

USING EVIDENCE CARDS TO IMPROVE EXPLANATIONS IN MIDDLE
SCHOOL CLAIM-EVIDENCE-REASONING (CER) STATEMENTS

by

Benjamin Andrew Pope

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2026

©COPYRIGHT

by

Benjamin Andrew Pope

2006

All Rights Reserved

ACKNOWLEDGEMENTS

First and foremost, I want to acknowledge the patient understanding and unwavering support of my wife Keren, and our four children, Ephraim, Mathias, Lillian, and Julliette. All the many hours that my attention was pulled away towards research and writing were hours given to me by the five of you. There was not a moment these past two years when I was on my own. This accomplishment is as much yours as it is mine.

I also wish to acknowledge the dedicated science teachers and supportive administration of Bemidji Middle School and the shared desire you all have for continual improvement. I particularly want to thank Dana Smith, who first introduced me to Three-Dimensional science teaching six years ago, and has ever since patiently provided support as I learn. Dana, you have answered every question, enthusiastically provided feedback, and served as a leader in science education among us all.

Finally, I am grateful to Montana State University and the MSSE program for providing a flexible, truly meaningful graduate degree. To all my professors, as well as my Science Reader Kenneth Taylor, I thank you for the constructive feedback and direction that made the once seemingly impossible task of this paper possible. To my fellow MSSE students, I am grateful for the camaraderie and collaborative spirit in our discussions and shared experiences. Your questions, ideas, and mutual support made the challenges of this program more manageable and often inspiring.

Thank you all.

TABLE OF CONTENTS

1. INTRODUCTION AND BACKGROUND.....	1
Context of the Study	1
Focus Statement/Questions	2
2. CONCEPTUAL FRAMEWORK	4
Direction for the Study.....	4
Theoretical Framework	7
Methodologies.....	8
Classroom Articles	10
3. METHODOLOGY	12
Demographics	12
Treatment	13
Data Collection and Analysis Strategies	19
Attitude Surveys.....	21
Evidence Surveys.....	21
CER Rubrics	21
4. DATA ANALYSIS: CLAIMS AND EVIDENCE.....	23
Approach.....	23
Results.....	24
Evidence Survey and CQ Responses	31
5. VALUE: REASONING AND REFLECTION	35
Claims From the Study	35
Middle School Students Can Reason about Evidence	35
Evidence Cards Provide a way for Students to think about	
Evidence More Concretely	36
Critical Questions Improved Student Ability to Assess the	
Quality of Evidence	37
Benefits for Formative Assessments	38
Effect of Evidence Cards on the Teacher	39
Value of the Study and Consideration for Future Research	40
Impact of Action Research on the Author	42
REFERENCES CITED.....	43
APPENDICES	46

APPENDIX A: IRB EXEMPTION	47
APPENDIX B: PRE-TREATMENT EVIDENCE PACKET	49
APPENDIX C: EVIDENCE GRADIENT TOOL	51
APPENDIX D: ATTITUDE SURVEY	53
APPENDIX E: EVIDENCE SURVEY.....	55
APPENDIX F: CER RUBRIC.....	57

LIST OF TABLES

Table	Page
1. Table 1. Data Collection Matrix.....	20
2. Table 2. Attitude Survey Likert Items by topic with comparison of a results from pre-treatment ($N=63$) and post-treatment ($N=60$)	25
3. Table 3. Chi-Square test for statistical significance ($N=123$), $df=4$, $p<.001$	28
4. Table 4. Comparison of Good / Specific verses Vague / Non-Specific explanations in the pre- treatment ($N=65$) and post-treatment ($N=60$) Evidence Surveys.....	32
5. Table 5. Depth of responses to CQs on evidence cards ($N=57$) compared to post-treatment CER scores.....	33

LIST OF FIGURES

Figure	Page
1. Figure 1. Standard example of a CER template.....	1
2. Figure 2. Example of an Evidence Card from my post-treatment CER assignment.....	3
3. Figure 3. Students sorting evidence on a gradient during the research treatment	10
4. Figure 4. Note catcher for planning CERs during evidence card sorts.....	11
5. Figure 5. Student ethnicities from combined three-class sample, ($N=65$).....	12
6. Figure 6. Introductory Evidence Cards without CQs.....	14
7. Figure 7. Sample Evidence Card “A” with pre-completed CQs.....	16
8. Figure 8. Box plot comparison of pre-treatment and post-treatment CER scores ($N=61$).....	24
9. Figure 9. Pre-treatment Attitude Survey percentages by item ($N=63$).....	27
10. Figure 10. Post-treatment Attitude Survey percentages by item ($N=60$).....	28

ABSTRACT

One of the most difficult tasks for middle school science students is writing Claim Evidence Reasoning statements in response to scientific questions. The difficulty lies in the reasoning portion of these statements, in which students employ the science practices of Constructing Explanations and Arguing from Evidence to demonstrate how evidence supports their claim. Many young students fail to grasp that it is not enough to simply say that a claim is correct or that the evidence is strong. They must also communicate how the reliability, pertinence, and strength of the evidence make it adequate for the task of supporting the claim. I sought to address this problem by introducing evidence cards as a tool to be used prior to writing these Claim Evidence Reasoning statements. Each card features a piece of evidence gathered during classroom coursework. Alongside the evidence is one or more Critical Questions that prompt students to consider the meaning and value provided by the evidence. Students answer these questions with a partner and then sort the cards in order from weakest evidence to strongest. In a sample of 61 sixth grade students, scores on the reasoning portion of CER statements increased from a mean of 5.38 out of 12 in the pre-treatment to a mean of 7.42 in the post-treatment. The perceived ability and readiness of students to use evidence also saw gains when pre-treatment and post-treatment survey responses were compared. This was demonstrated by the portion of students assessing themselves in the highest two levels of mastery increasing by 23.7%. These results show that using evidence cards with critical questions resulted in a significant improvement in the reasoning section of these middle school students' Claim Evidence Reasoning statements.

CHAPTER ONE

INTRODUCTION AND BACKGROUND

Context of the Study

I teach Earth and Space science to approximately 110 sixth graders at Bemidji Middle School in northern Minnesota. As I converse with other science teachers in my own school and father abroad, I repeatedly hear the same sentiment being voiced – that science students struggle with the Reasoning portion of Claim Evidence Reasoning (CER) statements. CERs are meant to be templates that aid students in writing scientific explanations and arguments by breaking the statement down into the separate parts of claim, evidence, and reasoning (see Figure 1).

Name: _____		Hour: _____	
Driving Question: Are there volcanoes at Location 2?			
Claim Answers the question in a complete sentence.			
Evidence #1 1 st piece of data used to support the claim			
Reasoning #1 Explanation connecting the evidence to the claim, so as to explain the significance of the evidence.			
Evidence #2 2 nd piece of data used to support the claim			
Reasoning #2 Explanation connecting the evidence to the claim, so as to explain the significance of the evidence.			

Figure 1. Standard example of a CER template

While presenting evidence can itself be difficult for students and depends upon their ability to analyze and interpret the data gathered in course work, the substantial disconnect for middle school students exists in communicating the reasons why they interpret the evidence as being in support of their claims. The dozens of middle and high school science teachers I discussed CERs with at the 2025 NSTA conference in Minneapolis almost universally testified that the reasoning portion of CERs proves the most difficult for students in their classrooms. Any method of instruction that effectually improves the reasoning performance of students in written CERs would provide valuable benefit to myself and other teachers looking for solutions.

Focus Statement/Questions

Although I have attempted a variety of lessons and approaches in hopes of improving student reasoning in CERs, no definite innovations had thus far been discovered. That being said, several years ago I had a brief but positive experience demoing a science curriculum from Amplify that made use of scaffolding tools known as Evidence Cards to facilitate oral debates known as science seminars (see Figure 2). I also happened across mention of similar Evidence Cards in educational literature. Since these cards have been used as tools by other science teachers to facilitate the consideration and deliberation of evidence, I chose to make the use of these Evidence Cards the focus of my action research project.

Evidence Card D: The Activity Status of this Volcano is _____					
CQ #1: What does this activity status mean? 					
CQ #2: Where did you get this information from? 					
CQ #3: What makes this source reliable? 					
Evidence Rank:	1 (weaker)	2	3	4	5 (stronger)

Figure 2. Example of an Evidence Card from my post-treatment CER assignment

My primary research question was: How will the use of evidence cards affect the quality of students' explanations in the reasoning portion of their written CERs?

My sub-questions include the following:

1. What effect will student use of evidence cards have on their perceived readiness to evaluate and use evidence?
2. How will evidence cards incorporating critical questions impact students' ability to assess the quality of each piece of evidence?
3. What effect will the use of evidence cards have on my role as the teacher?

CHAPTER TWO

CONCEPTUAL FRAMEWORK

Direction for the Study

To provide direction for my study, I began with an examination of available literature. The first question requiring an answer was whether middle school aged minds are developmentally capable of thinking abstractly enough to reason about evidence. To be sure, the extent of the problem that students have with reasoning statements in CERs is significant enough that, in the past, I heard multiple educators chalk it all up to a limitation in cognitive development. One educator told me that 6th grade students are only capable of the more abbreviated Claim and Evidence (CE) statements, and that the reasoning section requires a potential that does not develop in kids until 7th grade at the earliest. If a definite limit to increasing the reasoning practices of my 6th grade students exists, then I needed to know what that point of diminishing returns was before extending substantial effort on attempts to improve performance.

My review of available research quickly revealed that increasing levels of complex reasoning are certainly possible in students even younger than middle school. In one professional study of 135 children divided equally in grades 3 through 5, the percentage of students using complex reasoning in problem situations was able to increase from 22% in 3rd grade to 51% in 5th grade (Sickler et al., 2024). One of the criteria used to discern whether these students engaged in complex reasoning was whether they “connected the dots between the evidence sources and their choices” (p. 1744). To ensure that no other variables other than age and normal cognitive

development could be considered responsible for these results, the study took careful precautions to eliminate other potential barriers that require scaffolding. The problems were intentionally ‘non-mathematised,’ meaning that mathematical literacy would not be a limiting factor for younger students. The study also ensured that the problems were in an accessible context, meaning they would include highly familiar and observable phenomena for all students in the study. In this case, the phenomenon and questions pertained to images of empty drinking glasses filling with varying amounts of water when placed in different locations during varying intensities and lengths of rainfall, which was something within the realm of what all students had experienced. Finally, the phenomenon was designed so that the questions were complex and without a single certain solution. This was done to elicit demonstrations of complex reasoning for students in each grade. Although grasping the scenario and question was simple and required no mathematical literacy, students needed to be able to “reason through complex variables, conflicting information, limitations of the data source or collection method, and draw interpretations” (p. 1739). With other barriers out of the way, any differences in reasoning ability could more easily be linked to age related cognitive development and problem-solving experience.

