Scaffolding Elementary Students' Scientific Evaluations of Model-Evidence Relationships About Fossils

Timothy G. Klavon[1], Sydney S. Haugland[1], Nancy Gans[2], & Melissa Schweisow[3]

1. Black Hills State University 2. University of Maryland- College Park 3. Douglas School District, Box Elder, SD.

Subject

The development of students' scientific evaluative skills, particularly those of elementary students, has been an ongoing focus for science education reforms and research for some time. Michael Ford (2015) emphasized the foundational role of critical evaluation embedded in the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). Recent science education reforms have espoused the need for elementary students to learn more than just science facts, but to also develop explanations for the scientific models they encounter (NRC, 2012). Additionally, research has shown that students require extended practice and experience to develop their abilities to evaluate scientific evidence and to create evidence-based arguments about scientific phenomena (Crowell & Kuhn, 2014).

To develop these skills, researchers (Chinn & Buckland, 2012; Lombardi, 2016) have developed Model-Evidence Link diagrams (MELs) to scaffold students' evaluation of the relationships between scientific evidence and explanatory models. When using the MEL scaffold, students first rate the plausibility of two competing explanatory models about a scientific phenomenon, one model is the scientifically accepted model, and the other is a plausible alternative. Once they have completed this rating, students then complete the MEL scaffold where they are presented with the two models and four lines of scientific evidence. Students work in small groups to evaluate the evidence and its relation to the models, construct their own MEL, then create a consensus MEL for the group. Upon completion of the MEL, students are then asked to re-appraise their plausibility of each model and explain their evaluations of the most influential relationship they formed on their MEL.

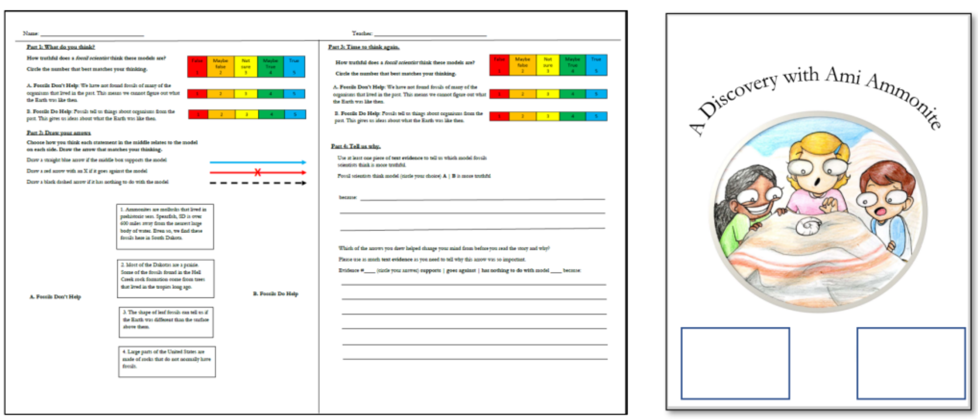
Previous research (Lombardi, Bailey et al., 2018; Lombardi, Bickel et al., 2018, Dobaria et al., 2022; Medrano et al. 2020) has shown that middle and high school students' re-appraisal of their plausibility judgements after completing the MEL diagram shifts these judgments towards the scientifically accepted model. In doing this, students develop their evaluative skills and learn more deeply about the topic at hand. Although the MEL has been used mostly with middle school and high school students, researchers have questioned if the MEL could potentially be useful for elementary students. To this end, this project seeks to investigate the development of a MEL activity designed for elementary students (i.e., eMEL).

The underlying framework for this project was the Plausibility Judgements for Conceptual Change model as described in Lombardi, Nussbaum, et al. (2016). Plausibility is defined as a "judgment of potential truthfulness when evaluating explanations" (Lombardi, Nussbaum et al., 2016; p. 35). Continuing research into MEL activities has consistently supported that when middle and high school students are given the opportunity to revisit their plausibility judgments after evaluating the relationship between lines of evidence and explanatory models, they learn more deeply about science academic content (Bailey et al., 2021; Dobaria et al., 2022; Klavon et al., 2023; Lombardi, Bailey et al., 2018; Lombardi et al., 2013; Medrano et al., 2020).

