeling examples before carrying out an inquiry task provides learners with more working memory capacity to focus on problem-solving strategies and construct useful cognitive schemata for solving subsequent (virtual) inquiry tasks (Kant et al., 2017). The goal of the present study ($N = 174$ 6/7th graders) is to investigate the benefits of combining example-based learning with physical, hands-on investigations in inquiry-based learning for acquiring scientific reasoning skills. The study followed a 2 (video modeling example vs. no example) $\times$ 2 (guided vs. structured inquiry) $\times$ 2 (retention interval: immediate vs. delayed) mixed-factorial design. In addition, the students' need for cognition (Preckel, 2014), cognitive abilities (Heller and Perleth, 2000) (intrinsic, extraneous, and germane) cognitive load (Cierniak et al., 2009) and performance success were measured. Although the results of an intermediate test after the first manipulation were higher among students who watched a video modeling example ($d = 0.97$), combining video modeling examples with inquiry was not found to benefit performance success. Furthermore, regardless of manipulation, all students achieved equal results on an assessment immediately following the inquiry task. Only in the long run did a video modeling example prove to be advantageous for guided inquiry ($\eta_p^2 = 0.023$). A video modeling example turned out to be a crucial prerequisite for the long-term effectiveness of guided inquiry because it helped create stable problem-solving schemata; however, the long-term retention of structured inquiry did not rely on a video modeling example.

Keywords: inquiry(-based) learning, example-based learning, scientific reasoning skills, control of variables strategy, video modeling example, prior knowledge, cognitive load

INTRODUCTION

Scientific reasoning is an essential component of science education standards in many countries (OECD, 2007; National Research Council, 2013). Two distinct teaching approaches have been employed to foster scientific reasoning skills in school that appear contradictory at first glance: inquiry-based learning (see section Inquiry-Based Learning) and example-based learning (see section The Relevance and Effectiveness of Example-Based Learning).

In inquiry-based learning, learners actively construct knowledge by investigating scientific phenomena (Klahr and Dunbar, 1988; Hmelo-Silver et al., 2007; Mayer, 2007). Although meta-analyses have revealed (relatively modest) benefits of inquiry-based learning in science (Furtak et al., 2012), other studies have revealed an overload of working memory capacity (e.g., Kirschner et al., 2006). High levels of inquiry, such as open inquiry, are highly cognitively demanding and can overstrain working memory resources, particularly among novice students.

In contrast, in example-based learning, students simply receive an example illustrating how a specific model can be used to solve a scientific problem. This approach is rooted in the notion that learners are more likely to focus on crucial aspects and procedures when they observe examples containing helpful strategies before encountering problems they must solve themselves. However, passively studying examples to reduce the cognitive load might create illusions of understanding, which might in turn inhibit the learning process (Baars et al., 2018) or even result in the *expertise reversal effect* (Kalyuga et al., 2003) when learners' level of expertise is already high (see section The Roles of Cognitive Load and Prior Knowledge). Thus, along with their many advantages, both approaches have limitations that can be explained with reference to cognitive load theory (see section The Roles of Cognitive Load and Prior Knowledge).

According to van Gog et al. (2011), the provision of an example before a problem-solving task is more effective than problem-solving alone. Kirschner et al. (2006) recommend the use of worked examples as effective methods for guided learning. However, only a few studies have analyzed the effect of example-based learning on a special form of problem-solving, inquiry-based learning (Mulder et al., 2014; Kant et al., 2017). The present study investigates the need for video modeling examples (combining features of modeling examples and worked examples, Leahy and Sweller, 2011) prior to participation in two different levels of inquiry involving less (guided inquiry) or more guidance (structured inquiry). In addition to the effect of the combination of video modeling examples and inquiry on short-term retention (immediate performance), the potential long-term benefit (7 days after the inquiry task) is particularly interesting.

Inquiry-Based Learning

Previous research has found that inquiry-based learning can be more effective than direct instruction (Alfieri et al., 2011). In inquiry-based scientific investigations, students solve authentic scientific problems (e.g., investigating the impact of light on the growth of plants) in a collaborative form of learning in which they apply both content-related knowledge and methodological skills (inquiry skills/scientific reasoning skills). After generating hypotheses and planning appropriate experiments, students actively conduct these experiments and analyze the results to answer their scientific questions (Klahr and Dunbar, 1988; Klahr, 2000; Mayer and Ziemek, 2006; Mayer, 2007). The degree of activity or open-endedness in both the methodological and content phases is associated with students' autonomy and the amount of instructional support or teacher input (**Table 1**). In *open inquiry,* the students themselves manage their learning process, like real scientists (Bell et al., 2005). They independently formulate research questions, design and conduct investigations, and analyze their results. At the second highest level, *guided inquiry*, students investigate a teacher-provided question using an experimental plan they develop themselves. They also conduct the investigations and interpret their results with teacher guidance and support (e.g., scaffolding and feedback). In *structured inquiry*, both the research question and an appropriate experimental plan are provided by the teacher, but students are asked to generate their own explanations for the results they obtain. In *verification inquiry*, students are provided with the maximum level of guidance and instructional support; they merely conduct the experiment to verify already known results. Thus, at a low activity level, students primarily passively receive instructions, whereas a high activity level involves many different prompts for students to generate new knowledge and thus a maximum level of student output. Based on the results of a meta-analysis by Lazonder and Harmsen (2016), students must be adequately supported to achieve higher performance success ($d = 0.71$, 95% CI [0.52, 0.90]) and learning outcomes ($d = 0.50$, 95% CI [0.37, 0.62]) and to increase learners' involvement in learning/learning activities ($d = 0.66$, 95% CI [0.44, 0.88]). Guidance and support are needed to compensate for learners' low prior knowledge or poor scientific reasoning skills. Therefore, guided and structured inquiry are the most common, powerful and effective inquiry levels used in practice (Hmelo-Silver et al., 2007).

The inquiry level can vary both with respect to the content phases, which convey domain-specific concepts, and the methodological phases, which promote scientific reasoning skills. A focus on scientific reasoning is a key recommendation of international science education standards (OECD, 2007; National Research Council, 2013) to promote students' understanding of scientific and technical issues in our society and their active participation in society. Scientific reasoning involves hypothesizing, planning, experimenting, evaluating and

TABLE 1 | Levels of inquiry (Abrams et al., 2008) adapted from Schwab (1962) and Colburn (2000).

Phases	Levels of inquiry			
	Verification	**Structured**	**Guided**	**Open**
Source of the question	Given	Given	Given	Open
Data collection methods	Given	Given	Open	Open
Interpretation of results	Given	Open	Open	Open

Given, Given by teacher; Open, Open to student.

communicating the results of investigations (National Research Council, 2013). Insights into the basic rules of unconfounded evidence and their value are a crucial element of the inquiry process and scientific reasoning (Chen and Klahr, 1999; Kuhn and Dean, 2005). This essential scientific reasoning skill has a critical contribution to science education and is known as the *control of variables strategy (CVS)* (Linn et al., 1981; Chen and Klahr, 1999). It refers to one's ability to plan a controlled experiment by holding exogenous variables constant and examining one or more factor(s) of interest. The application of this strategy substantially curtails the number of options available from the *experiment space*, which consists of all experiments that could potentially be performed (Klahr and Dunbar, 1988). Moreover, the use of this strategy requires an ability to differentiate between confounded and unconfounded experiments in order to evaluate the evidence for and against scientific propositions (Zimmerman et al., 1998). Debate and controversy exist regarding the most effective approach to use in teaching CVS. In some studies, learners are allowed to obtain more knowledge about a system's function through unguided exploration, as is typical in open inquiry, leading to higher learning outcomes (Vollmeyer and Burns, 1996), while other studies show that unguided discovery methods are less effective in teaching CVS (Klahr and Nigam, 2004; Alfieri et al., 2011). Furthermore, the principles of unconfounded evidence are not learned automatically; explicit practice is needed (Sneider et al., 1984; Schwichow et al., 2016).

Regardless of the inquiry level at which investigations are conducted, inquiry-based learning is characterized by active engagement. Nevertheless, dynamic, effortful active learning techniques, such as generating knowledge in a hands-on inquiry-based learning environment, require a considerable investment of cognitive effort and time, as they are characterized by a high degree of complexity (Clark and Linn, 2003). Generation requirements such as those found in authentic learning settings impede learning, as their greater open-endedness correlates with a higher cognitive burden (Kirschner et al., 2006; Chen et al., 2016). Receiving instructional guidance via examples on how to solve an inquiry task can reduce the degree of complexity and result in better performance than solving problems without any examples (e.g., Aleven, 2002; McLaren et al., 2008; van Gog et al., 2009), a learning approach referred to as *example-based learning*. According to the borrowing and reorganizing principle, highly structured problem-solving strategies are best learned from other people (Sweller and Sweller, 2006). This approach prevents learners from overstraining their cognitive resources with incorrect problem-solving strategies (Sweller and Sweller, 2006).

The Relevance and Effectiveness of Example-Based Learning

Example-based learning distinguishes between two forms of examples (van Gog and Rummel, 2010; Renkl, 2014): *worked examples* (Sweller and Cooper, 1985; Cooper and Sweller, 1987; Sweller et al., 1998; Schwonke et al., 2009), in which each step of the procedure used to solve a problem is explained in

a text-based manner, and *modeling examples* (Bandura, 1977, 1986; Collins et al., 1989), in which a model demonstrates and/or explains how to complete a problem-solving task. Worked examples are effective in promoting problem-solving strategies and integrating new with prior knowledge (Roth et al., 1999). They are one of the most time-efficient, effective and widely used instructional learning strategies, particularly in the initial stages of skill acquisition (vanLehn, 1996; Salden et al., 2010). Experiments have repeatedly demonstrated the *worked example effect* (e.g., Renkl, 1997; Atkinson et al., 2000; Sweller et al., 2011), mainly in fields such as algebra (Sweller and Cooper, 1985) and computer programming (Kalyuga et al., 2001)—domains that are clearly defined, well-structured (mostly iterative), and can be investigated in laboratory studies. More recently, positive effects have also been observed on scientific reasoning (Mulder et al., 2014; Kant et al., 2017). The basic structure of a worked example typically includes three crucial components: (1) examining the key problem to raise awareness of the problem to be solved, (2) explaining the procedure for solving the problem through the completion of a certain number of steps in a specific order to promote the construction of appropriate schemata, and (3) describing the final solution to the problem (Renkl, 1997). After completing all three steps, learners are asked to solve a similar problem on their own to enhance the automation of their problem-solving skills and ensure transfer (Atkinson et al., 2000).

The effect of worked examples is rooted in cognitive load theory (see section The Roles of Cognitive Load and Prior Knowledge). Worked examples provide learners with full guidance concerning the key steps required to solve a problem, thus automatically drawing learners' attention to relevant aspects that form a basis for subsequent problem-solving. These examples allow appropriate cognitive schemata to be developed (Crippen and Earl, 2007; Schworm and Renkl, 2007) before learners are confronted with actual problem-solving demands and information. Sweller and Cooper (1985) claim that worked examples lead to better learning of solution procedures. While studying problems with detailed solutions provides learners with a basic understanding of domain-specific principles, the conventional problem-solving method focuses on searching for processes rather than on aspects crucial to the acquisition of cognitive schemata (Sweller and Cooper, 1985).