The study suggests that when barriers are removed, students in each of these three elementary grades are all capable of reasoning in complex ways involving critical thinking. And although the older 5th grade students performed notably better in their use of critical thinking, the results clearly demonstrate that quality reasoning is possible even at younger ages when certain pieces of scaffolding are in place. This provided direction for my study by demonstrating that

alterable approaches to instruction, despite younger ages, can be the determining factor affecting how well students can engage in complex reasoning about evidence.

More evidence that intentional dedicated instruction can improve the reasoning performance of middle school students comes from research done by Kuhn, (2010). After heavily emphasizing the integration of scientific argument in the classroom, Kuhn found that students improved their use of reasoning in both classroom discussion and in writing tasks such as CERs. In a post-survey, students in Kuhn's study responded to what they had learned through dedicated argumentation instruction with statements such as "we should think a little more before we respond so that we can have better constructed responses" and "I've learned to make strong arguments and be sure to use evidence" (p. 819). This suggests that with dedicated instruction student ability to discern differences in evidence or argument quality can improve. And although the focus of Kuhn's program was improving oral argumentative discussions, it was suggested that argumentative writing such as CERs can also be improved "when the writer accesses an internal argumentative discourse with an imagined interlocutor who the writer then attempts to convince" (p. 819). This suggests that a lack of thinking about evidence from a mindset of argumentation may be another potentially removable barrier to complex reasoning in my classroom. So then, complex reasoning is possible for upper elementary and early middle school students with the correct scaffolding, and students can improve in their ability to discern the quality of evidence and reasoning when thinking in terms of argument. The next step was to investigate theoretical framework for engaging in students in scientific arguments, to find whether this and the use of some sort of scaffolding could yield promising results in my action research project.

Theoretical Framework

Depending upon what a claim is responding to, CERs either lend themselves more to Science and Engineering Practice (SEP) #6 – Constructing Explanations, or #7 - Arguing from Evidence. The practice of Constructing Explanations uses the CER to explain the facets of an observed phenomenon or model, while the practice of Arguing from Evidence employs CERs to state or respond to a contested interpretation of data (Bell & Shouse, 2014). According to framework from NGSS (2013), Middle school learners should develop in the practice of constructing explanations “based on valid and reliable evidence,” in which they also should “apply scientific reasoning to show why the data or evidence is adequate for the conclusion” (p. 11). When it comes to Arguing from Evidence, middle schoolers should be able to develop a “written argument supported by empirical evidence and scientific reason” which includes “citing relevant evidence, and posing or responding to questions that elicit pertinent elaboration and detail” (p. 13). Putting those together, I have come to hold the expectation that the reasoning portion of a middle schooler’s CER statement should always reference how the chosen evidence is both reliable and relevant for the task of supporting a claim. During a presentation at the NSTA 2025 Minneapolis National Conference, Dr. Robert Vossel, the K-12 Science and STEM Coordinator at Waukegan Community Unit School District, built upon what relevance in reasoning means by suggesting that student reasoning sections should connect the evidence to the claim using scientific definitions and concepts (personal communication, November 2025). On numerous occasions I have heard other educators suggest the reasoning section of CERs be a reference to scientific laws and principles that bring meaning and validity to the evidence. While I would agree that this is the pattern seen in professional scientific journals and therefore the end

goal for science educators, the fact is that many middle school students do not have a collection of scientific laws and principles ready in their arsenal for reference. Instead, in 6th grade terms, I expect the reasoning section of CERs to explain what the student knows about the significance of the evidence and how it connects to and supports their claim. Just as in the scientific community, student claims and reasonings should be “evaluated and critiqued in light of their ability to explain nature” (Ford, 2015, p. 1045). Furthermore, students should make some mention of the strength of their evidence and its ability to support the claim just as when “scientists consider alternative interpretations of the same observations, they argue to identify weaknesses in various explanations” (Reiser et al., 2012). Even if formal debates about competing claims are rarely facilitated in a 6th grade science classroom, the practice of explaining or arguing the strength of the data-claim connection is a vital part of science. So, although the collection of natural laws familiar to my 6th grade students are extremely limited, I can still evaluate each student’s ability to recall the concepts covered in coursework for use in explaining how a piece of evidence adequately supports their claims. Let us look now at examples of research methodologies that attempted to enhance this element of CERs.

Methodologies

CERs, although popular in science classrooms today, often do not portray scientific argumentation and explanation as it is naturally used in the scientific community, where counterarguments, rebuttals, and detailed defenses of why a particular explanations were reached are as valued a part of communication as the claim itself. Recognizing many of the same struggles that I have in the quality of CER completions, Nussbaum et al. (2024) conducted a study with the aim of improving student reasoning, stating that “students should provide a

rationale for their evidence and grapple with counterarguments and alternative explanations versus simply providing evidence and support for their claims” (p. 1421). Several alternatives or addendums to CERs were considered by the study before detailing a plan of using Critical Questions (CQs) as an accessible prompt for students to “assess the strength and cogency of an argument” (p.1424). The CQs are question prompts the student can consider when preparing for either oral argument or a written CER. Although a nearly infinite number of unique CQs could be generated and used depending on the age of students and context of the argument, the study narrowed the prompts down to nine questions which would guide students to consider things such as the structure of their argument, the reliability in the data’s source, the meaning behind terms being used, etc. An example of a CQ directed at evaluating the meaning of terms in the argument might be worded as, “can I (or you) explain the meaning [of terms/concepts in the study]” (p. 1425). Such questions, although simple, slow a student down and guide them to consider whether they really understand what they are arguing. For example, if a student is trying to figure out whether an extrusive igneous rock sample is evidence of volcanic activity, that will be impossible if they do not understand the term ‘extrusive igneous.’ If a student cannot adequately answer a CQ, “this lowers the strength and cogency of an argument” (p. 1425). The unanswerable CQ alerts the student to the fact that they are not yet ready to evaluate the evidence, let alone use it in argument. In short, the CQ reveals that the student should spend more time with other science practices such as Analyzing and Interpreting Data before they proceed to construct explanations or arguing from evidence. I chose to include CQs on the evidence cards in my action research to ensure that students have an adequate understanding of

concepts related to the phenomenon and the significance of the evidence before they attempt to use the information in written CER statements.

Classroom Articles

I wish to finish my literature review with an examination of how these methodologies can look in a 6th grade classroom using evidence cards. González et al. (2018) provides a glimpse of how a middle school teacher used evidence card sorts to remedy a problem similar to what I have faced, where students find it challenging to express “why evidence does or does not support a claim” (p. 74). This teacher used evidence card sorts and an introductory lesson developed by a partnership of Boston College and the Lawrence Hall of Science in a project known as ‘The Argumentation Toolkit.’ In the lesson, students were given separate cards for each piece of evidence, which they sorted and placed with peers along a gradient sheet developed for distinguishing between weak and strong evidence (Figure 3). “The physical act of moving cards provides students with a way to consider each piece of evidence in relation to others [...] helping them differentiate between evidence that is of high or low quality” (González et al., 2018).

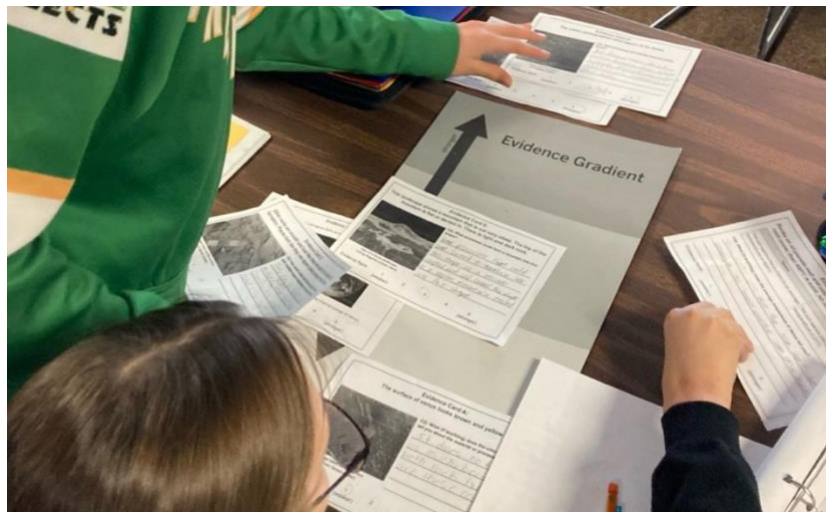


Figure 3. Students sorting evidence on a gradient during the research treatment

Of course, while this activity can help students simply by increasing awareness of the idea that not all evidence is equivalent in value, what constitutes ‘high’ or ‘low’ quality evidence is difficult to judge objectively, especially in 6th grade. Discernment of what is ‘strong’ or ‘decisive’ evidence requires significant subject matter knowledge (Sickler et al., 2024). For example, students attempting to make claims about whether rock samples are evidence of volcanic activity are limited by their understanding of how various types of rocks form and by their ability to identify the rock samples. Once again, this is a reason I chose to augment the evidence cards with critical questions. The CQs prompt students to fully understand the evidence before evaluating how well it supports the claim. One 6th grade student mentioned that “the gradient tool was pretty easy to understand, but you had to have, um, more specific evidence, or else you didn’t know exactly where to put it” (The Lawrence Hall of Science, n.d., 5:20). Including CQ prompts about the reliability or meaning of the evidence on each card enables student partner groups to better sort of evidence based upon strengths. Using a simple note catcher (see Figure 4) in conjunction with the evidence cards preserves a written record of their evaluations after the card sort is over. Students then review written responses to the CQs and note catchers later when writing CERs to ensure memory of the evidence’s quality and meaning.