With the historical success of the MEL activities in secondary level science classes and the continuing need to develop elementary science students' evaluative practices, this pilot study asks the research question: *What are the relationships between elementary students' plausibility* *judgements, knowledge, and scientific evaluation of evidence about the use of fossils as indicators of past environments on the Earth's surface?*

Procedure

Instrument development took place in three phases, the scaffold phase, the book phase, and the knowledge instrument phase. The foundational MEL activities upon which this study is based included three scaffolds: the model plausibility rating (MPR), the MEL, and the explanation task. Additionally, each activity included four informational texts. In order to provide a concise and accessible process for elementary education students, all three scaffolds were converted into an interactive informational text serving as expository literature (Figure 1).



(a) (b)

Figure 1. The eMEL scaffolds. (a) The interactive student worksheet. (b) The eMEL children's book.

The knowledge instruments (pre-instruction, post-instruction, and delayed post-instruction) included two parts, a four question Likert scale asking how truthful a "fossil scientist" would find statements about fossils to be true (Table 1.) and a free response question asking about the students' self-reported knowledge about fossils. The Likert scale ran from 1 to 5, denoting False, Mostly False, Unsure, Mostly True, to True, respectively.

Included in the worksheet scaffold were the explanatory models, the pre and post MPR, the MEL diagram with lines of evidence (Table 1), and space to complete the explanation task. The storybook scaffold included guidance for completing the activity and a narrative context to provide the information found in the secondary-level MEL's evidence texts.

Table 1

Knowledge Survey Statements

1. There are equal numbers of fossils found in all kinds of rocks.
2. We can learn about past climates by examining fossilized leaves.
3. There are places on land that have fossils from animals that used to live underwater.
4. Fossils of tropical trees exist in rocks in South Dakota.

Note: Item 1 is negatively stated and was properly coded upon data entry.

Once the instruments were developed, third (3) and fourth (2) grade teachers were recruited to implement the activity in their classrooms. The teachers participated in a 3-hour professional development workshop to learn how to implement the lesson. The lesson consisted of reading the story and completing the accompanying eMEL scaffold as directed in the story. Students also completed the knowledge instruments prior to the lesson, directly after the lesson, and two weeks later to gauge knowledge retention.

Table 2

Explanatory Models and Lines of Evidence for the eMEL Lesson

Explanatory Models	Lines of Evidence
A: Fossils Don't Help: We have not found fossils of many of the organisms that lived in the past. This means we cannot figure out what the Earth was like then. (Alternative)	1. Ammonites are mollusks that lived in prehistoric seas. Spearfish, SD is over 600 miles away from the nearest large body of water. Even so, we find these fossils here in South Dakota.
B: Fossils Do Help: Fossils tell us things about organisms from the past. This gives us	2. Most of the Dakotas are a prairie. Some of the fossils found in the Hell Creek rock

ideas about what the Earth was like then. (Scientifically Accepted)	formation come from trees that lived in the tropics long ago. 3. The shape of leaf fossils can tell us if the Earth was different than the surface above them. 4. Large parts of the United States are made of rocks that do not normally have fossils.

Findings and Analysis

A convenience sample (N = 71) of 3rd (n = 38) and 4th (n = 33) grade students was selected from an elementary school located in the upper Midwest of the United States. The sample students were predominantly white (88.2%), with 8.8% students reporting to be of mixed ethnicity, 1.5% reporting as Native American, and 1.5% reporting as Asian. Additionally, four students did not report their ethnicity.

The pre-instruction to post-instruction changes in plausibility ratings and student knowledge about fossils have been previously presented (Gans et al., 2023). The students' plausibility judgements about the explanatory models moderately shifted towards the scientific model, $F(1, 70) = 4.6$, $p = .036$, $\eta^2 = 0.060$. The pre-instruction to post-instruction knowledge gains and pre-instruction to delayed-post instruction knowledge gains were both significant and quite larger, $F(1, 70) = 59.3$, $p < .001$, $\eta^2 = 0.455$ and $F(1, 70) = 62.2$, $p < .001$, $\eta^2 = 0.467$, respectively.

In addition to the knowledge and plausibility values, we also coded for students' level of scientific evaluation as evidenced in their explanation task writings. The students were asked to describe the connection from their MEL diagram that was, in their estimation, the most important to their final plausibility judgements. Using the rubric developed for earlier MEL activities

(Lombardi, Brandt et al., 2016), the lead investigator for the study trained the co-investigators in coding for students' levels of evaluation ranging from incorrect/nonsensical (1), descriptive (2), relational (3), to critical evaluation (4). The training consisted of discussing approximately 10% of the sample together to reach consensus for each participant. The second phase of training consisted of each coder doing an additional 10% of the sample independently and coming together to reach consensus. Finally, each coder was assigned two thirds of the remaining explanation items, overlapping with each of the other coders, to evaluate and then meeting with the other coders separately to reach consensus of their shared items. The third coder was available to mediate any major disagreements between coding pairs.

