Journal of Science Education and Technology
https://doi.org/10.1007/s10956-023-10046-z

Socio-Scientific Learning During the COVID-19 Pandemic: Comparing In-person and Virtual Science Learning Using Model-Evidence Link Diagrams

Nancy Gans[1] · Vivian Zohery[2] · Joshua B. Jaffe[1] · Anissa Ahmed[1] · Luke Kim[1] · Doug Lombardi[1]

Accepted: 5 April 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Science learning is an important part of the K-12 educational experience, as well as in the lives of students. This study considered students' science learning as they engaged in the instruction of scientific issues with social relevance. With classroom environments radically changing during the COVID-19 pandemic, our study adapted to teachers and students as they were forced to change from more traditional, in-person instructional settings to virtual, online instruction settings. In the present study, we considered science learning during a scaffold-facilitated process, where secondary students evaluated the connections between lines of scientific evidence and alternative explanations about fossil fuels and climate change and gauged the plausibility of each explanation. Our investigation focused on the relations between students' levels of evaluations, shifts in plausibility judgments, and knowledge gains, and examined whether there were differences in these relations between in-person classroom settings and virtual classroom settings. The results revealed that the indirect relational pathway linking higher levels of evaluation, plausibility shifts toward a more scientific stance, and greater knowledge gains was meaningful and more robust than the direct relational pathway linking higher levels of evaluation to greater knowledge gains. The results also showed no meaningful difference between the two instructional settings, suggesting the potential adaptiveness and effectiveness of properly-designed, scaffolded science instruction.

Science learning happens every day and is a ubiquitous part of the K-12 educational experience. To facilitate science learning, students should engage in scientific reasoning and evaluations, which might involve evaluating connections between scientific evidence and alternative explanations about a phenomenon (NRC, 2012). In such situations, students can extend their understanding of the process of making such evaluations through culturally common and relevant scientific issues (which some label as socioscientific issues or SSIs; Sadler et al., 2011). However, scientific issues with social relevance can often be controversial and complex, which may challenge students' reasoning and evaluations (Owens et al., 2017). Controversial scientific topics are characterized as issues where conflicting perspectives arise from distinct sources, "each with a distinct set of assumptions, points of view, target audiences, and goals" (Medrano et al., 2020).

When learning about scientific issues with social relevance, students may encounter alternative, but non-scientific explanations that conflict with consensus claims held by a particular community of scientists. Such competing, non-scientific explanations may challenge students during their science learning (Sinatra & Lombardi, 2020). These issues can include universal everyday issues such as vaccine efficacy, climate change, the effects of fracking, and the consequences of a lack of freshwater. Although there is scientific consensus on the topics we use in our studies, most are considered to be both controversial and complex. In our current six-year design and development project, middle and high school students used instructional scaffolds, called Model-Evidence Link (MEL) diagrams, to facilitate evaluations about lines of scientific evidence, scientific claims, and alternative explanations (Lombardi et al., 2016b). Based on earlier research, we consider explanations to be accounts of how phenomena unfold that may lead to a feeling of understanding (Braaten

✉ Nancy Gans
ngans@umd.edu

1 Department of Human Development and Quantitative Methodology, University of Maryland, 3942 Campus Drive, College Park 20742-1131, USA

2 Department of Teaching, Learning, Policy, and Leadership, University of Maryland, 3942 Campus Drive, College Park, MD 20742-1311, USA

Published online: 05 May 2023

& Windschitl, 2011; Brewer et al., 1998). Lombardi, et al. (2016) also write that "explanations may not include all facets of a fully developed scientific theory; rather, individuals more commonly explain by using parts of scientific theories (Giere, 1990; Salmon, 1994)." With this definition in mind, previous studies showed that MEL scaffolds can help students shift their plausibility judgments toward a more scientific stance when evaluating competing explanations and deepen their scientific knowledge about SSIs (see, for example, Lombardi et al., 2018, 2022; Medrano et al., 2020). King and Kitchener (2004) posited that more reflective judgements, involving "views of knowledge" and "concepts of justification," may be "initiated when an individual recognizes that there is controversy or doubt about a problem that cannot be answered by formal logic alone, and involve careful consideration of one's beliefs in light of supporting evidence" (p. 6). We have designed the MEL scaffolds to make these reflective judgments more explicit, purposeful, and scientific during classroom instruction.

Of course, classrooms changed radically during the COVID-19 pandemic. Teachers and students in classroom settings all over the world were forced to make drastic changes to their teaching and learning in order to adapt from more traditional, in-person to virtual, online learning environments (Horowitz & Igielnik, 2020). During this time, our research group also adapted to the COVID-19 pandemic, and in response, we developed and tested virtual versions of the MEL scaffolds. During the height of the pandemic, our research team was able to collect data in both in-person and virtual instructional settings. In-person settings used our original paper–pencil version of the MEL materials and virtual settings piloted our rapidly developed electronic MEL materials using Google documents and forms. In previous studies, our team specifically investigated how MELs facilitated shifts in plausibility toward a scientific explanatory model, in light of an alternative, and how such shifts might act as a mediating factor between scientific evaluations and knowledge (see, for example, Lombardi et al., 2018, 2022; Medrano et al., 2020). With the different ways that schools responded to the COVID-19 pandemic, the present study compared the MEL activities in these two instructional settings (in-person and virtual). We specifically asked: *What are the direct and indirect relations between students' levels of evaluation about the connections between lines of scientific evidence and alternative explanations, the shifts in plausibility judgments, and their gains in knowledge about scientific topics of social relevance? And are there differences in these relations between in-person classroom settings and virtual classrooms settings?* Prior to discussing the context and methods used to investigate this question, as well as the results and potential implications of the finding, we first turn to the theoretical framework that supported our research design.

Theoretical Background

Our theoretical framework study pulls from literature related to science learning, conceptual change, and educational psychology. Furthermore, because this research was conducted within secondary school settings (middle and high school classrooms), we considered extant research related to effective scaffolding for adolescent learning.

Scientific Evaluations and Science Learning

Much of our research has been conducted under the well-developed thesis stating that evaluations between lines of scientific evidence and alternative explanations of phenomena are critical to the processes of scientific knowledge construction and science learning (see, for example, Lombardi et al., 2016b, 2022; Ford, 2015). Specifically, our team used the theoretical work of Lombardi et al. (2016b) as a basis for understanding how students' evaluations impact their plausibility judgments about scientific explanations and influence their science learning. When learning about various scientific topics, students often come into class having some prior knowledge related to the subject matter, including knowledge about alternative explanations not consistent with scientific consensus (e.g., that the current climate crisis is caused by increased amounts of the Sun's energy being received by the Earth). In many cases, students can use their previously held knowledge and revise it such that their reconstructed conceptions align with scientifically-accepted understandings (Sinatra et al., 2008). The process of scientific knowledge construction may be facilitated via purposeful and explicit evaluations of both scientific evidence and explanations (Lombardi et al., 2016b; Medrano et al., 2020). With controversial scientific topics, deep learning may require such evaluations to consider connections between lines of scientific evidence and alternative explanations that might be encountered both within the classroom (e.g., the scientific consensus explanation being presented by the teacher) and outside the classroom (e.g., plausible, but non-scientific explanations being presented by media pundits) (Lombardi et al., 2022).

Scientific topics that are socially relevant may afford students the opportunity to reflect on their understanding and judgments about explanations. King and Kitchener (2015) said that when adolescents reflect on topics that are ill-structured and contextual, they may come to a better understanding of how scientific evaluation can result in an explanation being "more plausible than others" (p. 112). However, the complexity of some science topics may make such reflectivity and reasoning challenging, particularly when thinking about connections between

lines of scientific evidence and alternative explanations (Kuhn, 2011). For example, the phenomenon of extreme weather events and their potential relation to the climate crisis is complex. Through the media, peer, and/or family networks, students may be exposed to explanations suggesting that the number and strength of extreme weather events varies naturally based on periodic oscillations in oceans and plant absorption of atmospheric carbon. Such explanations are counter to the scientific explanation that increases in extreme weather events are linked to climate change caused by human activities, such as fossil fuel use (Schiermeier, 2018). Because students may face challenges related to being reflective and purposeful when making judgments about alternative explanations (e.g., relative plausibility judgments about the scientific explanation that human activities are the underlying cause of the climate crisis), instructional scaffolding may be required to facilitate their learning (see, for example, Bielik et al., 2021; Gobert & Pallant, 2004).

Plausibility Judgments about Scientific Explanations and Science Learning

Plausibility judgements have long been theoretically and empirically tied to students' science learning. For example, Strike and Posner (1992) theorized that the construction and reconstruction of scientific concepts involve consideration that the explanations "must at least appear as a candidate for the truth," especially in comparison to "well-established [personal] beliefs" (i.e., scientific explanations must be plausible in light of alternative explanations based on beliefs and prior experiences; p. 148). Similarly, Dole and Sinatra (1998) posited that to-be-learned conceptions must not only be comprehended by students but must also be plausible in a way that renders the topic to be, at least somewhat, believable. In the 1990s and early 2000s, past empirical studies used such perspectives to investigate students' science knowledge reconstruction. For example, Özkan et al. (2004) designed and tested instructional texts to increase the plausibility of "scientifically acceptable explanations" of ecological concepts (p. 99). However, plausibility was only considered as one of many factors considered in this conceptual development process, and the importance of these reflective judgments, particularly in relation to controversial science topics, was largely overlooked.

More recently, however, Lombardi and colleagues have been conducting many empirical studies using a robust theoretical model focused on plausibility judgments' role in science learning and teaching (see, for example, Lombardi et al., 2013, 2018, 2022; Medrano et al., 2020). This theoretical model—based on philosophical, developmental, and scientific foundations—posited that the perceived plausibility of a scientific claim can be thought of as a judgment of potential truthfulness. If a student thinks that a claim is highly plausible, they are tentatively accepting that it is true and worthy of some level of acceptance. This theoretical framework predicts how science learning may be facilitated in the form of developing and changing knowledge, specifically when one makes more explicit evaluations about the connections between scientific evidence and claims. In such situations, students may shift their plausibility judgments toward a more scientific stance and this influences knowledge gains above and beyond the explicit evaluation alone. For example, Lombardi et al. (2022) used MEL scaffolding to facilitate plausibility shifts toward the scientific consensus models that indicate that increases in extreme weather events are linked to human fossil fuel use, while also deepening students' knowledge of Earth's climate. However, to date, these studies have been conducted using pencil and paper forms of scaffolding. With the advent of the COVID-19 pandemic, we created electronic forms of these scaffolds to be used in virtual instruction. With theoretical and empirical studies suggesting that there might be differences in students' scientific thinking and reasoning when using virtual scaffolds (see for example, DeCoito & Estaiteyeh, 2022; Gerard et al., 2022; Graham, 2018; Singer & Alexander, 2017), the aim of the present study was to compare electronic MEL scaffolds used during virtual instructional settings with pencil and paper MEL scaffolds used during in-person instructional settings during the height of the COVID-19 pandemic.

The Present Study

In the present study, we considered science learning during a scaffold-facilitated process, where secondary students first evaluated the connections between lines of scientific evidence and alternative explanations, and then gauged the plausibility of each explanation. The outcome of learning was students' science knowledge after doing these evaluations and explicitly considering plausibility. From our past similar studies, shifts in students' plausibility towards a more scientific stance have mediated the positive direct effect between their evaluation and knowledge gains (see, for example, Bailey et al., 2022; Dobaria et al., 2022; Klavon et al., 2023; Medrano et al., 2020; Lombardi et al., 2018, 2022). However, what is novel about the present study is the context of the COVID-19 pandemic, which greatly impacted students' formal learning settings because of social isolation, including distancing and quarantining (Engzell et al., 2021). During this pandemic, where in-person interactions have become limited for some students, the instruction setting (in-person vs. virtual), as well as the instructional tools (pencil and paper handwritten activities vs. electronic typewritten activities) used by students may interact with their development of scientific understanding. Specifically, making scientific

evaluations that involve complex reasoning and reflective processes (e.g., inferential comprehension and critical thinking), can make learning challenging, even in the most optimal settings (Nückles et al., 2020; Wäschle et al., 2015). We also recognized the many studies stating positive outcomes of online learning, including increased autonomy of students, increased flexibility in studies and lifestyle, and increased access to online resources (see, for example, Adedoyin & Soykin, 2020; Basir et al., 2021). The different features of in-person and virtual instructional settings and tools could therefore interact with students' reasoning, either positively or negatively, because of the different contextual and technological features (Gnesdilow & Puntambekar, 2022; Graham, 2018).

Using data collected from classrooms during the height of the COVID-19 pandemic, this investigation specifically compared (a) in-person instructional settings using more traditional pencil and paper forms of the MEL scaffolds, completed using handwriting; and (b) virtual instructional settings using electronic forms of the MEL scaffolds, completed using typewriting. In this comparison, we examined two hypothesized pathways in students' knowledge construction process, the (a) direct pathway relating levels of evaluation to knowledge gains and (b) the indirect pathway relating levels of evaluation to shifts in plausibility judgements to knowledge gains. We hypothesized that shifts in plausibility judgments would have an indirect and positive effect on the relation between levels of evaluation and knowledge gains per Lombardi et al. (2016b) theoretical framework and past empirical studies showing such relations (see, for example, Lombardi, et al., 2018, 2022; Medrano et al., 2020). Furthermore, we hypothesized that this indirect effect would be above and beyond the direct and positive effect on the relation between levels of evaluation and knowledge gains.

Methods

Participants and Context

Our team collected data for this study from seventy-seven secondary science students (grades 8–10) located at three schools: two in the mid-Atlantic US and one in Southeastern US. These schools were located in suburban settings (predominantly White, 47% and female, 51% participants in either in-person ($n = 26$) or virtual ($n = 51$) classrooms) experiencing MEL instruction on geology topics of social relevance, including fossils and fossil fuel impacts on environmental sustainability and climate change (Lombardi et al., 2016a; Medrano et al., 2020; Governor et al., 2020; Hopkins et al., 2016). Overall, the MEL instruction lasted approximately 90 min (just under two traditional class periods). During this time, students completed the knowledge surveys, plausibility ratings, diagram construction, and explanation tasks. These data were collected during the 2020–2021 academic year (the height of the COVID-19 pandemic), with different community responses to formal school instruction. The one Southeastern school met in-person throughout the academic year, and the two Mid-Atlantic schools met virtually throughout much of the academic year. There were no major differences in instruction and delivery between the two settings, with students completing the same set of materials using pencil-paper (in-person) or electronic Google forms and documents (virtual). Video conferencing software was used during virtual instruction, with students working in breakout rooms to complete their MELs in small groups. Overall instruction time (~90 min) was essentially the same for in-person and virtual classrooms.

MEL Instructional Materials

Each classroom engaged in a lesson involving the MEL activities. The socio-scientific topics covered in these scaffolds focused on fossils and fossil fuel impacts on environmental sustainability and climate change. Depending on the curricular scope and sequence imposed by the local school setting, the exact topics varied somewhat in specific content, but all materials were aligned were each school's curricular requirements and performance expectations in the *Next Generation Science Standards* (e.g., HS-ESS3-4: Earth and Human Activity: Evaluate or refine a technological solution that reduces impacts of human activities on natural systems; NGSS Lead States, 2013). Figure 1 shows an example of a completed virtualMEL scaffold used in the study, with paper–pencil pcMEL and baMEL scaffolds shown in Supplementary Materials 1. Also, depending on the decisions of each classroom teacher, students used either the preconstructed or build-a-MEL form of the scaffold diagram. In the preconstructed form, students are presented with two alternative explanatory models of a phenomenon in the center of the diagram. For example, in the Extreme Weather MEL, the diagram shows two alternative explanations for increases in extreme weather events over the last 50 years. One of these explanatory models is the scientific consensus explanation (e.g., Increases in extreme weather events are linked to climate change. Current climate change is mainly caused by human activities, such as fossil fuel use). The other explanatory model is a non-scientific explanation. (e.g., The number and strength of extreme weather events vary naturally. Human activities release carbon into the atmosphere. Yet, plants and oceans absorb any carbon increases.) Teachers do not reveal the scientific explanatory models during MEL instruction, but later discuss the scientific consensus with their students. The build-a-MEL form allows students to construct their own MEL diagrams by selecting two explanatory models from three possible

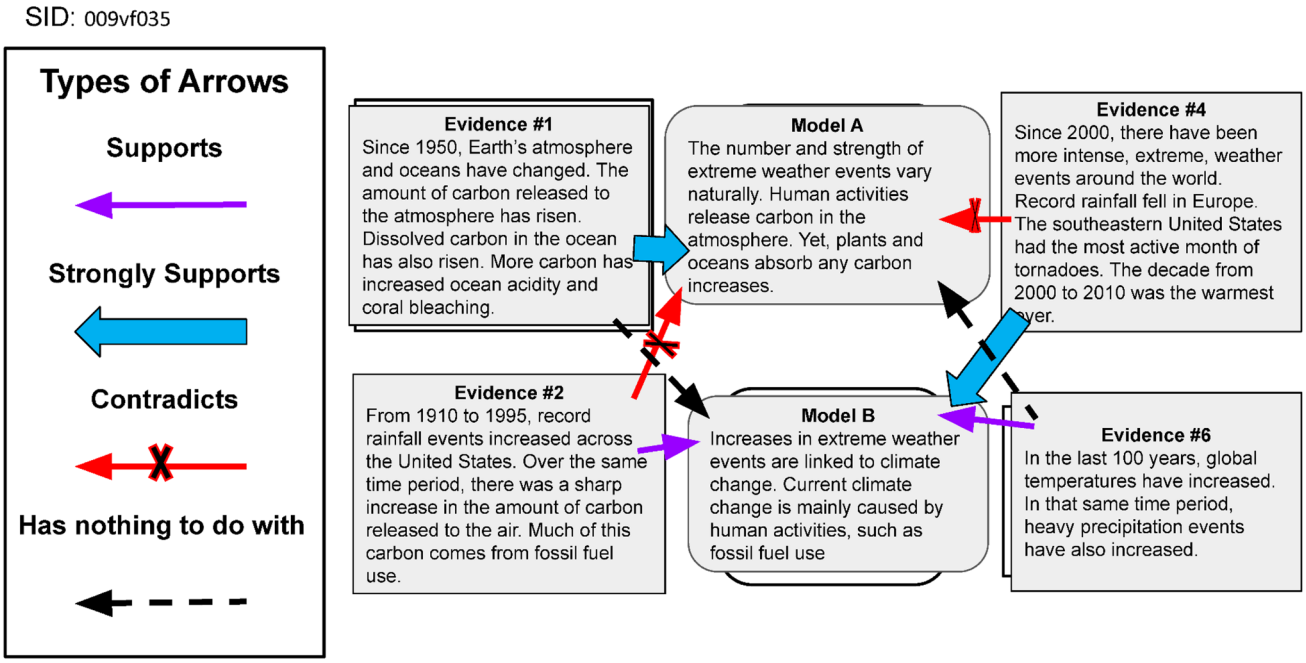

Fig. 1 *Participating Student Example of a Completed Extreme Weather MEL Diagram. Note.* This student example is an electronic form of the MEL diagram that was used in a virtual classroom setting during the height of the COVID-19 pandemic

choices (one scientific consensus and two non-scientific, alternative explanations).

Students are also presented with several lines of scientific evidence in the MEL lesson. In the preconstructed form, students are presented with four lines of scientific evidence, which appear in short boxed statements on the left and right sides of the explanatory models. In the build-a-MEL, students select four lines of scientific evidence from a possible eight choices to finalize their diagram construction. Students were also given "evidence texts," which were one-page summaries for each line of evidence that included expository text, graphs, and/or diagrams reviewed and validated by scientific experts. After reading the texts, the teacher instructed students to draw (in-person classrooms) or select (virtual classrooms) different types of arrows from each evidence text to both models based on how well they thought the evidence supported the model. Students were allowed to draw or select four different types of arrows: a squiggle or thick blue arrow indicated that the evidence strongly supports the model, a straight or thin purple arrow indicated that the evidence supports the model, a dotted line and black arrow indicated the evidence had nothing to do with the model, and a line with an "X" in the middle of it indicated that the evidence contradicts the model. Overall, the participants drew or selected eight arrows in total.

Explanation Task: Evaluation Scores

After completing their MEL Diagram, students completed an "Explanation Task". Participants picked two of the connections that they drew from the MEL activity and wrote a written response explaining why they drew a particular type of arrow and their evaluation of the strength between a particular line of evidence and a particular model (in-person responses were handwritten and virtual responses were typewritten). Using a scoring system and rubric developed by Lombardi et al. (2016a), coders rated explanations for different levels of evaluation using a rubric: 1 = erroneous, 2 = descriptive, 3 = relational, or 4 = critical. These categories represent levels of evaluation based on the scientific accuracy and reasoning present in students' written responses. Few students reached level 4 (critical evaluation), which would involve making a scientifically accurate connection between evidence and claim, as well as a thorough explanation of the weakness or strength of the causal connection present in light of the competing model. More commonly, we scored students at level 2 (descriptive) or 3 (relational) evaluation. To establish coding reliability, two raters independently coded participants' explanation tasks. They then met and resolved all differences in scoring via discussion, at times with a third coder to assist with that resolution, with a full consensus reached after consultation. We used both evaluation scores when constructing the levels of evaluation latent variable used in the analyses.

Model Plausibility Rating Task: Plausibility Judgment Scores

Students were instructed to rate the plausibility of all explanatory models, immediately before and immediately

after working on their diagrams. For the preconstructed scaffold form, students recorded their plausibility judgments for both explanatory models, while for the build-a-MEL form, students recorded their plausibility judgments for all three explanatory models. Students gauged the plausibility of each model using a 1–10 scale (1 = highly implausible and 10 = highly plausible), based on methods developed by Lombardi et al. (2013), where plausibility scores were calculated as the rating of the scientific consensus explanatory model minus either (a) the non-consensus explanatory model (preconstructed form) or (b) minus the average of the two non-consensus explanatory models. Scores could range on a scale from -9 to + 9, where positive scores indicated that students judged the scientific consensus explanation as more plausible than the non-consensus model(s). We also calculated shifts in judgments as the plausibility judgments score after completing the diagram minus the plausibility judgment score prior to completing the diagram, and used this calculation as the basis for our plausibility shift latent variable in our analyses.

Knowledge Survey and Scores

Students completed a multi-item knowledge survey at pre- and post-instruction. These instruments were aligned to the specific scientific topic, with items that measured understanding of the phenomenon and the scientific evidence and at least one question in each set addressed to each line of scientific evidence in a particular MEL. Some questions' statements were negatively worded (i.e., in effect scientists would disagree with these knowledge statements) and we reverse-coded these statements prior to calculating knowledge scores. Students rated each item on a 5-point Likert scale (1 = strongly disagree and 5 = strongly agree) on their knowledge of how scientists would agree with each item statement per the methods outlined in Lombardi et al. (2013). We used McDonald's omega (ω) coefficients to gauge the internal consistency (reliability) of the knowledge scores, with pre-instruction knowledge $\omega = 0.827$ and post-instruction knowledge $\omega = 0.720$, with both internal consistencies being acceptable (Hayes & Coutts, 2020). We acknowledge that the internal consistency values went down from pre to post-instruction, but any potential decrease is probably not meaningfully different. We summed student ratings for each item, and to account for varying item difficulty due to our alignment of all knowledge survey items to a specific MEL lesson, we calculated a simple normalization of these sums to come up with all final scores on a common scale that ranged on a continuum from 0 (minimum score in a class) to 1 (maximum score in a class) (Freedman et al., 2007). Finally, we calculated gain as the post-instructional normalized knowledge score minus the pre-instructional knowledge score and used this as the basis for our knowledge gain score in our analyses.

Results

We used structural equation modeling (SEM) to test our hypothesis about the relations between students' levels of evaluation, shifts in plausibility judgments, and knowledge gains. We analyzed these relations using WarpPLS 8.0 (Kock, 2022), a program that employs allows "warping" of latent relations (i.e., statistically fitting the data, without the assumption of linearity) to afford greater accuracy by not assuming linear relations between variables (Kock, 2016). WarpPLS also uses partial-least squares methods to reduce standard error and increase statistical power for relatively small sample sizes compared to ordinary-least squares methods employed by many other SEM programs (Kock, 2022). We used fit and quality indices to gauge the validity of our structural equation models. These indices include overall goodness-of-fit (Tenenhaus GoF) and average path coefficient (APC). Tenenhaus et al. (2004) proposed that researchers use Tenenhaus GoF as a criterion for the overall model prediction performance when using partial-least squares methods. A model has a large explanatory power when Tenenhaus GoF is greater than 0.36, with unacceptable explanatory power when Tenenhaus GoF is less than 0.1 (Wetzels et al., 2009). APC provides further information about model adequacy to gauge the predictive and explanatory power of the model (analogous to the total variance explained).

Preliminary Screening

Prior to conducting our SEM analyses, we screened the data to gauge any differences between various grouping variables using analyses of variance (ANOVAs). The ANOVAs revealed no significant differences between levels of evaluation, plausibility shifts, and knowledge gains when participants were grouped by specific instructional topics (e.g., extreme weather events vs. fossil evidence of paleoclimates), level (e.g., grade 8 vs. 9 vs. 10), and MEL scaffold form (preconstructed vs. build-a-MEL), with all p-values greater than 0.05 and η^2 values less than 0.03 (small effect sizes). In terms of the comparison of interest for the present study (e.g., in-person classroom settings using handwritten writing tools vs. virtual classrooms settings using typewritten writing tools), ANOVAs indicated there was also no significant difference between these two groups (Table 1) in levels of evaluation, plausibility shifts, and knowledge gains. Table 1 also lists the means and standard deviations for the variables. For both groups, levels of evaluation were low (i.e., less than 2 or less than a descriptive level of evaluation). The

Table 1 Means, standard deviations, and differences by instructional setting

Variable	Instructional setting		*F* ratio	*p*	η^2
	In-person	Virtual			
	M (*SD*)	*M* (*SD*)			
Levels of Evaluation	1.64 (0.64)	1.49 (0.54)	1.08	.303	.014
Plausibility Shift	-0.21 (1.42)	0.70 (2.91)	2.24	.139	.029
Knowledge Gain	.098 (.098)	.074 (.150)	0.56	.458	.007

Note. The possible range for levels of evaluation scores was 1 (low level of evaluation) to 4 (high level of evaluation). The possible range of plausibility shifts was -9 (least scientific stance) to + 9 (most scientific stance). The possible range of knowledge gains scores was normalized on a scale from 0 to 1, and in terms of practical significance can be converted to percent gain by dividing by 100, with .098 = 9.8% gain and .074 = 7.4% gain

plausibility shift for the in-person group was just under zero, indicating no real change in scientific stance. However, for the virtual group, the plausibility shift was just over one-half category toward a more scientific stance. Both groups experienced about 7–10% gains in knowledge. Such knowledge gains are practically significant considering the lessons took about 90 min of instructional time. Again, there was no significant difference between the groups, which gave us some confidence in conducting our comparison of hypothesized models using SEM.

Structural Equation Models

The structural equation model (SEM) for the in-person instructional group (Fig. 2a) was statistically significant and of good quality, with Tenenhaus Goodness of fit (GoF) = 0.371, and average (standardized) path coefficient (APC) = 0.288, $p = 0.025$. Individual standardized path values indicated moderate relations among the variables, with slightly greater weight for the indirect pathway of levels of evaluation to plausibility shift to knowledge gains compared to the direct pathway of levels of evaluation to knowledge gains. Some of the pathways for the in-person instructional group (Fig. 2a) had *p*-values slightly exceeding the 0.05 level, a cutoff commonly considered to ascertain significance. However, recent guidance suggests that *p*-values alone should not be used to evaluate potentially meaningful relationships, in the light of additional pathway strength indices, such as individual standardized path values (Amrhein et al., 2019). Furthermore, Wasserstein et al. (2019) asked researchers not to "believe that an association or effect is absent just because it was not statistically significant" (p. 1). This model explained about 30.2% of the total variance in knowledge gains ($R^2 = 0.302$), where the indirect pathway (evaluation → plausibility → knowledge) accounted for about 20.3% of the variance explained ($R^2 = 0.203$), and the direct pathway (evaluation → knowledge) accounted for about 9.9% of the variance explained ($R^2 = 0.099$).

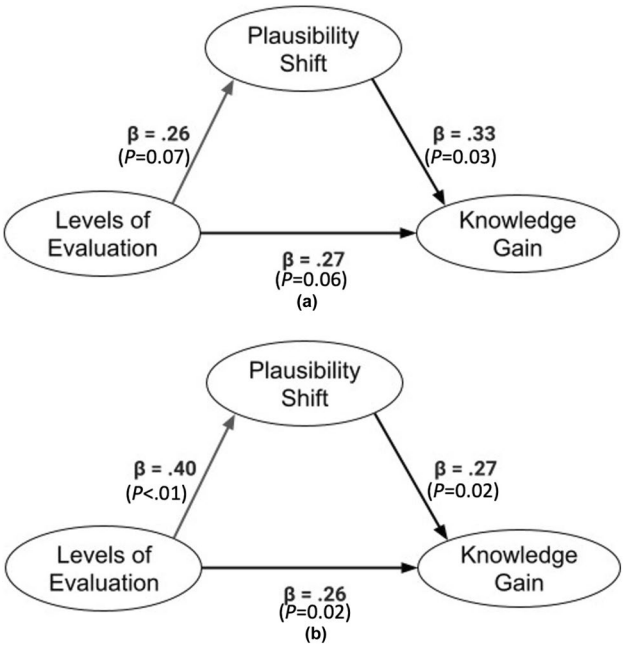

Fig. 2 *Structural Equation Models Showing Relations Among Study Variables. Note.* These figures show structural relations between the study variables for the two instructional settings: in-person (**a**) and virtual (**b**)

The SEM for the virtual instructional group (Fig. 2b) was also statistically significant and of good quality, with Tenenhaus Goodness of fit (GoF) = 0.387, and average (standardized) path coefficient (APC) = 0.313, $p = 0.004$. Individual standardized path values indicated moderate relations among the variables, with slightly greater weight for the indirect pathway of levels of evaluation to plausibility shift to knowledge gains compared to the direct pathway of levels of evaluation to knowledge gains. All of the pathways for the virtual instructional group (Fig. 2b) had *p*-values less than the 0.05 level. This model explained about 32.2% of the total variance in knowledge gains ($R^2 = 0.322$), where the indirect pathway (evaluation → plausibility → knowledge) accounted for about 24.3% of the variance explained ($R^2 = 0.243$), and the direct pathway (evaluation → knowledge) accounted for about 7.9% of the variance explained ($R^2 = 0.079$).

Discussion

The present study investigated the relations between students' levels of evaluation about the connections between lines of scientific evidence and alternative explanations, their shifts in plausibility when considering these alternative

explanations, and their knowledge gains about topics related to fossils and fossil fuel impacts on environmental sustainability and climate change. The novel aspect of the present study was conducting a comparative examination of these relations in two types of instructional settings that occurred during the height of the COVID-19 pandemic: (a) more traditional in-person settings, doing pencil and paper lesson activities (in-person/traditional/handwritten), and (b) settings, doing electronic lesson activities (virtual/electronic/typewritten). We found no meaningful difference between the two settings, with results supporting our hypothetical model. Results suggest that students' evaluation levels have an indirect and positive effect on the relation to their knowledge gains, by way of plausibility shifts towards the scientific explanation. This relation mediated by plausibility shift is above and beyond the direct relations between levels of evaluation and knowledge gains for both in-person and virtual settings.

Levels of evaluation scores in both settings were between the lowest (erroneous) and second lowest (descriptive) levels. These means are somewhat lower than in our many other studies involving the MEL scaffolds, with scores falling between the second lowest (descriptive) and second highest (relational) levels of evaluation (see, for example, Lombardi et al., 2018, 2022; Medrano et al., 2020). Similarly, plausibility shifts were also lower than in our past studies, with no meaningful shift for the in-person setting and about half a category shift toward the scientific for virtual settings. These scores may have been lower than previous studies due to the nature of the pandemic. Teachers were having a generally difficult time instructing students during this time.

However, knowledge gains, albeit slightly lower than what we have found in past MEL studies, were still relatively robust 7–10%, especially with the MEL activities constituting about two traditional class periods of instruction time (~90 min). Given that some were worried about decreased learning with virtual instruction during the COVID-19 pandemic (Horowitz & Igielnik, 2020), the present study showed that although in-the-moment engagement in scientific evaluation and reflective plausibility reappraisal was lower than before the pandemic, the learning outcome was still meaningful.

Both instructional settings also showed that students who had higher levels of evaluation also had associated greater shifts in plausibility toward the scientific and greater knowledge gains, above and beyond the direct association between higher levels of evaluations and greater knowledge gains. Thus, students in both in-person and virtual settings benefitted from instruction that facilitated their evaluations between lines of scientific evidence and alternative explanations. In considering the many challenges that the pandemic imposed on classroom science instruction both in-person and virtual (Gillespie et al., 2021; Reuge et al., 2021; Tomasik et al., 2021), we were somewhat heartened by the effectiveness—although muted—of scaffolded instruction about scientific topics with social relevance. However, informal discussions with participating teachers, which were held via Zoom at roughly 6-week intervals during the study, suggested that it was especially challenging to use the MEL scaffolding in both the traditional pencil and paper and electronic forms. These teachers also said that it was challenging to use other NGSS-aligned and reform-based instructional techniques during the pandemic. The dedication of these three teachers to provide quality instruction for their students was uplifting and inspiring to their entire research team but does suggest that other teachers may have shifted to a less effective, lecture-based type of instruction during the pandemic (Pressley, 2021).

Research conducted during the pandemic has shown that well-planned virtual instruction, which has undergone many iterations of development and testing, was more effective than hastily constructed electronic tools (Adedoyin & Soykan, 2020). Therefore, we approach the results with some caution because of the exploratory nature of the present study and because students in the virtual classrooms used a first-cycle iteration of our electronic MEL scaffolds. However, the form and structure of the MEL scaffolds have been tested extensively over the past decade, similar to other design-based instruction using other types of scaffolding (see, for example, Darner, 2019; Dauer et al., 2022; Ke et al., 2021) Thus, design-based research, involving practicing science teachers who can help researchers know what works and does not work in their classrooms, is essential for development and implementation of effective instructional tools (Gerard et al., 2022). Furthermore, the science education research community, indeed the entire education community, should shift from the mindset of technology in the classroom being something that is "nice to have" to something that is "mission critical" (Ribiero, 2020). This suggests that the infrastructure behind effective scaffolding development and testing should be in place to support adaptive and effective instruction that can transition between in-person and virtual settings (Hodges et al., 2020).

Limitations and Future Directions

Conducting classroom research, in both in-person and virtual settings, was extremely difficult due to the many challenges imposed on our educational system, indeed society as a whole, during the height of the COVID-19 pandemic. We are therefore extremely grateful for the three teachers who invited us into their classrooms and for the secondary students who participated in our study. Understandably, the difficulties imposed by the pandemic made it quite difficult to conduct classroom-based research and our sample size was limited. The limited sample may have resulted in decreased power to detect more pronounced effects. Within

this sample, we also were unable to compare students in the same classrooms, so we were comparing students taught by different teachers in different school settings. Although we were able to employ robust statistical techniques to increase the power of our analyses and found no meaningful differences in school settings, we do suggest that future studies involving larger sample sizes are warranted. Individual teaching differences may also have influenced the results; however, results of past studies using the MEL scaffold indicate virtually no teaching effects due to the robust nature of the design (Lombardi et al., 2018, 2022; Medrano et al., 2020). Furthermore, the virtual instruction group had twice as many students, and although we did not have any evidence of statistical irregularities between groups (e.g., the homogeneity of variance assumption was verified for all group comparisons), we acknowledge that larger samples may give more confidence to the broader community when comparing virtual and in-person science instruction.

We recognize that virtual learning is a tricky and confusing term that needs to be more clearly defined in the literature with the proliferation of online teaching and learning that has occurred in light of the pandemic (Hodges et al., 2020). Virtual learning can be done in a way that is well-planned and validated, but it can also be done as an emergency measure, as was the case in the schools that participated in this study (Gerard et al., 2022; Hodges et al., 2020; Ribiero, 2020). Online learning can also be defined as learning that is happening completely in the virtual sphere, as well as learning online while physically embedded in the structure and scaffolding of the classroom (DeCoito & Estaiteyeh, 2022). In future studies, there should be considerations of how virtual learning formats might change the way the MEL and similar instructional scaffolds are crafted to ensure optimal effectiveness. This would also help more clearly distinguish if virtual learning was truly a buffer during the pandemic times or if it is a learning modality that is helpful for all students at all times.

We also acknowledge that the educational system was widely affected by the pandemic and the results of the present study were also likely to suffer during this time. Teachers and schools faced issues with the logistics of teaching, technological gaps, and socio-economic challenges that could have factored into the results (Adedoyin & Soykan, 2020). To the extent that some of those issues have been mediated as schools have returned to more pre-pandemic functioning, we might see a change in the way that the virtual version of the MEL and other instructional scaffolds focusing on SSIs are used. For example, many have cited inequity issues with Internet access and how the lack of internet in communities, such as low SES communities and highly populated Black and Brown communities, negatively impacted students' ability to learn online, simply because they did not have stable access to the tools being used (Fishbane & Tomer, 2020). To effectively use virtual instructional scaffolds in the future will require addressing inequities in Internet access, as well as maintaining relevance to the context of local communities and the socio-scientific challenges being experienced (Fishbane & Tomer, 2020).

Conclusion

Our research team recognized the great privilege of working with teachers and students during the tumultuous learning environment imposed by the COVID-19 pandemic. This research led us, in some ways, to more questions than answers. The results support our earlier studies (Lombardi et al., 2018, 2022; Dobaria et al., 2022), suggesting that students' plausibility shifts have an indirect effect between levels of scientific evaluation and their knowledge gains, above and beyond the direct effect of evaluation on knowledge—particularly when simultaneously considering alternative explanations about scientific topics of social relevance. Furthermore, instructional scaffolding, such as the MELs, can facilitate such shifts in both in-person and virtual settings. Although we found no difference between these two settings, we did see decreased performance from our previous classroom-based investigations, particularly from making more scientific evaluations about the connections between lines of scientific evidence and alternative explanations. We speculate this performance reduction was due to lower reflective thinking during the COVID-19 pandemic affecting students in both in-person and virtual instructional settings (Adedoyin & Soykan, 2020; Horowitz, & Igielnik, 2020).

As we continue to brave the new frontier that is life beyond the COVID-19 pandemic, we have to ask ourselves what knowledge was gained about science teaching and learning. We are humbled by the science teachers and students who learned how to teach and learn in this challenging environment. Even when faced with challenges, the teachers and students were able to adapt to an online version of the instructional scaffold that helped them gain knowledge about controversial scientific topics of social relevance. This gives us hope that instructional scaffolding born out of necessity for effective science teaching and learning promises to be a technology that benefits classroom instruction in the future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10956-023-10046-z.

Acknowledgements Our many thanks to the teachers who allowed us into their in-person and virtual classrooms during the height of the COVID-19 pandemic, as well as their students who participated in this research study.

Author Contribution Conceptualization: N. Gans, V. Zohery, A. Ahmed, & D. Lombardi; Funding acquisition: D. Lombardi; Methodology: N. Gans, V. Zohery, & D. Lombardi; Project administration &

supervision: D. Lombardi; Visualization: N. Gans; Writing (original draft): N. Gans, A. Ahmed, & D. Lombardi; Writing (review & editing): N. Gans & D. Lombardi.

Funding This research was funded, in part, by the US National Science Foundation (Grant No. 2027376).

Availability of Data and Materials Data and materials are available from the sixth author upon request.

Declarations

Ethical Approval This study was approved by the University of Maryland's Institutional Review Board under Protocol No. 1581814.

Informed Consent All minor participants included in this study provided informed assent and their parent or legal guardian provided informed consent.

Statement Regarding Research Involving Human Participants and/or Animals This research involved only human participants (i.e., no animals) and was guided by the University of Maryland's Institutional Review Board under Project No. 1581814.

Consent to Participate All minor participants included in this study provided informed assent and their parent or legal guardian provided informed consent.

Consent to Publish Not applicable.

Competing Interests The authors declare no competing interests.

References

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments*, 1–13. https://doi.org/10.1080/10494820.2020.1813180

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305–307. https://www.nature.com/articles/d41586-019-00857-9?fbclid=IwAR258v566K49c_cig0f9FhPov_UnuGFt3DZ0YbM095emw4iK87Zmz1ovY0s

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

Basir, A., Kamaliah, K., Harahap, A., Fauzi, A., & Karyanto, B. (2021). How Universities Entrust Digital Literacy to Improve Student Learning Outcomes During the COVID-19 Disruption. *Jurnal Iqra': Kajian Ilmu Pendidikan*, *6*(1), 235–246. https://journal.iaimnumetrolampung.ac.id/index.php/ji/article/download/1146/765

Bielik, T., Fonio, E., Feinerman, O., Duncan, R. G., & Levy, S. T. (2021). Working together: Integrating computational modeling approaches to investigate complex phenomena. *Journal of Science Education and Technology, 30*(1), 40–57. https://doi.org/10.1007/s10956-020-09869-x

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education, 95*, 639–669. https://doi.org/10.1002/sce.20449

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (1998). Explanation in scientists and children. *Minds and Machines, 8*, 119–136. https://doi.org/10.1023/A:1008242619231

Darner, R. (2019). How can educators confront science denial? *Educational Researcher, 48*(4), 229–238. https://doi.org/10.3102/0013189X19849415

Dauer, J. M., Sorensen, A. E., & Jimenez, P. C. (2022). Using structured decision-making in the classroom to promote information literacy in the context of decision-making breadcrumb. *Journal of College Science Teaching, 51*(6), 75–82. https://www.nsta.org/journal-college-science-teaching/journal-college-science-teaching-julyaugust-2022/using-structured

DeCoito, I., & Estaiteyeh, M. (2022). Transitioning to online teaching during the COVID-19 pandemic: An exploration of STEM teachers' views, successes, and challenges. *Journal of Science Education and Technology, 31*(3), 340–356. https://doi.org/10.1007/s10956-022-09958-z

Dobaria, A., Bailey, J. M., Klavon, T. G., & Lombardi, D. (2022). Students' scientific evaluation of astronomical origins, *Astronomy Education Journal, 02*(1), 1–16. https://doi.org/10.32374/AEJ.2022.2.1.032ra

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*(2–3), 109–128. https://doi.org/10.1207/s15326985ep3302&3_5

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences, 118*(17). https://doi.org/10.1073/pnas.2022376118

Fishbane, L. & Tomer, A. (2020). *As classes move online during COVID-19, what are disconnected students to do?* Brookings. https://www.brookings.edu/blog/the-avenue/2020/03/20/as-classes-move-online-during-covid-19-what-are-disconnected-students-to-do/

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics: Fourth international student edition*. New York, New York: W. W. Norton & Company.

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

Gerard, L., Wiley, K., Debarger, A. H., et al. (2022). Self-directed science learning during COVID-19 and beyond. *Journal of Science Education and Technology, 31*(2), 258–271. https://doi.org/10.1007/s10956-021-09953-w

Giere, R. N. (1990). Explaining science: A cognitive approach. Chicago, Illinois: University of Chicago Press.

Gillespie, D. L., Meyers, L. A., Lachmann, M., Redd, S. C., & Zenilman, J. M. (2021). The experience of 2 independent schools with in-person learning during the COVID-19 pandemic. *The Journal of School Health, 91*(5), 347–355. https://doi.org/10.1111/josh.13008

Gnesdilow, D., & Puntambekar, S. (2022). Comparing middle school students' science explanations during physical and virtual laboratories. *Journal of Science Education and Technology, 31*(2), 191–202. https://doi.org/10.1007/s10956-021-09941-0

Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology, 13*(1), 7–22. https://doi.org/10.1023/b:jost.0000019635.70068.6f

Governor, D., Strickland, K., & Bailey, J. M. (2020). Climate changes of the past: Engaging in evidence-based argumentation. *The Earth Scientist, 36*(3), 13–17. https://par.nsf.gov/servlets/purl/10282182

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist, 53*(4), 258–279. https://doi.org/10.1080/00461520.2018.1481406

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). *The difference between emergency remote teaching and*

online learning. Educause. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Hopkins, J., Crones, P., Burrell, S., Bailey, J. M., & Lombardi, D. (2016). Evaluating the connections between fracking and earthquakes. *The Earth Scientist, 32*(2), 23–30. https://d32ogoqmya1dw8.cloudfront.net/files/mel/teaching_resources/evaluating_connections_between_frac.pdf

Horowitz, J. M., & Igielnik, R. (2020). *Most parents of K-12 students learning online worry about them falling behind*. Pew Research Center. https://www.pewsocialtrends.org/2020/10/29/most-parents-of-k-12-students-learning-online-worry-about-them-falling-behind

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

King, P., & Kitchener, K. (2015). Cognitive development in the emerging adult: The emergence of complex cognitive skills. In J. J. Arnett (Ed.), *Oxford handbook of emerging adulthood,* 105–125. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199795574.013.14

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist, 39*(1), 5–18. https://doi.org/10.1207/s15326985ep3901_2

Klavon, T. G., Mohan, S., Jaffe, J. B., Stogianos, T., Governor, D., & Lombardi, D. (2023). Scientific Evaluations and Plausibility Judgements in Middle School Students' Learning about Geoscience Topics, *Journal of Geoscience Education*. https://doi.org/10.1080/10899995.2023.2200877

Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods, 15*(1), 299–315. https://doi.org/10.22237/jmasm/1462076100

Kock, N. (2022). *WarpPLS user manual: Version 8.0*. ScriptWarp Systems. https://scriptwarp.com/warppls/UserManual_v_8_0.pdf

Kuhn, D. (2011). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (2nd ed., pp. 397–523). Blackwell.

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., Brandt, C. B., Bickel, E. S., & Burg, C. (2016a). 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, M. 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. (2016b) Plausibility judgements in conceptual change and epistemic cognition. *Educational Psychologist, 51*(1), 1–22. 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

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

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

NRC [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

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, 1–38. https://doi.org/10.1007/s10648-020-09541-1

Owens, D. C., Sadler, T. D., & Zeidler, D. L. (2017). Controversial issues in the science classroom. *Phi Delta Kappan, 99*(4), 45–49. https://doi.org/10.1177/2F0031721717745544

Özkan, Ö., Tekkaya, C., & Geban, Ö. (2004). Facilitating conceptual change in students' understanding of ecological concepts. *Journal of Science Education and Technology, 13*(1), 95–105.https://doi.org/10.1023/b:jost.0000019642.15673.a3

Pressley, T. (2021). Factors contributing to teacher burnout during COVID-19. *Educational Researcher, 50*(5), 325–327. https://doi.org/10.3102/2F0013189X211004138

Reuge, N. Jenkins, R., Brossard, M., Soobrayan, B., Mizunoya, S., Ackers, J., Jones, L., & Taulo, W. G. (2021). Education response to COVID 19 pandemic, a special issue proposed by UNICEF: Editorial review. *International Journal of Educational Development,* 87. https://doi.org/10.1016/j.ijedudev.2021.102485

Ribiero, R. (2020). How university faculty embraced the remote learning shift. *Ed Tech Magazine.* https://edtechmagazine.com/higher/article/2020/04/how-university-faculty-embraced-remote-learning-shift

Sadler, T. D., Dawson, V., Klosterman, M. L., Eastwood, J., & Zeidler, D. L. (2011). Metalogue: Issues in the conceptualization of research constructs and design for socio-scientific department of issues. In T. D. Sadler (Ed.), *Socio-scientific issues in the classroom: Teaching, learning and research* (pp. 79–87). Springer. https://doi.org/10.1007/978-94-007-1159-4_5

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61*, 297–312.

Schiermeier, Q. (2018). Droughts, heatwaves and floods: How to tell when climate change is to blame. *Nature, 560*, 20–22. https://doi.org/10.1038/d41586-018-05849-9

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach, 1*(2), 189–195. https://doi.org/10.1007/s12052-008-0037-8

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

Singer, L. M., & Alexander, P. A. (2017). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research, 87*(6), 1007–1041. https://doi.org/10.3102/2F0034654317722961

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice,* 147–176. SUNY Press.

Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). A global goodness-of-fit index for PLS structural equation modelling. In *Proceedings of the XLII SIS scientific meeting 1*(2), 739–742.

Tomasik, M. J., Helbling, L. A., & Moser, U. (2021). Educational gains of in-person vs. distance learning in primary and secondary schools: A natural experiment during the COVID-19 pandemic school closures in Switzerland. *International Journal of Psychology, 56*(4). https://doi.org/10.1002/ijop.12728

Wäschle, K., Gebhardt, A., Oberbusch, E. M., & Nückles, M. (2015). Journal writing in science: Effects on comprehension, interest, and critical reflection. *Journal of Writing Research*, *7*(1). https://doi.org/10.17239/jowr-2015.07.01.03

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$". *American Statistician, 73*(S1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS quarterly*, 177–195. https://doi.org/10.2307/20650284

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