Name: _____	Hour: _____
• Which two evidence cards will best support your claim? _____ and _____ • Are there any other cards / pieces of evidence you will mention in your reasoning? • Are there any pages in our binders with information you should use while writing the Reasoning of your CER? • What other information (if any) do you want to remember to include in your CER?	

Figure 4. Note catcher for planning CERs during evidence card sorts

CHAPTER THREE

METHODOLOGY

Demographics

Bemidji Middle School has an approximate total student population of 1,035 sixth through eighth grade students, 47% of whom are considered economically disadvantaged (U.S. News & World Report, n.d.). I conducted my research in three 6th grade Earth and Space science classes, with a combined total of 65 students who consented to take part in the study. This sample included 29 boys and 36 girls. Ten students in this group had Individualized Education Programs (IEPs), which provide specialized instruction and related services for students with identified disabilities. A breakdown of student ethnicities gathered from the school's student information system is shown in Figure 5. This project was granted an exempt status by Montana State University's Institutional Review Board, and all applicable requirements for research involving human participants was followed throughout. (see Appendix A).

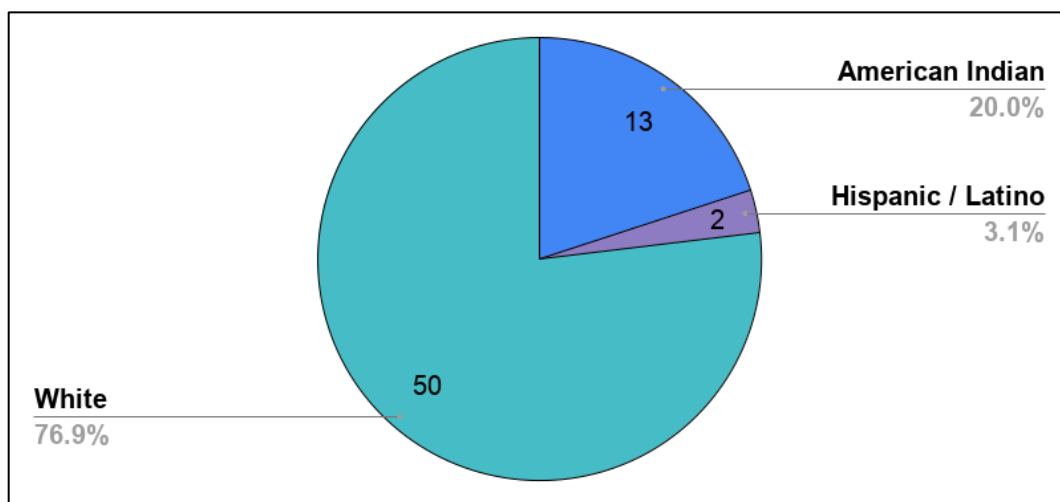


Figure 5. Student ethnicities from combined three-class sample, ($N=65$)

Treatment

The purpose of this study was to determine whether the use of physical evidence cards might affect the ability of 6th grade students to reason about the evidence they present in written CER statements. Although the customary practice in my classroom was for students to assemble data in their science binders for use as a resource to review during CER writing, the use of evidence cards was new for this cohort. Treatment, which involved introduction to evidence cards and instruction in their use, began in the middle of my two-month long volcano / rock cycle unit, immediately after collecting initial data from pre-treatment surveys and a CER.

Introduction of Evidence Cards

In a whole-class setting I explained the concept of evidence cards, relating them to the nearest equivalent we had used up until then, which was to have a collection of data in our science binders. For the past few years, I had also given students a page on which to answer questions and thereby summarize the data prior to writing their CERs (see Appendix B). I explained to students that a benefit of having each piece of evidence on its own card is that it allows us to tangibly think in terms of “this is what I have to support my claim. It is all in front of us. I can spread the evidence out, consider whether I understand each piece, talk about it with my classmates, and even sort it from weakest to strongest.” I then gave each table group of two students evidence cards that went along with the context of the recently completed pre-treatment CER, which had involved answering whether evidence of volcanic activity could be seen in a particular location. Although the students had previously analyzed data from four locations, the pre-treatment CER prompt only had asked about the existence of volcanic activity at location one. The evidence cards I handed out pertained to the second of the four locations, making it data

and concepts that the students were all very familiar with; but they now had to interpret evidence from a different location, with different rock types, tectonic boundaries, and landforms.

This first set of evidence cards (see Figure 6) were very simplified and did not include CQs for students to respond to. The reason for this was twofold. First, I wanted to begin by only introducing the concept of evidence cards without the additional confusion of CQs. Second, I wanted to work towards emphasizing the value of CQs. The pre-cut cards were given to students along with an evidence gradient sheet printed on a 12x18 inch glossy cardstock with a gradient of grayscale colors ranging from light at the bottom to dark at the top, similar to the one in the Argumentation Toolkit offered by The Regents of the University of California (Appendix C). On the Evidence gradient was a double-sided arrow pointing up and down, with simple instructions to place weaker evidence cards towards the bottom and stronger evidence towards the top.

Evidence Card A: The rocks from location 2 are Andesite	Evidence Card B: Location 2 is an Island, surrounded by water
Evidence Card C: Photos show a cone-shaped dark-colored mountain with a light-grey peak	Evidence Card D: The map shows that Location 2 is on a convergent boundary between the Australian and Pacific tectonic plates.
Evidence Card E: Bougainville Island is located perfectly on the "Ring of Fire"	Evidence Card F: Photos do not show any snow on top of the mountain

Figure 6. Introductory Evidence Cards without CQs

I told students to discuss the evidence cards with their partners and make these assessments as if they were preparing to write another CER answering whether evidence of volcanic activity could be identified at the second location. I walked around as students engaged in discussion. Most spent a few minutes with highly engaged discussion about which cards deserved to be placed higher on the gradient. A few groups quickly compromised and placed cards without much discussion until I encouraged them to take the time to place the cards where they truly felt each belonged based upon how much the evidence could help their argument. I explained that the cards they placed at the top would likely be the pieces of evidence used to write CERs to explain whether the given location had volcanic activity, so they should only place the best evidence near the top of the gradient.

I made sure to walk around the room during the card sort and check in with each of the groups once they had made their final placements on the gradient sheet. I then returned to the front and noted to the class how it was interesting that not all groups made the same placement. I shared my observation that the cards some groups proposed as being the strongest evidence were determined by other groups to be the weakest evidence. I asked students why that might be. In each of the three classes, the answer eventually came out that perhaps these differences in evaluation were because not all groups shared the same level of understanding about the concepts referred to in the cards. For example, not all students realized the significance of a rock being Andesite, making Card A of little value. Others remembered that Andesite was extrusive igneous and that extrusive igneous rocks form from volcanoes. This provided the perfect opportunity to introduce the real evidence cards with critical questions.

These new evidence cards had the same information as the simplified ones, but with the addition of critical questions and a scale on the bottom for students to use in ranking the evidence from 1 (weaker) to 5 (stronger). For the sake of providing students with examples of ideal-length responses, I pre-answered each of the CQs on these cards, but explained that for future CERs the students would need to use their binders and work with a partner to answer the critical questions. The CQs included had to be unique for each card, so I came up with them by contemplating what students would need to know about the evidence to handle it accurately when writing a CER statement. For example, Evidence Card A (Figure 7) included CQs asking about the significance of rocks being Andesite, a term not all students comprehended with the initial sort of the simplified cards.

Evidence Card A: The rocks from location 2 are Andesite					
CQ #1: How can you be sure the rock from location 2 is Andesite ?					
We looked carefully with hand lenses to check that it matched the description of Andesite, which is a gray scratch-resistant rock with white or black glassy flecks.					
CQ #2: Which of our rock type vocab terms describes Andesite ?					
Extrusive Igneous					
CQ #3: Why is that vocab term significant? (why does it matter?)					
Extrusive igneous rocks only form from volcanic activity. They cool quickly on the surface, which is why their background color has such a small grain size.					
Evidence Rank:	1 (weaker)	2	3	4	5 (stronger)

Figure 7. Sample Evidence Card “A” with pre-completed CQs

Partner groups then sorted these new cards without removing the original simplified cards from the gradient. This allowed students to see how their assessments of the strength of each piece of evidence were often altered by the additional information from answered Critical questions. While the simplified Card A had been placed by many groups on the weaker end of the gradient sheet, the newer evidence card A with completed Critical Questions was more often placed near the top as a “stronger” piece of evidence. The revaluation was dramatically displayed as both sets of cards sat beside the gradient sheet together. The CQs on the new cards ensured that students understood what the evidence on the card meant, which made all the difference. I then pointed out how the answers to the CQs on Card A came directly from information they all had in their science binders from an earlier lesson. I informed them that in the future; they would need to use the information collected in their science binders to answer CQs before sorting them and using them to write a CER. If the pages in their binder were incomplete, they would need to return to the original sources of information, such as scientific texts we read in class. This scenario is one I saw play out multiple times during treatment and post-treatment CQ answering. I should note that for the remainder of treatment, these partner groups worked together so long as both students were in attendance each day. After sorting was over and I had checked in with each group to hear the rational for their placements of each card, we came back as a whole class to discuss the process, again making note of how the ranking of evidence on the gradient was strongly altered in some instances because of the addition of CQs, making these something to take seriously. I pointed out to students how the word “critical” in critical questions meant that these could play a decisive role in whether they succeed or fail in their handling of the evidence during CER writing.