To answer the research question for this study, what are the relationships between elementary students' plausibility judgements, knowledge, and scientific evaluation of evidence about the use of fossils as indicators of past environments on the Earth's surface, we employed partial-least squares structural equation modeling (PLS-SEM) using WARP-PLS 8.0 (Kock, 2022). The choice for using PLS-SEM was due to its increased stability when analyzing a relatively small sample size of this pilot study. We also employed the jackknifing resampling technique in order to mitigate the impact of outliers in such a sample size. As we had removed students with missing data from the sample, there was no need to impute any missing values.

The student data was entered into WARP-PLS 8.0. The initial model was drawn from previous research associated with MEL activities (Bailey et al., 2021; Dobaria et al., 2022; Klavon et al., 2023; Lombardi, Bailey et al, 2018; Medrano et al., 2020). Constructs were organized chronologically, with Pre-instruction Knowledge (PrK) and Pre-instruction Plausibility (PrP) coming first, followed by Evaluation (Eval), Post-instruction Plausibility (PoP), with Post-instruction Knowledge (PoK) or Delayed Post-instruction Knowledge (DPoK) completing the structure. Once constructed, we ran WARP-PLS to determine the relationship values (Table 3). Both the PoK (Figure 2a) and DPoK (Figure 2b) models exhibited a high Tenenhaus Goodness of Fit (GoF = 0.386 for both), which indicates that each model is highly representative of the data (Kock, 2022).

Table 3

Structural Equation Model Relationship Values

		Lead Construct											
		PrK			PrP			Eval			PoP		
Model	End Construct	β	*p*	ES	β	*p*	ES	β	*p*	ES	β	*p*	ES
Post-Instruction Knowledge	PrP	**0.34**	**<.01**	**0.115**	-	-	-	-	-	-	0.22	.31	0.046
	Eval	**0.19**	**.07**	**0.035**	0.38	.44	.143	-	-	-	-	-	-
	PoP	-	-	-	**0.56**	**<.01**	**0.331**	**0.14**	**.05**	**0.035**	-	-	-
	PoK	**0.38**	**.25**	**0.141**	-	-	-	0.04	.43	0.005	-	-	-
Delayed Post-Instruction Knowledge	PrP	**0.26**	**<.01**	**0.068**	-	-	-	-	-	-	0.19	.33	0.045
	Eval	**0.17**	**.05**	**0.025**	0.39	.47	0.149	-	-	-	-	-	-
	PoP	-	-	-	**0.56**	**<.01**	**0.331**	**0.14**	**.05**	**0.035**	-	-	-
	DPoK	**0.44**	**<.01**	**0.212**	-	-	-	0.14	.46	0.028	-	-	-

Notes. PrK- Pre-instruction Knowledge, PrP- Pre-instruction Plausibility, Eval- Evaluation, PoP- Post-instruction plausibility, PoK- Post-instruction Knowledge, DPoK- Delayed Post-instruction knowledge. **Bold** relationships are included in the final model.

When considering which relationships to include in the model, we use a holistic approach that considers the significance of each pathway, the standardized path values, and the effect size of said pathway (Bailey et al., 2021; Dobaria et al., 2022; Klavon et al., 2023). Overall, each model (Figure 2) was driven by the PrK construct. Past research (Braasch & Goldman, 2010, Klosterman & Sadler, 2010) has shown correlations between students' prior knowledge and post-instruction learning gains. The PrK-PoK relationship is interesting due to the unusual dynamic between its β value and effect size and the rather high *p*-value. This relationship is one that we will consider for future investigation. Of note, the Eval-PoP relationship was significant though not particularly powerful. This is the relationship that is supported by the PJCC framework.

While it is not as strong as the PrK relationship between PoK and DPoK in their respective models, we are confident that further refinement of our scaffolds can improve the strength and importance of this relationship.

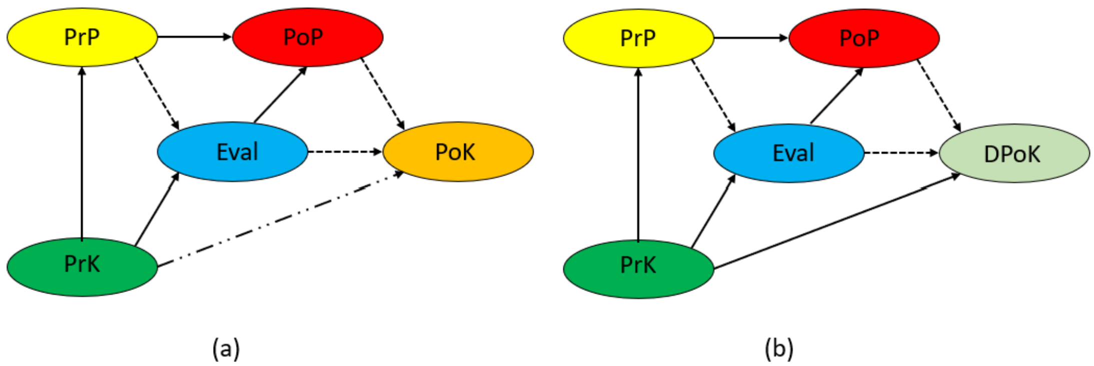


Figure 2. Structural Equation Modeling for Elementary MEL Diagram (a) Post-instruction Knowledge and (b) Delayed Post-Instruction Knowledge. See Table 3 for relationship values.

When looking at the structural equation models on the whole, we do find that this activity is currently driven heavily by prior knowledge about fossils. We also found in our data that we may have over-scaffolded the explanatory models. The "fossils help" and "fossils don't help" may have oversimplified the choices for the students and that is reflected in their plausibility ratings. As this is a pilot study, we allow that instrument refinement will be necessary. We also recognize that this study leans heavily on white perspectives, not just the participants but also the classroom teachers and the research team. We will seek to diversify our project in all three categories.

Contribution to Science Education

This project contributes to elementary science education by providing elementary science students with developmentally appropriate science readings that are integrated with scaffolding for critical evaluation of the relationship between explanatory models and lines of scientific evidence. The eMEL provides students with yet another mode through which students may develop their critical thinking skills and scientific literacy of which they need multiple and diverse opportunities to do so (Marques Vieira & Tenreiro-Vieira, 2016).

Impact For NARST Membership

This pilot study provides a potential roadmap for science education researchers to reimagine and reformat successful secondary science scaffolds into ones that are engaging and potentially highly instructional for elementary students. By creating an interactive scaffold with an expository fictional story book, our project provided elementary science students with a developmentally accessible lesson that activated their critical evaluation of scientific models and evidence and provided the opportunity to learn deeply about fossils.

References Cited

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.

Braasch, J. L., & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes*, *47*(6), 447-479.

Chinn, C. A., & Buckland, L. A. (2012). Model-based instruction: Fostering change in evolutionary conceptions and in epistemic practices. In K. S. Rosengren, E. M. Evans, S. Brem, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 211–232). Oxford University Press.

Crowell, A. & Kuhn, D. (2014). Developing dialogic argumentation: A 3-year intervention study. *Journal of Cognition and Development*, *15*(2), 363–381. https://doi.org/10.1080/15248372.2012.725187

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

Ford, M. J. (2015). Educational implications of choosing "practice" to describe science in the next generation science standards. *Science Education*, *99*(6), 1041-148. https://doi.org/10.1002/sce.21188

Gans, N., Klavon, T.G., Haugland, S.S., & Schwiesow, M. (2023). A discovery with Ami Ammonite: Scaffolding elementary students' scientific evaluations. Presentation, at the 2023 National Consortium for Instruction and Cognition Annual Conference, Chicago, IL. (refereed)

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*, 1-15. https://doi.org/10.1080/10899995.2023.2200877

Klosterman, M. L., & Sadler, T. D. (2010). Multi-level assessment of scientific content knowledge gains associated with socioscientific issues-based instruction. *International Journal of Science Education, 32*(8), 1017-1043.

Kock, N. (2022). *WarpPLS User Manual: Version 8.0*. ScriptWarp Systems.

Lombardi, D. (2016). Beyond the controversy: Instructional scaffolds to promote critical evaluation and understanding of Earth science. *The Earth Scientist*, *32*(2), 5–10.

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.

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. (2016) 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 Instructions*, *27*, 50-62.

Marques Vieira, R. & Tenreiro-Vieira, C. (2016) Fostering scientific literacy and critical thinking in elementary science education. *International Journal of Science and Mathematics Education*, *14,* 659-680. https://doi.org/10.1007/s10763-014-9605-2

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

National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press.

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