A main difference between worked examples and modeling examples concerns attentional focus (Hoogerheide et al., 2014). Modeling examples provide learners with the opportunity to observe a model solving a task without explicitly focusing on relevant aspects or dividing the procedure into individual steps. This approach requires learners to selectively focus on the most critical elements of the demonstrated behavior. The observed information is actively organized and integrated with the learner's prior knowledge during a constructive process. However, the nature of learners' cognitive representations and the level at which they possess the component skills determines whether learners are able to effectively apply the observed strategies (Bandura, 1986). Previously, modeling examples have mainly been used to convey (psycho) motor skills (e.g., Blandin et al., 1999) and skills with low levels of structure (e.g., Braaksma et al., 2002; Zimmerman and Kitsantas, 2002: writing; Rummel

and Spada, 2005; Rummel et al., 2009: collaboration). However, over the last few years, new variants of modeling examples have been established in online learning environments that combine features of both worked and modeling examples. For instance, the steps of a problem-solving procedure are shown or/and illustrated on a model's computer screen while a non-visible model explains the relevant actions (e.g., McLaren et al., 2008; van Gog et al., 2009, 2014; Leahy and Sweller, 2011). These new formats (known as "video modeling examples") combine the advantages of both forms of examples. They employ the audiovisual method of modeling examples and the structured, step-wise procedure of worked examples. By structuring the problem-solving procedure into separate steps and dispensing with a visible model, learners' attention can be focused on task performance and not distracted by task-irrelevant information, e.g., other people's faces, gestures, clothes, and movement (see van Gog et al., 2014). The replacement of written text of worked examples with spoken text leads to a division of information processing into two working memory systems (Baddeley, 1986). Learners direct their visual attention to the images while simultaneously listening to the explanation of the non-visible model. According to the *modality effect* (Mousavi et al., 1995; Mayer and Moreno, 1998; Kühl et al., 2011), this strategy helps reduce the working memory load (Ginns, 2005; Leahy and Sweller, 2011; Sweller et al., 2011). In addition, learners' attention can be guided to the most relevant elements by highlighting, coloring and zooming in on important aspects.

The Roles of Cognitive Load and Prior Knowledge

An unguided problem provides no indication of which elements should be considered, in contrast to a worked example. Therefore, the study of worked examples reduces the number of elements that must be processed by the working memory (Chen et al., 2016). Since the cognitive architecture is restricted by the working memory capacity, *element interactivity*—or the degree of complexity of learning content within the framework of cognitive load theory that depends on the learner's prior knowledge (Sweller, 2011; Chen et al., 2016), may not exceed a certain amount if the goal is to promote effective learning. A higher level of element interactivity requires a greater working memory capacity, resulting in a high *intrinsic cognitive load*. Approaches that guide learners in the right direction removes the need to employ trial and error strategies (Renkl, 2014). Thus, learners can apply their full working memory capacity to construct a problem-solving schema to use in future problem-solving tasks (Cooper and Sweller, 1987). According to the information store principle, knowledge borrowed from others (i.e., instructors) can be reorganized and transferred to long-term memory for storage (Sweller and Sweller, 2006).

The way instructional material is presented also affects working memory, which is referred to as *extraneous cognitive load*. Both high intrinsic and high extraneous cognitive load might restrict long-term learning outcomes (e.g., Klahr and Nigam, 2004; Kirschner et al., 2006). This influence should be considered when deciding on an appropriate level of instructional guidance. In particular, learners with little expertise or little prior knowledge in the relevant content domain do not benefit from being confronted with too much information and opportunities for active participation at one time. Providing those learners with more instructional guidance before a problem-solving task (in the form of an example) and/or during the task (e.g., via guided or structured inquiry) can reduce mental exertion, thus ensuring that learners' cognitive resources are focused on the most relevant aspects (Sweller et al., 2011; Chen et al., 2016). This approach in turn increases the *germane cognitive load*, which promotes learners' understanding and the transfer of newly acquired knowledge to long-term memory (Paas and van Merriënboer, 1994; van Merriënboer and Sweller, 2005). On the other hand, the long-term retention and transfer of acquired skills were recently shown to only be achieved through active knowledge construction/generation (Bjork and Bjork, 2014), and thus require high levels of inquiry.

Indeed, an investigation of the active generation of scientific reasoning skills revealed a long-term benefit when a high level of generation success was ensured during inquiry (Kaiser et al., 2018). Students who (successfully) generated plans for scientific investigations (scientific reasoning skills) were at an advantage compared to a matched group that simply followed provided instructions. This phenomenon is referred to as the *generation effect* (Jacoby, 1978; Slamecka and Graf, 1978). It arises when items are better remembered when they are generated rather than simply read. It is considered an indication that active knowledge construction leads to a higher level of retention than passive observation. On the one hand, direct instruction that completely explains the underlying principles and procedures promotes effective learning, particularly for novel information with high element interactivity—as is usually the case in structured inquiry (Kirschner et al., 2006). On the other hand, the generation effect indicates that active knowledge construction leads to higher retention than passive observation, which favors guided inquiry. However, only a few studies have reported a positive generation effect on complex educationally relevant science material (e.g., Foos et al., 1994; Richland et al., 2007; Kaiser et al., 2018). As shown in the study by Foos et al. (1994), the effect is masked in applied settings because overall test performance is examined instead of performance on (successfully) generated items alone. A generation effect does not exist for non-generated items and is only observed for (successfully) generated items (Foos et al., 1994). Thus, the effectiveness of active generation in an authentic and complex learning environment, such as inquiry-based learning, relies on high generation success during the inquiry session, which in turn depends on prior knowledge (Kaiser et al., 2018). According to Kaiser et al. (2018), immediate performance (success) and the retention of scientific reasoning skills in guided inquiry are primarily influenced by prior knowledge provided through video modeling examples. Thus, learners who aquire a certain amount of (prior) knowledge via a video modeling example are more likely to profit from active generation.

Little research has been conducted on complex curriculum-based material and the impact of prior knowledge on active generation. Most previous studies on the generation effect have considered rather simple material (e.g., synonyms and rhymes) in controlled laboratory settings. They have mainly included non-curricular material for which no preexisting knowledge is required. Moreover, the studies that have examined the influence

of prior knowledge by employing educationally relevant material tend to focus on mathematics. For instance, the study by Rittle-Johnson and Kmicikewycz analyzed the effect of prior knowledge on generating or reading answers to multiplication problems. Third graders with low levels of prior knowledge profited from self-generating answers to the problems. These students had better performance on the post-test and retention test than their peers subjected to the reading condition, even on problems they had not practiced (Rittle-Johnson and Kmicikewycz, 2008). Thus, learners' prior knowledge and intuitions often contravene new knowledge (Bransford et al., 2000). In contrast, the effect of active generation tends to be much more muted for the retrieval of unfamiliar material, such as nonwords, or new material, such as unfamiliar sentences from textbooks or experimental plans (Payne et al., 1986; McDaniel et al., 1988; Lutz et al., 2003; Kaiser et al., 2018). Therefore, the generation effect only applies to information rooted in preexisting knowledge (Gardiner and Hampton, 1985; Nairne and Widner, 1987). The results reported by Chen et al. (2016) confirm these findings and explain the discrepancy with the findings described by Rittle-Johnson and Kmicikewycz (2008) by showing that the generation effect only occurs for material with low element interactivity. Element interactivity, in turn, depends not only on the complexity of the material but also on learners' prior knowledge. Learners with a low level of prior knowledge have more problems generating correct information and procedures when faced with highly complex material, resulting in poor performance compared to high-knowledge learners (e.g., Siegler, 1991; Shrager and Siegler, 1998). Learners with a higher level of relevant prior knowledge face a lower element interactivity and require less guidance to successfully solve a problem due to the low intrinsic cognitive load (Sweller, 1994). In contrast, a high intrinsic cognitive load must be reduced to prevent the learner from exceeding his/her working memory limits. However, reducing cognitive load is unnecessary or even counterproductive when the intrinsic cognitive load of the relevant content is low due to the learner's high level of expertise (Chen et al., 2016). High-knowledge learners even tend to face disadvantages above a certain level of guidance and receipt of **Supplementary Information**—known as the *expertise reversal effect* (Kalyuga et al., 2003). Thus, the role of guidance in teaching remains an important and controversial issue in instructional theory (Craig, 1956; Ausubel, 1964; Shulman and Keisler, 1966; Mayer, 2004; Kirschner et al., 2006). Mulder et al. (2014) found that heuristic worked examples (Hilbert et al., 2008; Hilbert and Renkl, 2009) enhanced students' performance success but did not result in higher post-test scores. However, they recommended further research on the delayed effects of worked examples in the area of inquiry-based learning, consistent with the findings reported by Hübner et al. (2010) of a worked example effect on a delayed transfer task using strategies for writing learning journals. Kant et al. (2017) observed higher learning outcomes for students who watched a video modeling example before solving an inquiry task than for students who were provided with an example after the inquiry task. The authors compared four groups (example-example, example-inquiry task, inquiry task-example, and inquiry task-inquiry task) with regard to their learning outcomes, perceived difficulty and mental effort, judgments of learning, and monitoring accuracy in a simulation-based inquiry learning environment. The learners in the example groups were provided with a video modeling example in which two models solved an inquiry task—the same task the learners were required to solve on their own in the control condition. Studies on the necessity of combining example-based learning with different levels of inquiry-based learning for the acquisition of scientific reasoning skills are still outstanding. Overall, long-term investigations are lacking.

RESEARCH QUESTIONS

The present study aims to investigate the necessity of a video modeling example for the development of scientific reasoning skills, determine the extent to which different inquiry levels (guided and structured inquiry) benefit from example-based learning, and identify the role of learners' cognitive load in the long-term retention of scientific reasoning skills. An experiment with students in Grades 6 and 7 was conducted that compared the active generation of scientific reasoning skills in guided inquiry to an inquiry task in which learners simply read instructions on experimental design (structured inquiry) with or without a video modeling example to achieve these aims.

Consistent with recent findings reported by Kant et al. (2017) and Chen et al. (2016), we expected that watching a video modeling example of a method to solve a scientific problem by following the inquiry cycle and using the CVS would positively affect learning outcomes in guided but not structured inquiry (H1). We further expected an interaction between the inquiry level and the presence or absence of a video modeling example such that watching a video modeling example would be more effective when combined with generating answers (in guided inquiry) than reading answers (in structured inquiry), particularly in the long term (Hübner et al., 2010) (H2). Furthermore, we hypothesized that the perceived cognitive load during the learning process would differ across the four conditions (video modeling example vs. no example x guided vs. structured inquiry). According to Kirschner et al. (2006), sturctured inquiry with a video modeling example should result in the lowest cognitive load, while guided inquiry without a video modeling example should result in the highest load on working memory capacity. In contrast, guided inquiry with a video modeling example should reduce learners' intrinsic and extraneous cognitive load, increase the germane load, and promote the learning process (H3). Generation success has been reported to be a reliable predictor of learning outcomes (Foos et al., 1994; Kaiser et al., 2018). Based on these findings, we assumed that students would achieve higher performance during guided inquiry when a video modeling example is provided (H4).

METHODS

Participants

We conducted an a priori power analysis using G*Power (Software G*Power; Faul et al., 2007) with a significance level of

$\alpha = 0.05$, a medium effect size of $f = 0.25$ and a desired power of 0.8; the results indicated a recommended sample size of $N = 179$. Two hundred and fifteen German students in Grades 6 and 7 from 9 classes in five different schools participated in the present study. A total of 174 students (M = 12.05 years, SD = 0.629) completed all tasks and the first and second post-test. Forty-one students were excluded due to illness or failure to consent to data usage. All data were collected and analyzed anonymously. A subsample of this dataset was already used by Kaiser et al. (2018) to analyze the role of generating scientific reasoning skills in inquiry-based learning in a 2 × 2-mixed-factorial design. In the present study, we used the total sample in an extended 2 × 2 × 2-mixed-factorial design and with (partially) different test instruments. Thus, new data were analyzed. Since the goal of our study was to analyze whether an example is actually needed to achieve a long-term benefit from inquiry-based learning, the control condition was not provided with any form of example. We based our design on the study by Mulder et al. (2014), who also withheld access to worked examples among students in the control condition.