Practice Treatment CER

The introduction to Evidence cards had taken only a single class period. The next day I presented students with a new set of evidence cards pertaining to the question of whether the rock forming processes on Venus are forming primarily igneous or sedimentary rocks. Years ago, I taught a science unit from Amplify that used science cards on this topic to prepare students for an oral discussion referred to as a science seminar. I adapted these cards to better fit our unit and added critical questions. This nicely extended the material covered in our volcano unit. The concepts of rock-forming processes remained the same; but none of the evidence on these new cards had previously been considered in the classroom. With this new question, student partner groups worked more independently as they answered the CQs and sorted the evidence cards based upon their understanding of the reliability and relevance of the data. I observed while periodically stepping in to provide necessary feedback on how students discussed and evaluated the evidence. This practice sort provided me with an excellent opportunity to formatively assess student understanding and use of the cards. In one instance, I discovered that most of my students held an unforeseen misconception about one of the evidence cards. In response, I called the class back together and gave them an additional CQ to write on the back of that card. They answered the new CQ using information in their science binders, which helped them resolve the misconception. With smaller or more isolated misunderstandings, I simply provided necessary feedback and guiding questions to individual groups during the sort. Finally, the students used these cards to complete a second CER. Once again, I provided feedback to both individual groups. I checked in with each group before giving them the note catcher from figure 4, which I referred to as a “shopping list” for what they would want to include in their CERs. Quite a few groups tried to rush through the process and complete the CER without spending adequate time

answering the CQs or sorting the evidence cards. I discovered it was necessary to hold students back from proceeding with CER writing until they could demonstrate they had adequately thought through the evidence and CQs. This became a practice I continued post treatment. I did not give students their gradient sheets until they had answered all CQs. Likewise, I did not provide them with the note catcher until I had checked their sorting of the cards. Lastly, I did not allow students to begin writing their CER until they had completed their note catchers. This process of checks meant that some students proceeded much more quickly than others. Some finished writing their CERs before the more off-task students had even finished answering their CQs. Even the quickest students took a full 60-minute class period to answer the CQs. Sorting the cards and writing the CERs then took another full class period. After two full class periods, all but a few absent students had at least finished answering the CQs with their partners. A few then had to finish writing their CERs as homework, which proved somewhat awkward because they had shared evidence cards with their partner. This led to me adopting the practice post treatment of each student owning their own set of cards, which they would still discuss, answer CQs on, and sort with a partner, but would then have to themselves during the independent writing stage. Once all CERs were in, the final step of treatment was to discuss how it went as a class, and to inform students that in the coming weeks they would be writing another CER with the use of new evidence cards. Aside from each student having their own set of evidence cards, students completed this final CER in the same way they had completed the practice one. The results from these final CERs served as post-treatment data for my research. In total, the in-classroom treatment required four full 60-minute class periods.

Data Collection and Analysis Strategies

To investigate the impact of evidence cards on students' CER reasoning, I used three primary data collection methods, which were each employed both before and after treatment: Attitude Surveys, Evidence Surveys, and a CER Rubric. Additionally, my own teacher journal and the student-completed evidence cards for the final post-treatment CER were both used to gather qualitative data. These methods are detailed below along with how they each align with my research questions. See Table 1 for an overview of each instrument.

Table 1. Data Collection Matrix

	Primary Research Question “How will use of Evidence Cards affect the quality of students’ explanations in the reasoning portion of their written CERs?”	Subquestion #1 What effect will student use of evidence cards have on their perceived readiness to evaluate and use evidence?	Subquestion #2 How will evidence cards incorporating critical questions impact students’ ability to assess the quality of each piece of evidence?	Subquestion #3 What effect will the use of evidence cards have on my role as a teacher?
Initial Attitude Survey		X		
Pre-treatment CER Rubric	X			
Pre-treatment Evidence Survey			X	
Post-treatment Evidence Survey			X	
Post-treatment CER Rubric	X			
Final Attitude Survey		X		
Completed post-treatment Evidence Cards	X		X	
Teacher Journal	X	X	X	X

Attitude Surveys

What effect will student use of evidence cards have on their perceived readiness to evaluate and use evidence? Prior to beginning treatment, I collected initial data for this first subquestion using a survey combining ten 5-point Likert items and four long-response prompts (See Appendix D). Students responded to each of the ten Likert item by selecting the statement that best describes their current estimation of themselves, ranging from (1) “I’m just starting,” to (5) “I’ve Mastered this.” The lowest possible value for all Likert items is 10, and the highest possible value is 50. This allowed me to quantify responses comparing attitudes pre-treatment and post-treatment using split bar graphs. The four long-response questions on the survey provided further qualitative insight into student selections.

Evidence Surveys

Both pre- and post-treatment, I administered these surveys after students had selected two pieces of evidence to use in their CERs but just prior to beginning the actual CER writing. The surveys (see Appendix E) included four questions for each piece of evidence, asking students why they chose the evidence they did and how ready they felt to write about the quality of that evidence. The qualitative data gathered by these surveys helped provide answers for Subquestion #2, by revealing whether the thought processes behind evidence choice changed significantly after students were introduced to evidence cards and began using them.

CER Rubrics

I used a rubric to score the reasoning section of students’ pre- and post-treatment CER responses (See Appendix F). This data provided most my study’s quantitative data, answering the

primary research question of how using evidence cards will affect the quality of students' explanations in the reasoning portion of their written CERs. My CER rubric was shaped in part by the framework of an example rubric from an article on evidence-based writing strategies (Esparza, 2024). My rubric was also influenced by the Reasoning Rubric from the Argumentation Toolkit (Argumentation Toolkit, n.d.), which scores student reasoning based upon its ability to explain the pertinent scientific concept and how it relates to the claim. Additionally, I sought to ensure the validity of my rubric by using feedback from professors and colleagues in the MSSE program, as well as insight from an experienced colleague at my school who has taught the same volcano and rock cycle units.

In my rubric, students were scored by two metrics. First, the rubric assessed how well student reasoning connected the evidence to the claim and offered support. This metric ranged from the reasoning contradicting or explaining no connection up to the reasoning explaining in detail how the claim is better supported by the evidence than other potential competing claims. Secondly, the rubric assessed student reasoning based on the detail and accuracy with which students explained the pertinent concepts to the data. For example, if the evidence stated that the rocks from a location were Andesite, students would need to expound upon the meaning of that term in enough detail to demonstrate understanding of the evidence's ability to support the claim. The maximum score for the rubric was 12, with 6 points possible for each evidence's reasoning. The minimum score possible was 0.

CHAPTER FOUR

DATA ANALYSIS: CLAIMS AND EVIDENCE

Approach

I used a mixed-methods approach to analyze and interpret the qualitative and quantitative data for this research. I used the same rubric to score both the before and after treatment CERs in order to answer my primary question of how evidence cards might affect the quality of student explanations in the reasoning portion of their CERs. Pre- and post-treatment attitude surveys measured students' perceived readiness and proficiency with selecting and using evidence. Pre- and post-treatment Evidence Surveys gathered insight into the basis of student evidence selection. These, when combined with an analysis of the completed post-treatment evidence cards, provided insight into how the incorporation of critical question on the evidence cards impacted students' ability to assess the quality of each piece of evidence. Finally, the notes added to my teacher journal throughout the entire process provided further clarification and insight to accompany each of these methods, as well as allowing an answer to the question of what effect the use of evidence cards had on my role as the teacher.

Although I focus in some sections on the findings from a particular data collection method, my overall aim is to present the data in a way which accurately responds to the research questions. I therefore interspersed data from all collection methods throughout this chapter as necessary to demonstrate a comparison of where my students were pre- and post-treatment. I reserve further interpretation of these values for Chapter 5.

Results

There was a significant improvement in the reasoning ability of students when using evidence cards. The reasoning section of students CER statements were graded using the same rubric pre- and post-treatment, with a maximum possible score of 12 and a minimum possible score of 0. The results from the students who completed both CERs ($N=61$) increased from a pre-treatment mean of 5.38 to a post-treatment mean of 7.42 (see Figure 8).

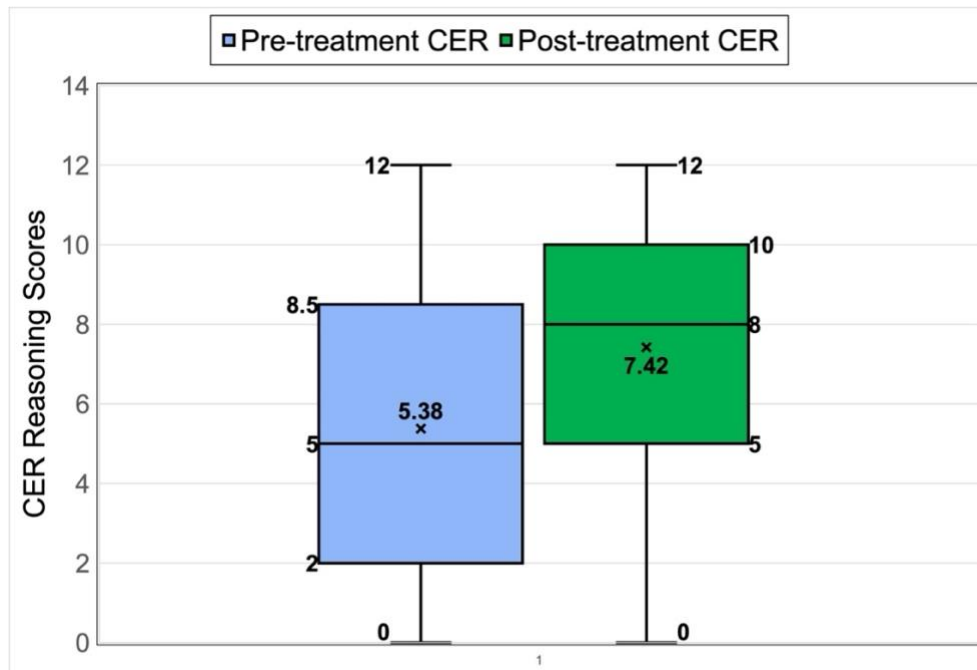


Figure 8. Box plot comparison of pre-treatment and post-treatment CER scores ($N=61$)

All but the lowest scoring 25% of post-treatment CERs scored at or above the median score of 5 from the pre-treatment CER. Mean scores rose from 5.38 ($SD = 3.73$) to 7.42 ($SD = 3.21$), which a paired-samples t-test concluded to be significant ($t(60) = 5.29$, $p < .001$).

Because $p < .001$, which was less than the alpha level of .05, the null hypothesis was rejected.

The calculated effect size (Cohen's $d = 0.68$) indicates a medium-to-large practical significance, suggesting the improvement derived from my treatment was substantial.

Student perceptions of their proficiency and practice using evidence also improved. The ten Likert items in the Attitude Survey, although similar in their asking, were sorted into three topics. Items 1, 6, and 7 connected with the topic of students' "Experience" using evidence. Items 2, 4, 9, and 10 pertained more to students' "Skill / Approach" in using evidence; and items 3, 5, and 8 looked more into student "Understanding" of what makes evidence strong/weak, helpful / unhelpful, and more clearly or less clearly connected to the claim (see table 2). Students responded to each of the ten items in the survey by selecting the most fitting response on a scale ranging from (1) "I'm just starting," to (5) "I've Mastered this". For each of the topics and individual Likert items, the average of student responses increased by a value of 0.4 or more.

Table 2. Attitude Survey Likert Items by topic with comparison of a results from pre-treatment ($N=63$) and post-treatment ($N=60$)

Topic	Likert Survey Item		Pre-treatment mean (M)	Post-treatment mean (M)
Experience	1.	I can pick evidence that strongly supports my claim.	2.9	3.6
Skill / Approach	2.	I take time to think about whether evidence is strong or weak before using it.	3.1	3.7
Understanding	3.	I understand what makes evidence strong vs. weak.	3.2	3.7
Skill / Approach	4.	I have a clear system that I use for deciding if evidence is strong enough for my claim.	2.6	3.3
Understanding	5.	I can connect my evidence clearly to my scientific claim.	2.9	3.6
Experience	6.	I use numbers or measurements as evidence when they help prove my claim.	2.6	3.3
Experience	7.	I use observations (what I see, hear, smell, or feel) as evidence when they help.	3.3	3.7
Understanding	8.	I can spot evidence that does NOT help my claim.	3.2	3.8
Skill / Approach	9.	I have the skills to make sure the evidence I use comes from trustworthy / reliable sources.	2.9	3.5
Skill / Approach	10.	When I'm unsure about what a piece of evidence means, I check to make sure I understand it.	3.2	3.7

Although the Likert data is ordinal, with no real measurable value between a response of a 2 or a 3, the increased means demonstrate that on average students are considering themselves more ready to make use of evidence in science class. For example, a mean of 2.6 for Likert item four of the pre-treatment survey indicates that on average students were initially solidly perceiving themselves as having “beginner skills” when it comes to having a system to use for deciding if evidence is strong enough for a claim. That average was almost just as close to perceiving themselves at the lowest readiness of “just starting” as they were to perceiving themselves as “really starting to get this.” Such an explanation is confirmed by the long response follow-up prompt in the survey asking students what system they currently were using to decide if evidence is strong enough for a claim. In the pre-treatment survey, nearly two thirds of students ($n=40$) said that they either had no system for deciding whether evidence was strong enough to support their claim, or their only system was to ask Google or Artificial Intelligence. Only one third of students in the pre-treatment survey ($n=21$) referred to a specific system or method. In contrast, more than half of students in the post-treatment survey ($n=31$) were able to mention a specific system they used, such as using evidence cards ($n=5$), turning to their notes ($n=6$), checking with their group ($n=2$), or asking themselves critical questions such as “do I understand it?” or considering “what detail can explain and support it”.

The percentage of students having a system for discerning the strength of evidence was not the only area to see improvements. For the Attitude Survey the combined percentage of students who felt they had “mastered” items or were able to do them “really well” improved dramatically. Before treatment (see figure 9), the items with the lowest percentage of students perceiving themselves in these upper two values was 20% for both items four and six. In the

post-treatment survey (see figure 10) these percentages jumped to 43% for item four and 45% for item six. The items with the lowest percentage of mastery in the pre-treatment survey were now on par with the highest percentage for any item in the pre-treatment (46% with item seven). The largest increase in mastery percentage belongs to item one, which went from 21% of students to 65% of students determining themselves to be in these top two readiness levels. The mean increase for all ten items combined was +23.7%, with the new highest item now being item eight with 69% of students claiming to have “mastered” it or to be able to do it “really well”. And while in the pre-treatment, each item had more than 20% in the lowest two readiness levels, this was only true in the post-treatment survey for items four and six.

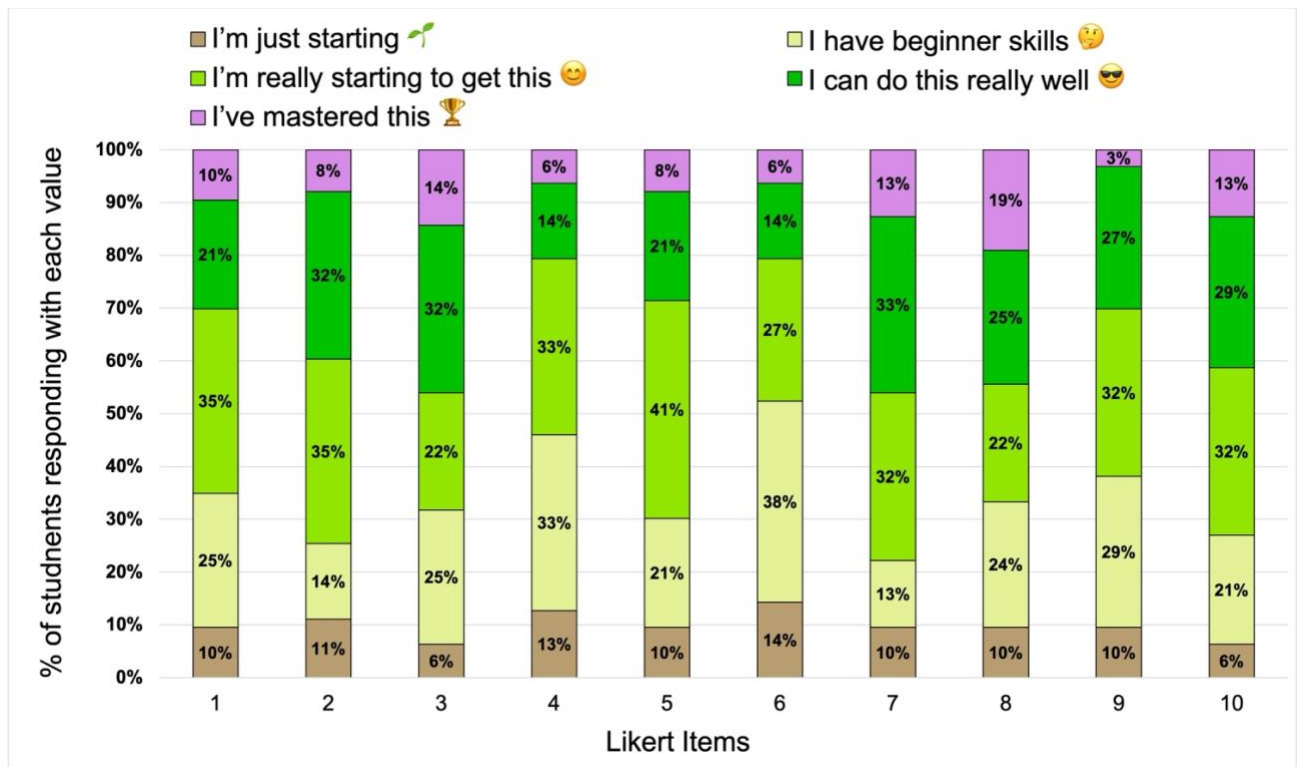


Figure 9. Pre-treatment Attitude Survey percentages by item (N=63)

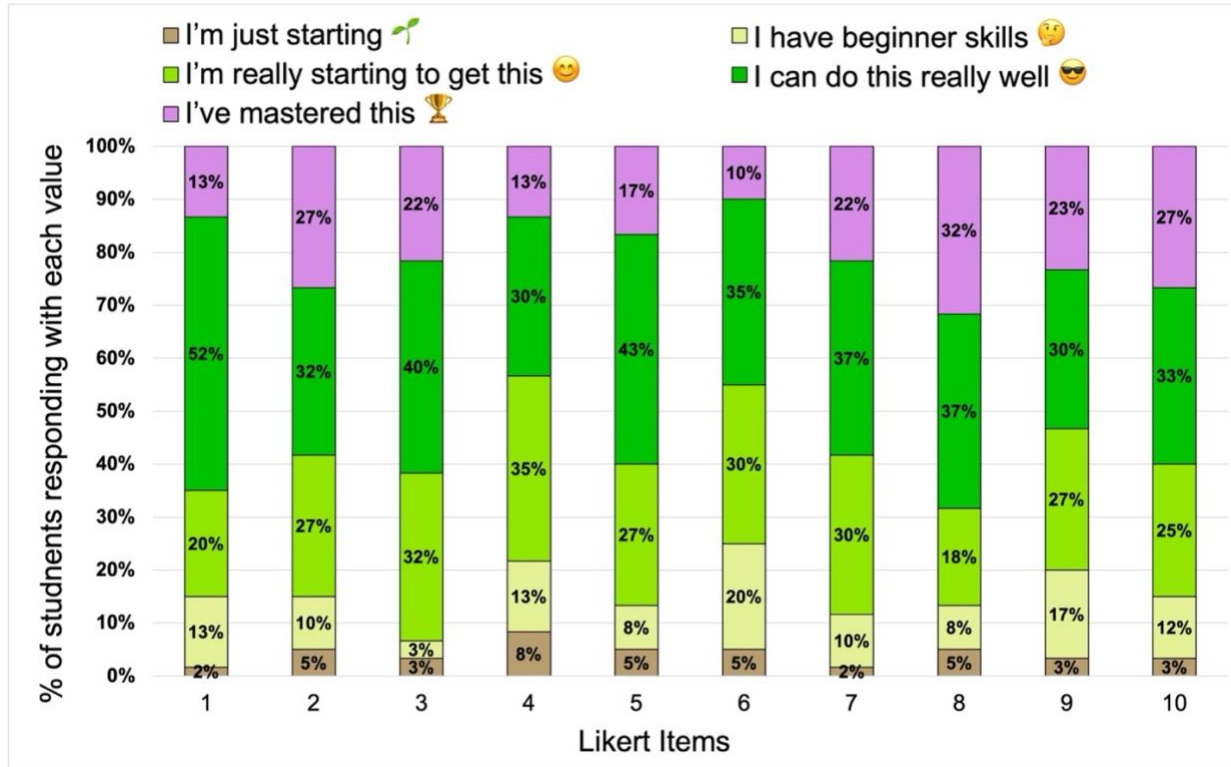


Figure 10. Post-treatment Attitude Survey percentages by item ($N=60$)

I conducted a Chi-Square test which verified that Likert increases were statistically significant for each of the three Likert categories ($N=123$), $df=4$. Cramer's V indicates a medium effect for the topic of "Experience", and a large effect for both "Understanding" and "Skill / Approach".

Table 3. Chi-Square test for statistical significance ($N=123$), $df=4$, $p < .001$

Topic	Chi-Square Value (χ^2)	p-value	Cramer's V
Understanding	29.53	$p < .001$ (significant)	.25 (large effect)
Skill / Approach	33.22	$p < .001$ (significant)	.26 (large effect)
Experience	27.33	$p < .001$ (significant)	.24 (Medium effect)

Although some of these gains no doubt resulted from the increased practice students received using evidence during treatment, the observed increases in each Likert item cannot be solely attributed to simply getting a bit of extra practice. In the pre-treatment survey, when students were asked how much experience they had using evidence to support an answer or claim, there were few who mentioned any specific experience prior to entering 6th grade. Most survey responses ($n=38$) simply claimed to have “none”, “a little”, or “not much.” Multiple students claimed to have just started learning about claims and evidence this year. One student mentioned a 5th grade debate about whether animals should be kept in zoos. Another stated, “I have a lot of experience with using evidence to support my claim because I am in robotics and part of our score is a presentation where we create something and we try to convince the judges that our idea is a good one.” Such responses were the exception in the pre-treatment attitude survey, which did not change much by the post-treatment survey. A smaller, but still large subset of students in the post-treatment survey ($n=24$) still considered themselves to have “little” or “not much” experience. Roughly one week of treatment and a couple extra CERs did not change much in that regard. Only eight additional students considered themselves as having “a lot” of experience following the treatment.

Some of the greatest improvements revealed in the Likert surveys pertain to the students’ ability to recognize what makes evidence strong or weak. In a long response survey prompt following up Likert item three, I asked students what they thought makes some evidence better for supporting ideas in science. Almost half of responses ($n=30$) in the pre-treatment survey gave specific answers about what makes some evidence better. Key themes in student responses included mentioning that evidence is better for supporting a claim when the two correlate well

($n=9$), when the evidence can be referenced such as with pictures, videos, or articles from reliable sources ($n=12$), and when detailed terms and well written responses are included ($n=9$). The other 52% of respondents ($n=33$) offered a vaguer explanation of what make some evidence better. Two students, who both selected “I have beginner skills” to Likert item three, simply said “thinking” in response to the prompt asking what makes some evidence better for supporting claims. Another student, one who identified themselves as having mastered Likert item three, stated “What I think makes some evidence better for supporting an idea is if you know it’s right and it gives you more in detailed information.” While this student hit on the theme of better evidence including detailed information, their inclusion of simply knowing “it’s right” reveals an unfamiliarity with the expectation to reason about the strength of evidence by constructing scientific explanations and arguments from evidence. More clarity was added to student responses to this prompt in the post-treatment survey, with one student responding with, “something that makes evidence better is information to go along with the evidence because it will make the evidence stronger and the claim seem more like it is a correct hypothesis.” Another said, “some evidence is better by including where you got the information or stating a few facts or models to help the reader get a better understanding.” There were of course still students with vague answers to the prompt, such as one simply saying that you know evidence is better for supporting a claim “if it is true.” Still, there was a noticeable increase in the depth of student’s answers relating how strong supporting evidence does not just correlate with the claim, it must be robustly connected to the claim with an explanation of the evidence’s significance. This is the type of understanding that I designed the evidence cards with CQs to elicit.

Similar revelations emerged in long-response answers connected to Likert item eight's ask about whether students could spot evidence that does not support a claim. I should note that this was the Likert item with the highest percentage (19%) of students selecting "I have mastered this" in the pre-treatment survey. A follow-up survey prompt asked students to explain why a piece of evidence might not help their claim. Matching the higher percentage of perceived mastery, most respondents ($n=38$) in the pre-treatment survey replied sensibly that evidence is only helpful if it is pertinent to the claim, with one student writing, "a piece of evidence may not support your claim because it may not directly correlate with your evidence and reasoning." This theme of non-correlating evidence being weak and unable to support a claim continued to the post-treatment Attitude Survey; but there were more students in the post-treatment who described in depth what the reasoning process should look like. One student said that, "a piece of evidence that would not support my claim is evidence that does not make sense, gives little information, and that does not connect to the claim." Another student responded with "A piece of evidence might not help my claim if I can't explain it, it doesn't help me, or it could be used against me." All of these are things the CQs on evidence cards were designed to make students check for. To confirm whether this is a growing understanding, possibly caused by the treatment, I will now examine the Evidence Survey that students completed just before their CERs both pre- and post-treatment, as well as the responses to CQs on the post treatment evidence cards.

Evidence Survey and CQ Responses

Before writing their CERs, students selected the best two available pieces of evidence for supporting a claim and then completed the Evidence Survey to explain why they made the choice they had. Out of all the students who completed the pre-treatment survey ($N=65$), 44% gave

specific reasons for their choices and 56% gave vague explanations such as “it shows good evidence” or “it shows good information and clues”. In contrast, out of the students who completed the post-treatment Evidence Survey ($N=60$), 61% gave good / specific explanations for their evidence choices, while only 39% gave vague / nonspecific explanations (see table 4).

Table 4. Comparison of Good / Specific versus Vague / Non-Specific explanations in the pre-treatment ($N=65$) and post-treatment ($N=60$) Evidence Surveys

Quality of reasons provided to explain evidence choices	Pre-treatment ($N = 65$)	Post-treatment ($N = 60$)
Good / Specific	44%	61%
Vague / Non Specific	56%	39%

I attribute the more specific responses in the post-treatment evidence survey to the use of evidence cards with CQs; and I believe the depth of CQ answers on cards reveals the extent of preparation and consideration made by students. In the pre-treatment, more students completed the evidence survey and moved on to write their CERs before fully understanding their evidence. One student in the pre-treatment had chosen to use the rock samples from a location as evidence for whether volcanic activity exists there. Then in his response to the Evidence survey prompt asking if there was any reason why that might not be strong evidence, the students stated, “I need to see what type of rocks are made after a volcano erupts.” Essentially, they went into a CER with chosen evidence not knowing what their evidence meant. More of the post-treatment responses provided specific explanations for why they selected each piece of evidence. The depth of some of these explanations were nearly full CERs themselves, demonstrating a much higher level of understanding of the selected evidence as well as the available evidence on other

cards. One student, when noting that their selected evidence had some weaknesses included, “but I chose to use it anyway because my other evidence card will help back it up”. Other students highlighted the value of CQs by referring specifically to their CQ answers when explaining why they selected their evidence cards, demonstrating that they saw their CQ answers and the evidence at the top of the card as providing a unified value to their arguing support for their claim.

When examining the sets of evidence cards students used to write their post-treatment CER, I saw a strong correlation between the depth CQ answers, and the score received on the CER rubric. Out of those who turned in their completed evidence cards ($N=57$), 18 students answered the CQ prompts with what I considered to be vague answers (see table 5). These CQ responses ranged from being unanswered to being unanswered incorrectly or answered without adding any beneficial information to the card.

Table 5. Depth of responses to CQs on evidence cards ($N=57$) compared to post-treatment CER scores

Depth of answers to Critical Questions on post-treatment Evidence Cards ($N = 57$)	Vague ($n = 18$)	Neutral ($n = 15$)	Good ($n = 24$)
Students with post-treatment CER rubric scores in the <u>4th quartile (10-12)</u>	1 (6%)	2 (13%)	13 (54%)
Students with post-treatment CER rubric scores inside the <u>middle two quartiles (5-9)</u>	9 (50%)	10 (67%)	10 (42%)
Students with post-treatment CER rubric scores in the <u>1st quartile (0-4)</u>	8 (44%)	1 (7%)	1 (4%)
Students who answered CQs on evidence cards but did not complete the post-treatment CER		2 (13%)	

At first glance it may seem concerning that roughly 32% of students who turned in their evidence cards provided only vague responses to the CQs following all that was taught in treatment. Further examination shows a promising pattern though. Of those who did provide good, detailed, and specific responses to the CQs (n=24), 54% scored in the highest quartile on the post-treatment CER rubric, with grades ranging from 10-12. In comparison, only 1 student out of the 18 who gave vague CQ responses scored within that top quartile. A near equal mix of students scoring inside the middle two quartile range gave vague, neutral, and good responses to CQs. Out of students in the lowest quartile of post-treatment CER scores, only 1 had given neutral depths of CQ responses and only 1 had provided good CQ responses. Essentially, making correct, highly detailed, and specific answers to CQs on the evidence cards strongly correlated with performing well on the CER.

CHAPTER FIVE

VALUE: REASONING AND REFLECTION

Claims From the Study

The purpose of this study was to gain insight into improving the performance of middle school students in the reasoning portion of their CERs. My primary research question was: How will the use of evidence cards affect the quality of students' explanations in the reasoning portion of their written CERs?

My sub-questions include the following:

1. What effect will student use of evidence cards have on their perceived readiness to evaluate and use evidence?
2. How will evidence cards incorporating critical questions impact students' ability to assess the quality of each piece of evidence?
3. What effect will the use of evidence cards have on my role as the teacher?

Middle School Students Can Reason about Evidence

Although it was not one of my research questions, going into this I had my doubts about whether I was fighting against immovable limits of cognitive development when I sought to improve the reasoning of 6th grade students in CERs. I wondered whether any but the highest-performing students could explain the value and significance of their evidence at the level of my expectations. Upon analyzation of the data, I must conclude that even 6th grade students who have had little previous experience writing CERs can, within less than the scope of a year, develop to write exceptional CER papers. After teaching the use of evidence cards, the mean

scores in the reasoning portion of my students' CERs rose from a 5.38 to a 7.42 with a possible 12 points maximum. Those statistically significant gains outperformed my expectations of what was possible. The fact that only the lower quartile of post-treatment rubric scores were at or below the median of pre-treatment scores was astounding. None of my previous approaches to CER instruction over the years have produced gains anywhere close to this. Consideration of the data has allowed me to determine with high confidence that the answer to my primary research questions is that the use of evidence cards does beneficially affect the quality of students' explanations in the reasoning portion of written CERs.

Evidence Cards Provide a way for Students to think about Evidence More Concretely

On more than one occasion in my teacher journal I noted how students were holding evidence cards in their hands discussing them as if they were physical items to be weighed and bartered with their partner. Although students had always been allowed to converse before treatment when completing the evidence page in their binders, I had never before witnessed such an interactive evaluation of evidence like I observed while partners were sorting evidence cards. The fact that the cards could be placed and rearranged on a gradient of stronger to weaker seemed to help them see that each piece of evidence had unique value that it could add to an argument. With the cards, students seemed to better comprehend that what they had was not merely an answer from a page in their binders, it was evidence that could be used to structurally support their claims. I did note in my journal that even with evidence cards, CER performance will always be limited by student concern and effort for doing well; but I also noted that students expressed a preference for the physical and more interactive approach of using the cards and

gradient as opposed to using the old standard of evidence packets, which likely will help improve student effort.

Critical Questions Improved Student Ability to Assess the Quality of Evidence

In general, the purpose of writing CERs is to provide students with the opportunity to engage in the science and engineering practices of Arguing from Evidence and Constructing Explanations. I incorporated the critical questions on the cards to serve as an interlocutor to improve students' awareness of the details necessary for a good argument or explanation. Such improvements were observed. When using evidence cards, the percent of students who specifically explained why they selected particular evidence increased by 17%, as revealed in the results of the Evidence Surveys. I also found a strong correlation between higher scores on the post-treatment CER and the extent to which students engaged the evidence card CQs with good specific responses.

One reason the students were able to improve their explanations and arguments is no doubt also tied to how the CQs also provided a check on the science practice of analyzing and interpreting data. I noted in my teacher journal during the pre-treatment CERs, that several students seemed to make up evidence on the fly while writing. In contrast, the critical questions on evidence cards forced students to slow down and consider whether they truly understood the data they were about to write about. The conversations they had with their partner or the looks back into their science binders during the answering of the critical questions served as a valuable review of the specifics related to piece of evidence. Students took this renewed and refined understanding of the concepts with them as they wrote their CERs. I still needed to place checkpoints throughout the process to ensure students actually took time to answer the CQs and

understand the evidence; but this doesn't demonstrate a flaw in the design of the cards themselves. When students answered the CQs according to expectations, they did well in their CERs. Going forward, I expect students will do better with their CQ answers from the beginning, now that they know I have set up checkpoints.

Benefits for Formative Assessments

The CQs were not just prompts for students to return to their binders and make sure they understood and had collected all their best details before writing a CER. They also provided me with improved opportunity to formatively assess this understanding. I was able to move around the room during card sorts to listen as students considered each CQ. Even what I missed during student discourse, I easily caught prior to CER writing thanks to the various checkpoints. This again provided me with exceptional opportunity to formatively assess the real-time readiness of each student. I could instantly spot any abnormal placement in the ranking of evidence cards on the gradient sheet. This alerted me to take a deeper look at the CQ answers so that I could identify the misconception or lapse of understanding. There was no such way to quickly spot something being off with the evidence packets I used to use, for with them I needed to invest enough time to read multiple sentences to gauge student understanding. Now, a quick glance at the placement of cards on the gradient provided an indicator of what students thought of each piece of evidence.

Even when students sorted the cards reasonably on the gradient, I always looked at the answers to CQs on the highest ranked cards – those which students were likely to use when writing their CERs. Incorrectly or incompletely answered CQs revealed elements of the evidence that the student did not quite grasp. When this was the case, I directed students back towards

binder pages and the concepts we had covered in class. On several occasions I directed students to reread scientific texts previously used in class to ensure understanding of concepts. The use of evidence cards improved my opportunity to assess and highlighted misconceptions whether held by individual students or the entire class. Of course, I eventually allowed all students to proceed past my checkpoint even if they did not make use of constructive prompts. Nevertheless, there certainly was increase opportunity for timely formative assessment and feedback.

Effect of Evidence Cards on the Teacher

The final sub-question of my research asks how the use of evidence cards affects my role as the teacher. The answer is that in every discernible way I benefited by their use, and in no discernible way did this process make my job any more difficult. To be sure, the teaching of how to use evidence cards took some time, as did the practice Venus CER completed during treatment. Both took a couple days, resulting in a total time investment of about four 60-minute class periods. This amount of instructional time is no more than what I generally invested in teaching CER writing. Few of my students had ever written or even heard of CERs prior to entering my 6th grade classroom. If I must provide dedicated instruction each year on how to write CERs, then beginning with evidence cards from the onset does nothing to increase the necessary allotted class time.

The planning and preparation time required by evidence cards was also unexceptional in comparison to the evidence pages I used prior, such as the one in Appendix B. The card template was simple to create and easy to alter for each new CER. Once the template was made, a full set of cards could be developed within a single one-hour prep period. For the Critical Questions, I only needed to consider what had been covered in class that students should mention in their

reasoning statements of their CERs. This may be more difficult for a first-year teacher or one who has never taught the material the CER pertains to; but I was well familiar with these units and found the creation of evidence cards to be simple and quick.

For what it is worth, it was also my informal assessment that students enjoyed the process more. When I asked a few kids during the treatment whether they prefer the card sorts or the old way, they all mentioned preferring the evidence cards because it gave them more of a chance to work with others and talk about the evidence. I too noted that some of the students who frequently “miss” concepts as we go over them in class, found a benefit in the ability to go back over the data and concepts with others who might be able to fill in the gaps in their understanding.

Value of the Study and Consideration for Future Research

By no means do I consider this research to be an ending point. Though gains were made in student CER performance, I look forward to seeing how the increases seen from using evidence cards continues long term. It is possible that this particular cohort of students had not yet plateaued in their CER performance prior to the introduction of evidence cards. Since CERs in general were so new to them, it is reasonable to assume that increased practice of the older method of preparing for CERs would have also produced some gains. I am interested to observe whether year after year the use of evidence cards continues to result in better CERs than what I have seen in the past years.

I also am curious in pinning down what differentiation and special education accommodations may look like when using evidence cards. For this research project, I chose not to include my first hour science class because it is an “inclusion hour” in which more than 30%

of the students have significant special education needs. Some in that inclusion hour could barely read or write. I determined that it was out of the question for these students to participate in the research since their survey responses might have been too heavily influenced by the statements necessarily fed to them by myself or the paraprofessional in the room. Despite not being in the research study, they still went through the process of using evidence cards. I learned that they needed (and benefitted from) accommodations such as limiting the number of cards and providing answer options for the CQs. These accommodations were not systematically carried out as part of this research project, so additional research in using evidence cards as accommodations could prove highly beneficial.

Finally, future research could be conducted to refine the use of evidence cards and CQs. I attempted to make evidence cards for my CERs with a variety of evidence. Some cards included nothing more than charts and uninterpreted quantitative data. Some evidence cards and CQs prompted students to discern the value of evidence by considering the reliability of source material, such as the article we read that was written by a lady with a PhD in volcanology. Other evidence pertained more to terms taught within the unit with definitions that were essential for concept comprehension. For each of the CERs I attempted to have this mix of evidence types; but I was very much figuring this out as I went. In the future, I would like to research the effectiveness of different types of evidence cards. For example, are some types of evidence cards better suited to instances where the objective for students is to engage in the practice of arguing from evidence as opposed to constructing explanations? Perhaps it would be beneficial to include evidence cards that students make themselves using data which they collected through their own observations and experiments. These questions await the inquiry of future research.

Impact of Action Research on the Author

The implementation of this research in my classroom will hopefully be long lasting. First, I plan to use what I have discovered in future years. Just as this year's students benefited, so too I think future classes will benefit by using evidence cards with critical questions. I also hope what I have learned this year may continue to benefit this particular cohort as they move on to seventh grade. Since evidence cards are part of a process of a CER writing process that must continue being used, there needs to be buy-in from the science teachers in upper grades. I am confident enough in the results of this research to share it with the eight other science teachers in my building. Each of them has expressed frustrations with the quality of CER writing, and especially the reasoning portion of CERs. While more room for research remains, I am convinced that using evidence cards with critical questions does provide a benefit to students. I fully intend to use them myself moving forward and would happily teach the process to my colleagues near at hand. As I refine this approach and learn more, I can envision possibly sharing this farther abroad, such as at a National Science Teachers Association (NSTA) conference.

REFERENCES CITED

- Argumentation Toolkit. (n.d.). *How can you support student use of reasoning in a scientific argument?* Lawrence Hall of Science.
<https://argumentationtoolkit.lawrencehallofscience.org/teacher-learning/advanced-evidence-reasoning/how-can-you-support-student-use-of-reasoning-in-a-scientific-argument/>
- Bell, P., & Shouse, A. (2014). *Is it important to distinguish between the explanation and argumentation practices in the classroom?* STEM Teaching Tool #1. University of Washington Institute for Science + Math Education. <https://stemteachingtools.org/brief/1>
- Delen, I., & Krajcik, J. (2015). What Do Students' Explanations Look Like When They Use Second-Hand Data? *International Journal of Science Education*, 37(12), 1953–1973.
<https://doi.org/10.1080/09500693.2015.1058989>.
- Esparza, M. S. (2024, August 25). *The power of CER: Strategies to engage middle school students in evidence-based writing*. BrainPOP.
<https://www.brainpop.com/discover/post/the-power-of-cer>
- Ford, M. J. (2015). Educational Implications of Choosing “Practice” to Describe Science in the Next Generation Science Standards. *Science Education (Salem, Mass.)*, 99(6), 1041–1048. <https://doi.org/10.1002/sce.21188>.
- González-Howard, M., Marco-Bujosa, L., McNeill, K. L., Goss, M., & Loper, S. (2018). Teacher's Toolkit: The Argumentation Toolkit. *Science Scope (Washington, D.C.)*, 42(3), 74–78. https://doi.org/10.2505/4/ss18_042_03_74.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education (Salem, Mass.)*, 94(5), 810–824. <https://doi.org/10.1002/sce.20395>.
- The Lawrence Hall of Science. (n.d.). *Strategy: Discussing Quality of Evidence, Using the Gradient Tool* [Video]. Argumentation Toolkit. <https://argumentationtoolkit.lawrencehallofscience.org/argument-elements/evidence/>.
- Next Generation Science Standards (NGSS). (2013, June 5). *Science and Engineering Practices in the NGSS* (Appendix F). Next Generation Science Standards.
- Nussbaum, E. M., Van Winkle, M. S., Tian, L., Putney, L. G., Huerta, M., Perera, H. N., Dove, I. J., Herrera, A. N., & Carroll, K. R. (2024). Extending science instruction beyond the CER: Use of critical questions in the argumentation of middle school science students. *Science Education (Salem, Mass.)*, 108(5), 1420–1447.
<https://doi.org/10.1002/sce.21877>.
- Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging Students in the Scientific Practices of Explanation and Argumentation. *Science Scope (Washington, D.C.)*, 35(8), 6-.
https://doi.org/10.2505/3/ss12_035_08.
- Sickler, J., Lentzner, M., Goldsmith, L. T., Brase, L., & Kochevar, R. (2024). Reasoning about data in elementary school: student strategies and strengths when reasoning with multiple

variables. *International Journal of Science Education*, 46(16), 1736–1756. <https://doi-org.proxybz.lib.montana.edu/10.1080/09500693.2023.2298214>.

U.S. News & World Report. (n.d.). Bemidji Middle School, Minnesota – Education K-12. <https://www.usnews.com/education/k12/minnesota/bemidji-middle-260498#data>

APPENDICES

APPENDIX A: IRB EXEMPTION

****External Sender****

Hello Pope, Benjamin,

Your protocol was reviewed by the IRB and has been approved.

PI: Pope, Benjamin

Approval Date: 12/1/2025

Title: "Using Evidence Cards in Middle School CERs"

Protocol #: 2025-2381-EXEMPT

Review Type: Exemption

Expiration Date: [12/1/2030](#)

APPENDIX B: PRE-TREATMENT EVIDENCE PACKET

Names: _____

Hour: _____

Volcano or Not?**Question:** Which of the four locations are volcanic?

Assemble evidence from the information you have gathered so far.

Location: #1 Kathmandu

1. **Pictures / Description:** From what you can see in the pictures (and what you know of volcano shapes, does the location look like it might have a volcano? **TAG.**

2. **Maps / Tectonic Plates:** Compare the location map with the tectonic plate map

- Is this location near a plate boundary? ☐ **Yes** ☐ **No**
- If the location is near a plate boundary, select which type of boundary seems most likely for that location.

☐ **Oceanic Convergent**
☐ **Continental Convergent**

☐ **Hot Spot**
☐ **Divergent**
- Why does this type of tectonic plate boundary matter? **TAG.**

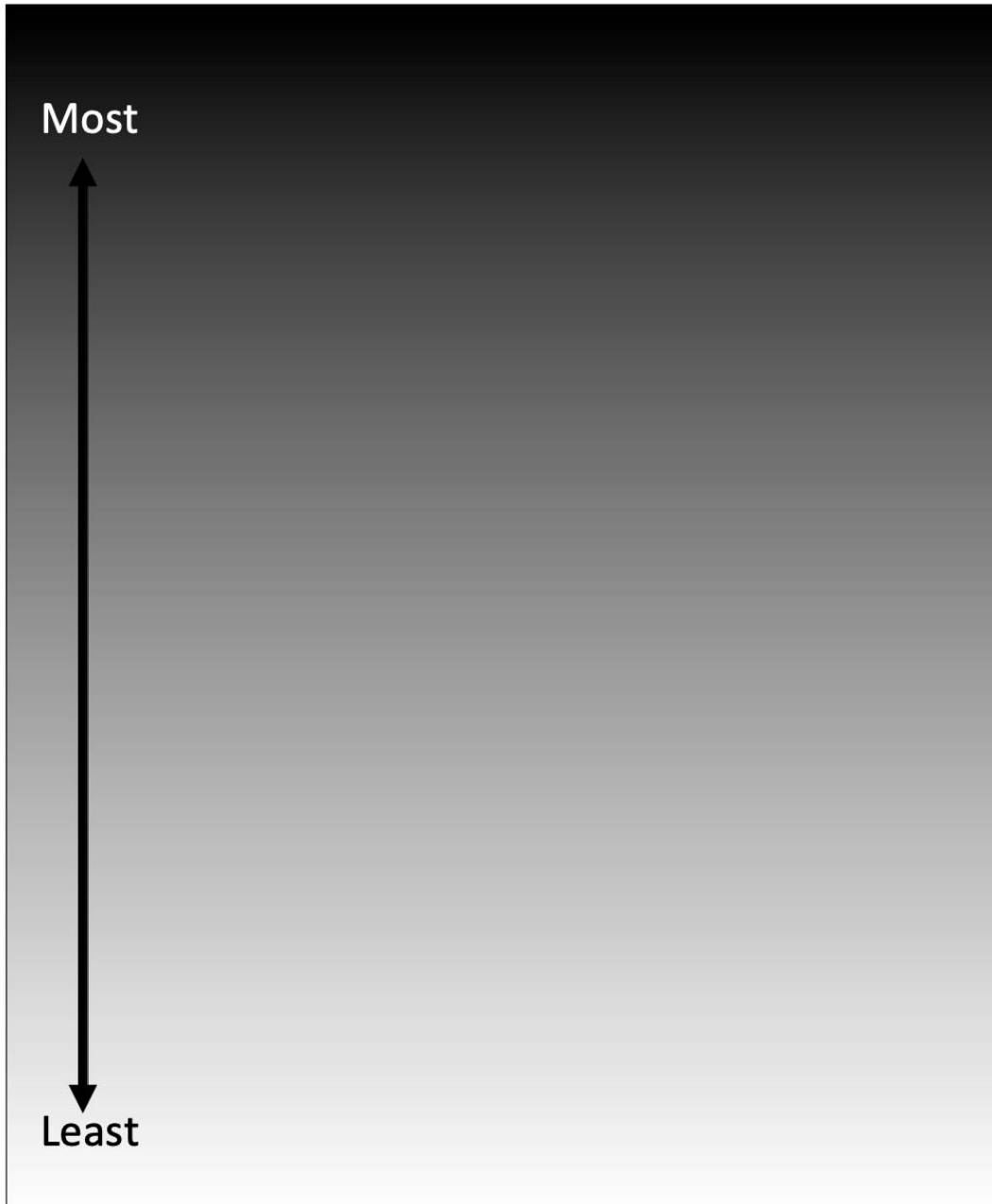
3. **Rocks:** what type of rock(s) can be found at your location?

_____ and _____, which are.... (circle one)
 extrusive igneous intrusive igneous sedimentary metamorphic

- Is that type of rock evidence of volcanic activity? **TAG.**

Claim: Is there volcanic activity at this location?

APPENDIX C: EVIDENCE GRADIENT TOOL

Evidence Gradient Tool

APPENDIX D: ATTITUDE SURVEY

Name: _____

Answer the following questions by selecting the statement that best describes you currently.
There are no right or wrong answers—just tell me what you think!

1. I can pick evidence that strongly supports my claim.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

2. I take time to think about whether evidence is strong or weak before using it.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

3. I understand what makes evidence strong vs. weak.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

→ What do you think makes some evidence 'better' for supporting an idea in science?

4. I have a clear system that I use for deciding if evidence is strong enough for my claim.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

→ What is the system you currently use?

5. I can connect my evidence clearly to my scientific claim.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

→ How much experience do you have using evidence to support an answer or claim?

6. I use numbers or measurements as evidence when they help prove my claim.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

7. I use observations (what I see, hear, smell, or feel) as evidence when they help.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

8. I can spot evidence that does NOT help my claim.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

→ Write one sentence explaining why a piece of evidence might not help your claim:

9. I have the skills to make sure the evidence I use comes from trustworthy / reliable sources.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆

10. When I'm unsure about what a piece of evidence means, I check to make sure I understand it.

- ☐ I'm just starting 🌱
- ☐ I have beginner skills 😊
- ☐ I'm really starting to get this 😊
- ☐ I can do this really well 😊
- ☐ I've mastered this 🏆



APPENDIX E: EVIDENCE SURVEY

Reasoning About Evidence

Please answer in complete sentences

Evidence #1

What is the 1st piece of evidence you selected to support your claim?

Why did you decide this was a good strong piece of evidence?

What are the weaknesses of this evidence? Why did you choose it anyways?

How prepared are you to describe "the science" of why this evidence matters?

(using words or ideas we have learned in class)

1	2	3	4	5
(I am not ready at all)				(I am completely ready)

Evidence #2

What is the 1st piece of evidence you selected to support your claim?

Why did you decide this was a good strong piece of evidence?

What are the weaknesses of this evidence? Why did you choose it anyways?

How prepared are you to describe "the science" of why this evidence matters?

(using words or ideas we have learned in class)

1	2	3	4	5
(I am not ready at all)				(I am completely ready)

APPENDIX F: CER RUBRIC

Name: _____

Pre-Treatment Reasoning Rubric

REASONING #1	0	1	2	3
Connects Evidence to Claim	Contradicts the claim or makes no connections	The evidence is suitable for the claim, but the connection is weakly explained	The evidence is clearly connected to the claim	The student explained how their claim is better supported by the evidence than competing claims
Scientific Accuracy and Detail	No key science concepts accurately explained	Pertinent concepts are mentioned; but the accuracy or detail of the explanations are very limited	Concepts linked to the chosen evidence were explained in moderate detail or with only minor inaccuracies	Accurately explains all key science concepts with appropriate detail

Reasoning 1 Total: ____ / 6

REASONING #2	0	1	2	3
Connects Evidence to Claim	Contradicts the claim or makes no connections	The evidence is suitable for the claim, but the connection is weakly explained	The evidence is clearly connected to the claim	The student explained how their claim is better supported by the evidence than competing claims
Scientific Accuracy and Detail	No key science concepts accurately explained	Pertinent concepts are mentioned; but the accuracy or detail of the explanations are very limited	Concepts linked to the chosen evidence were explained in moderate detail or with only minor inaccuracies	Accurately explains all key science concepts with appropriate detail

Reasoning 2 Total: ____ / 6

Pre-Treatment Reasoning Total: ____ / 12

Notes: