Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Learning about science topics of social relevance using lower and higher autonomy-supportive scaffolds

Eric C. Schoute [a,*], Janelle M. Bailey [b], Doug Lombardi [a]

[a] Department of Human Development and Quantitative Methodology, University of Maryland, United States
[b] Department of Teaching and Learning, Temple University, United States

ARTICLE INFO

Keywords:
Plausibility
Scientific evaluation
Earth science learning
Scaffolding
Autonomy
Mediation model

ABSTRACT

Evaluation of plausible alternative explanations of scientific phenomena is an authentic scientific activity. Instructional scaffolding can facilitate students' engagement in such evaluations by facilitating their reflections on how well various lines of scientific evidence support alternative explanations. In the present study, we examined two forms of such scaffolding, with one form providing more autonomy support than the other, to determine whether any differential effects existed between the two. Nearly 300 adolescent students in middle school, high school, and university courses completed two activities on scientific topics of social relevance (e.g., the climate crisis, fossils and fossil fuel use, water resources, and astronomical origins), with the less autonomy-supportive form being completed prior to the more autonomy-supportive form. In line with prior pilot studies, both scaffold types demonstrated significant pre- to post-instructional shifts in plausibility judgments toward the scientific model and gains in knowledge with small to medium effect sizes. A mediation model provided a robust replication of previous findings showing that the indirect path meaningfully linked greater levels of evaluation to more scientific plausibility judgments and topic knowledge, above and beyond the direct relational path linking greater levels of evaluation to topic knowledge. However, we found no difference in relations between the two scaffold types, counter to our hypothesis that the more autonomy-supportive version would lead to better outcomes. This suggests that the implementation of more autonomy-supportive learning environments may be conditional, opening up a promising avenue for additional research, such as looking at specific contexts and how activities could be sequenced to optimize learning.

1. Introduction

Many scientific topics of social relevance, such as the current climate crisis and the availability of freshwater resources, are complex and present considerable learning challenges. These topics, sometimes aptly referred to as socioscientific issues (SSIs), can also lead to controversy that can arise from conflicting perspectives, each with a distinct set of assumptions, points of view, target audiences, and goals (Dawson & Carson, 2020; Sinatra & Lombardi, 2020; Zeidler & Sadler, 2008). Among those perspectives are claims that are based on non-scientific information, presenting alternative claims that are in conflict with the scientific consensus—perspectives that students may encounter in and outside the science classroom (Lombardi, Nussbaum et al., 2016).

Instructional scaffolds that promote students' scientific evaluations may be one way to overcome barriers and facilitate students' SSI learning. Over the past decade, our research team has developed several instructional scaffolds—called Model-Evidence Link (MEL) activities—covering many scientific topics of social relevance (e.g., causes of climate change; availability of freshwater resources; impacts of hydraulic fracturing) in Earth and environmental sciences. A well-developed line of empirical research suggests that these MEL scaffolds facilitate middle and high school students' evaluations of connections between lines of scientific evidence and alternative explanatory models, help students shift students toward a more scientific stance (i.e., judging scientific explanations as more plausible than non-scientific ones), and deepen their knowledge of disciplinary core ideas and scientific practices (see, e.g., Bailey et al., 2022; Dobaria et al., 2022; Klavon et al., 2024; Lombardi, Bailey et al., 2018; Lombardi, Bickel et al., 2018; Lombardi et al., 2013; Medrano et al., 2020).

Above and beyond the benefits of scaffolded science learning,

* Corresponding author at: Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742-1131, United States.
E-mail address: edu@schoute.org (E.C. Schoute).

https://doi.org/10.1016/j.cedpsych.2024.102284

researchers have theorized and conducted empirical studies suggesting that students may be more engaged with the learning process when they perceive their classroom environments to be more autonomy supportive. For example, from a self-determination theory (SDT) perspective, Deci and Ryan (2000) posited that social contexts that support the innate psychological need for autonomy, as well as the need for perceptions of competence and relatedness, increase both intrinsic and extrinsic motivation and yield higher learning outcomes (Okada, 2023). Similarly, Collie and Martin (2017) found that students perceived more adaptable teachers to be more autonomy supportive and have higher levels of achievement than students who perceived their teachers to be less autonomy supportive.

With the intention of facilitating higher autonomy in our SSI learning scaffolds, we modified the first form of the MEL scaffold, which we call the preconstructed MEL (pcMEL), to a more autonomy-supportive form, which we call the build-a-MEL (baMEL; Bailey et al., 2020). Earlier pilot studies suggested that, upon engagement with the baMEL as compared to the pcMEL, students demonstrated significantly but moderately greater pre-to-post shifts in the demonstrated quality of their plausibility judgments toward the scientific claim, as well as modestly greater knowledge gains (Bailey et al., 2022; Klavon et al., 2024; Medrano et al., 2020). In the present study, we more systematically compared the differential effects of the two scaffolds across all the topics on which MEL interventions have been developed to investigate if these promising pilot study results could be more robustly corroborated.

2. Theoretical framework

The educational psychology, developmental psychology, learning sciences, and science education research communities have theorized and empirically tested how the learning of complex scientific concepts and problems can be optimized. From these various perspectives, we relied on the knowledge construction metaphor to guide our inquiry and scaffold design, where individuals build their knowledge and understanding by encoding information (e.g., sensory perceptions) into mental representations for later use (Mayer, 1992; Ormrod, 2017). This knowledge construction metaphor is integral to the cognitive component of scaffolding (Wood et al., 1976), as discussed below in more detail. In addition to this process of cognitive knowledge construction, some researchers have proposed that knowledge is constructed during social and cultural interactions, and as such, knowledge construction is often an externalized process (e.g., Bakhurst, 1995; Blumer, 1986). Importantly, it is this view on knowledge construction that is integral to the scientific process of evaluation by which scientists advance understanding and knowledge of scientific topics of social relevance by systematically and normatively evaluating evidence and explanatory models for observations (Ford, 2015). We considered four foundational aspects crucial to the effectiveness of facilitating students' science construction in developing the theoretical framework for the present study: (a) the process of scientific claim evaluation, (b) the making of plausibility judgments, (c) the scaffolded construction of scientific knowledge, and (d) the importance of autonomy support in scaffolded science learning.

2.1. Scientific evaluations

Members of scientific communities evaluate and assess evidentiary connections that favor or refute certain explanatory and predictive models, hypotheses, and propositions. The effectiveness of this evaluation process is not limited to the process of scientific knowledge generation by members of respective scientific communities but is also relevant to students' knowledge construction about scientific findings and processes (Ford, 2015). Specifically, within the classroom, students may simulate these practices from the research community by evaluating evidence provided to them in light of scientific hypotheses. Evaluation, such as occurs through the simulation of scientific research practices (National Research Council, 2012; NGSS Lead States, 2013), is at the heart of authentic scientific and engineering practices that should be part of classroom learning environments (Ford, 2015; Lombardi et al., 2013). Moreover, extant research suggests that students' learning benefits from the weighing of evidence in light of genuinely competing claims—a situation that more closely resembles knowledge generation as it occurs in real scientific conduct (Duschl et al., 2007). Lombardi, Nussbaum et al. (2016) suggest that situations in which students are presented with credible strands of evidence or premises as well as multiple plausible, competing claims may be more beneficial for their learning, as this better represents the degrees of freedom and uncertainty faced by scientists. However, students typically have difficulty distinguishing evidence from scientific claims (Kuhn & Pearsall, 2000). As such, explicit scaffolding may be necessary to facilitate the critical evaluation of competing propositions in light of specific evidence and, ultimately, to allow for effective learning from this evaluation (Lombardi et al., 2013).

2.2. Plausibility judgments

Within the context of science *learning* or science *instruction,* there is no expectation for students to create or advance scientific knowledge and theory. Rather, the goal is for them to come to a deeper understanding of knowledge and consensus that ensue from the relevant scientific communities, as well as the grounds upon which such determinations are based. Thus, although students can engage in the process of evaluation akin to the process of scientific knowledge generation within expert communities, they are not responsible for collecting primary evidence (e.g., climate data), nor are they expected to generate plausible explanatory models (i.e., hypotheses, theories) from these lines of evidence. However, students can come to a greater understanding of scientific topics of social relevance by engaging with data or premises in relation to plausible, competing, non-scientific explanatory models (Ke et al., 2021). When reasoning with scientific claims and evidence, students often make plausibility judgments—either implicitly or explicitly—regarding specific claims. Plausibility is an often informal and, importantly, tentative judgment of the potential truthfulness of a claim, and plausibility reappraisal can lead to greater knowledge construction and reconstruction (see Lombardi, Nussbaum et al., 2016, for an extensive discussion). Because of the tentative nature of these evaluations, judgments toward multiple explanations can be made, where students may consider more than one explanation to be plausible at one time (Lombardi, 2019). This degree of freedom allows students to suspend commitment to one specific claim as wholly truthful, enabling scientific inquiry into the evidence on which this claim is based, exercising their scientific reasoning abilities, and increasing their scientific knowledge (Lombardi et al., 2022), as well as their understanding of the conditions on which that knowledge is based (Alexander & Schoute, 2022).

Throughout this reasoning process, there may be gaps between the explanations that scientists find plausible and those that students and teachers find plausible, particularly when they are confronted with competing claims that are non-scientific (Lombardi et al., 2013). Importantly, several studies have suggested that the scaffolding of scientific reasoning and learning may mitigate this so-called "plausibility gap" by facilitating explicit evaluation of the connections between lines of scientific evidence and alternative explanatory models to promote plausibility reappraisal and conceptual learning (for an overview, see Lombardi, Bailey et al., 2018; Lombardi et al., 2013).

2.3. Scaffolding science knowledge construction

The concept and practice of scaffolding to help non-experts learn complex problems are well-established within educational psychology, learning sciences, and science education. The term "scaffolding" coined by Wood and colleagues (1976)—yet sometimes incorrectly attributed to Vygotsky (see Smagorinsky, 2018)—describes the interaction between a more knowledgeable other and a novice in the solving of a complex problem. According to Wood et al., scaffolding "enables a… novice to solve a problem, carry out a task or achieve a goal which would be beyond [their] unassisted efforts" (p. 90). For scientific topics of social relevance (i.e., SSIs), such as causes of climate change, instruction to foster scientific understanding cannot take place in the absence of requiring students to engage with scientific models and theories, which requires rather sophisticated strategies (Alexander, 1997). Thus, coming to an understanding of such topics is, by all means, a complex task for students who are non-expert, acclimating learners in the relevant scientific domains, such as climatology orgeoscience.

To overcome limitations in their original understanding of scientific topics of social relevance, as well as their rudimentary strategy use, students' learning may benefit from scaffolded learning. Such limitations may be due to students' own empirical sensory observations that might be at odds with scientific models of the phenomenon (e.g., it snowed last winter, so a model of human-driven climate change is implausible; Chinn et al., 2020; Woolf, 2015), while they may be similarly unwilling or unable to accept empirical evidence forwarded by experts (e.g., large-scale, long-term climate data) that they might not be able to corroborate using their own perceptive faculties. As a result, many students and non-experts may implicitly or explicitly operate using non-scientific explanatory mechanisms for such topics. As an educational tool to help students overcome unscientific ways of thinking, instructional scaffolding has been found to be effective in fostering scientific thinking and reasoning, as well as critical thinking within the domain in which the scaffold is situated (Quintana et al., 2004).

Importantly, although science learning in the classroom attempts to convey scientific understanding generated within specific scientific communities, students benefit from experiencing knowledge construction rather than merely absorbing scientific explanations or facts (e.g., Alexander, 2018). Rather, through reasoning with scientific evidence, students may come to a deeper understanding than would be accomplished with mere rote memorization (Ford, 2015). Such deep learning has immediate implications for students' epistemic dependence on experts. Although there is no bypassing expert consensus when learning about socioscientific phenomena, one goal of science learning in the classroom is to reduce epistemic dependency and to enable students to understand and judge scientific topics for themselves (Kienhues et al., 2020). An avenue to facilitate science learning may be providing students with epistemic roles when engaging in science learning (Kirch, 2009), as students can be given the responsibility to learn about controversial scientific claims by engaging with its premises in a scaffolded fashion. For example, educational interventions, such as refutation text interventions (Sinatra & Broughton, 2011) or problem-based learning (Loyens et al., 2015), have reported mixed but overall positive effects in helping students grapple with understanding complex science problems. Mere engagement with premises and models does not automatically lead to increased scientific understanding (Bae et al., 2022). Rather, students require assistance and guidance in distinguishing claims from evidence (Kuhn & Pearsall, 2000) and require tasks that explicitly demand the weighing of evidence in light of plausible claims. Our research efforts have been directed for some time at crafting and validating tasks that endeavor to foster such scientific evaluations, plausibility reappraisal, and science knowledge construction.

2.4. Model-Evidence Link (MEL) diagrams

A specific approach to scaffolding students' science learning is the use of MEL diagrams (Figs. 1 & 2). Within a MEL activity, students are explicitly instructed to evaluate lines of given evidence in relation to multiple plausible explanatory models using a semi-constructed diagram and materials expanding upon the lines of evidence. Effective science learning and building an understanding of scientific topics of social relevance (e.g., climate change, freshwater resource security; Sadler et al., 2017) requires students to critique and evaluate scientific evidence in light of alternative plausible explanations (Lombardi, Nussbaum et al., 2016). Specifically, such critical evaluations may help students make more science-guided judgments when evaluating evidence and explanations (Ford, 2015) and having the opportunity to engage in plausibility reappraisal (Lombardi, Nussbaum et al., 2016). Crucially, scaffolds, such as the MEL, may be effective in promoting science learning because scientific topics of social relevance are often beyond students' prior knowledge or ability to allow for meaningful and lasting learning (Pea, 2004).

In the context of promoting science learning, we use the MEL to bring students' understanding of important scientific topics of social relevance closer to the consensus of the relevant expert communities (e.g., climatologists on the causes of current climate change) by presenting them

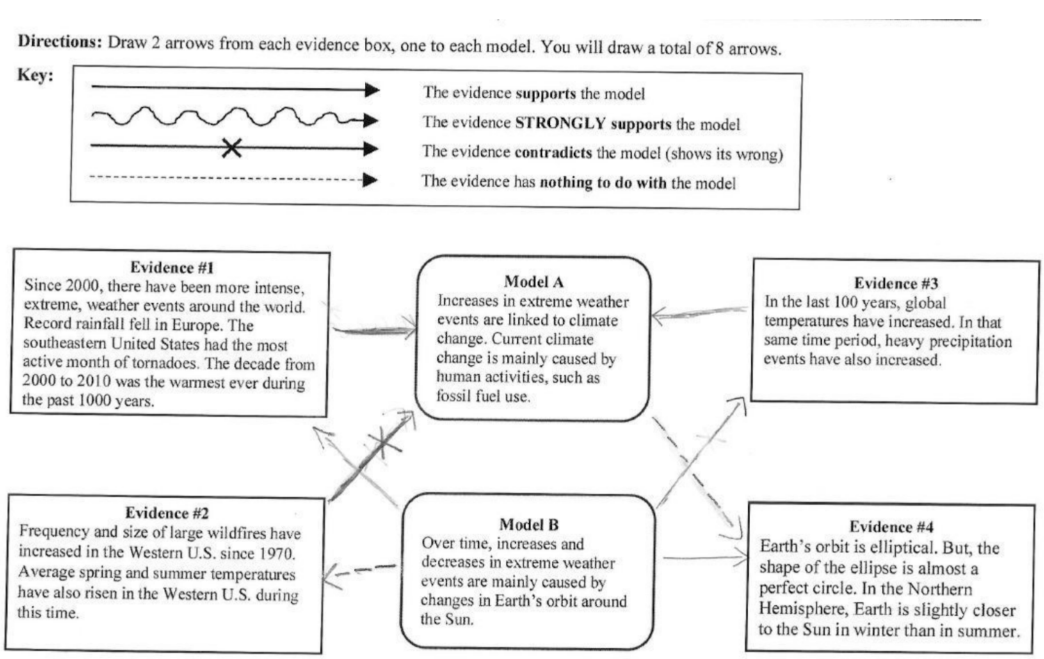

Fig. 1. Example of Extreme Weather Preconstructed Model–Evidence Link (pcMEL) Diagram.

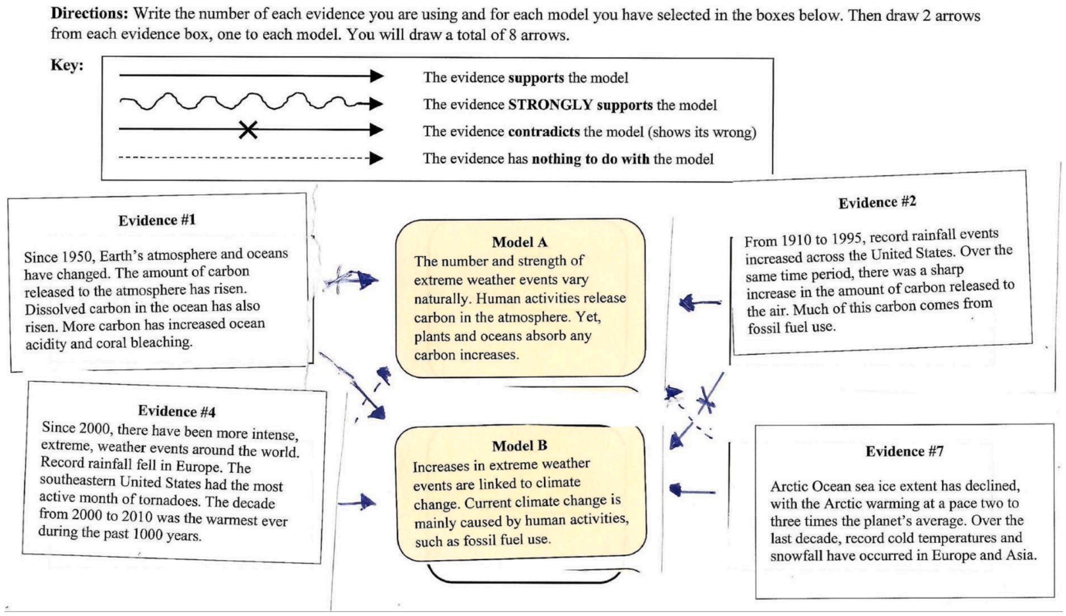

Fig. 2. Example of Extreme Weather Build-a-Model–Evidence Link (baMEL) Diagram.

with scientific and plausible non-scientific claims, as well as several lines of scientific evidence that in varying degrees provide support for both kinds of claims (Fig. 1). By presenting students with evidence and claims—all the while carefully distinguishing the two (Kuhn & Pearsall, 2000)—the MEL scaffolding can support evaluation that results in meaningful science knowledge gains and shifts of students' plausibility judgments toward the scientific consensus (Lombardi, Nussbaum et al., 2016; Quintana et al., 2004).

In our earlier studies (Lombardi, Bailey et al., 2018; Lombardi et al., 2013), students engaged with a preconstructed MEL (pcMEL; Fig. 1) in which they were to critically evaluate the relation between four lines of evidence and two competing explanatory models, judging whether each line of evidence supports, strongly supports, conflicts with, or has nothing to do with each model. These studies all suggest that students benefit from the scaffolding activity, given increases in knowledge as well as shifts in plausibility ratings toward the scientific consensus (e.g., Lombardi et al., 2013).

Notwithstanding meaningful effect sizes, students are limited in the freedom to choose what claims to evaluate and with what evidence to consider connections. More recently, we have crafted and implemented so-called build-a-MELs (baMELs; Fig. 2), which afford students more degrees of freedom. Before we compare the workings of the pcMEL versus the baMEL, we introduce the hypothesized role of autonomy support in scaffolded science learning, which underlies this newer MEL format.

2.5. Autonomy support

While engaging in learning, students have a need to perceive autonomy (Deci & Ryan, 2000). Importantly, they are likely to pursue goals or tasks that satisfy this need for autonomy. Empirical studies have corroborated this theoretical perspective, as students' learning efforts are typically higher and yield greater outcomes when the environment is highly autonomy supportive as compared to highly controlling (e.g., see Okada, 2023; Reeve, 2009). The degree of autonomy afforded by the learning environment is affected by both teacher behaviors and pedagogy, as well as the nature of a task or assignment. For instance, taking and acknowledging the perspective of the student; welcoming their thoughts, feelings, and behaviors; and supporting the development of students' motivation are practices that provide autonomy support (e.g., Reeve, 2009). Particularly in classrooms or while engaging with academic tasks that are perceived to be challenging or difficult, such as science learning, autonomy-supportive strategies and practices can be more beneficial to learning than an educational environment that is generally more controlling (Hagger et al., 2015; Patall et al., 2019; Reeve et al., 2004).

In the context of scaffolded learning, tasks may facilitate autonomy support by providing students with degrees of freedom in how to bring the task to completion (Reeve & Cheon, 2021). Offering students choice in what elements of the task to engage with has been suggested to facilitate greater levels of agentic engagement (Reeve et al., 2020), which is students' deep engagement with a task as a result of perceptions of conceptual agency. Degrees of freedom or choice in a task may be provided to the student by permitting them to select elements to (or not to) engage with for task completion.

In the specific context of using MELs to promote learning about scientific topics of social relevance, the role of autonomy support has become increasingly intriguing to us. A pcMEL provides students with two plausible explanatory models—one scientific and one non-scientific—for a phenomenon and a fixed set of lines of evidence to evaluate in relation to those two models. Although significant and meaningful knowledge gains and plausibility shifts have been realized through these MELs (e.g., Lombardi, Bailey et al., 2018), the pcMELs do not permit students a choice regarding what elements of the task to engage with. To afford students greater degrees of choice in learning about scientific topics of social relevance by interacting with a scaffold, we created the baMEL, in which students select and evaluate two out of three possible explanatory mechanisms. Similarly, students select four out of eight provided lines of evidence that they can use in the evaluation task. It is this greater level of choice that we hypothesize leads to greater plausibility shifts and knowledge gains as a function of higher perceived autonomy support.

3. The present study

The purpose of the present study was to examine differences between two different MEL formats, namely the less autonomy-supportive pcMEL and the more autonomy-supportive baMEL. In the pcMEL, students evaluated given lines of scientific evidence and two competing explanatory models, whereas the baMEL guides students to select relevant scientific evidence from a larger set and to evaluate two self-chosen competing models (Bailey et al., 2020), providing more choice and autonomy support. Although initial pilot studies considered the differential effect for topics separately (i.e., the current climate crisis, Bailey et al., 2022; fossils and fossil fuel use, Klavon et al., 2024; water resources, Medrano et al., 2020; and astronomical origins, Dobaria et al., 2022), we wanted to see if these findings could be corroborated by a systematic approach in which we consider the overall omnibus effects as well as the effect of different topics and scaffold approaches. Given our interest in the potential effect of higher versus lower autonomy support in the MEL, we set out to answer two research questions:

1. How do pre- to post-instructional differences compare between the less and more autonomy-supportive MEL scaffold forms (pcMEL and baMEL, respectively), specifically for
   a. students' levels of evaluation about the connections between lines of scientific evidence and alternative explanations;
   b. plausibility judgments about competing explanations of phenomena; and
   c. knowledge about scientific topics of social relevance?
2. How do post-instructional plausibility judgments mediate the relationship between levels of evaluation and post-instructional knowledge when controlling for MEL scaffold form?

Based on the results of our earlier pilot studies, we hypothesized that both forms of the MEL scaffold would result in both significant and meaningful plausibility shifts toward the scientific consensus, as well as gains in knowledge, from before to after the interventions (H1a; Bailey et al., 2022; Dobaria et al., 2022; Klavon et al., 2024; Medrano et al., 2020). A novel part of the present study was investigating the various Earth and environmental science MEL topics together (i.e., climate change, water resources, fossils and fossil fuels, and astronomical origins) to assess the agglomerative effects of these instructional scaffolds. In doing so, we applied min–max normalization (Jain et al., 2005) to linearly transform knowledge scores to systematically account for topic differences in difficulty. We also hypothesized that the more autonomy-supportive baMEL would have stronger shifts toward scientific judgments and deeper knowledge gains than the less autonomy-supportive pcMEL (H1b; Collie & Martin, 2017).

Our previous pilot studies, as well as earlier investigations involving the MEL scaffolds (Lombardi, Bickel et al., 2018; Lombardi et al., 2013), further suggested that post-instructional plausibility judgments partially mediate the relation between levels of evaluation and post-instructional knowledge, supporting Lombardi, Nussbaum et al.'s (2016) theoretical model. Therefore, when controlling for scaffold form, we hypothesized that the direct and indirect effects between evaluation and knowledge would still be meaningful (Fig. 3) and that relations would also suggest an advantage for the more autonomy-supportive baMEL (H2; Reeve & Cheon, 2021).

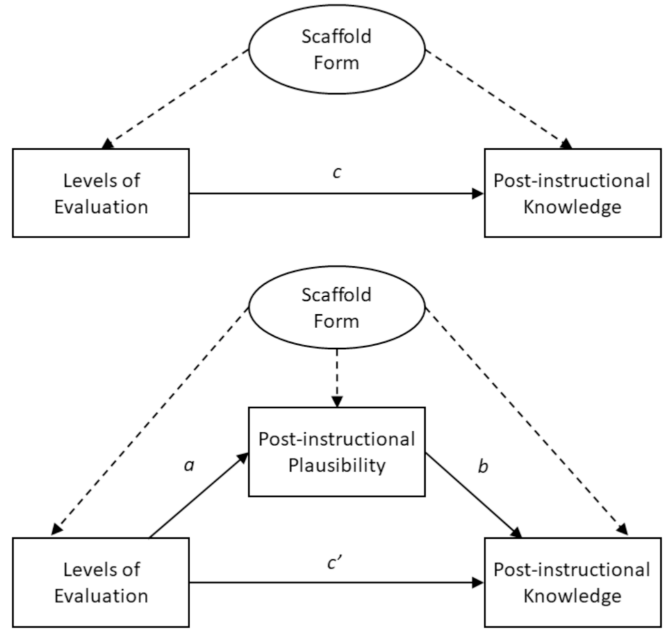

Fig. 3. Hypothesized Relations between Study Variables *Note.* The top image shows the hypothesized direct relational pathway (c) between levels of evaluation and post-instructional knowledge, when controlling for scaffold form (dashed lines); and the bottom figure image shows the hypothesized direct (c') and indirect relational pathways (a + b) between levels of evaluation and post-instructional knowledge, when controlling for scaffold form (dashed lines).

4. Method

4.1. Participants and context

The present study was part of a multi-year project involving data collected in several middle and high school classrooms in the Middle Atlantic and Southeastern US and one undergraduate course at a Middle Atlantic university. Two hundred ninety-seven ($N = 297$) participants were involved in the present study and completed all instructional tasks and measures. Middle school ($n = 82$) and high school ($n = 197$) participants were enrolled in Earth and environmental science courses and used the MEL during their normal curricular scope and sequence, covering scientific topics of social relevance topics, including the climate crisis, fossils and fossil use, water resources, and astronomical origins. Undergraduate ($n = 18$) participants were enrolled in a pre-service science teaching methods course. The classrooms and university were located in urban and suburban settings, with the majority of participants identifying as male (52.6 %) and White (59.2 %), and the remainder identifying themselves as Hispanic (21.5 %), Black (8.2 %), Asian (7.4 %), and Other (3.7 %).

Each teacher whose classes used the MEL activities selected the number (between two and four) and topics that best fit their curricular needs, although all were focused on Earth and environmental science themes. The materials, as described below, did not differ based on grade level but teachers may have adjusted the implementation somewhat to better support their students. For example, the university course instructor may have asked students to read all of the evidence texts individually while a middle school teacher may have used a jigsaw approach (Aronson, 1978) or classwide read-aloud protocols for the same documents. Prior work, using variable-centered analyses, has not found any "teacher effects" (e.g., Bailey et al., 2022; Dobaria et al., 2022; Klavon et al., 2024; Medrano et al., 2020), suggesting that aggregating the different implementation formats for this study was reasonable.

4.2. Materials and measures

Before students engaged in any of the scaffolds for any of the scientific topics of social relevance, they participated in a short learning task about how scientists connect lines of evidence and explanatory models, first defining plausibility as "a judgment we make about the potential truthfulness of one model compared to another and the judgment may be tentative (not certain)." After doing this initial task, each student participant sequentially completed a pcMEL and baMEL covering a particular topic area (e.g., the climate crisis). Spacing between each scaffold may have been as little as one or two days or as much as one or two weeks, depending on the classroom's curriculum and pacing. The sequence of pcMEL first and baMEL second follows a time-honored and historical approach of introducing scaffolded instruction, where students learn both concepts (e.g., topic knowledge about scientific topics of social relevance) and skills of inquiry (e.g., evaluating evidence to explanation connections, in light of alternatives) (Bruner, 1966; Posner & Strike, 1976). Specific to the present study, participants first learned how to use the pcMEL scaffold by reasoning with four pre-selected lines of evidence in light of two alternative plausible explanations. In doing so, students learned about relevant scientific topics. After engaging with pcMELs, students were familiarized with baMELs. The baMEL, which affords more choice, was more complex than the pcMEL because participants considered eight lines of scientific evidence and three alternative models, from which they chose to build their diagrams.

Both the pcMEL and baMEL followed a similar instructional progression. First, participants completed a knowledge pre-survey to gauge their topic knowledge. Second, they rated the plausibility of the scientific and alternative explanatory model(s); note that the models are simply labeled by letters (A, B, and where applicable, C; see Figs. 1 and 2) and are not identified as scientific or alternative. Third, participants read one-page expository texts elaborating on each line of evidence. Fourth, participants completed and/or constructed their MEL diagrams. Fifth, participants gave plausibility ratings for the same explanatory models as before the intervention, and then wrote written reflections on two links between lines of evidence and an explanatory model as they determined in their MEL. Sixth, they completed the post-instruction knowledge survey. The entire time of each MEL lesson was approximately 90 min (about two traditional class meetings or one block class meeting). In what follows, we briefly describe the characteristics and procedures of pcMEL and baMEL scaffolds, the procedure of determining the levels of evaluation exhibited in students' explanations, and the plausibility judgment and topic knowledge measures that are embedded within and accompany the scaffolds.

4.3. MEL scaffolds

The pcMEL scaffold provided participants with a preconstructed diagram featuring two alternative explanatory models for a scientific phenomenon of social relevance, with one model being the scientific consensus model and the other being a plausible but non-scientific alternative (Fig. 1). For example, in the Climate Change pcMEL, Model A declares that the current climate change is caused by human activities (the scientific consensus explanation) while Model B states that current climate change is caused by an increasing amount of energy received by the Sun (a non-scientific but plausible explanation). The activity presents these two explanatory alternatives without any indication of the veracity or validity of either. The pcMEL also presented four lines of scientific evidence in the form of one-to-three-sentence declarative statements. One-page evidence texts accompanied each evidence line to provide more detail for students to use in the activity.

Participants constructed their own diagrams in the baMEL activity by selecting two explanatory models from a choice of three, with one of these being the scientific model and the other two being plausible but non-scientific alternatives (Fig. 2). Participants also constructed their diagrams by selecting four lines of scientific evidence from a choice of eight. Their selection of lines of scientific evidence and alternative models renders the baMEL as the more autonomy-supportive scaffold in relation to the pcMEL, wherein participants do not have a choice of what evidence and models they wish to reason with.

Participants completed their diagrams by drawing one of four different arrow types between the lines of scientific evidence and each of the explanatory models. A straight, solid-line arrow indicated that the participants thought a line of scientific evidence *supported* a model; an arrow with a squiggly, solid line indicated that they thought the evidence *strongly supported* the model; an arrow with a straight, dotted line indicated that they thought the evidence *had nothing to do* with the model; and an arrow with a straight, solid line that had an "X" marked through its middle indicated that they thought the evidence *contradicted* the model.

4.4. Evaluation scores

Participants constructed their written responses after completing their MEL diagrams. Following a prompt, students indicated the model-evidence link they wanted to discuss, iterating the strength and type of relation they identified earlier (i.e., strongly supports, etc.), after which they wrote about their reasoning for the arrow that they drew. To gauge participants' level of evaluation, Lombardi, Brandt et al. (2016) developed a rubric for scoring written responses about two of the arrows that the participants drew on their diagrams. Two independent scorers read each written explanation and used this rubric to score participants' levels of evaluation, with 1 indicating an erroneous evaluation, 2 indicating a descriptive evaluation, 3 indicating a relational evaluation, and 4 indicating a critical evaluation, with intraclass correlation coefficient (ICC) of 0.72, an acceptable level of coder reliability. After scoring explanations, the scorers met to discuss any differences and reached a full consensus on all scores. We used the consensus-based scores of participants' evaluation quality for analysis purposes.

4.5. Plausibility judgment scores

Participants rated the plausibility of each explanation model on a scale of 1 (greatly implausible) to 10 (highly plausible) before and after completing their MEL diagrams (Lombardi et al., 2013). We calculated the plausibility judgment score as the difference between a participant's scientific model plausibility rating and alternative model rating for the pcMEL, while computing the difference between the rating for the scientific model and the average of the alternative models' ratings for the baMEL. Scores could range from -9 to $+9$, with positive plausibility judgment scores indicating a more scientific stance; that is, rating the scientific model as more plausible than the alternative(s) in accordance with the scientific consensus.

4.6. Knowledge scores

Students completed a multi-item knowledge survey before and after instruction. Based on the methods used in earlier MEL studies (see, for example, Lombardi, Bailey et al., 2018), students rated each item on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree), indicating the extent to which they believed scientists would agree with the statement. Having students rate their level of agreement from a scientist's point of view reflected their knowledge of scientific phenomena rather than their own personal beliefs or opinions on the topic (Lombardi et al., 2013). The MEL project team developed these statements from information on which there is clear scientific consensus, with at least one question addressing each evidence statement and at least one question addressing each explanatory model.

In order to account for differences in the four topics (climate change, fossils and fossil fuel use, water resources, and astronomical origins), we linearly transformed knowledge scores using min–max normalization (Jain et al., 2005). This transformation technique is a relatively simple method to account for differences in topic and topic difficulty, where the minimum in a range of scores is scaled to zero, and the maximum is scaled to one. This approach is commonly used in medical, computing, and environmental research (Mazziotta & Pareto, 2022). Using such a method maintains the original distribution of the data as long as there are no outliers in the sampled scores (Mu, 2020). Therefore, prior to employing this technique, we screened for any score outliers and found none. We calculated the internal consistency of the knowledge items, with McDonald's $\omega = 0.787$, indicating acceptable reliability.

5. Results

Table 1 shows means, standard error of the means, bivariate correlations, skewness, and kurtosis for the variables of level of evaluation, pre- and post-instruction plausibility rating, and pre-and post-instruction knowledge. For both scaffolds, all scores were significantly and positively correlated at low to moderate strength and normally distributed, with all absolute values of skewness and kurtosis ≤ 1 (Nussbaum, 2014).

Table 1
Descriptive Statistics and Bivariate Correlations.

Variable	Mean	SE	Skewness	Kurtosis	1	2	3	4	5
1. Evaluation	1.951	0.030	0.741	0.100	–				
2. Plausibility Pre	1.365	0.145	−0.533	−0.095	.127*	–			
3. Plausibility Post	2.604	0.135	−0.720	0.517	.220*	.390*	–		
4. Knowledge Pre	0.556	0.006	−0.057	−0.131	.181*	.060	.133*	–	
5. Knowledge Post	0.608	0.007	−0.214	0.394	.241	.072	.173*	.636*	–

Note. Significant bivariate correlations are indicated with an asterisk (*).

5.1. Research Question 1: scaffold score comparisons

An analysis of variance (ANOVA) revealed no statistically significant difference in levels of evaluation scores by scaffold type (pcMEL and baMEL), with $F(1, 592) = 0.86$, $p = .354$. Similarly, repeated measures ANOVA revealed no significant interaction over time (pre- to post-instruction) between plausibility scores by scaffold type, with $F(1, 592) = 0.46$, $p = .494$, or knowledge scores by scaffold type $F(1, 592) = 0.60$, $p = .440$. However, there were significant increases over time (i.e., pre- to post-instruction) in both plausibility scores, $F(1, 592) = 63.8$, $p < .001$, $\eta^2 = .032$ (small to medium effect size; Cohen, 1988), and knowledge scores, $F(1, 592) = 83.2$, $p < .001$, $\eta^2 = .025$ (small to medium effect size). In terms of practical significance, plausibility scores shifted about one category, and knowledge scores increased by about 10% for both forms of the MEL scaffold, a relatively robust educational effect given the short instructional time of approximately 90 minutes (Kraft, 2020). A follow-up simple-effects analysis showed that baMEL plausibility scores were significantly higher at both pre- and post-instruction compared to the pcMEL, $F(1, 592) = 24.1$, $p < .001$, $\eta^2 = .026$ (small to medium effect size), but that knowledge scores were significantly lower, $F(1, 592) = 104$, $p < .001$, $\eta^2 = .12$ (medium to large effect size).

5.2. Research Question 2: score relations

Prior to addressing our second research question, we screened for differences by level (middle school, high school, and undergraduate) in evaluation, plausibility, and knowledge scores. An ANOVA revealed a significant difference in evaluation scores by level, with $F(2, 591) = 7.31$, $p < .001$, partial $\eta^2 = 0.024$ (small to medium effect size; Cohen, 1988). Further, a repeated measures ANOVA did not reveal a significant interaction in plausibility scores by level over time (pre- to post-instruction), with $F(2, 591) = 1.99$, $p = .14$, but did reveal a significant interaction between knowledge scores by level over time (pre- to post-instruction), with $F(2, 591) = 6.66$, $p = .001$, partial $\eta^2 = 0.022$ (small to medium effect size). Therefore, with two of the three variables being significantly different by level, we included level as a covariate in our mediation analysis.

To answer our second research question, we conducted a mediation analysis to gauge how levels of evaluation predicted post-instructional knowledge mediated by post-instructional plausibility, controlling for scaffold type (pcMEL, coded as 1, and baMEL, coded as 2) and level (middle school, coded as 1, and high school/undergraduate, coded as 2). The direct effect of evaluation on knowledge, controlling for scaffold type and level, was statistically significant ($b = 0.17$, $z = 4.58$, $p < .001$, 95% CI 0.10, 0.25). The indirect effect of evaluation on knowledge, mediated by plausibility and controlling for scaffold type and level, was also statistically significant ($b = 0.037$, $z = 3.44$, $p < .001$, 95% CI 0.02, 0.06). These results suggest that plausibility partially mediated the relation between evaluation and knowledge when controlling for scaffold and level, accounting for about 17% of the total effect ($b = 0.21$, $z = 5.58$, $p < .001$, 95% CI 0.14, 0.29). The model explained about 20% ($R^2 = .199$) of post-instructional knowledge, with $F(4, 593) = 36.5$, $p < .001$, indicating a moderate effect size.

The mediation path analysis (Fig. 4) showed significant and robust relations along the indirect pathway, meaning the relations between evaluation and post-instructional plausibility ($b = 0.21$, $z = 5.30$, $p < .001$, 95% CI 0.14, 0.28) and between post-instructional plausibility and post-instructional knowledge ($b = 0.17$, $z = 4.52$, $p < .001$, 95% CI 0.10, 0.26), as well as the direct pathway between evaluation and post-instructional knowledge ($b = 0.17$, $z = 4.58$, $p < .001$, 95% CI 0.10, 0.25). These findings largely support Lombardi, Nussbaum et al.'s (2016) theoretical framework on plausibility judgment and scientific knowledge construction. Results also showed that scaffold type (pcMEL or baMEL) was a significant predictor of both post-instructional plausibility ($b = 0.33$, $z = 4.16$, $p < .001$, 95% CI 0.18, 0.48) and post-instructional knowledge ($b = -0.69$, $z = -9.21$, $p < .001$, 95% CI -0.83, -0.53), but not levels of evaluation ($b = -0.08$, $z = -0.94$, $p = .35$, 95% CI -0.23, 0.09). These results corroborate the findings suggested by the ANOVA results reported in research question 1, where the baMEL scaffold had plausibility judgments reflecting a more scientific stance than the pcMEL, even though overall knowledge scores were lower.

6. Discussion

We set out to investigate whether there were differential effects for the two MEL scaffold forms, the pcMEL and the baMEL. We hypothesized that both scaffolds would generally be effective, given that their designs and implementation facilitate students' learning about socio-scientific issues by supporting their evaluation of evidence in light of two or more competing explanatory models, as emulated from real scientific practice (Ford, 2015; Lombardi, Nussbaum et al. 2016). Yet, important to the present investigation were the potential differential effects between the two scaffolds, as the baMELs are crafted to afford greater autonomy support in accordance with students' need for self-

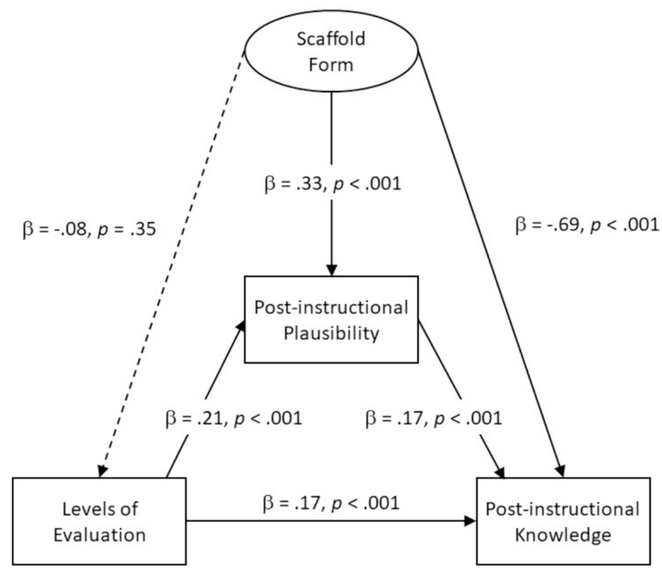

Fig. 4. Results of the Mediation Analysis *Note.* Pathway relations are shown as standardized values, with solid lines indicating statistically significant relations and the dashed line indicating a statistically insignificant relation.

determination (Reeve, 2009; Reeve et al., 2020). Such greater perceived autonomy support is suggested to allow students to learn more effectively (Patall et al., 2019).

Importantly, we tested our hypotheses using linearly transformed within-person data, which allowed us to more systematically model the interrelations of instructional scaffolds designed to promote science learning across four scientific topics of social relevance (i.e., the climate crisis; fossils and fossil fuel use; water resources; and astronomical origins). To our initial surprise, regarding the hypothesized differential effect in favor of the scaffold that afforded greater autonomy support (i.e., baMEL), our robust analyses could not corroborate the findings of our earlier studies. That is, the repeated-measures ANOVA and mediation analysis did not reveal any appreciable differences between the two scaffold forms in terms of levels of evaluation, shifts in plausibility judgment toward the scientific, or topic knowledge gains. However, we have come to understand that this non-significant finding is quite meaningful for at least three reasons.

First, it is worth noting that both scaffold forms resulted in plausibility shifts toward a more scientific stance (approximately 1 category on a ten point Likert-scale) and knowledge gains (10%). Although these are rather modest effect sizes in terms of standard rules of thumb (Cohen, 1988), they do represent practically significant results, especially when considering that the MEL scaffolds can be used at relatively low cost and convenience, being easily inserted into the standard middle and high school science curriculum and taking only about 90 min of instructional time. In light of this great cost-effectiveness, relatively modest effect sizes may have appreciable meaning in terms of practical significance (Kraft, 2020).

Second, the main aim of the present study was to use a more systematic approach considering the overall omnibus effects of the two scaffold forms. Earlier initial pilot studies separately investigated the differential effects of pcMEL and baMEL for specific topics, including climate change (Bailey et al., 2022), water resources (Medrano et al., 2020), fossils and fossil fuel use (Klavon et al., 2024), and astronomical origins (Dobaria et al., 2022). One purpose of these previous pilot studies was to inform the design-based research approach that we used during the project. The results from these studies were somewhat mixed but did suggest a modest effect favoring the more autonomy-supportive baMEL. However, a major limitation of the previous studies was not looking at the overall performance of the scaffolds across various topics. In the present study, we linearly transformed knowledge scores using min–max normalization to account for topic and topic difficulty (Jain et al., 2005). This more robust analysis showed that the mixed effects of the pilot study largely disappeared across the various topics but that both scaffold forms resulted in equally positive learning outcomes. Similar trends of finding mixed effects that disappear in robust replication studies can be found in related educational interventions (e.g., see Sinatra & Broughton, 2011). To further explore the role of scaffolded science instruction using the MELs, ongoing work uses a person-centered analysis to look at potential differential effects—including age level, sequencing, and more—in greater detail (Robertson et al., 2024).

Whereas the present study corroborated our earlier studies in terms of supporting the important indirect relation between levels of evaluation, post-instructional plausibility, and post-instructional knowledge above and beyond the direct relation between evaluation and knowledge, the present study did not corroborate an increased benefit to a more autonomy-supportive instructional scaffold. This elevates the value of replication studies, a rarity in both education and psychology research, in showing theory building (Plucker & Makel, 2021). In other words, the present study supports Lombardi, Nussbaum et al.'s (2016) basic theoretical premise that scaffolded instruction can facilitate learning about complex issues, while providing a greater understanding of when such a theory may be relevant to classroom practice. On one hand, more autonomy-supportive learning environments are not necessarily a motivational boost when learning about complex scientific topics of social relevance. On the other hand, scaffolded instruction can provide learners with the opportunity to deepen their scientific judgments and knowledge when making more reasoned claims about evidence to model connections. Debates that have a long history in educational psychology about direct versus inquiry-based instruction (see, for example, Kirschner et al., 2006) often oversimplify and leave out the necessary instructional balance and benefits of scaffolded instruction that includes elements of both direct instruction and scientific inquiry (Hmelo-Silver et al., 2007).

Third, an interesting result of the study was that, although knowledge gains between the two scaffolds were very similar, knowledge scores were lower at both pre- and post-instruction for the baMEL. At first glance, this may seem fully consistent with the greater complexity of the baMEL compared to pcMEL, where students consider eight lines of scientific evidence (baMEL) rather than four (pcMEL) and three alternative models (baMEL) rather two (pcMEL). However, although shifts in plausibility were similar, plausibility judgment scores were greater at both pre- and post-instruction for the baMEL. Therefore, despite the increased complexity of the baMEL scaffold, students had a more scientific stance in terms of plausibility judgments. Of course, this may merely reflect the instructional sequence in which participants first engaged with the pcMEL and only then the baMEL. This order may suggest that participants learned the process of evaluation supported by the pcMEL's relatively easy task of evaluating four pre-selected lines of scientific evidence in light of two alternative explanatory models before moving on to the more ill-structured, complex baMEL (Bruner, 1966; Posner & Strike, 1976). Indeed, with the baMEL, students had to grapple with more information by choosing among eight lines of scientific evidence to evaluate two of the three explanatory models and construct their explanatory diagrams, potentially increasing cognitive load compared to the pcMEL (Kalyuga & Singh, 2016), which may explain the lower pre- and post-instructional knowledge scores. However, an alternative explanation could be that the greater autonomy found in the baMEL increased student engagement, overcoming topic complexity (Patall et al., 2019). In other words, the baMEL, with its greater degrees of freedom and higher autonomy support, could have permitted learners to construct knowledge more like scientists (Alexander, 2018; Ford, 2015; Reeve, 2009; Reeve et al., 2020), and thus may have been advantageous as suggested by the previous pilot studies. We do acknowledge, however, that this autonomy support explanation may not be as plausible as the instructional sequencing explanation.

6.1. Limitations

Although the sample size was somewhat robust for classroom-based studies that have a high degree of ecological validity, such as the present study, we acknowledge that the results reflect two specific regions (the Middle Atlantic and Southeastern US). The present study is consistent with previous MEL studies in terms of the relations between levels of evaluation, plausibility judgments, and topic knowledge, which have also occurred in the Middle Atlantic and Southeastern US, as well as those that have occurred in the Southwestern US (Lombardi, Bailey et al., 2018; Lombardi et al., 2013). These regions do differ in their natural environments, which may have influenced prior knowledge and learning about scientific topics of social relevance. The participant sample also reflects demographics typical of some urban and suburban settings in these regions and is not necessarily characteristic of widespread situations and contexts of all learners. Thus, we approach the results with some caution in terms of generalizability.

We also acknowledge the limitations of the sequence of instruction, with the pcMEL administered first and the baMEL second. Some limitations are discussed in the results above, but we would also add that some classroom instruction may have occurred due to spacing of the scaffold administration. Additional instruction between scaffolds may have resulted in the increased pre-instructional plausibility scores with the baMEL. We do note, however, that any beneficial effects from additional instruction are not reflected in levels of evaluation scores, which were similar between the two scaffolds, or knowledge scores, which were lower for the baMEL than the pcMEL. Classrooms are complex learning environments, and conducting ecologically valid studies in such settings is often challenging. Sequencing of instruction is one such challenge and does not reflect less authentic but more experimental conditions of random assignment, where students might have been administered the baMEL first and the pcMEL second. Therefore, we approach the results of the present study with additional caution in terms of generalizability.

The project team designed the baMEL to support greater student autonomy. Although we did not use a measure to investigate students' perceptions of autonomy, which could be considered a limitation of the present study, observations by team members and anecdotal reports from teachers using both forms of the scaffold report that students' responses to the baMEL were positive and that they liked having more choice and control over it compared to the pcMEL. This suggests that students felt they had more autonomy working with the baMEL. Future work should explicitly probe the extent to which and how students perceive greater autonomy, if at all, using the baMEL, and what further opportunities for autonomy may not have yet been implemented.

Finally, although prior work (Bailey et al., 2022; Dobaria et al., 2022; Klavon et al., 2024; Lombardi, Bailey et al., 2018; Medrano et al., 2020) suggested that aggregating these data across multiple classrooms and regions was reasonable, it is possible that the types of analyses in these studies could not detect differences that might have been present. Furthermore, we also detected differences in evaluation and knowledge by level (middle, high, and undergraduate) that had not been seen in prior work. Therefore, we decided to conduct a study using person-centered analyses to provide additional insights into any differences across grade levels, instructional sequence, and region (Robertson et al., 2024).

6.2. Implications and conclusion

As is typical with empirical studies, we are left with more questions than answers. The results are meaningful in terms of partially supporting prior empirical studies and corroborating the theoretical framework suggesting that increased levels of evaluation result in stronger shifts in plausibility toward the scientific and deeper knowledge gains (see, for example, Lombardi, Bailey et al., 2018; Lombardi, Nussbaum et al., 2016). Although more autonomy-supportive scaffolding did not result in more robust relations between these variables, they also did not have reduced effectiveness. Furthermore, while the sequencing of the scaffolds in order of less to more autonomy supportive is in accordance with curricular principles to facilitate progressions of learning scientific concepts (Barnes et al., 2020), it simultaneously is a limitation of the study in reaching a causal conclusion regarding the differential effects. Future research that incorporates a counter-balanced, repeated measures, quasi-experimental design, with more iterations of less and more autonomy-supportive scaffolding, may be able to tease out the effect of sequencing on learning outcomes. Anecdotal evidence suggests that the baMEL may also contribute to students' perception of autonomy in, enjoyment of, and engagement with the activity, providing an additional future research area.

From an instructional standpoint, this sequencing of pcMEL before baMEL may help maintain an appropriate cognitive load by offering less choice when first learning about the activity and increasing autonomy after students have gained experience with evaluating alternative explanations via the MEL. Additionally, because the activities come in pairs with related content (e.g., causes of climate change in a pcMEL and relations between climate change and extreme weather events in a baMEL), this sequencing may better align with existing curricula that address the cause of a phenomenon before its effects.

In conclusion, we have learned much about the differential effects of an instructional scaffold that provides lower autonomy support and few degrees of evaluative freedom (i.e., pcMEL) versus a scaffold that was specifically designed to provide higher autonomy support through greater degrees of evaluative freedom (i.e., baMEL). Our results suggest that the baMEL was more effective in bringing about desirable learning outcomes, such as plausibility reappraisal and knowledge gain. This shift in students' reasoning toward well-established scientific consensus across the four scientific topics of social relevance (i.e., water resources, origins, climate change, & geology), paired with greater topic knowledge, is in line with the educational goals as forwarded by the *Next Generation Science Standards* (NGSS Lead States, 2013). Given that these instructional scaffolds have been implemented with high ecological validity in existing curricula in several schools, both forms can be seen as viable tools to promote science learning in the classroom.

CRediT authorship contribution statement

Eric C. Schoute: Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization. **Janelle M. Bailey:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition. **Doug Lombardi:** Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Data curation.

Acknowledgements

This research project was supported in part by the US National Science Foundation (NSF) under Grant Nos. 2201012 and 2201015. Any opinions, findings, conclusions, or recommendations expressed are those of the authors and do not necessarily reflect the NSF's views.

References

Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213–250). JAI.

Alexander, P. A. (2018). Information management versus knowledge building: Implications for learning and assessment in higher education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. Pant, C. Lautenbach, & C. Kuhn (Eds.), *Assessment of learning outcomes in higher education* (pp. 43–56). Springer. https://doi.org/10.1007/978-3-319-74338-7_3.

Alexander, P. A., & Schoute, E. C. (2022). Knowledge, knowing, and information: Their meaning and meaningfulness to learning and development. In T. Good & M. McCaslin (Sect. Eds.), and D. Fisher (Gen. Ed.), *Online encyclopedia on education*. Routledge. https://doi.org/10.4324/9781138609877-REE203-1.

Aronson, E. (1978). *The jigsaw classroom*. Sage Publications, Inc.

Bae, C. L., Sealy, M. A., Cabrera, L., Gladstone, J. R., & Mills, D. (2022). Hybrid discourse spaces: A mixed methods study of student engagement in U.S. science classrooms. *Contemporary Educational Psychology, 71*, Article 102108. https://doi.org/10.1016/j.cedpsych.2022.102108

Bailey, J. M., Jamani, S., Klavon, T. G., Jaffe, J., & Mohan, S. (2022). Climate crisis learning through scaffolded instructional tools. *Educational and Developmental Psychologist, 39*(1), 85–99. https://doi.org/10.1080/20590776.2021.1997065

Bailey, J. M., Klavon, T. G., & Dobaria, A. (2020). The origins build-a-MEL: Introducing a scaffold to explore the origins of the Universe. *The Earth Scientist, 36*(3), 7–12. https://www.nestanet.org/resources/Documents/TES/2015-2020/Fall20public.pdf.

Bakhurst, D. (1995). On the social constitution of mind: Bruner, Ilyenkov, and the defence of cultural psychology. *Mind, Culture, and Activity, 2*(3), 158–171. https://doi.org/10.1080/10749039509524697

Barnes, N., Fives, H., Mabrouk-Hattab, S., & SaizdeLaMora, K. (2020). Teachers' epistemic cognition in situ: Evidence from classroom assessment. *Contemporary Educational Psychology, 60*, Article 101837. https://doi.org/10.1016/j.cedpsych.2020.101837

Blumer, H. (1986). *Symbolic interactionism: Perspective and method*. University of California Press.

Bruner, J. S. (1966). *Toward a theory of instruction*. W. W. Norton.

Chinn, S., Hart, P. S., & Soroka, S. (2020). Politicization and polarization in climate change news content, 1985–2017. *Science Communication, 42*(1), 112–129. https://doi.org/10.1177/1075547019900290

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Collie, R. J., & Martin, A. J. (2017). Teachers' sense of adaptability: Examining links with perceived autonomy support, teachers' psychological functioning, and students' numeracy achievement. *Learning and Individual Differences, 55*, 29–39. https://doi.org/10.1016/j.lindif.2017.03.003

Dawson, V., & Carson, K. (2020). Introducing argumentation about climate change socioscientific issues in a disadvantaged school. *Research in Science Education, 50*(3), 863–883. https://doi.org/10.1007/s11165-018-9715-x

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dobaria, A., Bailey, J. M., Klavon, T. G., & Lombardi, D. (2022). Students' scientific evaluations of astronomical origins. *Astronomy Education Journal, 2*(1), 1–16. https://doi.org/10.32374/AEJ.2022.2.1.032ra

Duschl, R. A., Schweingruber, H. A., & Shouse, A. E. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press. https://doi.org/10.17226/11625.

Ford, M. J. (2015). Educational implications of choosing "practice" to describe science in the Next Generation Science Standards. *Science Education, 99*(6), 1041–1048. https://doi.org/10.1002/sce.21188

Hagger, M. S., Sultan, S., Hardcastle, S. J., & Chatzisarantis, N. L. D. (2015). Perceived autonomy support and autonomous motivation toward mathematics activities in educational and out-of-school contexts is related to mathematics homework behavior and attainment. *Contemporary Educational Psychology, 41*, 111–123. https://doi.org/10.1016/j.cedpsych.2014.12.002

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition, 38*(12), 2270–2285. https://doi.org/10.1016/j.patcog.2005.01.012

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831–852. https://doi.org/10.1007/s10648-015-9352-0

Ke, L., Sadler, T. D., Zangori, L., & Friedrichsen, P. J. (2021). Developing and using multiple models to promote scientific literacy in the context of socio-scientific issues. *Science & Education, 30*(3), 589–607. https://doi.org/10.1007/s11191-021-00206-1

Kienhues, D., Jucks, R., & Bromme, R. (2020). Sealing the gateways for post-truthism: Reestablishing the epistemic authority of science. *Educational Psychologist, 55*(3), 144–154. https://doi.org/10.1080/00461520.2020.1784012

Kirch, S. A. (2009). Identifying and resolving uncertainty as a mediated action in science: A comparative analysis of the cultural tools used by scientists and elementary science students at work. *Science Education, 94*(2), 308–335. https://doi.org/10.1002/sce.20362

Kirschner, P. A., Sweller, J., & 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. *Educational Psychologist, 41* (2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klavon, T. G., Mohan, S., Jaffe, J. B., Stogianos, T., Governor, D., & Lombardi, D. (2024). Scientific evaluations and plausibility judgements in middle school students' learning about geoscience topics. *Journal of Geoscience Education, 72*(2), 170–184. https://doi.org/10.1080/10899995.2023.2200877

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189x20912798

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development, 1*, 113–129. https://doi.org/10.1207/S15327647JCD0101N_11

Lombardi, D. (2019). Thinking scientifically in a changing world. *Psychological Science Agenda*. https://www.apa.org/science/about/psa/2019/01/changing-world.aspx.

Lombardi, D., Bailey, J. M., Bickel, E. S., & Burrell, S. (2018). Scaffolding scientific thinking: Students' evaluations and judgments during Earth science knowledge construction. *Contemporary Educational Psychology, 54*, 184–198. https://doi.org/10.1016/j.cedpsych.2018.06.008

Lombardi, D., Bickel, E. S., Bailey, J. M., & Burrell, S. (2018). High school students' evaluations, plausibility (re) appraisals, and knowledge about topics in Earth science. *Science Education, 102*(1), 153–177. https://doi.org/10.1002/sce.21315

Lombardi, D., Brandt, C. B., Bickel, E. S., & Burg, C. (2016). Students' evaluations about climate change. *International Journal of Science Education, 38*(8), 1392–1414. https://doi.org/10.1080/09500693.2016.1193912

Lombardi, D., Matewos, A. M., Jaffe, J., Zohery, V., Mohan, S., Bock, K., & Jamani, S. (2022). Discourse and agency during scaffolded middle school science instruction. *Discourse Processes, 59*(5–6), 379–400. https://doi.org/10.1080/0163853X.2022.2068317

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. (2016). Plausibility judgments in conceptual change and epistemic cognition. *Educational Psychologist, 51*(1), 35–56. https://doi.org/10.1080/00461520.2015.1113134

Lombardi, D., Sinatra, G. M., & Nussbaum, E. M. (2013). Plausibility reappraisals and shifts in middle school students' climate change conceptions. *Learning and Instruction, 27*, 50–62. https://doi.org/10.1016/j.learninstruc.2013.03.001

Loyens, S. M. M., Jones, S. H., Mikkers, J., & van Gog, T. (2015). Problem-based learning as a facilitator of conceptual change. *Learning and Instruction, 38*, 34–42. https://doi.org/10.1016/j.learninstruc.2015.03.002

Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology, 84*(4), 405–412. https://doi.org/10.1037/0022-0663.84.4.405

Mazziotta, M., & Pareto, A. (2022). Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min-Max method. *Environmetrics, 33* (5), Article e2730. https://doi.org/10.1002/env.2730

Medrano, J., Jaffe, J., Lombardi, D., Holzer, M. A., & Roemmele, C. (2020). Students' scientific evaluations of water resources. *Water, 12*(7), 2048. https://doi.org/10.3390/w12072048

Mu, A. Y. (2020). A hybrid machine learning model with cost-function based outlier removal and its application on credit rating. *Journal of Physics: Conference Series, 1584*(1), Article 012001. https://doi.org/10.1088/1742-6596/1584/1/012001

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. https://doi.org/10.17226/13165.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. https://doi.org/10.17226/18290.

Nussbaum, E. M. (2014). *Categorical and nonparametric data analysis: Choosing the best statistical technique*. Routledge. https://doi.org/10.4324/9780203122860.

Okada, R. (2023). Effects of perceived autonomy support on academic achievement and motivation among higher education students: A meta-analysis. *Japanese Psychological Research, 65*(3), 230–242. https://doi.org/10.1111/jpr.12380

Ormrod, J. E. (2017). *How we think and learn: Theoretical perspectives and practical implications*. Cambridge University Press. https://doi.org/10.1017/9781316691458.

Patall, E. A., Pituch, K. A., Steingut, R. R., Vasquez, A. C., Yates, N., & Kennedy, A. A. (2019). Agency and high school science students' motivation, engagement, and classroom support experiences. *Journal of Applied Developmental Psychology, 62*, 77–92. https://doi.org/10.1016/j.appdev.2019.01.004

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist, 56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796

Posner, G. L., & Strike, K. A. (1976). A categorization scheme for principles of sequencing content. *Review of Educational Research, 46*, 665–690. https://doi.org/10.3102/00346543050001055

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist, 44*(3), 159–175. https://doi.org/10.1080/00461520903028990

Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist, 56* (1), 54–77. https://doi.org/10.1080/00461520.2020.18626

Reeve, J., Cheon, S. H., & Yu, T. H. (2020). An autonomy-supportive intervention to develop students' resilience by boosting agentic engagement. *International Journal of Behavioral Development, 44*(4), 325–338. https://doi.org/10.1177/0165025420911103

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28* (2), 147–169. https://doi.org/10.1023/B:MOEM.0000032312.95499.6f

Robertson, J. R., Logan, M. W., Rosenberg, J. M., & Lombardi D. (2024). *Profiles of scientific thinking* [Manuscript submitted for publication]. Department of Human Development and Quantitative Methodology, University of Maryland.

Sadler, T. D., Foulk, J. A., & Friedrichsen, P. J. (2017). Evolution of a model for socio-scientific issue teaching and learning. *International Journal of Education in Mathematics, Science and Technology, 5*(2), 75–87. https://doi.org/10.18404/ijemst.55999

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly, 46*(4), 374–393. https://doi.org/10.1002/RRQ.005

Sinatra, G. M., & Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist, 55*(3), 120–131. https://doi.org/10.1080/00461520.2020.1730181

Smagorinsky, P. (2018). Deconflating the ZPD and instructional scaffolding: Retranslating and reconceiving the zone of proximal development as the zone of next development. *Learning, Culture and Social Interaction, 16*, 70–75. https://doi.org/10.1016/j.lcsi.2017.10.009

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Woolf, N. (2015). Republican Senate environment chief uses snowball as prop in climate rant. *The Guardian.* https://www.theguardian.com/us-news/2015/feb/26/senate-james-inhofe-snowball-climate-change.

Zeidler, D. L., & Sadler, T. D. (2008). Social and ethical issues in science education: A prelude to action. *Science & Education, 17*, 799–803. https://doi.org/10.1007/s11191-007-9130-6