Participants in all classes were randomly assigned to one of two inquiry conditions: guided inquiry, $n = 68$ with a video modeling example and $n = 22$ without an example; or structured inquiry, $n = 64$ with a video modeling example and $n = 20$ without an example.

The limited number of participants assigned to the control conditions was based on decisions by the participating classes. Classes were able to choose between an additional computer-based introduction to inquiry-based learning in the form of a video modeling example 1 week before completing the experimental unit or a short briefing (without an explicit example) on the same day the experimental was conducted. Most classes selected the extended version. However, students' level of experience in inquiry-based learning was not the reason for their decision. All students had the same low level of expertise.

Research Design

The study used a 2 (video modeling example vs. no example) × 2 (guided vs. structured inquiry) × 2 (retention interval: immediate vs. delayed) mixed-factorial design. Two levels of inquiry, guided inquiry (GI) vs. structured inquiry (SI) and with (+VME) vs. without a video modeling example (-VME), served as the independent variables. As dependent variables, scientific reasoning skills were tested at two different measurement points: post-test performance immediately after the intervention and a follow-up test 1 week later. This approach allowed us to compare the learning and transfer effects on the CVS resulting from guided or structured inquiry with or without a worked example in the short- and long-term. The tests were constructed by applying an *equating facet design* to control for item difficulty and avoid unanticipated test effects (see section Scientific Reasoning).

Materials
Learning Content
The students were to learn procedures and strategies for holding variables constant (CVS), as well as the fundamental scientific reasoning skills of hypothesizing (searching the hypothesis

space), experimenting (testing hypotheses), and evaluating evidence. The learning environment consisted of two different student experiments: a virtual experiment with a computer-based learning program and a real experiment in an inquiry-based student lab. Both experiments analyzed the concept of behavioral adaptations among animals living in and around a pond.

Video modeling example
In the first session, all students briefly discussed the purpose and intent of scientific inquiry with a specially trained instructor, who subsequently introduced them to the topic of "animals of the pond." Afterwards, one group of the students was taught the CVS in a uniform computer-based introductory session in the new format of a video modeling example (+VME), which combines the benefits of worked examples and modeling examples (see section The Relevance and Effectiveness of Example-Based Learning). The session was designed to develop the students' scientific thinking and understanding of the reason for holding all variables constant across experimental conditions while varying the one variable being investigated. After a short introduction to the discipline-specific methods employed by scientists, a virtual professor ("Professor Plankton") familiarized the students with the inquiry cycle and the learning content of the unit (the concept of behavioral adaptations among animals living in and around a pond) by guiding them through eight video units corresponding to the steps of an illustrative experiment about dragonfly (*Anisoptera*) larvae hunting their prey: *phenomenon, research question, hypotheses, plan, investigation, analysis, interpretation,* and *discussion*. The example of dragonfly larvae hunting their prey was used to introduce the students to the crucial phases of scientific inquiry: (1) formulating research questions, (2) inferring one or more hypotheses, (3) planning and conducting an experiment, and (4) analyzing the experiment (describing the data, interpreting the data, and critically evaluating the methods used). The students were shown the steps of the procedure on the Professor's computer screen while a non-visible speaker explained the Professor's actions. Hence, the students were able to study the example in a step-by-step procedure by directing their visual attention to the images while simultaneously listening to an explanation by a non-visible model (see the **Supplementary Material**: *Screenshots VME*).

Inquiry tasks
In the laboratory sessions, all students completed a scientific experiment using the CVS entitled "The Mystery of Water Fleas' Migration" (Meier and Wulff, 2014), which focused on the daily vertical migration of water fleas (*Daphnia magna).* This phenomenon was related to the initial example in the learning program, as it also involves a biological adaptation, or structural or behavioral changes that help an organism survive in its environment. Biological adaptation is considered a core disciplinary concept in leading science standards (National Research Council, 2013), which none of the participating classes had covered previously in class.

The module aimed to teach scientific thinking and scientific reasoning skills via guided experimentation. All students received a research workbook (see **Supplementary Material**: *Research*

Workbooks in Kaiser et al., 2018) to support the students' learning process and provide guidance across all phases of the inquiry cycle (hypothesis generation, designing and conducting an experiment, and interpreting the results). Students in the "Guided Inquiry" (GI) condition received 13 short prompts that helped them plan an appropriate experiment by identifying the independent and dependent variables, control variables (see **Supplementary Material**: *Example Inquiry task*), and confounding variables (short answer tasks), as well as a cloze (consisting of 130 words and 15 prompts) that asked them to retrieve information about the CVS immediately following the experimental session. The students in the "Structured Inquiry" (SI) condition received research workbooks with direct instructions for conducting an experiment instead of generation prompts, and a reading text rather than a cloze at the end.

The content of the research workbooks was structured in a similar manner across conditions to ensure comparability. All prompts and feedback material in the GI condition were derived from the text material in the SI condition. Moreover, the students were provided the same amount of time for cognitive processing.

Instruments

Three assessment time points were integrated into the experimental design: the first test was administered prior to the inquiry task or after the video modeling example, the second was administered after the inquiry task, and a final test was administered after a retention interval of 1 week. In addition to scientific reasoning skills (see section Scientific Reasoning), the students' success in generation (see section Learners' Performance Success in Guided Inquiry) and perceived cognitive load (Cierniak et al., 2009) (see section Learners' Cognitive Load) were assessed during the experimental task. Data on the students' demographics; grades in biology, math, and German; need for cognition (Preckel, 2014) and cognitive abilities Heller and Perleth, 2000 (see section Learners' Prerequisites) were collected at each of the three assessment time points. All measurements were paper-based.

Scientific Reasoning

Three questionnaires assessing the acquisition and retention of scientific reasoning skills were developed to evaluate the learning outcomes. After conducting statistical item analyses, the final assessment tests consisted of 6 to 10 items, both single choice and open-ended (Janoschek, 2009; Hof, 2011; Wellnitz and Mayer, 2016; modified). All single-choice items had four possible answer options. In contrast to Kaiser et al. (2018), we also tested the students' inquiry skills in an open-ended format, which allowed us to examine higher levels of competence in inquiry skills.

Immediately after the video modeling example or immediately before the inquiry session, depending on the condition, students completed an intermediate assessment test consisting of six items to identify individual differences in scientific reasoning skills. The assessment test comprised four open-ended items and two single-choice items. Item difficulty was appropriate ($p = 0.56$), a moderate level of difficulty, and the test indicates an acceptable level of reliability ($\alpha = 0.60$) for comparing groups (Lienert and
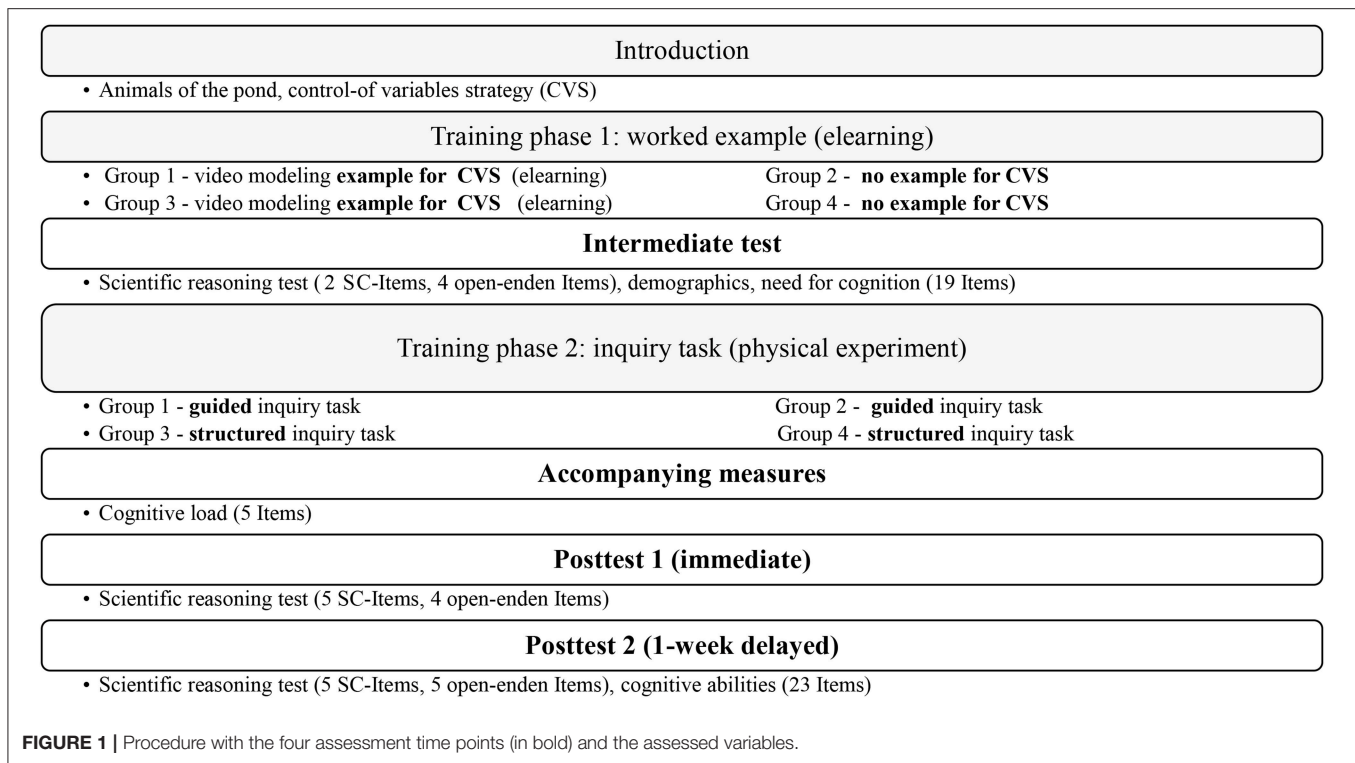
Raatz, 1998). Furthermore, the discrimination parameters were all above rit > 0.30.

The following scientific reasoning tests were completed 10 min after the inquiry task and 1 week later (five single-choice items and four or five open-ended items, respectively) (**Figure 1**). All tests required students to demonstrate their understanding of CVS. They were either asked to select the appropriate design from a set of confounded and unconfounded experiments, amend a confounded experiment, or identify the independent and dependent variables in an unconfounded experiment. We incorporated anchor items into the two post-tests to ensure comparability and provide a baseline for an equating analysis. The construction of the anchor items was based on an *equating facet design* with three dimensions to ensure systematic variation (**Table 2**). Each anchor item provided a uniform description of an experimental design (*task context*) in each post-test, followed by a prompt to either complete *Task (1), (2), or (3)* in one or two *task formats* (single choice and/or open-ended item). The use of the same *task context* ensured the comparability of the two post-tests and sought to focus students' attention on inquiry skills rather than distracting them with excess content-related information. The three different *tasks* invited students to evaluate the quality of others' research—to identify the independent and dependent variable (searching the hypothesis space), select an appropriate experimental design (testing a hypothesis) or evaluate appropriate measurements (analyze scientific evidence). One of six *task contexts* was allocated to each task. In addition, some anchor items encompassed two different *task formats*: single choice (SC) and open-ended (O) counterpart items. Thus, two to six versions of each task context appeared in the test, with varying variables to be defined (see the **Supplementary Material**: *Example Anchor Item*). Three task contexts were used in all three tests, while five contexts were used in post-tests 1 and 2 only. Thus, students were tested with 19 ($3 \times 3 + 5 \times 2$) anchor items referring to the same scientific knowledge construct and skills across the three measurement points.

Item difficulty, internal consistency, and discrimination parameters were analyzed for post-tests 1 and 2. Item difficulty was appropriate ($p = 0.50–0.58$) and the tests were reliable ($\alpha = 0.70–0.72$) for comparing groups (Lienert and Raatz, 1998). Furthermore, the discrimination parameters were all above rit > 0.30.

Learners' Cognitive Load

The students' perceived cognitive load was assessed under all conditions immediately after the inquiry session. Since the main focus of the study was the learning outcomes (see section Scientific Reasoning) and student performance (see section Learners' Performance Success in Guided Inquiry), we sought to keep the questionnaire brief to avoid overtaxing our sample of young learners and decreasing their motivation. The instrument comprised five items (after excluding one) to which the students responded on a six-point Likert scale (ranging from 1 = low to 6 = high) ($\alpha = 0.66$, rit > 0.20, Cierniak et al., 2009, modified). Cierniak et al. (2009) used this instrument to analyze how different cognitive load types mediate the split attention effect (e.g., Chandler and Sweller, 1991, 1992, 1996)

FIGURE 1 | Procedure with the four assessment time points (in bold) and the assessed variables.

in a learning environment with biological content that included complex figures with accompanying texts. Their measure was chosen because their learning environment was similar to our environment and their questionnaire was shorter than more frequently used scales, such as the scale used by Leppink et al. (2014). In our study, intrinsic and extraneous cognitive load were measured with 4 items, for IL: (1) "How difficult was it for you to understand the experiment?" and (2) "How difficult was it for you to work like a research scientist?," and for EL: (3) "How difficult was it for you to work with the research workbook?" and (4) "How difficult was it for you to understand the work instructions in the research workbook?." A single item was used to assess germane load, GL: (5) "How strongly did you concentrate while learning today?." One item "How much effort did you need to invest into learning today?," was excluded due to insufficient item properties (see the **Supplementary Material**: Questionnaire for Cognitive Load in Kaiser et al., 2018). Items (1), (3), and (5) were adopted from Cierniak et al. (2009); items (2) and (4) were new **Supplementary Items**.

Learners' Abilities
Learners' prerequisites
Two questionnaires with good validity (NFC: $p = 3.58$, $\alpha = 0.89$, rit > 0.30; CA: $p = 0.48$, $\alpha = 0.91$, rit > 0.30) were included in the study design to assess students' prerequisites, namely the need for cognition (Preckel, 2014) and cognitive abilities (Heller and Perleth, 2000). The questionnaire for the need for cognition comprised 19 items, with responses indicated on a five-point Likert scale. The Questionnaire for Cognitive Abilities for 6th Graders measured the students' figural inductive

reasoning skills by asking them to identify figural analogies (KFT 4-12+ R, Subtest N, Heller and Perleth, 2000). It comprised 24 items (after excluding one). Each item had five answer options and only one correct answer. The students were tasked with answering as many items as they could within 9 min (see the **Supplementary Material**: Questionnaire for Cognitive Abilities in Kaiser et al., 2018).

Learners' performance success in guided inquiry
We further collected qualitative data in the form of all student responses to the generation prompts in the students' research workbooks under the inquiry condition, including the students' proposed experimental designs, discussions of research methodology and the final cloze. This made it possible to confirm the effect of the treatment and examine the role of generation success in short-term and long-term retention. The data were coded on a scale with a potential range of 0 to 33 points. The following components of the *experimental design* were assessed (each on a 0–2-point scale): identifying the independent and dependent variables; designing a controlled experiment in which one independent variable is varied and all other relevant variables are held constant, thus controlling for potential biases and confounding factors; and specifying the measurement time points and number of animals (water fleas) in the experiment. With respect to the *methodological discussion,* the following factors were evaluated (also on a 0–2-point scales): ensuring equal control conditions and describing its importance, using an LED light and more than 10 water fleas and describing their importance, avoiding external confounders (light pollution, bumping into the desk,

TABLE 2 | Equating facet design with the three dimensions task, task context, task format (SC, O).

Task	Task context	Intermediate test	Post-test 1	Post-test 2
(1) Choose an appropriate design (AD) from a set of confounded and unconfounded experiments	*Anchor* Factors influencing the growth of beans	AD: clay vs. soil SC –	AD: Sunlight vs. no sunlight SC O	AD: Water vs. no water SC –
(2) Identify the independent variable (IV) and the dependent variable (DV) in an unconfounded experiment	*Anchor* Factors influencing the sugar production of sugar beets		IV: temperature DV: sugar production of sugar beet SC O	IV: care DV: sugar production of sugar beet SC O
	Anchor Factors influencing fish's breathing in an aquarium		IV: number of fishes in an aquarium DV: fish breathing SC O	IV: temperature DV: fish breathing SC O
	Anchor Factors influencing woodlice's habitat selection	IV: humidity DV: preferred habitat of woodlice SC O	IV: darkness DV: preferred habitat of woodlice SC O	IV: temperature DV: preferred habitat of woodlice SC O
	Non-anchor Factors influencing backswimmers' hunting for prey	IV: visual stimulus DV: reaction of backswimmers SC O		
	Non-anchor Factors influencing dragonfly larva's hunting for prey	IV: size of prey DV: reaction of the dragon fly SC –		
	Non-anchor Factors influencing effervescent tablets' release of CO_2			IV: temperature – O
(3) Correct a confounded experiment/identify the disturbance variable (DI)	*Anchor* How light influences water fleas' behavior		DI: aquatic plant on one side of the aquarium SC –	DI: feeding of a number of experimental animals SC –

and noise) and describing its importance, and the necessity and duration of a habituation period for the water fleas (for further information, see the **Supplementary Material**: Coding scheme in Kaiser et al., 2018).

Interrater reliability was calculated using the Kappa statistic to evaluate the consistency of the two independent raters. The Kappa value was 0.94 ($p < 0.001$), indicating almost perfect agreement (Landis and Koch, 1977).

The research workbooks and the complete coding scheme are published in the study by Kaiser et al. (2018).

Procedure

The experiment consisted of three phases: an introductory video modeling example with a subsequent intermediate test (see section Computer-Based Introduction via a Video Modeling Example), an inquiry-based learning session with a subsequent post-test (see section Inquiry Task), and a second post-test. One hundred and thirty-two students engaged in all three sessions (+VME), which were scheduled over 3 weeks. The other 45 students did not participate in the first computer-based session (-VME).

Computer-Based Introduction via a Video Modeling Example

The first session required ∼60 min to complete and was performed at school. A group of students (+VME) received guided instruction in a computer-based learning environment and then individually worked through a brief learning session on computers. Each student had a headset that allowed them to explore the learning program, which consisted of videos and short reading passages, at their own pace. A video modeling example familiarized the participants with fundamental scientific reasoning skills. A virtual figure called Professor Plankton led the students through the learning program. The students were introduced to all experimental phases and the specific terminology associated with them. This instruction lasted 30 min. Immediately afterwards, the students completed a paper-based intermediate assessment test, which sought to identify individual differences in scientific reasoning skills. The students required an average of ∼25 min to complete the test; a time limit was not established. The students who did not work through the computer program (-VME) were asked to complete the assessment test items immediately before the inquiry task (in the second session).

We also collected data on the students' demographics, cognitive abilities, need for cognition, and grades in math, German and biology. All students were also asked to indicate whether they had previously attended an inquiry course in our student lab. Students who attended this course were excluded from the calculations. Students were clearly informed that the learning program was in preparation for a subsequent inquiry module at the university.

Inquiry Task

The inquiry module, a scientific experiment on water fleas' vertical migration, took place 1 week after the computer-based introduction. It was conducted in an inquiry-based learning environment in a university lab tailored to work with school students.

During this learning phase, individual students in each class were randomly assigned to the two conditions [guided (GI) vs. structured inquiry (SI)] and separated into small groups (up to five students). They received instruction from trained supervisors. Thus, the students in each group knew one another before the start of the inquiry activity. Intermixing students across classes was not feasible because we only had access to one student lab, a limited number of experimental materials, rooms and supervisors were available, and for other organizational reasons. The supervisors received scripts with detailed information about each inquiry phase to assist them in providing uniform guidance to all groups during the inquiry activity. Supervisors at both inquiry levels were prohibited from answering questions on scientific reasoning to ensure that we collected accurate data on students' inquiry skills. The key difference between the two inquiry levels was the amount of information and instructional support provided; however, the total instructional time remained the same across conditions. The students in each condition were allowed ∼180 min to complete the inquiry task in two separate rooms after receiving

uniform (general) instructions from their supervisor. Each task was assigned a certain maximum duration (see the **Supplementary Material**: Research Workbook in Kaiser et al., 2018).

The main differences between the conditions are listed below. Students in the SI condition were provided with a detailed experimental plan and a discussion of the method that would be used, whereas students in the GI condition were required to actively generate their own experimental plan and discuss the data they collected using the inquiry skills acquired in the introductory section. They first generated information individually by identifying independent and dependent variables and jotting down ideas for experimental procedures (*scientific reasoning skills*: inferring hypotheses, *aspects*: independent variable and dependent variable; Arnold et al., 2014) (individual work). After discussing their preliminary ideas with one another, the students in each group worked together to develop a detailed experimental plan that operationalized the dependent variable, appropriately varied the independent variable, identified and controlled for biases and confounders, and specified the measurement intervals and number of measurement points (*scientific reasoning skills*: planning experiments, *aspects*: independent variable, dependent variable, confounding/nuisance variables, measurement points, and repeated measures; Arnold et al., 2014) (team work). The second phase proceeded in the same manner. First, the students individually analyzed the biases for which they had controlled in the experiment by completing a corresponding checklist (see *Example 3*) (individual work); then, they discussed their data in groups (team work). The students followed the same procedure and used the same terminology presented in the video modeling example.

As students have been shown to perform better during inquiry when provided more specific guidance (Johnson and Lawson, 1998; Borek et al., 2009; Lazonder and Harmsen, 2016), the students received corrective feedback from their supervisor after both phases to ensure that the students had access to a sufficient amount of information. However, the information the supervisors were permitted to provide was limited to the material defined in a workbook of instructions (see the **Supplementary Material**: Workbook of Instructions for Generation Group in Kaiser et al., 2018), which all supervisors were required to use. Supervisors provided the students with correct responses or instructed them on how to supplement and/or revise their proposed experimental plans to help the students dismiss incorrect ideas and identify new ideas by following the provided cues. In contrast, students in the SI condition were explicitly informed about which variables to investigate and were provided a series of prescribed steps to follow, similar to a recipe. Instead of completing a checklist and discussing bias after the experiment, the students were simply informed about possible confounders that may have influenced the dependent variable.

Apart from these differences, the procedure was identical under all conditions. Students in both groups completed the physical hands-on activities involved in conducting the experiment, because practice is necessary for learners to develop an understanding of the principles of unconfounded evidence

(Sneider et al., 1984; Schwichow et al., 2016). Moreover, no students were asked to generate any content-related information. Thus, they stuck to appropriate interpretations of their experimental data.

Immediately after the inquiry-based learning session, students in all treatment groups completed the same questionnaire about cognitive load, followed by an assessment test measuring scientific reasoning skills (with five SC and four open-ended items). Students were not informed in advance that they would be taking these tests to prevent them from studying for the tests and to increase the probability that post-test scores would reflect knowledge acquired during the experiment. One week later, all students completed a second, comparable post-test with five SC and five open-ended items. The students required an average of ∼30 min to complete each test; again, no time limits were imposed.

Data Analysis

We conducted statistical analyses from the paradigm of classical test theory using SPSS software to identify differences between groups and among students with different abilities, as well as to detect the influence of students' characteristics on their learning outcomes.

All results were significant at the 0.05 level unless indicated otherwise. Pairwise comparisons were Bonferroni-corrected to the 0.05 level. The partial eta squared (η_p^2) value is reported as an effect size measure for all ANOVAs, while Cohen's d is reported as an effect size measure for all t-tests.

RESULTS

No significant differences were observed between conditions in students' demographic data, grades, need for cognition or cognitive abilities, indicating that randomization was successful. Additionally significant differences were not observed between the classes that participated in the computer-based introduction and classes that did not, with the sole exception of biology grades. The Bonferroni-adjusted *post-hoc* analysis revealed that students in the SI+VME condition achieved better grades in biology than students in the SI-VME condition (0.59, 95% CI [1.09, 0.09], p = 0.011). However, biology grades did not significantly affect the learning outcomes.

We also monitored the data for CVS experts, or students who answered all items on the intermediate test correctly without receiving a video modeling example. However, no such experts who might have distorted the results were identified.

Descriptive results for the learners' performance in all test sessions are shown in **Table 3**.

Learning Outcome—Video Modeling Example vs. No Example in Guided or Structured Inquiry on Short- and Long-Term Retention (H1) and (H2)

The results were analyzed using a 2 (video modeling example vs. no example) x 2 (guided vs. structured inquiry) x 2 (retention interval: immediate vs. delayed) ANOVA with

TABLE 3 | Means and standard deviations (in parentheses) of performance assessed in post tests 1 and 2.

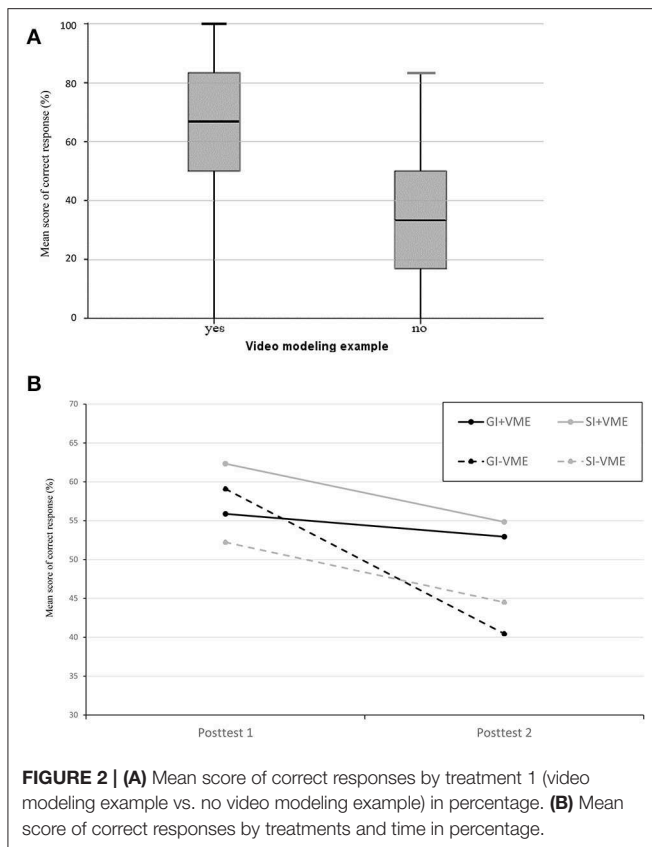|	GI+VME	GI-VME	SI+VME	SI-VME
Post-test 1 (%)	55.88 (25.32)	59.09 (27.51)	62.33 (23.48)	52.22 (27.00)
Post-test 2 (%)	52.94 (24.19)	40.45 (24.59)	54.84 (26.73)	44.50 (23.95)
Generation success	17.30 (6.16)	14.59 (4.38)	–	–
N	68 (67[a])	22	64	20

[a]*Generation success could only be analyzed in 67 out of 68 research workbooks.*

repeated measures. This model yielded a significant main effect of time, F(1,170) = 18.54, p < 0.001, η_p^2 = 0.098, but no main effect of inquiry level, F(1,170) = 0.68, p = 0.412, and only a marginal significant effect of the use of a video modeling example, F(1,170) = 3.32, p = 0.070. Hence, students achieved higher results immediately after the inquiry task than 1 week later. Furthermore, we detected a significant interaction between the retention interval and receipt of a video modeling example, F(1,170) = 4.58, p = 0.034, η_p^2 = 0.026. The interaction between the retention interval and level of inquiry was not significant, F(1,170) = 0.06, p = 0.807, nor was the interaction between inquiry level and use of a video modeling example, F(1,170) = 0.48, p = 0.488. However, a significant three-way interaction was observed between the retention interval, receipt of a video modeling example and level of inquiry, F(1,170) = 3.96, p = 0.048, η_p^2 = 0.023. Thus, the usefulness of a video modeling examples depends on the level of inquiry and the measurement time point. Therefore, subsequent ANOVAs, *post-hoc* tests, t-tests and multilevel analyses were performed.

Consistent with our expectations, the results of the intermediate test after the first manipulation (video modeling example vs. no example) were higher among students who watched a video modeling example, t(172) = 5.48, p < 0.001, d = 0.97 (**Figure 2A**).

All students achieved equal results on the assessment immediately after the subsequent inquiry task, regardless of the manipulation. Students who watched a video modeling example before solving a guided or structured inquiry task only outperformed students who did not receive an example in the delayed tests, MD = 0.11, SE = 0.05, 95% CI [0.03,0.20], p = 0.011. No differences were observed at any time point between the levels of inquiry.

Post-hoc analyses (Bonferroni-corrected) of interaction effects revealed that the retention of scientific reasoning skills significantly decreased between the two measurement points in the GI-VME, MD = 0.19, SE = 0.05, 95% CI [0.09, 0.28], p < 0.001, and SI+VME groups, MD = 0.08, SE = 0.03, 95% CI [0.02, 0.13], p = 0.007, but remained stable in the GI+VME and SI-VME groups (**Figure 2B**). Furthermore, students who watched a video modeling example before solving a guided inquiry task (GI+VME) achieved higher learning outcomes in the second assessment test than students who did not receive an example before solving the same inquiry task (GI-VME), P2: MD = 0.13, SE = 0.06, 95% CI [0.03, 0.25], p = 0.045. No differences were observed in the results of both assessment tests for paired comparisons of structured inquiry (SI) (**Figure 2B**).

FIGURE 2 | (A) Mean score of correct responses by treatment 1 (video modeling example vs. no video modeling example) in percentage. (B) Mean score of correct responses by treatments and time in percentage.

In addition, multilevel analyses were conducted with the R packages lme4 (Bates et al., 2015), lmerTest and lsmeans (Lenth, 2016) in the R environment, version 3.4.4 (R Core Team, 2018) to determine differences between the inquiry levels and between groups provided with or without an example while controlling for class effects. The presence of a video modeling example (VME, no example) and the level of inquiry (guided inquiry or structured inquiry) were the independent variables; the dependent variable was scores on the two tests measuring students' achievement (P1 and P2). We controlled for classes to remove variation in the dependent variable resulting from class effects. Again, no significant differences were observed in the assessment performed immediately after inquiry, P1: $\beta = 0.029$ ($SE = 0.035$), and in the subsequent assessment measure, P2: $\beta = 0.019$ ($SE = 0.034$) between the treatments when controlling for class effects. However, the GI-VME group still produced the worst descriptive results for Post-test 2 compared to all other treatments.

Students' Cognitive Load (H3)

In univariate and multivariate analyses of variance, we did not observe a main effect of the video modeling example on overall cognitive load, and only marginally significant differences in germane load, $F(1,170) = 2.91$, $p = 0.090$. However, main effects of the inquiry level on overall cognitive load, $F(1,170) = 5.52$, $p = 0.020$, and extraneous load, $F(1,170) = 8.09$, $p = 0.005$, were observed. Overall cognitive load was lower in the SI+VME group

($M_{SI+} = 1.94$, $SD = 0.50$) than in the GI+VME group ($M_{GI+} = 2.26$, $SD = 0.69$), $MD = 0.315$, $SE = 0.11$, 95% CI [0.02, −0.61], $p = 0.028$, although both groups were exposed to the introductory video modeling example.

Pairwise comparisons of the two conditions (GI+VME vs. SI+VME; Bonferroni-corrected) revealed that this effect was due to an increased extraneous load caused by generation in guided inquiry, $MD = 0.52$, $SE = 0.13$, 95% CI [0.17, 0.86], $p = 0.001$. Only marginally significant differences were observed in the intrinsic load: $MD = 0.32$, $SE = 0.13$, 95% CI [−0.66, 0.019], $p = 0.076$, and no significant differences were observed in the germane load. Pairwise comparisons revealed no differences between the two inquiry levels when a video modeling example was not presented (GI-VME vs. SI-VME; Bonferroni-corrected): $M_{SI−} = 2.17$, $SD = 0.71$; $M_{GI−} = 2.38$, $SD = 0.75$.

Furthermore, detailed analyses of the two guided and structured inquiry conditions (GI+VME vs. GI-VME, SI+VME vs. SI-VME) revealed no significant differences in any of the three types of cognitive load. However, the GI-VME group exhibited the worst descriptive results for germane load (**Figure 3**).
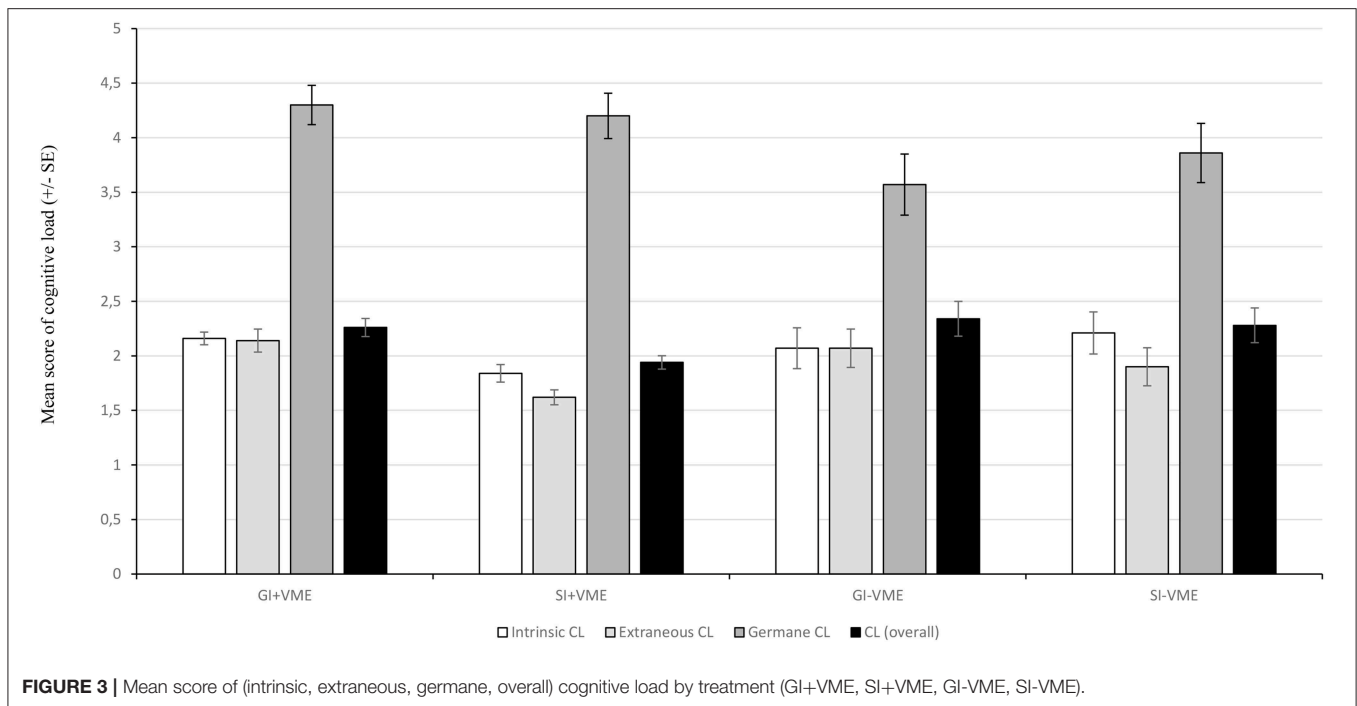
Students' Performance Success (H4)

The students' experimental plans and methodological discussions were investigated to assess how much information each individual student in the GI group was able to successfully generate and at what frequency (total score = 33). Thus, this assessment represented an analysis of the role of generation success.

No significant benefits of combining video modeling examples with guided inquiry were observed with respect to generation success, $t(87) = 1.91$, $p = 0.060$, although a clear descriptive difference was observed (**Table 3**).

DISCUSSION

The goal of the present study was to investigate the necessity of combining example-based learning with different levels of inquiry-based learning for the acquisition of scientific reasoning skills. Therefore, we analyzed the benefit of (a) presenting vs. (b) omitting a video modeling example before (1) an inquiry activity involving the generation of scientific reasoning skills (guided inquiry) vs. (2) an inquiry activity that had students simply read instructions for an experimental plan and an appropriate methodological discussion (structured inquiry). A computer-based learning program that contained a video modeling example of how to investigate an authentic scientific research question by following the inquiry cycle was developed for the purpose of the study as preparation for the subsequent inquiry task. Effects on the learning process, short-term and long-term learning outcomes in terms of scientific reasoning skills, and crucial prerequisites for effectiveness, such as performance success and perceived cognitive load, were measured.

Hypotheses (**H1**) and (**H2**) were partially verified, as watching a video modeling example of how to solve a scientific problem by following the inquiry cycle and using the CVS positively affected learning outcomes in guided, but not structured, inquiry (**H1**), particularly in the long term (**H2**). A significant decrease in

FIGURE 3 | Mean score of (intrinsic, extraneous, germane, overall) cognitive load by treatment (GI+VME, SI+VME, GI−VME, SI−VME).

retention was observed over a period of 1 week for guided inquiry when a video modeling example was not provided. However, the expected worked example effect for guided inquiry after a 1-week delay was not significant.

Consistent with our expectations, structured inquiry with a video modeling example resulted in the lowest cognitive load. However, in contrast to our hypothesis (**H3**), the provision of a video modeling example did not significantly reduce learners' intrinsic and extraneous cognitive load or increase germane load in guided inquiry.

Regardless of the treatment, students obtained equal results on assessments after and during the inquiry task (performance). Therefore, our hypothesis (**H4**) was not confirmed. Since the results of an intermediate test were higher among students who watched a video modeling example, the lack of differences between conditions during and after inquiry might be related to the fact that the inquiry task was designed in such a way that all students—regardless of whether they had been provided with a video modeling example—were able to plan, conduct and analyze a scientific experiment using the CVS.

Guided vs. Structured Inquiry

Consistent with the findings reported by Kaiser et al., extraneous load was significantly higher in the structured inquiry group (with a video modeling example) compared to the guided inquiry group (with a video modeling example). Nevertheless, both levels of inquiry were equally effective. No generation effect was observed after a 1-week delay. Students in the structured inquiry group still had higher performance in terms of absolute numbers. In contrast to Kaiser et al., we only identified a descriptive, insignificant short-term disadvantage among students who actively generated information in guided

inquiry. A potential explanation for this finding is that our short-term assessment used both open-ended items and single choice items, whereas Kaiser et al. only used a closed response format. According to Hirshman and Bjork (1988), a generation advantage or disadvantage is sensitive to different types of memory tests (recognition, cued recall, and free recall). Solving a generation task with an open-ended format in the inquiry-based learning environment may increase performance on open-ended retention test items. Conversely, students who passively receive information about the experimental plan and methodological discussion in structured inquiry may have an advantage in a recognition format (e.g., single choice items) (*transfer appropriate processing*, Morris et al., 1977). Therefore, an equal number of single choice and open-ended items was essential to ensure a fair comparison of both conditions. Furthermore, answering open-ended questions is a more demanding process for students, but enabled us to evaluate higher levels of competence in scientific inquiry (Mayer et al., 2008), which requires further analysis. Finally, although all students performed significantly better on single choice questions than open-ended questions, the difference between the two formats was indeed higher in the structured inquiry group.

We expected that students who engaged in guided inquiry, which required them to actively adopt the CVS, after watching a video modeling example would exhibit a lower forgetting rate than students who engaged in structured inquiry. In fact, students who had engaged in guided inquiry with a video modeling example exhibited the same performance on both tests, while retention significantly decreased among students who had engaged in structured inquiry. Based on these results, guided inquiry is potentially more effective in teaching students CVS in terms of memory and knowledge sustainability (*storage strength*,

Bjork and Bjork, 1992). However, further research controlling for generation success (Kaiser et al., 2018) is needed to confirm a long-lasting effect.

Guided Inquiry

Watching a video modeling example before completing an inquiry task was beneficial for students who were later asked to actively generate their own experimental design using the CVS, since retention in this treatment group did not decrease within a week. These results confirm our first two hypotheses (**H1** and **H2**), and are somewhat consistent with the findings reported by Kant et al. (2017) and Chen et al. (2016). However, a worked example effect did not arise. In contrast to the results presented by Kant and colleagues, in which a clear worked example effect was immediately observed for video modeling examples on virtual inquiry learning, video modeling examples only affect long-term retention in guided inquiry in the present study. When a video modeling example was omitted, retention significantly decreased over a period of 1 week. Our finding of a long-term advantage of watching a video modeling example for guided inquiry is consistent with the findings reported by Hübner et al. (2010) and Chen et al. (2016), who revealed the long-term effectiveness of worked examples.

According to our results, a video modeling example enabled students to borrow information from the non-visible model by utilizing the strategies discussed and applied in the video modeling example (Bandura, 1986). The video modeling example helped students focus on relevant aspects and procedures to acquire new cognitive schemata for planning and discussing scientific investigations during guided inquiry. These findings support the notion that learning through modeling is more than just simple imitation (Bandura, 1986). Reliance on observed strategies when solving a less structured inquiry task enabled the students to increase their working memory capacity during inquiry and helped foster their *storage strength* (Bjork and Bjork, 1992) for the observed strategies for up to 1 week. Thus, the generated information from the inquiry session was permanently integrated into the cognitive schemata acquired from the video modeling example, whereas new information generated during guided inquiry did not result in the same linkages with preexisting knowledge and thus did not exhibit the same storage strength in the absence of a video modeling example. Consistent with these results, participants who received a video modeling example before guided inquiry reported a higher germane cognitive load during inquiry than students who were not provided with an example. However, the difference was only marginally significant (**H3**). Nevertheless, since the retention of students who were provided with a video modeling example before guided inquiry did not decrease, a single video modeling example appears to be sufficient to guide students' attention to appropriate cognitive schemata, which fosters the long-term learning of inquiry skills (Scheiter et al., 2004; Crippen and Earl, 2007; Schworm and Renkl, 2007; Sweller et al., 2011; Chen et al., 2016). Unexpectedly and in contrast to the results from the study by Kant and colleagues on video modeling examples in virtual inquiry, one example did not appear to be sufficient to significantly reduce the intrinsic and extraneous load. A single example might be insufficient to significantly

reduce the cognitive load in physical, hands-on investigations. However, the lack of significant differences might also have been due to insufficient power for small effects (*post-hoc* power analysis: a significance level of a $= 0.05$ and a small effect size of $d = 0.2$ yielded a power of 0.2) and the fact that the test for cognitive load exhibited only an acceptable level of reliability ($\alpha = 0.66$) for comparing groups (Lienert and Raatz, 1998). Consequently, the results should be interpreted with caution.

Structured Inquiry

Students who watched a video modeling example before engaging in a structured inquiry task reported the lowest level of cognitive load. Consequently, participants who had received a video modeling example perceived the inquiry tasks as less cognitively demanding than students who did not watch an example or students who were provided with less instructional guidance during inquiry. However, the use of the borrowing and reorganizing principle to reduce the cognitive load and thus free more working memory capacity to focus on problem-solving strategies and construct useful cognitive schemata for solving the subsequent inquiry task (Sweller and Sweller, 2006) did not improve learning outcomes in structured inquiry. The students who completed a structured inquiry task achieved equal results, regardless of whether they were provided with a video modeling example. Additional guidance in the form of a video modeling example appears to have no long-term effect on inquiry tasks that are already strongly guided via direct instructions, as is typically the case in structured inquiry (Chen et al., 2016). A learner with a higher level of prior knowledge will perceive a lower element interactivity and require less guidance to solve a problem (Sweller, 1994; Chen et al., 2016). According to Chen and colleagues, the worked example effect only arises when element interactivity is high, resulting in a high intrinsic cognitive load. If the intrinsic cognitive load is already low, control of the extraneous cognitive load using worked examples is unnecessary because the total cognitive load does not threaten to overload the working memory capacity (*element interactivity effect*, Chen et al., 2016). Nevertheless, we did not observe an expertise reversal effect (Kalyuga et al., 2003). Based on these findings, solving an inquiry task at a low level of inquiry after watching a video modeling example is still challenging for students because, first, the video modeling example and the inquiry task (reading task) in structured inquiry were non-redundant. The strategies and procedures illustrated in the example were required to be applied to a completely new experiment. These conditions might have simultaneously challenged and motivated the students. Second, working memory is already taxed by physical, hands-on investigations (physical lab experiences), which require students to work with information with high element interactivity (Chen et al., 2016) and use a complex hypothetico-deductive procedure.

Further Limitations

Moreover, the following limitations must be considered when drawing conclusions from the experiment. First, the long-term disadvantage observed for the subsample of students who were not provided with a video modeling example might simply result from their spending less time with the learning material. Future research should compare groups of students who merely

study an example of how to solve a practice problem vs. actually solve a practice problem for the same amount of time to control for this limitation. Second, the intermediate assessment test and the test for cognitive load exhibited only an acceptable level of reliability ($\alpha = 0.60$ and $\alpha = 0.66$) for comparing groups (Lienert and Raatz, 1998). Consequently, the results should be interpreted with caution. Moreover, the subsample was too small for a detailed analysis. Due to the resulting small power, we were unable to apply techniques such as pathway analyses of the four individual conditions. An investigation designed to assess which and to what extent learner characteristics (cognitive load, NFC, KFT, grades, and generation success) affect the short-term and long-term retention of each treatment group would be interesting. Thus, replications are required. Furthermore, randomization within each class was confined to the second manipulation (inquiry level), while the first manipulation was conducted between classes. We were unable to intermix students within classes with respect to the first manipulation for organizational reasons. Third, the students participated in a physical inquiry-based lab experiment in all four conditions. These settings provide an authentic picture of scientific practice and support the application of authentic scientific procedures. On the other hand, higher authenticity is always sensitive to interferences and accompanied by a greater cognitive burden. The application of newly acquired inquiry skills and correct handling and manipulation of physical equipment might be very challenging for students. Moreover, authentic experimental settings include a large number of features that can cause a higher extraneous cognitive load and distractions, as students may focus on insignificant aspects. Hence, due to the reliance on physical experiments, the extraneous cognitive load was high in this study and might have obscured small differences between the treatments. Future research should analyze how to further reduce the extraneous cognitive load, particularly in guided inquiry, since structured inquiry (with a video modeling example) proved to be the least cognitively demanding condition. Consistent with the theory of transfer appropriate processing, the use of the same (digital) medium in both sessions—a learning program with a video modeling example in the introductory session and an accompanying digital scaffold for the hands-on inquiry-based learning environment instead of a human supervisor—might be beneficial.

IMPLICATIONS

In terms of the theoretical implications, this study broadens the research base on video modeling examples and the generation effect, as well as the unresolved didactic question of whether direct instruction or discovery-based methods deliver better learning outcomes and retention to a certain extent (Dean and Kuhn, 2007; Furtak et al., 2012).

In contrast to our expectations and recent findings on the generation effect (e.g., Chen et al., 2016), guided inquiry did not prove to be more beneficial than structured inquiry. As long as guided inquiry was preceded by a video modeling example, both levels of inquiry were equally effective. Consistent with recent studies on example-based learning (van Gog et al., 2011; Leppink et al., 2014; Kant et al., 2017), students who watched

a video modeling example in the present study benefitted from being provided with an indication of which elements should be considered when solving an inquiry task. They achieved the same performance results after a period of 1 week had elapsed, while retention was significantly decreased when a video modeling example was not provided in guided inquiry. Thus, a video modeling example affected how much mental effort students were able to invest in solving the inquiry task and promoted the integration of generated information into the cognitive schemata acquired from the example.

Generation in guided inquiry-based learning leads to better long-term learning outcomes when the germane cognitive load is increased through the use of a video modeling example. However, ultimately, higher learning outcomes are influenced either by providing a video modeling example or by directly providing a higher level of instructional guidance during inquiry.

CONCLUSIONS

Sufficient knowledge serves as a foundation for long-term retention by providing anchors to assimilate new information into preexisting cognitive schemata and facilitating retrieval. Guided inquiry does not automatically promote deeper learning and retention. Video modeling examples are required to provide a sufficient foundation in terms of scientific reasoning skills and increase working memory capacity. Ultimately, video modeling examples are effective for long-term learning gains in guided inquiry when teaching scientific reasoning skills in inquiry-based learning. In structured inquiry, they but have no significant benefit for long-term retention. But at least they can reduce the cognitive load.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript and **Supplementary Material** (computer-based introductory session, research workbooks, and instruments) are available to any qualified researcher: https://osf.io/uvrwn/?view_only= 4c8a7819fcae451a8ed9cdaef63f06f1.

ETHICS STATEMENT

No ethics approval was required for the present study according to national guidelines as well as the University of Kassel's own guidelines. The study was conducted in accordance with the recommendations of the University of Kassel's ethics committee and with the approval of the Ministry of Education and Cultural Affairs, Hesse, Germany (Hessisches Kultusministerium) (cf. Education Act of Hesse, section 84). The parents of all participants gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

JM developed the basic idea for the present study and supervised the project. He also took the lead on project

REFERENCES

Abrams, E., Southerland, S. A., and Evans, C. (2008). "Inquiry in the classroom: identifying necessary components of a useful definition," in *Inquiry in the Science Classroom: Challenges and Opportunities,* eds E. Abrams, S. Southerland, and P. Silva (Charlotte, NC: Information Age Publishing), 11–42.

Aleven, V. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cogn. Sci.* 26, 147–179. doi: 10.1207/s15516709cog2602_1

Alfieri, L., Brooks, P. J., Aldrich, N. J., and Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *J. Educ. Psychol.* 103, 1–18. doi: 10.1037/a0021017

Arnold, J. C., Kremer, K., and Mayer, J. (2014). Understanding students' experiments—what kind of support do they need in inquiry tasks? *Int. J. Sci. Educ.* 36, 2719–2749. doi: 10.1080/09500693.2014.930209

Atkinson, R. K., Derry, S. J., Renkl, A., and Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Rev. Educ. Res.* 70, 181–214. doi: 10.3102/00346543070002181

Ausubel, D. P. (1964). The transition from concrete to abstract cognitive functioning: theoretical issues and implications for education. *J. Res. Sci. Teach.* 2, 261–266. doi: 10.1002/tea.3660020324

Baars, M., van Gog, T., de Bruin, A., and Paas, F. (2018). Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks. *Stud. Educ. Evaluat.* 58, 51–59. doi: 10.1016/j.stueduc.2018.05.010

Baddeley, A. (1986). *Oxford Psychology Series, No. 11. Working Memory.* New York, NY: Clarendon Press/Oxford University Press.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bell, R. L., Smetana, L., and Binns, I. (2005). Simplifying inquiry instruction. *Sci. Teach.* 72, 30–33.

Bjork, E. L., and Bjork, R. A. (2014). "Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning," in *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 2nd Edn*, eds M. A. Gernsbacher and J. Pomerantz (New York, NY: Worth), 59–68.

Bjork, R. A., and Bjork, E. L. (1992). "A new theory of disuse and an old theory of stimulus fluctuation," in *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes,* Vol. 2, eds A. Healy, S. Kosslyn, and R. Shiffrin (Hillsdale, NJ: Erlbaum), 35–67.

Blandin, Y., Lhuisset, L., and Proteau, L. (1999). Cognitive processes underlying observational learning of motor skills. *Q. J. Exp. Psychol. Hum. Exp. Psychol.* 52 A, 957–979. doi: 10.1080/713755856

Borek, A., McLaren, B. M., Karabinos, M., and Yaron, D. (2009). "How much assistance is helpful to students in discovery learning?" in *Learning in the Synergy of Multiple Disciplines. EC-TEL 2009. Lecture Notes in Computer Science,* Vol. 5794, eds U. Cress, V. Dimitrova, and M. Specht (Berlin; Heidelberg: Springer).

Braaksma, M., Rijlaarsdam, G., and Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *J. Educ. Psychol.* 94, 405–415. doi: 10.1037//0022-0663.94.2.405

Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). *How People Learn.* Washington, DC: National Academies Press.

Chandler, P., and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cogn. Instr.* 8, 293–240. doi: 10.1207/s1532690xci0804_2

Chandler, P., and Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *Br. J. Educ. Psychol.* 62: 233–246. doi: 10.1111/j.2044-8279.1992.tb01017.x

Chandler, P., and Sweller, J. (1996). Cognitive load while learning to use a computer program. *Appl. Cogn. Psychol.* 10, 151–170. doi: 10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Chen, O., Kalyuga, S., and Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learn. Instr.* 45, 20–30. doi: 10.1016/j.learninstruc.2016.06.007

Chen, Z., and Klahr, D. (1999). All other things being equal: acquisition and transfer of the control of variables strategy. *Child Dev.* 70, 1098–1120. doi: 10.1111/1467-8624.00081

Cierniak, G., Scheiter, K., and Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Comput. Human Behav.* 25, 315–324. doi: 10.1016/j.chb.2008.12.020

Clark, D., and Linn, M. C. (2003). Designing for knowledge integration. The impact of instructional time. *J. Learn. Sci.* 12, 451–493. doi: 10.1207/S15327809JLS1204_1

Colburn, A. (2000). An inquiry primer. *Sci. Scope* 23, 42–44.

Collins, A., Brown, J. S., and Newman, S. E. (1989). "Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics," in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser,* ed L. B. Resnick (Hillsdale, NJ: Lawrence Erlbaum Associates), 453–494.

Cooper, G., and Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *J. Educ. Psychol.* 79, 347–362. doi: 10.1037/0022-0663.79.4.347

Craig, R. (1956). Directed versus independent discovery of established relations. *J. Educ. Psychol.* 47, 223–235. doi: 10.1037/h0046768

Crippen, K. J., and Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Comput. Educ.* 49, 809–821. doi: 10.1016/j.compedu.2005.11.018

Dean, D. Jr., and Kuhn, D. (2007). Direct instruction vs. discovery: the long view. *Sci. Ed.* 91, 384–397. doi: 10.1002/sce.20194

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Foos, P. W., Mora, J. J., and Tkacz, S. (1994). Student study techniques and the generation effect. *J. Educ. Psychol.* 86, 567–576. doi: 10.1037/0022-0663.86.4.567

Furtak, E. M., Seidel, T., Iverson, H., and Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Rev. Educ. Res.* 82, 300–329. doi: 10.3102/0034654312457206

Gardiner, J. M., and Hampton, J. A. (1985). Semantic memory and the generation effect: some tests of the lexical activation hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 732–741. doi: 10.1037//0278-7393.11.1-4.732

Ginns, P. (2005). Meta-analysis of the modality effect. *Learn. Instr.* 15, 313–331. doi: 10.1016/j.learninstruc.2005.07.001

Heller, K. A., and Perleth, C. (2000). *Kognitiver Fähigkeits-Test (Rev.) für 5.-12. Klassen (KFT 5-12+ R).* Göttingen: Beltz-Testgesellschaft.

Hilbert, T. S., and Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Comput. Human Behav.* 25, 267–274. doi: 10.1016/j.chb.2008.12.006

Hilbert, T. S., Renkl, A., Kessler, S., and Reiss, K. (2008). Learning to prove in geometry: learning from heuristic examples and how it can be supported. *Learn. Instr.* 18, 54–65. doi: 10.1016/j.learninstruc.2006.10.008

Hirshman, E., and Bjork, R. A. (1988). The generation effect: support for a two-factor theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 484–494. doi: 10.1037//0278-7393.14.3.484

Hmelo-Silver, C. E., Duncan, R. G., and Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educ. Psychol.* 42, 99–107. doi: 10.1080/00461520701263368

Hof, S. (2011). *Wissenschaftsmethodischer kompetenzerwerb durch forschendes lernen: entwicklung und evaluation einer interventionsstudie.* Ph.D. thesis, Universität Kassel, Kassel, Germany.

Hoogerheide, V., Loyens, S. M. M., and van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learn. Instr.* 33, 108–119. doi: 10.1016/j.learninstruc.2014.04.005

Hübner, S., Nückles, M., and Renkl, A. (2010). Writing learning journals: instructional support to overcome learning-strategy deficits. *Learn. Instr.* 20, 18–29. doi: 10.1016/j.learninstruc.2008.12.001

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *J. Verbal Learn. Verbal Behav.* 17, 649–667. doi: 10.1016/S0022-5371(78)90393-6

Janoschek, K. (2009). *Empirische studie zum kumulativen kompetenzaufbau des experimentierens mit lebenden tieren (asseln).* Ph.D. thesis, Universtät Wien, Wien, Austria.

Johnson, M. A., and Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *J. Res. Sci. Teach.* 35, 89–103. doi: 10.1002/(SICI)1098-2736(199801)35:1<89::AID-TEA6>3.0.CO;2-J

Kaiser, I., Mayer, J., and Malai, D. (2018). Self-generation in the context of inquiry-based learning. *Front. Psychol.* 9:2440. doi: 10.3389/fpsyg.2018.02440

Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. (2003). The expertise reversal effect. *Educ. Psychol.* 38, 23–31. doi: 10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. (2001). When problem solving is superior to studying worked examples. *J. Educ. Psychol.* 93, 579–588. doi: 10.1037/0022-0663.93.3.579

Kant, J. M., Scheiter, K., and Oschatz, K. (2017). How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learn. Instr.* 52, 46–58. doi: 10.1016/j.learninstruc.2017.04.005

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* 41, 75–86. doi: 10.1207/s15326985ep4102_1

Klahr, D. (2000). *Exploring Science: The Cognition and Development of Discovery Processes.* Cambridge, MA: MIT Press.

Klahr, D., and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cogn. Sci.* 12, 1–48. doi: 10.1207/s15516709cog1201_1

Klahr, D., and Nigam, M. (2004). The equivalence of learning paths in early science instruction: effect of direct instruction and discovery learning. *Psychol. Sci.* 15, 661–667. doi: 10.1111/j.0956-7976.2004.00737.x

Kühl, T., Scheiter, K., Gerjets, P., and Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Comput. Human Behav.* 27, 29–35. doi: 10.1016/j.chb.2010.05.008

Kuhn, D., and Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychol. Sci.* 16, 866–870. doi: 10.1111/j.1467-9280.2005.01628.x

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159. doi: 10.2307/2529310

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

Leahy, W., and Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Appl. Cogn. Psychol.* 25, 943–951. doi: 10.1002/acp.1787

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Softw.* 69: 1–33. doi: 10.18637/jss.v069.i01

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., and van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn. Instr.* 30, 32–42. doi: 10.1016/j.learninstruc.2013.12.001

Lienert, G. A., and Raatz, U. (1998). *Testaufbau und Testanalyse, 6th Auflage.* Weinheim: Beltz.

Linn, M. C., Pulos, S., and Gans, A. (1981). Correlates of formal reasoning: content and problem effects. *J. Res. Sci. Teach.* 18, 435–447. doi: 10.1002/tea.3660180507

Lutz, J., Briggs, A., and Cain, K. (2003). An examination of the value of the generation effect for learning new material. *J. Gen. Psychol.* 130, 171–188. doi: 10.1080/00221300309601283

Mayer, J. (2007). "Erkenntnisgewinnung als wissenschaftliches Problemlösen," in *Theorien in der Biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden,* eds D. Krüger and H. Vogt (Berlin: Springer), 177–186.

Mayer, J., Grube, C., and Möller, A. (2008). "Kompetenzmodell naturwissenschaftlicher erkenntnisgewinnung [Competence model for scientific inquiry]," in *Lehr- und Lernforschung in der Biologiedidaktik,* eds U. Harms and A. Sandmann (Innsbruck: Studien Verlag), 63–79.

Mayer, J., and Ziemek, H. P. (2006). Offenes experimentieren. forschendes lernen im biologieunterricht. *Unterr. Biol.* 317, 4–12.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *Am. Psychol.* 59, 14–19. doi: 10.1037/0003-066X.59.1.14

Mayer, R. E., and Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *J. Educ. Psychol.* 90, 312–320. doi: 10.1037/0022-0663.90.2.312

McDaniel, M. A., Waddill, P. J., and Einstein, G. O. (1988). A contextual account of the generation effect: a three-factor theory. *J. Mem. Lang.* 27, 521–536. doi: 10.1016/0749-596X(88)90023-X

McLaren, B. M., Lim, S., and Koedinger, K. R. (2008). "When and how often should worked examples be given to students? New results and a summary of the current state of research," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society,* eds B. C. Love, K. McRae, and V. M. Sloutsky (Austin, TX: Cognitive Science Society), 2176–2181.

Meier, M., and Wulff, C. (2014). "Daphnia magna as the ultimate classroom organism: Implementing scientific investigations into school practice," in *Daphnia: Biology and Mathematics Perspectives,* ed M. O. El-Doma (New York, NY: Nova Science Publishers Inc.), 225–244.

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *J. Verbal Learn. Verbal Behav.* 16, 519–533. doi: 10.1016/S0022-5371(77)80016-9

Mousavi, S. Y., Low, R., and Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *J. Educ. Psychol.* 87, 319–334. doi: 10.1037/0022-0663.87.2.319

Mulder, Y. G., Lazonder, A. W., and de Jong, T. (2014). Using heuristic worked examples to promote inquiry-based learning. *Learn. Instr.* 29, 56–64. doi: 10.1016/j.learninstruc.2013.08.001

Nairne, J. S., and Widner, R. L. (1987). Generation effects with nonwords: the role of test appropriateness. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 164–171. doi: 10.1037/0278-7393.13.1.164

National Research Council (2013). *Next Generation Science Standards.* Washington, DC: National Academies Press.

OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: OECD. Retrieved from http://www.nbbmuseum.be/doc/seminar2010/nl/bibliografie/opleiding/analysis.pdf. doi: 10.1787/9789264040014-en

Paas, F., and van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem solving skills: a cognitive-load approach. *J. Educ. Psychol.* 86, 122–133. doi: 10.1037/0022-0663.86.1.122

Payne, D. G., Neely, J. H., and Burns, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Mem. Cogn.* 14, 246–252. doi: 10.3758/BF03197700

Preckel, F. (2014). Assessing need for cognition in early adolescence. *Eur. J. Psychol. Assess.* 30, 65–72. doi: 10.1027/1015-5759/a000170

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cogn. Sci.*, 21, 1–29. doi: 10.1207/s15516709cog2101_1

Renkl, A. (2014). "Learning from worked examples: how to prepare students for meaningful problem solving," in *Applying Science of Learning in Education: Infusing Psychological Science into the Curriculum,* eds V. A. Benassi, C. E. Overson, and C. M. Hakala (Washington, DC: Society for the Teaching of Psychology), 118–130.

Richland, L. E., Linn, M. C., and Bjork, R. A. (2007). "Cognition and instruction: Bridging laboratory and classroom settings," in *Handbook of Applied Cognition, 2nd Edn.,* eds F. Durso, R., Nickerson, S., Dumais, S. Lewandowsky, and T. Perfect (West Sussex: John Wiley & Sons Ltd.), 555–583.

Rittle-Johnson, B., and Kmicikewycz, A. O. (2008). When generating answers benefits arithmetic skill: the importance of prior knowledge. *J. Exp. Child Psychol.* 101, 75–81. doi: 10.1016/j.jecp.2008.03.001

Roth, W.-M., Bowen, G. M., and McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *J. Res. Sci. Teach.* 36, 977–1019. doi: 10.1002/(SICI)1098-2736(199911)36:9<977::AID-TEA3>3.0.CO;2-V

Rummel, N., and Spada, H. (2005). Learning to collaborate: an instructional approach to promoting collaborative problem solving in computer-mediated settings. *J. Learn. Sci.* 14, 201–241. doi: 10.1207/s15327809jls1402_2

Rummel, N., Spada, H., and Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *Comput. Support. Learn.* 4, 69–92. doi: 10.1007/s11412-008-9054-4

Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., and McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educ. Psychol. Rev.* 22, 379–392. doi: 10.1007/s10648-010-9143-6

Scheiter, K., Gerjets, P., and Schuh, J. (2004). "The impact of example comparisons on schema acquisition: do learners really need multiple examples?," in *Proceedings of the 6th International Conference on Learning Sciences,* eds Y. B. Kafai, W. A. Sandova, N. Enyedy, A. S. Nixon, and F. Herrera (Mahwah, NJ: Erlbaum), 457–464. Retrieved from http://dl.acm.org/citation.cfm?id=1149182

Schwab, J. (1962). "The teaching of science as enquiry", in *The Teaching of Science*, eds J. Schwab and P. Brandwein (Cambridge, MA: Harvard University Press), 1–103.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., and Härtig, H. (2016). Teaching the control-of-variables strategy: a meta-analysis. *Dev. Rev.* 39, 37–63. doi: 10.1016/j.dr.2015.12.001

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., and Salden, R. (2009). The worked-example effect: not an artefact of lousy control conditions. *Comput. Human Behav.* 25, 258–266. doi: 10.1016/j.chb.2008.12.011

Schworm, S., and Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *J. Educ. Psychol.* 99, 285–296. doi: 10.1037/0022-0663.99.2.285

Shrager, J., and Siegler, R. S. (1998). SCADS: a model of children's strategy choices and strategy discoveries. *Psychol. Sci.* 9, 405–410. doi: 10.1111/1467-9280.00076

Shulman, L., and Keisler, E. (1966). *Learning by Discovery: A Critical Appraisal.* Chicago: Rand McNally.

Siegler, R. S. (1991). Strategy choice and strategy discovery. *Learn. Instr.* 1, 89–102. doi: 10.1016/0959-4752(91)90020-9

Slamecka, N. J., and Graf, P. (1978). The generation effect. Delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 592–604. doi: 10.1037/0278-7393.4.6.592

Sneider, C., Kurlich, K., Pulos, S., and Friedman, A. (1984). Learning to control variables with model rockets: a neo-piagetian study of learning in field settings. *Sci. Ed.* 68, 465–486. doi: 10.1002/sce.3730680410

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (2011). "Cognitive load theory," in *The Psychology of Learning and Motivation: Cognition in Education,* Vol. 55, eds J. P. Mestre and B. H. Ross (San Diego, CA: Elsevier Academic Press), 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive Load Theory.* New York, NY: Springer.

Sweller, J., and Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cogn. Instr.* 2, 59–89. doi: 10.1207/s1532690xci0201_3

Sweller, J., and Sweller, S. (2006). Natural information processing systems. *Evol. Psychol.* 4, 434–458. doi: 10.1177/147470490600400135

Sweller, J., Van Merrienboer, J. J. G., and Paas, F. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10. 251–296. doi: 10.1023/A:1022193728205

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., and Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Comput. Human Behav.* 25, 785–791. doi: 10.1016/j.chb.2009.02.007

van Gog, T., Kester, L., and Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemp. Educ. Psychol.* 36, 212–218. doi: 10.1016/j.cedpsych.2010.10.004

van Gog, T., and Rummel, N. (2010). Example-based learning: integrating cognitive and social-cognitive research perspectives. *Educ. Psychol. Rev.* 22. 155–174. doi: 10.1007/s10648-010-9134-7

van Gog, T., Verveer, I., and Verveer, L. (2014). Learning from video modeling examples: effects of seeing the human model's face. *Comput. Educ.* 72, 323–327. doi: 10.1016/j.compedu.2013.12.004

van Merriënboer, J. J. G., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0

vanLehn, K. (1996). Cognitive skill acquisition. *Annu. Rev. Psychol.* 47, 513–539. doi: 10.1146/annurev.psych.47.1.513

Vollmeyer, R., and Burns, B. D. (1996).Hypotheseninstruktion und zielspezifität: bedingungen, die das erlernen und kontrollieren eines komplexen systems beeinflussen. *Z. Exp. Psychol.* 43, 657–683.

Wellnitz, N., and Mayer, J. (2016). "Methoden der erkenntnisgewinnung im biologieunterricht," in *Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände,* eds A. Sandmann and P. Schmiemann (Berlin: Logos Verlag Berlin), 61–83.

Zimmerman, B. J., and Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *J. Educ. Psychol.* 94, 660–668. doi: 10.1037/0022-0663.94.4.660

Zimmerman, C., Bisanz, G. L., and Bisanz, J. (1998). Everyday scientific literacy: do students use information about the social context and methods of research to evaluate news briefs about science? *Alberta J. Educ. Res.* 44, 188–207.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership