

A PILOT STUDY OF EXPLICIT INSTRUCTIONAL SCAFFOLDS TO TEACH SCIENCE  
ARGUMENT WRITING

By

Cherish Marie Sarmiento

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Special Education – Doctor of Philosophy

2024

## **ABSTRACT**

An emerging body of research suggests that content-rich literacy instruction has the power to expand elementary children's domain-general and domain-specific literacy skills. For content teachers, such as science teachers, explicit instruction in disciplinary-language and literacy practices can support students reading and writing of school-based texts without sacrificing valuable, and limited, instructional time. The current pilot study tests the effectiveness of explicit instructional routines at the word-, sentence-, and discourse-level to support Grade 4 student science argument writing. Analysis of Co-variance was used to assess group differences of the intervention versus an active control on students' science knowledge and argument writing after controlling for the effects gender and topic prior knowledge.

For my grandparents, Ines Elaine and Henry Bueno Sarmiento, who gave me everything they ever had so that I could have everything I ever wanted.

## ACKNOWLEDGMENTS

The Kindergarten Me never would have believed she would be here today, and it is not with tremendous gratitude that Present Me acknowledges the community that has nurtured me through years of study.

I wish to give special thanks to my committee members, Drs. Adrea Truckenmiller, Eunsoo Cho, Troy Mariage, Gary Troia, and Amelia Gotwals who were instrumental in helping me mesh my varied interests and professional experiences into something that impacts science education. I am a better scholar for having been mentored by each of them.

I want to thank the veritable army of intelligent women that were there for key moments in my education. To Adrea, for helping to quiet any annoying little voices that said a first-generation girl couldn't cut it in ivory towers; for helping me persist. To Mrs. Guzman, who offered me a home when I did not have one, because someone believed she had a right to finish her education and so did I. To my mother, who did not let me intentionally fail Honors Geometry in the Sixth Grade because "smart girls figure out a way to get through difficult things." To my Grandmother who took me "into town" to go to the library all those years ago.

To the dearest friends in group chats with ridiculous names, they have been a welcome respite whenever my homesick soul needed a rest. I could not have navigated these five years of grad school without Lindy and Lo, Holly and Allison, and the WRITE lab. Their friendships go beyond this job and is lifelong. Finally, the friends I met at a little dance studio gave me the gift of colorful memories captured in the pages of so many journals. No goodbye will be harder, no impression more permanent.

Brian, my darling husband, we did it! Shall we start our next adventure? The weather looks spectacular.

## TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION .....	1
CHAPTER 2: LITERATURE REVIEW .....	5
CHAPTER 3: METHOD .....	25
CHAPTER 4: RESULTS .....	67
CHAPTER 5: DISCUSSION.....	99
REFERENCES .....	119
APPENDIX A: WRITING ARCHITECT SCIENCE PASSAGES .....	130
APPENDIX B: MORPHOLOGY ROUTINE .....	140
APPENDIX C: CER GRAPHIC ORGANIZER .....	142
APPENDIX D: CER SENTENCE STARTERS .....	144
APPENDIX E: TOPIC PRIOR KNOWLEDGE TEST ADMINISTRATIVE SCRIPT .....	146
APPENDIX F: SCIENCE KNOWLEDGE TEST.....	147
APPENDIX G: TEACHER CONSENT FORM .....	151
APPENDIX H: PARENT CONSENT FORM .....	154
APPENDIX I: STUDENT SOCIAL VALIDITY SURVEY.....	158
APPENDIX J: TEACHER ACCEPTABILITY SURVEY .....	160
APPENDIX K: PROFESSIONAL LEARNING AGENDA .....	165

## CHAPTER 1: INTRODUCTION

*Flatten the curve. Asymptomatic spread. Herd immunity. Pandemic.* These are just a few of many science-related terms that the average American household incorporated into their lexicon as the Covid-19 pandemic began to rock the nation in March 2020. There became an imperative need for the average American not only to understand what these terms meant conceptually but how these concepts suddenly had practical importance on their daily goings-on. Americans had to become fluent in using Zoom for homeschooling purposes because the *novel coronavirus*, which was *spread by respiratory droplets*, made large public gatherings a health risk. A nation's collective behavior was modified, in part, due to the convincingness of scientist's arguments.

The pandemic also highlighted how difficult it can be to enter scientific conversations when one is a novice (Lemke, 1990). Building scientific knowledge relies heavily upon sharing ideas through writing, which affords the writer the time to carefully construct and revise their message. Scientific language is rife with disciplinary vocabulary, dense and complex sentences, and are marked with unfamiliar text structures (Snow et al., 2009; Snow, 2010). These language features mark deliberate attempts to convey technical and abstract scientific concepts to other scientists (Halliday & Martin, 1993; Halliday & Hassan, 1976; Lemke, 1990). Entering a scientific conversation then becomes a matter of being not only familiar with the scientific concepts under study, but also with the language used to convey it.

### **Teaching Students to Write Like Scientists**

Writing is the primary communicative tool that scientists use to assert and disseminate their findings with the broader scientific community (Norris & Phillips, 2003). Engaging with scientific ideas can occur across cultures, time, and geographic location due to a shared vocabulary and a

common set of communicative practices (Halliday & Martin, 1993). Though school children are not full-fledged scientists, their initial exposure to scientific ideas and language in a school setting provides them with opportunities to establish a foundation of knowledge that they can build on in pursuit of personal ventures or later academic pursuits (Alexander, 2003; Halliday & Hassan, 1993). To summarize the Model of Domain Learning (MDL; Alexander, 2003; Alexander & Judy, 1988), for students to develop disciplinary expertise, they must transition from using general learning strategies to acquire content to using their broad knowledge base and applicable discipline-specific strategies to solve problems and generate new ideas. K-12 education endeavors to develop students who are competent in a range of subject areas and can utilize both general learning strategies (e.g., what makes a compelling argument) and discipline specific strategies (e.g., what makes a compelling scientific argument) (Alexander, 2003). For students to make the transition from novice scientists to competent ones, teachers have a role in supplying young scientists with the writing experiences and explicit instructional opportunities to become fluent in using the language practices of the scientific community (Lemke, 1990; Gotwals et al., 2012; Norris & Phillips, 2003).

Because reading has largely dominated the field of literacy research (Graham & Perin, 2007), there are numerous studies describing how teachers can bolster student's science content knowledge via reading instruction. Providing students with science-rich text instruction in the early grades can help build a foundation of background knowledge students can draw from to make sense of novel subject matter (Cabell & Hwang, 2020; Cervetti & Wright, 2020). Explicit instruction in vocabulary (Block et al., 2019; Truckenmiller & Petscher, 2020), connectives (Andreev & Uccelli, 2023; Cain & Nash, 2011; Crosson et al., 2008), and text structure (De La Paz & Graham, 2002; Hebert et al., 2018; Reynolds & Perin, 2009;) can support student's

comprehension of science-related informational texts, while the implementation of instructional routines and scaffolds can help students extract essential information and learn from them (Songer & Gotwals, 2012; McNeill & Krajcik, 2006). While research has shown that reading instruction can support children's writing abilities (Graham et al., 2018) and vice versa (Graham & Hebert, 2011), explicit writing instruction in schools is relatively uncommon (Coker & Lewis, 2008; Graham et al., 2014). In a nationally representative sample of 285 middle school teachers, Grades 6 to 8, only an average of 32.5 minutes of writing instruction occurs per week in the school subjects of language arts, social studies, and science (Graham et al., 2014).

As writing is an essential skill for civic and professional life (National Commission on Writing, 2003) and is emphasized as an important skill for promoting science learning (National Research Council [NRC], 2012), there has been a concerted effort to understand how more writing can be infused into the curriculum, especially for the benefit of students who experience difficulties in writing. A growing number of studies suggest that using writing as a reflective exercise is one valid means of enhancing knowledge (Hand et al., 2004; Ferretti et al., 2007; Klein & Rose, 2010). Other studies have demonstrated that curricular scaffolds (Bulgren et al., 2009; Songer & Gotwals, 2012) and explicit strategy instruction (Benedek-Wood et al., 2014; Mason et al., 2006) are also helpful for facilitating writing, and thus improving science knowledge. However, studies that have examined explicit writing instruction and its role in improving both content knowledge *and* writing quality are rare (Benedek-Wood et al., 2014; Bulgren et al., 2009; Lee & De La Paz, 2021b; Gillespie Rouse et al., 2017; Wright et al., 2019). If all students are to develop fluency in the language of science- if they are to be expected to write like scientists- many will need explicit writing instruction. While the body of research supporting writing as a valuable tool for learning is growing, there is a need to add to this



consensus via an examination using more rigorous methodologies. In addition, the field of writing research may also benefit from expanding our examination of interventions effects on important learning outcomes that might be targets for an instructional interventions, including general writing ability and writing quality.

## CHAPTER 2: LITERATURE REVIEW

### Writing-to-Learn in Science

The Common Core State Standards (CCSS) portray a vision of public-school graduates who approach an informationally laden environment with the critical lens of a scientist. By the end of high school, students are expected to be able to “question an author or speaker’s assumptions and premises and assess the veracity of claims and the soundness of reasoning (Council of Chief State School Officers [CCSSO], 2010).” These same students place a premium on evidence and use it to build knowledge via written byproducts that takes on the form of one of three macrogenres of text- narrative, informational/expository, or argument. Of the three macrogenres, argumentation is heavily emphasized within the Next Generation Science Standards (NGSS). Across eight science practices, “argument” is mentioned 72 times and is heavily emphasized in Practice 7- Engaging in Argument from Evidence. Across both set of standards, students are asked to submit their arguments in writing as part of a “writing-to-learn” process (CCSSO, 2010; NRC, 2012).

Writing has been conceptualized as a knowledge transforming process that requires writers to make connections between what they have learned previously and what they are learning currently through a series of inferences, insights, and commentary (Graham, 2020; Norris & Phillips, 2003; Klein & Rose, 2010). For this reason, writing activities in classroom settings can be effective ways to support children’s learning of academic content (Klein & Rose, 2010; Wright et al., 2019), with such activities ranging from information recording (i.e., note taking and summarization) to critical analysis (i.e., constructing explanations) in both narrative and informational genres (Graham et al., 2020). The utilization of writing as a medium to learn, organize, and evaluate information with respect to content is called writing-to-learn. There is a

growing body of empirical research to suggest that such activities produce a modest effect size (ES=0.30) on science learning (Graham et al., 2020).

Among the different types of writing-to-learn activities that might be implemented within a classroom environment, analytic writing is chief among them in promoting student's learning due to its potential to reorganize student's thinking around a subject. In fact, a meta-analysis that examined the effect of writing-to-learn activities on learning across three school subjects found that 70% of the studies ( $n=39$ ) incorporated analytic writing tasks (Graham et al., 2020).

Analytic tasks include those that ask students to integrate, compare, contrast, reinterpret, or argue using both newly learned and existing ideas (i.e., background knowledge). Together these types of tasks produce a moderate effect (ES=0.42) on student learning across school subjects, such as social studies or science, when compared to writing tasks with other communicative purposes (e.g., informational writing, journaling).

### **Analytic Writing**

Students may exhibit intra- and inter-individual performance on different genres of writing (Davidson & Berninger, 2016; Olinghouse & Wilson, 2013; Troia et al., 2019; Valentine et al., 2021), suggesting a need for instruction tailored to the unique features of various writing tasks. This may hold true for subgenres of analytic writing as well. Davidson and Berninger (2016) evaluated the variability of writing quality in essays produced by Grade 5 and Grade 7 students on three analytic subgenres of writing (i.e., informative, compare-contrast, and persuasive). With the subject matter held constant across prompts, essays were assessed for their quality of content and organization. Correlational analysis revealed that student essays exhibited only a modest amount ( $r=.36$ ) of shared variance in content and organization across text types. For both grades, the quality of content in their persuasive essays lagged that of their

informational and contrastive essays. Meanwhile, the organizational demands of contrastive essays proved to be more difficult for students in Grade 5 and, for Grade 7, persuasive organization was more difficult. Though some of the variability in scores both within and across grade-levels can be attributed to the evaluation tool used (e.g., the rubric used), much of it can be attributed to individual differences in student's knowledge of the elements specific to the different genres. This means that student's differential knowledge of the parts of persuasive or argumentative texts (e.g., claims, evidence, reasons) is malleable and can be targeted during instruction.

One way of providing students with instruction in a range of analytic writing subgenres is to situate the task within a content-area domain. Many writing-to-learn studies in science evaluate student's comprehension and knowledge, yielding a modest effect on student learning outcomes ( $ES=.31$ ; Graham et al., 2020). Argument writing, a form of analytic writing, is an effective writing-to-learn medium in the science classroom because it approximates professional practices and encourages students to interact with the subject matter beyond the level of rote memorization (Lemke, 1990; Graham et al., 2020; Hand et al., 2004; McNeill et al., 2006; Sampson et al., 2013). Research has shown that argument writing is particularly beneficial for middle school and high school students when science instruction involves inquiry activities, especially when students have consistent opportunities to process their learning via writing (Sampson et al., 2013; Wright et al., 2019). There are fewer studies on argument writing as a means of science learning in Grades K-4, but research is emerging regarding how literacy rich environments can support young children's science learning (Cabell & Hwang, 2020; Cervetti et al., 2016; Songer & Gotwals, 2012). Gotwals et al., (2012) demonstrate that scaffold-rich science units accompanied by similarly scaffolded assessment questions, can support G4-6 students

science arguments, and has positive impacts on overall science achievement. While writing activities were infused throughout the unit, there was no direct assessment of student's science writing

Struggling writers therefore may benefit from argument writing instruction being incorporated into science instruction, but especially when that instruction is explicit (Klein and Rose, 2010; Ferretti et al., 2007; Wright et al., 2019). One method of supporting student's science learning via argument writing is through an argument framework called Claim-Evidence-Reasoning (CER).

### **Structure of Science Argumentation: Claim-Evidence-Reasoning**

Scientific arguments represent a social practice whereby individuals attempt to make sense of natural events by constructing plausible explanations for how or why they occur via the interpretation of available data (Berland & Reiser, 2009). In its most simplistic form, scientific arguments follow a Claim-Evidence-Reasoning structure. This structure is used in educational settings to support both student's comprehension of scientific arguments when they read texts as well as helping them to structure their written compositions in science classes (Berland & Reiser, 2009; McNeill et al., 2006; McNeill & Martin, 2011; Songer & Gotwals, 2012).

Claims represent causal assertions or answers to critical science questions. Students conducting a science experiment may be asked to investigate the following question: Under what light conditions do mustard seedlings grow best? After the experiment, students might state that plants grown near a window grew more than those placed in a dark cabinet. Scientific writers attempt to build credibility for their claims by providing and explaining the evidence (e.g., observations from experiments, measurements or calculations, and textual evidence) that was used to arrive at the claim. Our students might cite the average rate of growth in centimeters that

they charted over the course of a week or cite observations that they made about their plant's growth. Finally, writers attempt to convince others of the connection between the evidence and claims by invoking scientific principles, concepts, and facts. This is reasoning. Students might use scientific terms like photosynthesis to describe the process that plants use to grow, of which a requirement is sunlight.

Some research has indicated that younger students struggle with the writing demands of argument texts compared to other genres (Davidson & Berninger, 2016; Kamberelis, 1999; Schleppegrell, 2008) and certain CER components seem to be easier for students to craft than others. In general, claims tend to be the easiest component of an argument as students are more familiar with generating answers to questions or identifying the main idea of a text (Klein & Samuels, 2010; McNeill & Martin, 2011). Articulating this response using causal reasoning and scientific language appropriate to the subject under study can increase the difficulty of crafting claims (Berland & Reiser, 2009). While what counts as scientific evidence can vary, students may have difficulty in selecting and interpreting data sources that are the most appropriate for answering the question (McCann, 1989; McNeill & Krajcik, 2006, 2007; McNeill & Martin, 2011). For example, students may overly rely on statements of plausibility rather than directly cite and interpret an available data source (i.e., explain what might happen given a set of conditions rather than directly cite and interpret observational data from an experiment). Students rarely provide reasonings in their arguments and when they do, they are often conflated with evidentiary statements and so are difficult to identify in writing (McCann, 1989; Sampson & Clark, 2009). Crafting reasons may be especially difficult as it requires students to make logical connections between claims and evidence using relevant content knowledge.

Though students tend to *describe* natural events when asked to *argue* how or why something occurs (Keys et al., 1999; Klein & Rose, 2010), children as young as Grade 5 are sensitive to the genre features and can produce argument texts with modest success (Davidson & Berninger, 2016; McCann, 1989). McCann (1989) asked ninety students (Grades 6, 9, and 12) and twenty-two adults to identify and rate the quality of written arguments. All four groups identified the same three passages, with 80% agreement, as being highly characteristic of argument texts. However, while younger students were as sensitive to argument structure as their older peers, they did not yet possess the same level of knowledge or skill when asked to write their own argument. McCann (1989) noted that the sixth graders produced fewer claims, had greater difficulty citing evidence, and tended to provide no reasons at all compared to older students. Given that a sophisticated knowledge of argument structure may be slow to develop (Davidson & Berninger, 2016), explicit instruction in the components of the CER framework could be productive at improving elementary students' overall science writing performance.

## **Supporting Student's CER Knowledge**

### ***The Role of Metacognitive Scaffolds***

One of the most well researched approaches used to support student's science argument knowledge using the CER framework is the Science Writing Heuristic (SWH; Hand et al., 2004; Hand et al., 2021; Lee & De La Paz, 2021a). The SWH incorporates writing-to-learn activities throughout an inquiry learning unit and behaves primarily as a metacognitive scaffold for reasoning through and connecting lab experiences to scientific cannon (Hand et al., 2004; Akkus et al., 2007). The SWH approach to science argumentation instantiates students within the social practices of the scientific community via student discussion that is supported by CER structured self-questioning (Hand et al., 2004; Hand et al., 2021). Together, students construct, evaluate,

and critique each other's arguments within the context of the lab activity during a unit of study. The writing activities that are embedded within the SWH ask students to summarize what they know and how they have come to know it, primarily by engaging in the cognitive process of argumentation.

High quality implementation of the SWH has been shown to support students' science knowledge, critical thinking skills, and summary writing as examined by the differences in pre and posttest conceptual and multiple-choice questions (Akkus et al., 2007; Hand et al., 2004; Hand et al., 2021). In fact, one study showed that high-achieving students receiving quality SWH instruction outperformed similar achieving peers receiving traditional instruction ( $ES=0.24$ ; Akkus et al., 2007). While discussion-based modes of instruction benefit student learning, it has a greater benefit for students who are already high achieving (Akkus et al., 2007; Rivard, 2004). The achievement gap between higher-achieving and struggling writers is narrower in classes with high implementation of the SWH approach (Akkus et al., 2007). Narrowing it further would require explicit and systematic language instruction for students who struggle to read and write in science.

The writing phase of the SWH requires students to summarize what they learned through discussion with their peers but does not emphasize many opportunities for students to write an argument (Hand et al., 2021). The SWH therefore is not a writing intervention as it does not provide students with enough explicit instruction, practice, or feedback in writing, and improvements in summary quality occur incidentally due to the CER structure and opportunities to write. Rather, the SWH assumes that student's argument-driven discussion and pre-existing writing knowledge are sufficient to meet any writing demands placed on them (Klein & Samuels, 2010). Both students with disabilities and students who are not proficient with writing (which is



80% of the students in the United States; National Commission on Writing, 2003) need explicit instruction to translate what they can say (discuss) to what they write (Berninger et al., 2002). Since most of these students receive their science education alongside their typically achieving peers (National Center for Education Statistics, 2023), explicit writing supports should be incorporated into any unit if the goal is to improve the quality of their written arguments.

### ***The Role of Graphic Organizers to Support Text Structure***

To better support student's science learning, researchers have explored the utility of written scaffolds to support student's argument writing using a CER structure (McNeill et al., 2006, 2007; Klein & Samuels, 2010; Songer & Gotwals, 2012). Scaffolds consist of "temporary supporting structures provided by people or tools to promote learning or complex problem solving" (McNeill et al., 2006). Scaffolds may include graphic organizers, mnemonics, procedural facilitators, and faded instruction. One form of scaffolding that can support student's performance in writing-to-learn activities are content enhancement routines which are structured organizers that facilitate the identification of ideas worth writing about (Bulgren et al., 2009). When content enhancement routines are used in the context of learning academic content, such as in the sciences, students with writing difficulties stand to benefit more readily from writing-to-learn activities since idea generation is not being constrained by working memory (Kim et al., 2019).

Bulgren and colleagues (2009) developed a Question Exploration Routine (QER) to assess whether scaffolded writing instruction could support high school special and general education student's acquisition of science content from an instructional video. A graphic organizer was designed to support student's ability to extract content from the video via the formulation of ancillary questions and answers to the critical question (also referred to as the

main idea). The researchers taught students how to use the QER to write an essay that answered the critical question. The graphic organizer that accompanies the QER reflects the thinking processes valued within science contexts in that it asks students to identify a critical science question and ancillary questions/responses that answer it. Students are encouraged to analyze an information source, and to complete the organizer collaboratively with their peers to answer the critical question. In a study consisting of 36 high school students, both with and without disabilities, treatment students received one 30-minute training in the QER and in a second session used the organizer to analyze a short science video. Students, after receiving instruction, used their organizer to independently write an essay that answers the question, “How do problems with the ozone layer teach us about human effects on our environment?” In contrast, control students were told that they could take notes in any fashion and use them to write a 5-paragraph essay following instruction.

Essays from both groups were evaluated holistically using the 6-Traits Model (IRA=99.1% agreement) and on content (IRA=98.3% agreement). After controlling for pre-test scores, results from ANCOVA show that there was a significant difference between treatment and control groups for both content, [ $F(1, 33) = 15.90, p < .001, d = 0.74$ ], and writing quality scores, [ $F(1, 33) = 17.14, p < .001, d = 1.44$ ]. With respect to writing quality, differences in pre and post test scores was significant, yielding a large effect,  $d = 1.32$ . This study demonstrated that brief explicit instruction in note taking, an important and common writing-to-learn activity in content classes (Graham et al., 2020), could support students in demonstrating what they have learned through writing. Others have also stated that explicit text structure instruction is beneficial for struggling writers in elementary grades who are crafting summaries from informationally laden texts (Mason et al., 2006).

Whether in elementary or secondary classrooms, science students are often required to compose texts using information from multiple sources, thus requiring more cognitive resources than what is required to summarize informational texts (Graham, 2020; Hebert et al., 2018; Klein & Samuels, 2010; Phillips Galloway et al., 2020). When composing a science argument, students may need to draw both from laboratory exercises, media, and text-based sources. Supporting students' ability to learn from such complex writing task requires not only explicit genre instruction, but instructional packages that incorporate explicit strategy instruction and ongoing opportunities to practice (Gillespie-Rouse et al., 2017). A failure to do may result in insufficient time to internalize the appropriate writing strategies that would otherwise enable students to demonstrate what they have learned in the expected disciplinary manner.

Klein and Samuels (2010) hypothesized that a multi-component instructional model that focused specifically on argument writing would increase middle school student's genre knowledge, thus positively affecting students' argument writing quality. Two teachers implemented a multicomponent intervention that explicitly taught students to write argument texts using a scaffolded approach that was faded over time. Past research has indicated that fading scaffolds overtime supports student's internalization of strategies and can improve students' ability to write arguments when scaffolds are eventually removed (McNeill et al., 2006, 2007; McNeill & Krajcik, 2006). Scaffolds included teacher modeling, shared and guided writing, analysis of model texts, frequent opportunities to write, peer to peer discussion, and use of graphic organizers. Treatment groups were characterized as either low or high implementation classrooms, with the low implementation classroom engaging in little content-area writing that required students to supply evidence for their claims. Both groups were compared to a control condition in which argumentation was not a focus.

The researchers hypothesized that texts that were high in argument structure, as indicated by the number of CER moves, would result in texts that were more linguistically/syntactically complex and rich in content (Klein & Samuels, 2010). Following a four-month instructional phase, students engaged in a multiple-source based writing activity describing the controversy around the acceptance of Continental Drift. Students were asked to construct their own argument that answered the questions, “Is Continental Drift theory true?” Students written responses were evaluated for several features including the number of argument moves, overall text quality, source and non-source knowledge units, word and sentence complexity. After controlling for argument and science knowledge, and pretest text quality, results from MANCOVA show that the quality of argument instruction affected the number of arguments moves that students made in their essays but that it did not affect overall writing quality. Subsequent discriminant analysis showed that posttest argument genre knowledge mediated other posttest variables and was correlated with posttest measures such as science knowledge ( $r=.55$ ), argument writing quality ( $r=.33$ ), and knowledge units ( $r=.32$ ). Though the number of arguments moves students included in their writing was correlated with better quality arguments, students use and elaboration of evidence statements most influenced quality ratings.

Typically, rubrics that are used to assess the quality of student’s argument writing using the CER framework fail to provide holistic scores of overall writing quality since the instructional focus is improvement on genre knowledge (McNeill et al., 2007; McNeill & Krajcik, 2006). However, CER specific rubrics can be used to assess the quality with which students execute each component of the CER framework, though not the overall accuracy of the argument (Berland & Reiser, 2009; McNeill et al., 2007). McNeill and colleagues (2007) used a rubric to evaluate students argument writing following an inquiry unit that consisted of

embedded writing instruction. Using a generic CER rubric, the researchers developed a specific rubric that guided raters' assessment of middle school students constructed responses on an end-of-unit assessment. Students were awarded points for accurately including components identified in the rubric. Results of ANOVA show that student gains in CER components had significant effects on their learning as measured by students' performance on multiple choice questions (ES=1.81), constructed response (ES=2.05), and overall posttest (ES=2.34) scores. The authors also demonstrate that students continued to struggle to select appropriate evidence and provide sound reasonings, and that students who had the highest claim scores also tended to include only appropriate forms of evidence in their arguments. The rubric used in this study was useful in assessing what components of CER structure students were improving upon from pre and posttest. It remains to be seen what linguistic and syntactic features, beyond the presence of CER structure, that may be contributing to evaluators ratings of argument quality.

Since students use their knowledge of genre to identify important textual information worthy for inclusion in their written compositions (Klein & Samuels, 2010), it is important for teachers to provide students with scaffolds that help them internalize this knowledge. Past studies (Klein & Samuels, 2010; McNeill et al., 2006) demonstrate that frequent opportunities to write in response to explicit instruction during inquiry and source-based learning is important for students' mastery of argument text structure. In this dissertation, elementary students receive explicit text structure instruction within two different contexts- content-area learning and literacy instruction- using a generic scaffold adapted from Bulgren's (2009) QER graphic organizer. The aim is to assess whether the use of the same graphic organizer across two different approaches to argument instruction- laboratory and text-driven- benefits elementary student's ability to express what they have learned through writing.

Previous research into student's science argument writing describes their arguments as being underdeveloped due to the difficulties in demarcating evidentiary statements and reasons (Berland & Reiser, 2009). Likewise, students need support in crafting arguments that are persuasive, which can be supported via the explicit referencing of data sources and the invocation of established scientific knowledge (Berland & Reiser, 2009). While researcher's have produced writing rubrics for CER (McNeill & Krajcik, 2006, 2007), and have described what constitutes "appropriate", "sufficient", and "complete" evidence and reasons, the rubrics themselves lack this clarity. Furthermore, previous rubrics (McNeill & Krajcik, 2006, 2007) fail to properly delineate CER components that are absent, in development, adequate, or superior due to these levels being clustered within the rubric. For example, the rubric from McNeill & Krajcik (2006) describes both a level 1 and level 2 evidence statement as "Provides appropriate, but insufficient evidence to support claim. May include some inappropriate evidence." The distinction between what constitutes a level 1 versus a 2 is unclear. It is possible that an evaluator's difficulty in distinguishing between a writer's claims and evidence (Sampson & Clark, 2009) is due to a lack of clarity between the components within evaluation tools themselves. Therefore, this dissertation attempts to make improvements upon past rubrics used to assess students CER responses.

While students' arguments will be evaluated for quality using a CER rubric, it is also worth exploring how linguistic features of writing below the textual level might affect rater's perceptions writing quality.

### ***The Role of Word and Sentence Level Supports***

Research evaluating the components of good argumentation have largely focused on improving student's genre knowledge through the implementation of various cognitive supports

used to identify ideas worth writing about (Lee & De La Paz, 2021a). Students who experience difficulties with writing benefit from instruction in lower order language skills that emphasize vocabulary or sentence construction (Block et al., 2019; Truckenmiller & Petscher, 2020; Truckenmiller et al., 2019), yet surprisingly few science writing interventions target these linguistic skills (Lee & De La Paz, 2021a). Due to the diversity of linguistic backgrounds that students bring with them during text structure instruction, it may be beneficial to also implement word and sentence level supports during CER instruction (Mason et al., 2006; Lee & De La Paz, 2021a, 2021b). As students with disabilities typically receive science education in a general education classroom, such linguistic instruction would support their ability to understand and communicate using language structures common in the disciplines (McNeill et al., 2007; Rivard, 2004; Sampson & Clark, 2009).

In addition to the overall argument structure, science writing involves unique language features that set it apart from everyday oral language. Scientific writing is densely populated with academic language features (Snow, 2010). Academic language, also referred to as the language of school subjects such as history or science (Schleppegrell, 2001), consists of structures at the word-, sentence-, and discourse-level that work to increase the density of information within the text as concisely as possible (Snow, 2010). Word-level academic language includes vocabulary that is less-frequently occurring in oral language. Academic vocabulary may consist of Tier 2 (e.g., *distinguish*, *analyze*), which have broad utility across subjects, and Tier 3 words (e.g., *photosynthesis*, *hibernation*), which tend to be discipline specific (Beck et al., 2013). These words may also be morphologically complex and conceptually abstract in nature (Nagy & Townsend, 2012).

Though the vocabulary common in scientific texts can be difficult to learn, students need broad word knowledge to become fluent readers and writers of science texts. Vocabulary instruction that contextualizes a word's meaning using concepts familiar to the student outside of the content area can facilitate students' discussion about science concepts, even as they acquire more robust knowledge of the term (Brown et al., 2010; Brown et al., 2019). In addition, isolated morphology instruction has been shown to be effective at supporting lower order skills associated with phonology, decoding, spelling, and vocabulary learning (Carlisle et al., 2010; Collins et al., 2020; Goodwin & Ahn, 2013) while multi-component interventions incorporating morphology have been shown to support higher order skills, such as reading comprehension (Goodwin & Ahn, 2013). In this study, I incorporate a routine that emphasizes word study across multiple contexts and various features of morphological instruction to support student's vocabulary development.

My previous work assessing various linguistic features of writing and their relation with informational writing quality (Sarmiento et al., 2022) found that long words, as indicated by words with 7 or more letters, was more predictive of Grade 5 ( $r=0.68, p<.01$ ) and Grade 8 ( $r=0.79, p<0.001$ ) informational writing quality than the Academic Word List (Coxhead, 2000), which was found to be insignificantly related to writing quality in both grades. Other studies (Hammill & Larsen, 2009; Hebert et al., 2018; Lee & De La Paz, 2021b; Sarmiento & Truckenmiller, 2024) have also explored the role of long words in writing. While work is still ongoing to unpack the nature of long words in writing (Sarmiento et al, in prep), I suspect that the inclusion of more complex words in writing is indicative of student's knowledge of word affixation. For this reason, students writing will also be assessed using automated scoring software to identify the number of long words that students include in their arguments.



Academic language includes more sophisticated sentence structures, which are necessary to increase the precision of meaning in a text (Troia, 2019; McNamara et al., 2014). And while oral language may follow a story grammar structure like narrative texts, which privilege descriptive or temporal formats that are easier for younger students to understand, academic texts may follow a wide variety of discourse structures (Berman & Nir-Sagiv, 2007). Students with reading and writing difficulties benefit from explicit text structure instruction (Hebert et al., 2018; Reynolds & Perin, 2009). Recognition of these varied structures is facilitated by devices called connectives, which help to explicate the relationships between ideas in a text (Crosson & Lesaux, 2013). Sentence writing interventions, including those focused on science writing, are rare in comparison to the numerous intervention studies aimed at developing students' word or text structure knowledge (Lee & De La Paz, 2021a, 2021b). However, some research has been conducted that suggests that explicit instruction in grammatical components associated with causal reasoning in science contexts can support struggling writer's explanations, at least when the intervention is coupled with other writing supports (Lee & De La Paz, 2021b).

Previous research suggests that children struggle to clearly articulate what constitutes scientific evidence or reasonings as associated with a claim (Berland & Reiser, 2009). While students may produce arguments that sufficiently answer a critical question under study, it can be difficult for readers outside of the immediate instructional context to identify CER components (Berland & Reiser, 2009; Sampson & Clark, 2010). This failure to clearly delineate essential components of an argument calls into question whether the writer is making claims based on inference or on scientific understanding. This can undermine the persuasive element of a scientific argument, thus making it a less effective form of communication.

Based on the idea that connectives play an essential role in “sign posting” different elements of informational text (Crosson & Lesaux, 2012; Uccelli et al., 2015), this study includes a CER Sentence Starters to support students’ production of clear scientific discourse. Because previous research suggests that students struggle with what constitutes appropriate evidence (Berland & Reiser, 2009), these sentence starters make explicit mention of what are good sources of data and how to attribute that information to the source. Students likewise struggle in explaining why their data make sense to answer the main science question. As a result, the reasoning component of the scaffold makes explicit mention that students should use science ideas and terms to make connections between their claims and reasons. The sentence starters also include logical and organizational connectives (Andreev & Uccelli, 2023) to help students interpret their evidence. By placing a focus on the language elements of an argument, the aim of this support is to help teachers explicitly model, and students produce, the discourse moves required to produce a good argument.

Though this study does not instruct students on the fine-grained grammatical components associated with CER, it does attempt to teach students to identify and use phrases that might be signpost these genre moves. One way to understand how students’ sentence construction abilities change over time is via automated software that captures information on the diversity of sentences. Sentence-length diversity is a syntactic complexity measure indicating the degree of variability of sentences in each passage and is reported as the standard deviation of the mean sentence length (McNamara et al. 2014; Wilson et al., 2017). In essence, it reports how dis/similar the sentences in a passage are. Good writers utilize their syntactic knowledge to compose sentences of varying lengths, thus expanding simple sentences into compound, complex, and compound-complex sentences (Troia, 2019). Wilson et al. (2017) found that

sentence-length diversity comprised part of a syntax similarity/variety factor in a three-factor model predicting Grade 6 expository and Grade 8 argumentative writing. Sentence-length standard deviation was found to be variable across the two models, indicating that it may be moderated by grade level though that conclusion may be confounded by the differences in genre. Finally, Truckenmiller & Bowles (2019) found that sentence length diversity was the only sentence-level academic language variable to differentiate between average and poor writers.

### **Purpose and Research Questions**

The purpose of the current study is to support general education teacher's explicit instruction of science argumentative writing through the provision of instructional routines that support the development of students' word-, sentence, and discourse-level language. A logic model for the study is detailed in Figure 1. Specifically, student's science argumentation is improved not only by increasing their genre knowledge through explicit instruction in CER components but also via practice in identifying such components in scientific texts. While possessing knowledge of the genre is necessary for constructing a science argument, it is insufficient when students require additional language supports. Morphology instruction is a research-based way to improve student's vocabulary (Carlisle et al., 2010) and is an essential part of supporting student's science argument writing. Science terms often consist of multiple affixations and knowledge of prefixes and suffixes becomes important in initial comprehension of unknown words (Nagy & Townsend, 2012; Snow, 2010). Vocabulary, morphology, spelling, and sentence instruction have had positive effects on students' reading and science knowledge for a range of student development (typically developing students, students with learning disabilities, and students at-risk of meeting grade level standards). By providing content teachers with instructional routines that support students' a) vocabulary and morphological knowledge, b)

sentence construction, and c) science argument genre-knowledge, I aim to demonstrate that explicit language instruction can have positive proximal effects on students' writing outcomes and science content learning. The current study is guided by the following research questions:

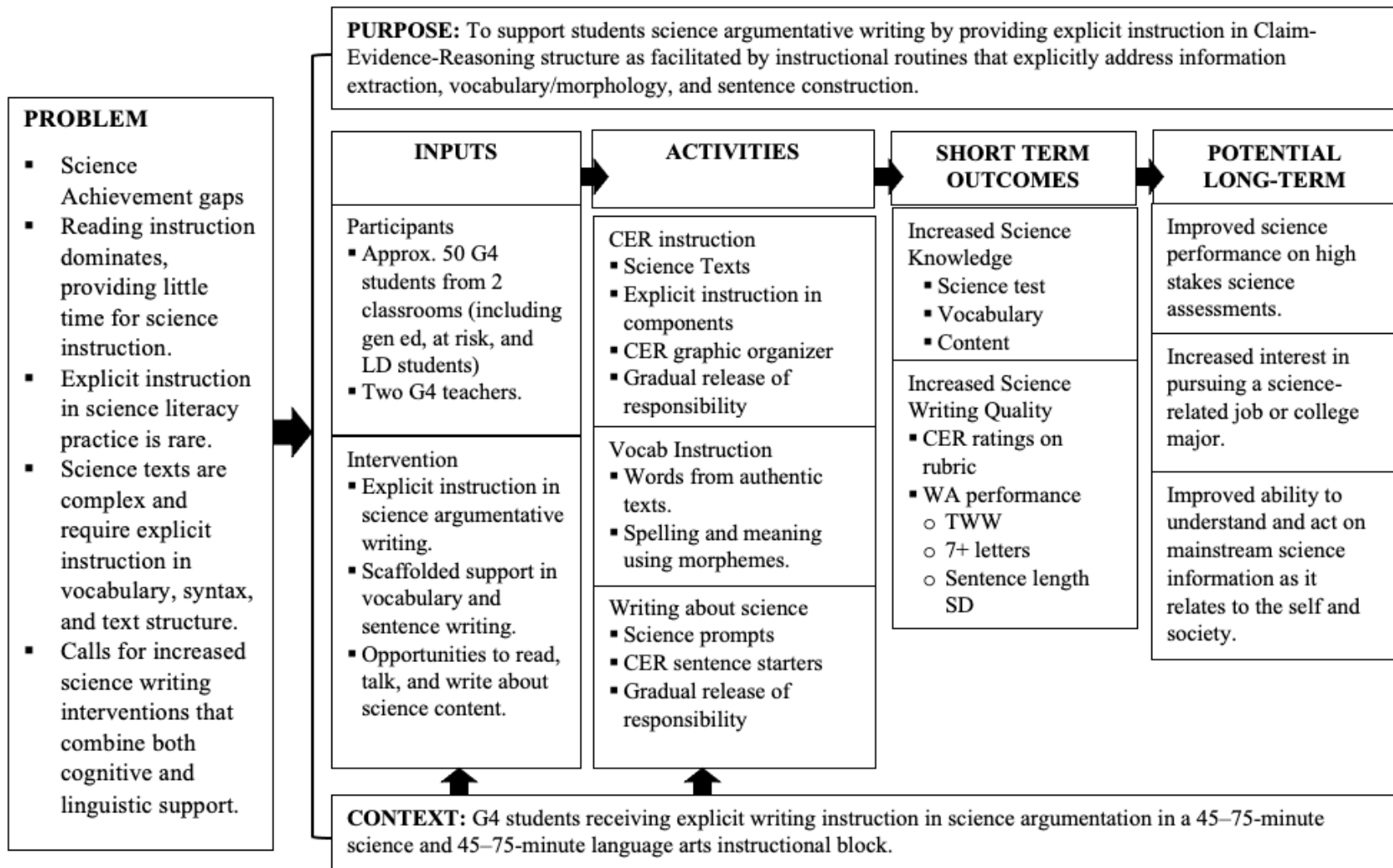
Research Question 1: Does a multi-component intervention incorporating explicit instruction at three-levels of language improve science argumentation (as measured by a CER rubric) of Grade 4 general education students when controlling for student's pretest science argumentation and any significant covariates?

Research Question 2: Does the intervention result in an increase in student's science knowledge as indicated by results on a science test assessing for vocabulary and content knowledge when controlling for students' prior science knowledge and any significant covariates?

Research Question 3: Does the intervention result in an increase in students' writing performance as indicated by the writing general outcome metric, CIWS, and when controlling for students' pretest CIWS and any significant covariate

**Figure 1.**

*Logic model for the Current Study*



## CHAPTER 3: METHOD

### Setting and Participants

This study took place in one elementary school in a Midwestern State. Note that all names are pseudonyms. Aster Elementary is a Grade 3-5 public school located in a rural district, servicing approximately 400 students (CCD Public School Data, 2023). Students at this school are predominantly white, accounting for 86.5% of the student body. 7% of students are Hispanic, 1.3% are Black, 0.5% are Indigenous, and 5% of students belong to two or more racial categories. 56% of students qualify for free or reduced-price lunch. During the prior school year, 44.6% of Grade 4 students within the district performed proficiently on the English Language Arts (ELA) end-of-year state assessment compared to 43.4% of Grade 4 students in the state. Only students in Grade 5 take the Science end-of-year state assessment. In this district, 41.7% of students scored proficient on this assessment compared to the 38.9% for the state. Per the school's Annual Education Report, Aster Elementary was not identified as a school target for improvement as its students perform well, and demonstrate growth year over year, on ELA and Math assessments.

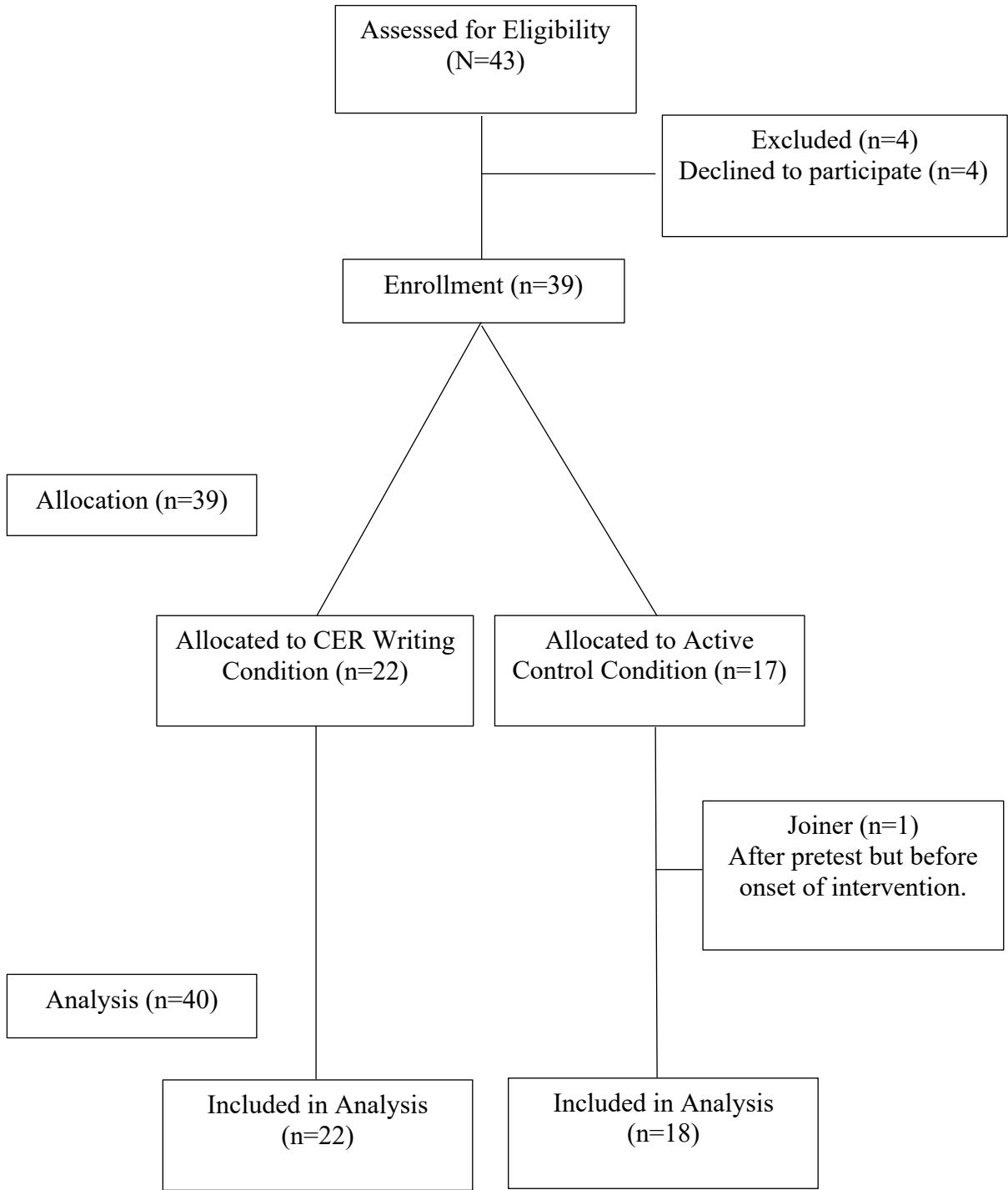
Two Grade 4 teachers, operating as a co-teaching pair, participated in this study. Both teachers identify as white. The first teacher, Mr. Frizzle, is primarily responsible for teaching history, mathematics, and science. Mr. Frizzle has twenty-one years of teaching experience with endorsements in all subjects for Grades K-5 and an endorsement in biology for Grades 6-8. He was also a professional biologist prior to his teaching career. His interests in biology and his teaching experience have resulted in extramural collaborations in curriculum design with universities and the National Park Service. Given his extensive professional experiences, Mr. Frizzle excels at designing problem-based learning opportunities for his students. Mrs. Honey is

primarily responsible for teaching ELA, including writing instruction. She has twenty-eight years of teaching experience. Her bachelor's degree is in Special Education with emphases on cognitive and emotional impairments. Mrs. Honey has attended numerous professional learning conferences that have focused on the incorporation of technology and project-based learning in the classroom. She has also participated in researcher-teacher summer institutes where her work included the integration of ELA standards with science curricula, a topic for which she wrote her thesis for her recently earned master's degree in literacy education. In summary, both teachers participating in this study are highly trained in science and literacy instructional practices.

This study also consists of two classrooms of student participants, referred to as Homerooms. A research brief was sent home to parents during parent teacher conference night explaining the study and 39 of the possible 43 students consented to participate. One student joined the study after administration of the pretest but prior to the start of the intervention and joined Mrs. Honey's Homeroom, which was allocated to the active control condition and described later in the Teacher Co-development section of this manuscript. There was no attrition in this study. 40 students participated in total. A CONSORT diagram depicting the flow of participants throughout the study can be found in Figure 2 (Moher et al., 2005).

**Figure 2**

*CONSORT Diagram Showing Passage of Participants in the Study*





The student participants in this sample ranged in age from 9 to 11 years old, with a mean age of 9 years and 10 months old. 55% of the participants were female, 97.5% spoke English as their native language, 87.5% did not have disability, and 45% of students did not receive free and reduced-price lunch. With respect to students' racial identity, 85% of students identified as white. Four students identified as biracial, one student identified as Asian, and one student identified as Hispanic. Student demographic information as broken down by Homeroom can be found in Table 1.

**Table 1.***Student Demographic Data by Condition*

	All Students <i>N</i> =40	Mr. Frizzle Homeroom <i>n</i> =22	Mrs. Honey Homeroom <i>n</i> =18
Gender			
Girl	22	10	12
Boy	18	12	6
Race			
Asian	1	0	1
Hispanic	1	0	1
Biracial	4	3	1
White	34	19	15
English Native			
Yes	39	21	18
No	1	1	0
Disability			
Yes	5	3	2
No	35	19	16
Free-Reduced Lunch			
Yes	22	14	8
No	18	8	10

*Note:* Mr. Frizzle Homeroom assigned to CER instruction condition; Mrs. Honey Homeroom assigned to Active Control condition.

## **Materials**

### *Assessment Materials*

**The Writing Architect Web Application.** The Writing Architect (WA) web application is a group-administered curriculum based measure for written expression designed to support teacher's instructional decisions with regard to student's current writing instructional needs (Truckenmiller et al., 2019). It consists of a web-based writing platform on the front end for students and a scoring platform on the backend for teachers and researchers (Truckenmiller et al., 2019). As an online tool, the WA is programmed with a series of source-based informational, narrative, or persuasive type writing tasks. During WA administration, students are given a paper packet containing a planning sheet and the writing passage that has been assigned to them. At the start of the WA administration, the students listen and follow along as the passage is read to them. Next, students are given three minutes to plan their response to the writing prompt provided. Once the planning period is over, the screen advances to a large text box where students have fifteen minutes to draft their response. Responses can be submitted early. They receive no automatic spelling or grammar support as that feature is disabled in the WA platform. Once students have submitted their responses, or the screen advances after 15 minutes, students complete a 90 second typing fluency task. This administration format was found to significantly predict students' writing achievement (Truckenmiller et al., 2019).

On the backend, passages are scored for a variety of expressive word- and sentence- level language features predictive of successful academic writing (Sarmiento et al., in review; Truckenmiller et al., 2022). While some scores are auto scored by the platform's programming (e.g., word and sentence complexity), other scores (e.g., word and sentence accuracy) must be scored by human raters.

**Science Argument Prompt.** In this study, the text-based writing prompts were adapted from Troia et al. (2020) which ask students to defend their position on a question. Science argumentative writing blends informational and persuasive writing since it is a sensemaking practice which uses writing to support the development of new knowledge based on the interpretation of evidence (Berland & Reiser 2009; NRC, 2012). Specifically, the reasoning component of CER serves to persuade the reader of the validity of their answer to a science question by rooting it and the writer’s interpretation of the available evidence in established scientific principles (Berland & Reiser, 2009). In this manner, writers are attempting to convey knowledgeability of a subject while simultaneously attempting to establish their own credibility. As a result, the opinion prompt by Troia and colleagues (2020) was adapted using principles of science argumentative writing (Berland & Reiser, 2009; McNeil & Krajcik, 2006, 2007; McNeil & Martin, 2011). The demands of this task are like the writing performance assessments on the MSTEP, an end of year assessment developed by the Smarter Balanced Assessment Consortium (Regents of the University of California, 2022), which asks students to write an essay incorporating information from a source text.

In the current study, students answer a question in response to a scientific article and they are prompted to respond argumentatively. For example, in response to the article How to Speed Up Extinctions, students are instructed to, “Write a scientific argument that answers the question below. Remember, a good science argument (1) clearly states your claim, (2) gives detailed facts to support your claim, (3) uses science ideas to persuade the reader that your facts support your claim, (4) has a conclusion that helps the reader understand why they should agree with your answer, and (5) follows the rules of writing.” In this example, students would be responding to

the question, “How have human activities caused many plant and animal species to go extinct?”

The writing prompts developed for this study can be found in Table 2.

**Table 2.**

*Writing Prompts Produced for the Science Texts*

Text	Prior Knowledge Prompt	Writing Prompt
How to Speed Up Extinctions <sup>a</sup>	Please tell me everything you know about <b>animal extinctions</b> . You may write your answer as bullet points.	Write a scientific argument that answers the question below. Remember, a good science argument (1) clearly states your claim, (2) gives detailed facts to support your claim, (3) uses science ideas to persuade the reader that your facts support your claim, (4) has a conclusion that helps the reader understand why they should agree with your answer, and (5) follows the rules of writing.  How have human activities caused many plant and animal species to go extinct?
A Diet for Invasive Carp <sup>a</sup>	Please tell me everything you know about <b>non-native plants and animals</b> . You may write your answer as bullet points.	How might an invasive species be successfully introduced to new environments?
CO <sub>2</sub> -loving Plants	Please tell me everything you know about <b>climate change</b> . You may write your answer as bullet points.	How are plants useful in the fight against climate change?
Fertilizer for Rooftop Gardens	Please tell me everything you know about <b>pollution in cities</b> . You may write your answer as bullet points.	How might people use plants to fight air pollution in cities?
Multi-tasking Windmills	Please tell me everything you know about <b>windmills</b> . You may write your answer as bullet points.	How do people use technology to solve difficult problems while also protecting the environment?

*Note:* The directions immediately following the unique context provided for Invasive Carp, Plants, Rooftop Gardens, and Windmills is the same as that provided for the text Extinctions. Directions have been removed from Invasive Carp, Plants, Rooftop Gardens, and Windmills to improve the readability of the table.

<sup>a</sup> Prompts selected for counterbalancing and subsequent administration in the Writing Architect platform.

**Science Passages.** Though the current iteration of the Writing Architect contains informational passages which could arguably be considered science-related (e.g., The passage Swat Up discusses the importance of flies), those passages were selected for their informational richness, not necessarily for their science curriculum-aligned content. For this project, science prompts related to academic standards for Grade 4-5, as outlined by The Next Generation Science Standards (NRC, 2012), were selected and adapted for use as pre and posttest measures. The standards alignment for the science passages used in this study can be found in Table 3. Passages were selected from ScienceNews.org, which is a common news source used by education sites like NewsELA.com or ReadWorks.com for adapting news reports for a child audience.

Science news articles were initially selected because they represented interesting topics and were related to science content addressed in academic standards. The automated scoring software, Coh-Metrix (McNamara et al., 2014) was used to assess the readability of the unmodified science passages. Before modification, the science passages had the following readability metrics: TWW= 589, sentence length (*SD*)=18.4(10.9), percent narrativity=19.5, Flesch-Kincaid grade level=10.35, Lexile range= 1210-1400. The passages were adapted so the average readability metrics of the science passages fell within the range of the average readability statistics of the Common Core State Standards grades 4-5 grade band (Nelson et al., 2012) and of the Coh-Metric Indices Norms for G4-5 Science texts (see Table 4; McNamara et al., 2014). The final texts used in this study are in Appendix A.

**Table 3***Science Passage Standards Alignment Using the Next Generation Science Standards*

Text	Science Area	Disciplinary Core Idea	Cross Cutting Concept
How to Speed Up Extinctions	Earth and Space Science	<p><b>ESS3.C: Human Impacts on Earth Systems</b> Human activities in agriculture, industry, and everyday life have major effects on the land, vegetation, streams, ocean, air, and even outer space. But individuals and communities are doing things to help protect Earth's resources and environments.</p>	<p>Systems and System Models A system can be described in terms of its components and their interactions.</p>
A Diet to Fuel Invasive Carp	Life Science	<p><b>LS1.C: Organization for Matter and Energy Flow in Organisms</b> Food provides animals with the materials they need for body repair and growth and the energy they need to maintain body warmth and for motion. (Secondary to 5-PS3.D)</p> <p><b>LS2. A. Interdependent Relationships in Ecosystems</b> ... Organisms can only survive in environments in which their particular needs are met. A healthy ecosystem is one in which multiple species of different types are each able to meet their needs in a relatively stable web of life. Newly introduced species can damage the balance of an ecosystem. (5-LS2-1)</p>	<p>Energy and Matter Energy can be transferred in various ways and between objects.</p> <p>Systems and System Models A system can be described in terms of its components and their interactions.</p>
CO <sub>2</sub> -loving Plants	Life Science	<p><b>LS1.C: Organization for Matter and Energy Flow in Organisms</b> Plants acquire their material for growth chiefly from air and water. (5-LS2-1)</p>	<p>Energy and Matter Matter is transported into, out of, and within systems.</p>



**Table 3 (cont'd)**

Fertilizer for Rooftop Gardens	Life Science	<p><b>LS1.C: Organization for Matter and Energy Flow in Organisms</b> Plants acquire their material for growth chiefly from air and water. (5-LS2-1)</p>	<p>Energy and Matter Matter is transported into, out of, and within systems.</p>
Fertilizer for Rooftop Gardens	Earth and Space Science	<p><b>ESS3.C: Human Impacts on Earth Systems</b> Human activities in agriculture, industry, and everyday life have major effects on the land, vegetation, streams, ocean, air, and even outer space. But individuals and communities are doing things to help protect Earth's resources and environments.</p>	<p>Systems and System Models A system can be described in terms of its components and their interactions.</p>
Multi-tasking Windmills	Earth and Space Science	<p><b>ESS2.A: Earth Materials and Systems</b> Earth's major systems are the geosphere, hydrosphere, atmosphere, and the biosphere. These systems interact in multiple ways to affect Earth's surface materials and processes. (5-ESS2-1)</p> <p><b>ESS3.C: Human Impacts on Earth Systems</b> Human activities in agriculture, industry, and everyday life have major effects on the land, vegetation, streams, ocean, air, and even outer space. But individuals and communities are doing things to help protect Earth's resources and environments.</p>	<p>Systems and System Models A system can be described in terms of its components and their interactions.</p>

**Table 4.***Readability Statistics for the Development of Science Passages*

Text Corpora	Text Name	Word Count	Sentence Length (SD)	Narrativity	Flesch-Kincaid Grade Level	Lexile Range <sup>b</sup>
Writing Architect Grade 5 Informational Passages <sup>a</sup>		622	11.89 (5.67)	37.51	7.02	770-960
Coh-Matrix Indices Norms G4-G5 Science New Grade 5 Science Passages <sup>a</sup>		273.18	11.03 (4.34)	40.81	4.84	
		597.2	11.77 (5.99)	34.59	5.89	810-1000
<b>WA Science Passages</b>						
Grade 5 Science Passages	How to Speed Up Extinctions	569	11.6 (5.83)	37.07	5.99	810-1000
Grade 5 Science Passages	A Diet for Invasive Carp	599	11.98 (5.73)	28.1	5.88	810-1000
Grade 5 Science Passages	CO <sub>2</sub> -loving Plants	609	11.94 (6.51)	44.83	5.76	810-1000
Grade 5 Science Passages	Fertilizer for Rooftop Gardens	626	11.18 (5.93)	20.9	5.88	810-1000
Grade 5 Science Passages	Multi-tasking Windmills	583	12.15 (5.99)	42.07	5.95	810-1000

<sup>a</sup> Values represent the average readability statistics for texts in the corpora,  $n=5$ .

<sup>b</sup> Lexile Range is calculated using the Lexile text analyzer tool from [lexile.com](http://lexile.com) and is based on the first 500 words of the passage.

### ***Instructional Materials***

**Morphology/Word Learning Graphic Organizer.** For a given word, students are asked to write the word down. Students are then asked to cover the word and sound it out. Once students have spelled the word, students are asked to identify any morphemes (root, prefix, and suffix) and to use the morphemes to predict the meaning of the word. They write their own definition and compare it to a dictionary definition. As students engage in the unit, they may encounter the target word across various texts. They are asked to examine how their word is used across various contexts, including whether it is used as a different part of speech (noun, verb, adjective, process). Students are asked to craft a definition for their word as used in context. The routine also asked students to identify synonyms and words related to it within a semantic network. Finally, after extensive interaction with the target word, students are asked to craft their own definition and to draw a picture that represents the word's meaning. The morphology routine is in Appendix B.

**Claim-Evidence-Reasoning Graphic Organizer.** In the current study, I adapted the organizer from Bulgren et al., (2010) to support elementary student's construction of a scientific argument using a claim-evidence-reasoning framework. The CER graphic organizer asks students to identify a central science question for a given text or scientific investigation as well as key vocabulary terms that may be key in answering the question. To support student's abilities to identify arguments in text or to craft their own in response to their own inquiries, the CER graphic organizer includes fields where students are prompted to identify claims, evidence, and reasons. Finally, at the conclusion of their text analysis or experimentation, students are asked to answer the main science question through the synthesis of information and vocabulary identified

in the organizer. The CER Organizer, entitled Let's Argue- For Science, can be found in Appendix C.

**Claim-Evidence-Reasoning Sentence Starters.** The CER Sentence Starters (see Appendix D) consist of four sections- claims, evidence, reasons, and an example response. For each component of CER, there is a child friendly definition, a series of sentence starter phrases, and examples of them in use. The final section includes an example of a response to a science question that makes heavy use of the sentence starters.

## **Measures**

### ***Covariate Measures***

**Gender.** One underpowered study found that the gender contributed 7% of the variance in Grade 4 informational writing quality and recommended controlling for it (Truckenmiller et al., 2021). A large meta-analysis (Reilley et al., 2019) analyzing gender differences from three decades of reading and writing data found that the gender difference in writing between boys and girls was greater (Cohen's  $d=-.42$ ) than that for reading (Cohen's  $d=-.19$ ). Riley and colleagues (2019) found that gender differences were smallest in Grade 4 and widened over time. With respect to science achievement, one meta-analysis (Voyer & Voyer, 2014) found that a "female advantage" in academic achievement existed, but it was smaller ( $d=.15$ ) compared to that in language arts ( $d=.25$ ). Given the nature of the writing prompts in this study, I also controlled for gender.

**Topic Prior Knowledge Task.** Prior knowledge plays a role in the comprehension of informational text (Cervetti et al., 2020), in writing (Phillips-Galloway et al., 2020), and in the selection of appropriate evidence when crafting an argument using CER (McNeil et al., 2007). High topical prior knowledge is predictive of reading comprehension and may operate as a

compensatory mechanism for students who struggle to understand what they read (Cervetti & Wright, 2020; Miller & Keenan, 2009). As a result, it is possible that student's posttest CER gains may be the result of their knowledge for the topic rather than due to the intervention itself. It therefore becomes necessary to control for a student's prior knowledge to determine the effect of an intervention on learning (Benedek-Wood et al., 2014; Phillips-Galloway et al., 2020; Gillespie Rouse et al., 2017; Samuel & Klein, 2010; Cervetti et al., 2016; Cabell & Hwang, 2020).

A knowledge unit is an accurate piece of information about a topic (Brown & Day, 1983). Students had five minutes to "tell me everything you know" about the given subject and were permitted to draft their responses in a note format (see Appendix E for administrative script). The prompts for the Topic Prior Knowledge questions can be found alongside the Writing Architect prompts in Appendix A. Using a previously validated codebook, the number of relevant knowledge units' students wrote in response to a topic was counted (Cronbach's  $\alpha = .70$ ; Cervetti et al., 2016, 2020; Cabell & Hwang, 2020). Past studies demonstrate that the reliability of this measure for exact rater agreement is adequate (Benedek-Wood et al., 2014; Cervetti et al., 2016) and within 1-point agreement is excellent (90.4%) (Benedek-Wood et al., 2014).

### ***Outcome Measures***

**Claim-Evidence-Reasoning.** The CER rubric for this study was modified and expanded upon from previous studies (Berland & Reiser, 2009; McNeill & Krajcik, 2006, 2007) and provides a distinction between evidence and reasons, and whether a component is absent, in development, adequate, or superior. For example, a student's evidence statement may receive a score of 1 (in development) if there was an evidence statement present but it utilized irrelevant

data or relied on logic statements of plausibility. A score of 2 (adequate) indicates that the student referred to an available data source but did not fully describe it or attribute the information to a specific source. A superior response (score of 3) indicates that the writer explicitly mentions, describes, and attributes evidence to a source. The distinctions between a 2 and 3 reflects the use of persuasive elements (e.g., source attribution, data description, and appeals to authority) necessary to distinguish a science argument from a science explanation (Berland & Reiser, 2009). The rubric can be found in Table 5. A previous study using a CER rubric to evaluate Grade 7 student science argument writing report excellent inter-rater reliability for assessing claims (IRR=0.98), evidence (IRR=0.94), and reasoning (IRR=0.98) components (McNeill & Krajcik, 2006). Past studies have demonstrated a significant and moderate relationship between middle school student's CER science content knowledge scores, with improvements in both multiple-choice (ES=1.81) and constructed response (ES=2.05) questions (McNeill & Krajcik, 2007).

**Table 5***Claim-Evidence-Reasoning Scoring Rubric*

Component	0- Absent	1- In Development	2- Adequate	3- Superior
Claim	Does not provide an answer to the science question under study OR provides an answer to an unrelated science question.	Provides an answer to the science question but it is inaccurate.	Students provide an answer to the scientific question under study, and it is partially accurate.	Students provide an answer to the scientific question under study, and it is accurate.
Evidence	Does not provide any evidence to support their answer to the science question under study.	Uses inappropriate evidence to support their claim; Relies on logic or plausibility to support their answer to the science question under study instead of available data sources.	Selection of evidence is appropriate but references to the data sources are vague with no explicit source attribution.	Explicitly mentions one or more appropriate data sources to support their answer to the scientific question under study; Presents numerical data similarly to source material to imply empiricism.
Reasoning	Does not provide any connection between their stated claim and evidence.	Provides a generalization of how the available evidence helps to answer the question but does not define or discuss any science terms or ideas.	Utilizes scientific ideas, vocabulary, and logical inferences to help explain why the available data counts as evidence and is important in answering the science question under study.	Utilizes persuasive elements (such as appeals to authority, scientific theory, contrasting alternative views) to convince the reader that their answer to the science question is the most plausible one of many

**Note:** Rubric adapted from McNeil & Krajcik (2006, 2007).

**Science Knowledge.** A curriculum-embedded science knowledge unit test was administered to students in both the treatment and control conditions at two time points, pre and posttest. The assessment was developed in partnership between the researcher and Mr. Frizzle, utilizing the web-based testing platform *Pear Deck* (formerly *Edulastic* during the time of this study). *Pear Deck* is an assessment platform that features ready-made standards-aligned assessments and a large data bank of questions to assist districts, schools, and teacher's progress monitoring in a range of school subjects (Pear Deck Learning, 2024). Questions can be sorted by grade, school subject, standard, and depth-of-knowledge to create assessments that are aligned with a current unit of study. The decision to administer the science test via *Pear Deck* was due to two factors: 1) the teacher's current science curriculum did not have a summative assessment available and 2) students did not typically take assessments in science class (pen and paper or otherwise) but were familiar with *Pear Deck* as a testing platform for math class and district-wide assessments. The assessment aligns closely with the unit of instruction planned for this study, which covers topics on weathering and erosion, and is comprised of questions from the *Pear Deck* question bank that address the relevant science standards. Originally, questions 14 and 15 were planned to be short-answer questions but, given that students do not tend to take assessments in science class and are not typically tasked with questions of this type, the teachers determined that these questions would be unduly burdensome to students and would take too much class time to complete. These questions were replaced with multiple choice questions that required greater knowledge to complete, including one item-sorting and one data-driven question. The teachers reviewed the assessment and determined it to be content valid.

This assessment, located in Appendix F, consists of 15 multiple-choice items representing a range of depth-of-knowledge levels; it is worth a maximum of 26 points. Internal



consistency of the assessment- the extent to which the test items collectively and reliably measure a construct (student's knowledge of weathering and erosion)- was calculated using Cronbach's alpha (Taber, 2018). The coefficient for the assessment at pretest was 0.42 and 0.69 at posttest, which I determine to be acceptable given that the average teacher-made assessment has a reliability of 0.5 (Frisbee, 1988).

**Correct Incorrect Word Sequence (CIWS).** Correct Incorrect Writing Sequences (CIWS) is a curriculum-based measure of writing achievement with moderate correlations ( $r=.55$ ) to criterion assessments of writing, such as the TOWL-3 or Woodcock-Johnson (Romig et al., 2017). It has been conceptualized as a writing fluency measure (Kim et al., 2019) whose development in late elementary, included Grade 4, shows strong correlations with reading fluency ( $r=.54$ ) and reading comprehension ( $r=.50$ ) (Tortorelli & Truckenmiller, 2024) . A correct word sequence is represented by the correct spelling or capitalization of adjacent-words or the correctness of adjacent word and punctuation marks and is denoted by carrots, such as in the following example, ^The^student^is^writing^. An incorrect word sequence results when an error is made between adjacent word strings (as denoted by an x) such as in the following example, xthe^student^is^ritingx. CIWS is calculated by subtracting the incorrect word sequences from the correct ones. In the above example, the CIWS score would be -1. CIWS is scored on the backend of the Writing Architect platform, using a manual, by trained coders who have achieved  $\geq 90\%$  interscorer reliability. The manual can be accessed via an Open Science Framework repository (<https://osf.io/tfvx2>).

### ***Potential Post-hoc Intervention Fidelity Measures***

The intervention package included instructional scaffolds to support student's abilities to use more complex academic vocabulary and to construct sentences that deployed science-like

language. Due to the slow developmental nature of writing (Troia et al., 2019; Valentine et al., 2021), it may be difficult to detect group differences on larger outcome measures. The following variables will be used to assess changes in more fine-grained writing skills that contribute to overall writing.

**7+ Letters.** In this study, students' responses will be automatically scored for the number of long words that students wrote by the Writing Architect computer application. Because it is automatically scored, interrater reliability is not needed. In a previous study, the correlation of long words with writing quality was high ( $r=0.68$ ) in Grade 5 (Sarmiento et al., 2022).

**Sentence-length diversity.** In this study, sentence-length diversity is automatically captured in the Writing Architect as the standard deviation of sentence length. This metric has been shown to be a key feature of academic text (McNamara et al., 2014), but has not been a significant predictor of writing quality for students in late elementary grade levels (e.g., Wilson et al., 2017; Sarmiento et al., 2022). In this study, it will be an exploratory outcome. Because it is automatically scored, interrater reliability is not needed.

## **Experimental Design**

This study is a pilot study using a quasi-experimental pre/posttest design with student homeroom assigned to treatment (science writing intervention package) and an active control (AC) condition. Both sections have the same teacher for ELA and the same teacher for science. Teachers will administer both the treatment and AC conditions to the assigned sections, thus preventing the confound of one teacher for the treatment and one for the control.

## ***Typical Instruction***

Typical instruction in ELA consists of two 1-hour blocks. During Block 1, Mrs. Honey provides ELA instruction on modules provided by EL Curriculum. EL Curriculum is a

curricular collaboration between Harvard Graduate School of Education and Outward-Bound USA (EL Education, 2024). EL Curriculum is designed to support children's content knowledge while learning to read, write, analyze, and discuss text. For example, the poetry module that preceded the intervention for this study was integrated with social studies content in that students analyze how the Civil Rights Movement inspired works by two poets. In Grade 4, there are four modules. Each module consists of multiple units that last 6-8 weeks. Each unit is guided by overarching questions that help to focus the learning for the unit. EL Curriculum provides the teacher with scripts, teaching resources, and student materials, including trade books and graphic organizers.

During ELA Block 1, instruction consists of whole group direct instruction before setting the day's work expectations. Students work in flexible groupings, ranging from small groups of 3-4 students to pair and individual work. As students work independently, Mrs. Honey circulates the room and provides individualized support. For example, she may direct students to personalized dictionaries or word walls to select "stronger" words that convey more meaning or imagery in their poems. Mrs. Honey indicates that she often creates her own graphic organizers to supplement the materials provided to her by her curriculum because students, especially those who receive special education services, need additional writing support despite the strong emphasis on teacher modeling.

ELA Block 2, also called All Block in the EL Curriculum, is an opportunity for students to receive Tier 2 services specifically in the areas of grammar and mechanics, vocabulary, and independent reading. For example, on one pre-intervention observation, Mrs. Honey started the period with a review of the /o/ and /ow/ sounds. After approximately 10 minutes of this warm-up exercise, Mrs. Honey gave students directions for the tasks they were to work on during their

centers. Centers include work that students do individually, with a table partner, and with Mrs. Honey herself. Students rotate centers about every 10-15 minutes. Work with Mrs. Honey usually involves classwork from the previous block.

At the end of ELA Block 2, students will go to art or PE classes and then lunch. At the end of lunch, Homerooms will switch. Mrs. Honey's class will receive math instruction with Mr. Frizzle while his homeroom receives both blocks of ELA instruction.

The last hour of the day is dedicated to science instruction and both classrooms merge during this hour. Groups sizes typically include four students, two from each Homeroom. The science curriculum used is Mystery Science. In Mystery Science, Grade 4 students are prompted through the investigation via a video. Learning is anchored by a phenomenon, an observable natural occurrence that prompts questions. Periodically, there are discussion opportunities for students to process their thinking with their peers and make predictions before the lesson advances. The investigation itself is very directive: students repeat the exact steps as they are prompted. During this block, Mr. Frizzle and Mrs. Honey's partnership roles become clear. They both facilitate the day's learning by asking probing or clarifying questions of the students. Mr. Frizzle takes point on the investigation and explanation of science concepts while Mrs. Honey spearheads the reading, writing, and elaborative discussions.

The teachers indicate that they usually do not have time at the end of the science hour to respond to the questions at the end of the science unit. Typically, students will do four science lessons a week but receive few opportunities to write. On a teacher survey of writing, Mr. Frizzle reports that he spends approximately 30 minutes a week on any type of writing instruction, 40 minutes a week on morphology instruction, and about one hour a week on revision. This writing instruction typically occurs in math. Mrs. Honey reports that she spends approximately 60-75

minutes on writing instruction in any given week. Depending on the phase of writing that students are in, a typical breakdown in time spent teaching writing skills breaks down as follows: 20 minutes of keyboarding, 10 minutes of spelling, 30 minutes of grammar/mechanics, 20 sentence combining or expanding, 60 minutes on vocabulary or morphology instruction, 10 minutes on goal setting, 30-45 minutes on planning, and 30 minutes on revising.

### ***Active Control***

Because typical instruction is not possible given the design of this study, students not receiving the CER intervention will be placed in an active control condition instead. To maintain equity between both classrooms and to introduce writing to the science block, all students regardless of condition will receive an additional 2-3 hours of writing time a week in science class. One science lesson will occur over two days: one day for the investigation and one day for the end of lesson questions. The Active Control (AC) condition will use a previously learned informational text structure from the EL Curriculum (RACES; **R**estate background, **A**nswer the question, **C**ite your evidence and **E**xplain, **S**ummarize) to respond to science questions posed to them in their ELA and Science classes. They will not have any of the science writing scaffolds developed for this study available to them.

### **Procedures**

#### ***IRB and Informed Consent***

After applying for exempt status with Michigan State University's Internal Review Board, forms for teacher consent and parental consent for their students to enroll in the study were distributed(found in Appendices G and H). Teachers received a \$350 honorarium following the completion of the study, with funds provided via Michigan State University's Hard Cost Dissertation and Practicum Support (HCDPS) fellowship.

### ***Implementation Fidelity***

To ensure that contamination across the two conditions did not occur, I observed each teacher whether the teachers utilized any of the science instructional scaffolds during the lesson. I used an observational tool that monitored for four key instructional practices including, explicit instruction, opportunities to respond, feedback, and student engagement (Truckenmiller et al., 2023). Each homeroom was observed four times throughout the course of the intervention (during weeks 2, 4, 6 and 7), with two observations occurring in science and ELA. Observations were audio recorded and detailed field notes were taken. I noted when students received instruction on vocabulary, sentence structure, and text structure as well as whether the science writing scaffolds, or some other instructional material was used.

Based on these observations, AC students did not receive instruction using the CER scaffolds. Instead, when AC students were in Language Arts, they used the RACES strategy to identify the main idea and supporting details for the texts that they read, which were related to the topic of animal defense mechanisms. Mrs. Honey provided extensive teacher modeling, call and response, and feedback when using the graphic organizer in a group setting. She likewise engaged in similar behaviors when working with students during Tier 2 instruction. When AC students were working with Mrs. Honey during the science block, they used the language of the RACES strategy to construct sentence frames about the purpose of the experiment they conducted the previous day, what they observed, and how that informed them about the concepts of weathering and erosion. Mrs. Honey relied on extensive teacher modeling, using an “I do-we do-you do” model to draft these sentence stems over the course of the science block. Students engaged in independent practice for the last 15 minutes of the science block to complete the end-of-lab questions. When AC students received science instruction with Mr. Frizzle (recall that

teachers provide instruction to both homerooms throughout the day and homerooms alternated teachers during the science writing block), they tended to not use either the RACES strategy at all. Rather, Mr. Frizzle was likely to engage in dialogic instruction where he posed a series of questions to students and the resulting classroom dialogue was recorded on the white board. Students would engage in a whole class discussion followed by individual or pair work to construct a response to a given science question. After a designated period, students would share out answers and the teacher would provide corrective feedback and/or praise. The class repeated this process until the science writing task, or the period, was complete.

It is worth noting that there is some overlap between the language used in the RACES and CER acronyms. The Active Control group, when in Mrs. Honey's classroom, made extensive use of the RACES graphic organizer. During an observation on Week 6, students in the AC condition were observed preparing written responses to an end of unit problem-based prompt. Students used the RACES organizer to help them map out their responses. The language used for the C+E portion of their organizer reads, "cite your evidence and explain it" which has parallels to what students are tasked with when "citing their evidence" from data sources and "provide reasons" to explain the how the evidence makes sense. Despite this overlap, there was no additional scaffolds to support Active Control students in determining what might count as evidence and from what knowledge sources they might build their rationale. No CER scaffolds were present in the room. When the Active Control condition was observed in Mr. Frizzles' classroom, extensive teacher modeling and group discussion were the primary tools used to support students' writing. The RACES strategy was not explicitly used. It was common practice for student's to be asked to explain their evidence for statements made during their discussions, but no CER scaffolds were used to support students in making them.

The Treatment group were observed using the CER sentence starters during all four observations, regardless of whether these observations were made in the ELA or Science classroom. Students had these supports in their science journals, which traveled with them as they switched Homerooms. Students were observed using them and teacher's made frequent references to them when modeling their writing. For example, Mrs. Honey modeled her think-aloud when selecting a sentence frame to start her response to a science question. Mr. Frizzle referred to the first page of the Sentence Starters when he needed to define a CER element that he was going to construct (e.g., "A claim is the answer to a science question. What's our question again?") before selecting his preferred sentence frame. This method is a more implicit form of instruction as students are relying on intuiting what constitutes a claim based off the sentence frame itself. Both teachers had students box their claims, underline their evidence statements, and circle their rationales, which demonstrates a clear example of teacher collaboration when using the scaffolds. Teachers also elicited student responses when constructing their own CER responses during science class. It should be noted that since the CER Graphic Organizer was not used, Mrs. Honey defaulted to RACES when providing Language Arts instruction to the Treatment group. However, since the RACES strategy includes similar language as CER (e.g., answer the question, cite your evidence, explain your reasons), she was able to integrate the CER Sentence starters with the RACES strategy.

In general, according to the observational tool, Mrs. Honey provided more explicit instruction and feedback across all observations. For example, during one observation, students were working on identifying (and writing) the main idea and supporting details of a passage about canyons. Mrs. Honey, when transcribing student's responses on the board, elaborates on the task at hand and says, "We are finding the main idea of the *whole* passage, not just a single



paragraph...I don't want skinny sentences. I want to expand them. What are some ways that we have learned to help us expand our sentences?" Mr. Frizzle scored higher in providing opportunities to respond. Opportunities to respond include instances in which students are explicitly engaged in a learning task, such as chorally responding and turn-and-talk, or where students are randomly selected to share. During the science block where students were engaged in an observational activity, students were tasked with making observations as a group before joining the whole class in a meeting circle at the front of the room. The dialogic part of science instruction often involves a series of teacher questions, student responding, and teacher follow up to deepen students thinking. During this lesson, 10 students shared their observations, many of whom extended points made by others.

Both teachers used the CER Sentence starters about twice every week and the morphology routine 1 time a week. He reported that he did not use the CER graphic organizer, primarily because "he did not have time to teach it" to students. Mrs. Honey echoes similar concerns. She reported that while she tried to use the graphic organizer at the very beginning of the intervention, it became apparent that more time would be needed for students to become fluent in using it. Given that teachers did not want to extend the duration of the experiment, teachers felt it was necessary to discontinue using the graphic organizer. They brought their concerns to me, and the decision was made to focus on the morphology and sentence starter routines for the remainder of the study.

### ***Interscorer Reliability***

In preparation for scoring, the responses from the Prior Knowledge task and constructed responses from the Science Knowledge Test were transcribed into a word processor. In addition, I conducted a preliminary scan of all the available writing passages to establish "anchors" for

each of the 12 fields in the CER rubric to facilitate both the training and scoring of items. This information was included in a codebook which was used to train the second rater in scoring. Responses to all assessments were blinded to student's identity and time of the assessment before scoring.

A second rater was trained to score CER components using 10% of the available samples, until greater than 90% exact agreement is established. On a training set of  $n=4$  passages, rater reliability was excellent,  $IRR(3,2)=0.96$ ,  $p<.01$ , 95% CI [.91-.98]. I scored all the writing samples for CER components while a second trained rater scored 25% ( $n=20$ ) of a randomly selected subset of each. The reliability of all CER scores was good to excellent,  $ICC(3,2)=.91$ ,  $p<0.01$ , 95% CI [.86-.94].

To assess student's prior knowledge, I counted the number of knowledge units that students wrote in response to Topic Prior Knowledge Task using a previously validated codebook, the Curriculum Specific Knowledge Test (contact authors for codebook; Cabell & Hwang, 2020; Cervetti et al., 2016). A second rater was trained to score Knowledge units on 10% of the available Prior Knowledge Task samples ( $n=4$ ), to establish IRR at 90% reliability. Rater reliability across the four passages was perfect and so a random sample of 20 responses was selected for double scoring. The reliability of the Prior Knowledge Task was excellent),  $ICC(3,2)=0.995$ ,  $p<0.01$ , 95% CI [.990-.997].

CIWS is scored on the back end of the Writing Architect using a codebook. Rater reliability for this measure was established previously for another study using CIWS to evaluate writing quality but was unrelated to the current study. Reliability coding had been established in September of 2023 and was high,  $ICC(3,2)=.87$ ,  $p<.01$ . For this study, I scored all the writing samples for CIWS while a second trained rater scored 25% ( $n=20$ ) of a randomly selected subset

The reliability of CIWS scores was good to excellent, ICC (3,2)= .990,  $p < 0.01$ , 95% CI [.970-.996].

### ***Social Validity***

At the end of the WA posttest period, I asked students a series of questions gauging how they viewed writing in science class and whether they felt they improved in this skill area. This student survey was administered as a pen and paper survey. Students responded after they completed their Students were also asked to identify all the emotions that they feel when writing. Finally, students were asked a series of open-ended questions regarding the perceived importance of science writing. All students, regardless of condition, responded to the student survey since all students received additional writing time during the science period. The student survey can be found in Appendix I.

The teachers also responded to a survey asking them to evaluate the implementation barriers and successes of the intervention (see Appendix J). They were also asked to identify aspects of the scaffolds that they liked, areas they would change, and what ideal instruction would look like if they were to utilize these scaffolds again outside of the context of a research study. This survey was distributed through RedCap (Harris et al., 2009, 2019) and teachers had the ability to save and return to the survey later. This was to encourage thoughtful and thorough feedback.

### ***Teacher Co-Development and Professional Learning***

Because the current study involves both teacher participants providing instruction to both Homerooms, two meetings occurred prior to the start of the intervention to prepare teachers to implement the instructional scaffolds to the treatment students within the framework of the research design.

The first meeting, occurred on a Saturday in early December, after teachers had provided affirmative consent to participate in the study. There were two sessions, one in the morning and one in the afternoon, each lasting approximately 4 hours. The agenda for this session is provided in Appendix K. The morning session was dedicated to helping teachers understand the theoretical orientation and purpose of the study. Each instructional scaffold- the morphology routine, the graphic organizer, and the sentence starters- was presented and I described how the material might be used in each classroom. The afternoon session was dedicated to helping teachers understand the rationale behind the experimental design- that the science instructional routines followed the treatment group as they switched homerooms throughout the day.

During this session, specific logistical issues that the teachers might encounter were addressed. The largest issue that teachers identified was the need for a plan to keep the amount of science instruction equal for both classrooms since they could not remain together during any science writing activities. Another issue was planning science instruction so that the unit did not take longer than 6 weeks since each lesson would now occur over two days rather than one. The teaching timeline that the team agreed upon is in Figure 3.

**Figure 3**

*Science Instruction Lesson Timeline*

	Monday	Tuesday	Wednesday	Thursday	Friday
<b>Week 1</b>	Jan. 8	Jan. 9	Jan. 10	Jan. 11	Jan. 12
<b>Mrs. Honey</b>	Grade 2 Lesson 1 - L1	Writing-RACES	Grade 2 Lesson 2 _ L2	Writing-RACES	
<b>Mr. Frizzle</b>		Writing-CER		Writing-CER	
<b>Week 2</b>	Jan. 15	Jan. 16	Jan. 17	Jan. 18	Jan. 19
<b>Mrs. Honey</b>		G2 Lesson 4 - L4	Writing-CER	G2 Lesson 4 - L4	
<b>Mr. Frizzle</b>		G2 Lesson 3 - L3	Writing-RACES	G2 Lesson 3 - L3	
<b>Week 3</b>	Jan. 22	Jan. 23	Jan. 24	Jan. 25	Jan. 26
<b>Mrs. Honey</b>	Writing-RACES	Grade 4 Lesson 1 - L5	G4 Lesson 2 - L6	Writing-CER	
<b>Mr. Frizzle</b>	Writing-CER		G4 Lesson 3 _ L7	Writing-RACES	
<b>Week 4</b>	Jan. 29	Jan. 30	Jan. 31	Feb. 1	Feb. 2
<b>Mrs. Honey</b>	G4 Lesson 2 - L6	Writing-RACES	Grade 2 Lesson 5 - L8	Grade 2 Lesson 5 - L8	
<b>Mr. Frizzle</b>	G4 Lesson 3 _ L7	Writing-CER			
<b>Week 5</b>	Feb. 5	Feb. 6	Feb. 7	Feb. 8	Feb. 9
<b>Mrs. Honey</b>	Grade 2 Lesson 5 - L8	Grade 2 Lesson 5 - L8	RACES Essay	RACES Essay	
<b>Mr. Frizzle</b>			CER Essay	CER Essay	
<b>Week 6</b>	Feb. 12	Feb. 13	Feb. 14	Feb. 15	Feb. 16
<b>Mrs. Honey</b>	RACES Essay	RACES Essay	RACES Essay	RACES Essay	
<b>Mr. Frizzle</b>	CER Essay	CER Essay	CER Essay	CER Essay	

*Note:* Orange boxes indicate lessons where both homerooms complete the science experiment together; no writing occurs. Blue boxes denote sessions where the Active Control receives instruction using RACES. Green boxes denote sessions where the Treatment Group receives instruction using the CER scaffolds

Teachers agreed that for the largest and most time-consuming lab activities, both Homerooms would complete the activity together in a shared space. Small groups would be comprised of students from the same Homeroom so that potential contamination across conditions would be minimized and I would observe the science block during these days. However, Homerooms would separate once again on the second day of the lesson to work on answering the science questions. On other weeks where the science experiments were not as laborious to set up, teachers would teach two lessons per week. The decision was made to keep students within their routines of switching homerooms. So, for example, if the treatment group received the first week's lesson with Mrs. Honey, the next lesson would be with Mr. Frizzle. It is in this manner that teachers taught both writing conditions. Teachers made sure to refer to the teaching timeline throughout the intervention.

The last two hours of the afternoon session involved mapping out the science unit that will be used during the intervention. The unit of instruction would be on topics related to weathering and erosion. The teachers noted that their students were still dealing with the effects of the pandemic on their learning and so it was important to incorporate weathering and erosion content from earlier grade levels with the content students were expected to learn this year. For this reason, teachers incorporated Grade 2 and Grade 4 content into this unit. It was agreed upon by the team that the unit should include an end-of-unit problem-based writing activity to help assess how much students had learned, particularly because such an activity was not available in their current curriculum. This activity is in Figure 4, but analysis of student responses to this question are beyond the scope of this dissertation. To ease the transition between teachers, students in both writing conditions would keep a lab folder that held their respective instructional

scaffolds and writing activities from the unit. I collected this lab folder at the end of the intervention.

## Figure 4

### *End of Unit Writing Prompt*

#### **The Trouble at Hill Crest**

The Hill Crest is a quiet neighborhood with the Gentle River nearby. On the other side of the river, there are some hills that are usually covered in trees. But last year, a forest fire burnt most of the forest down. Now, residents notice dirty brown water rushing into the river whenever it rains. They also notice that the once Gentle River floods more often, sometimes over the riverbanks. It is beginning to wear down the riverbanks near the neighborhood of Hill Crest. Residents aren't sure why the river is changing but they are concerned about the possibility of the neighborhood flooding during a big storm. Are the residents of Hill Crest right to be concerned? If so, what can they do to protect their neighborhood?



**Directions:** You are a local scientist who has been asked to help the neighborhood understand the problem and come up with a solution. They want to hear your thoughts at the next neighborhood meeting, especially on the topics listed below. Write a scientific report in 1 (minimum) to 3 (maximum) paragraphs that will help the neighbors understand the problem and come up with a solution.

- Explain how weathering and erosion have impacted the hilly forest and why that might change the river.
- Predict how and why the changing river might put their neighborhood at risk for flooding.
- Give the neighborhood some recommendations to help fix the problem.

Remember you are a scientist! You should use what you know from your experiences reading, writing, and experimenting with topics related to **weathering and erosion**. You should write your letter to the town *the way a scientist would*.



The last decision made during this meeting was to determine which Homeroom would receive the CER scaffolds. Because there was only one teaching pair included in this study, true randomization could not occur. Based on logistics, the team agreed that after pre-assessment took place, the teachers would take some time during their morning All Block period to pre-teach the routines to students. As a result, Mr. Frizzle's Homeroom was assigned to the CER condition and Mrs. Honey's Homeroom was designated Active Control.

The second meeting between the team occurred in early January, before school resumed from winter break. This Zoom meeting lasted approximately an hour and was dedicated to answering any remaining logistical questions, schedule planned observation, and review use of the instructional scaffolds.

### ***Pre-Intervention***

During this phase, pretest measures were administered to both the control and treatment conditions. Due to the team-teaching dynamic of the teacher participants, and to reduce the negative impact of instructional time being consumed by testing, measures were distributed between classroom instructional blocks.

The Science Knowledge Test was administered by each teacher to their Homeroom during the designated science instructional block. The assessment was administered online via *Pear Deck*. It took one class session, approximately 50-minutes, for the teachers to administer the assessment to the class.

The Topic Prior Knowledge Task was administered by the researcher to students during the start of their Language Arts instructional block, as part of the class's warm up, one day before the scheduled pretest Writing Architect administration. A script for administration of the Topic Prior Knowledge Task can be found in Appendix G, with the prompts used at pretest indicated.

I administered the Writing Architect during student's Language Arts block in Mrs. Honey's classroom. Prior to administering the assessment, I previewed the task that students would be asked to complete during our session, which was also facilitated by a video recorded set of instructions that are built into the WA platform. Students were given paper packets that correspond to the writing passages assigned to them in the WA platform. The pre- and posttest passages were counterbalanced across participants: one half of students in both conditions were randomly assigned to respond to *How to Speed Up Extinctions* or *A Diet for Invasive Carp*. After students submitted their responses, or after the screen automatically advanced after 15-min, students completed the Typing Fluency Task.

### ***Intervention***

In January and February of 2024, two teachers implemented the science writing routines with the Treatment group during science class and when reading about science in ELA. Teachers aimed to implement each science instructional scaffold with the treatment group twice weekly over the course of a 6-week instructional unit. Over the course of 6 weeks, students learned about weathering and erosion in science class. In ELA, students learned about animal defense mechanisms. The Active Control condition used a previously learned informational writing strategy, RACES, to support their writing during the same instructional units. In both conditions teachers incorporated explicit instruction (e.g., modeling and think-alouds), opportunities to respond, flexible groupings, and detailed feedback to support students' science writing. Below is an example of how the CER Scaffold was implemented in each classroom.

When reading a content-rich science text in language arts, students used the CER Sentence Starters to help them reconfigure the main idea of the text into a scientific question. Mrs. Honey drew the RACES organizer on a large piece of butcher paper while simultaneously

displaying a text on animal behavior on the board. As students read, they filled out the graphic organizer. Students summarized background information (R) from the text and used the main idea to answer a science question which could be answered by reading the text. For example, “How might animals blend into their environment to avoid predators?” Mrs. Honey, using the CER Sentence Starters, engaged in a think-aloud to select an appropriate sentence stem to start Answering the Question. Students provided feedback about what is and isn’t a good sentence stem, offering their own sentence starters using the science writing resources contained in their lab journals. Students used the sentence stems to identify relevant textual information that supports their Answer, sometimes working in small groups or pairs. As a class, as student pairs, or individually, students used the information extracted from the text and organized into the RACES structure to respond to the science question, again, with the aid of the sentence starters.

In science class, on the other hand, students used the CER graphic organizer to answer the driving scientific question posed at the end of an investigation. The teacher not only provided instruction on what a claim is in the context of science experimentation but also drew parallels to what was learned about crafting claims in Language Arts. Given the end of lab questions, Teachers referred to the Sentence Starters, contained in their lab notebook, and identified a starter to use. Through a combination of think-alouds, modeling, and classroom discussion, the teacher wrote a claim on the board. In response, students were asked to tell *how* they know the answer to the question. Students referred to the observations and measurements they made in the previous day’s science lab to make connections to the science investigation. Teacher-led discussion and clarifying questions were used to link student dialogue to the days’ science question, with the teacher recording student discussion points on the board. After a period, the teacher would ask, “So given the evidence from our experiments, how can we write an Evidence

statement that answers the question? Let’s look at our Sentence Starters.” The teacher would model how he might select a sentence starter particularly when discussing a lab experiment. Students would then complete the evidence statement using the sentence starter, “According to my data…” Reasoning statements also used a Sentence Starter and incorporated explanations that students generated during their small group or whole-class discussion. At the end of the period, the teacher would share the CER response that was generated by the class on the overhead board. Students would also share their responses, especially if they were different. The teacher provided feedback in the form of addendums or edits to the class response.

Through a gradual release of responsibility model (I do, We do, You do) teacher and peer support faded throughout the six-week instructional unit culminating into a problem-based writing prompt, *The Trouble at Hill Crest* (see Figure 4). Students spent six instructional days on this task. On Day 1, the students read through and annotated the prompt, specifically identifying and understanding the task requirements. On days 2-4, students made an outline of their response to each task (one day for each prompt), including the key vocabulary, concepts, and sentence stems that were relevant to each section of CER. The Teacher provided frequent class check-in at each section (C,E, or R) of the outline, soliciting student responses for the types of information that belong in each section of the response and writing them on the whiteboard for students to examine as a model during their own independent work. On days 5 and 6, students selected at least one prompt that they wanted to create a full draft for. Teachers provided individual-level support during this time in the form of spelling or sentence writing help. Students were prompted to identify their claims, evidence, and reasonings using the annotation strategy they learned (box their claims, underline their evidence statements, and circle their rationales). Student responses were collected at the end of the period on the sixth day.

### ***Post-Intervention***

During the post-intervention phase, the Science Knowledge Test, Topic Prior Knowledge Task, and the Writing Architect were administered. Due to the counterbalancing of the passages, students that responded to *How to Speed Up Extinctions* during the pretest responded to *A Diet for Invasive Carp* at posttest, and vice versa. Exit surveys soliciting student and teacher's perceptions of the intervention were also administered.

### **Analytic Plan**

#### ***Descriptive Statistics and Data Checking***

Data were analyzed using IBM SPSS Statistics (Version 27). Descriptive statistics were reported according to What Works Clearinghouse (WWC) guidelines for quantitative studies and therefore report the sample size for each measure, thus increasing transparency where missing data exist (WWC, 2021). A missing values analysis was conducted to examine the amount of missing data within the dataset. Little's (1988) chi square test was conducted to determine whether data was missing completely at random. While scoring student's CER data, scorers noted that there may be a difference in the student's ability to respond to the science writing prompt. Suspecting that one prompt may be more difficult to respond to than the other, despite attempts to make the readability of the texts similar, descriptive statistics per passage for the entire sample were computed. An independent samples t-test was conducted to determine if there were significant differences in pretest responding for the *Extinctions* and *Invasive Species* prompts for the following: gender, Total Words Written, Topic Prior Knowledge, CIWS, CER Total, Claims Score, Evidence Score, Reasons Score, Copied Elements, Long Words, and Sentence Length Diversity.

Prior to the main analysis, I checked for whether the statistical assumptions of normality, linearity, and homogeneity of variances were met (Mertler & Reinhart, 2017). The following posttest data exhibited excessive kurtosis, as indicated by scores  $\pm 3.00$  standard deviations from the mean: CIWS, Long Words, Sentence Length, Claims, CER Total Scores. Visual inspection of the corresponding box plots revealed two multivariate outliers. After determining that the corresponding values were not data entry errors, 7 data points were winsorized to the nearest value within the distribution (Dixon & Yuen, 1974) thus resolving issues of leptokurtosis. Normal Q-Q plots and corresponding significant Komogorov-Smirnov statistical suggest that the following variables had some departures from normality: posttest Prior Knowledge, pre- and posttest Sentence Length Diversity, pre-test Evidence Scores, and pre- and posttest Reasoning Scores. Data were square root transformed in response, thus normalizing the data. An examination of residual scatterplots produced by linear regression yielded no clustering and so I concluded that data met the assumption of linearity. Finally, because homoscedasticity is not fatal to analysis nor is it affected by assumptions of normality (Mertler & Reinhart, 2017), the homogeneity of variances was assessed using the Levene's test statistic with a correction when applicable to aid in the interpretation of inferential statistics supplied by t-tests and ANCOVA.

### ***Baseline Equivalence***

To evaluate the effect of an intervention of student learning, even small differences between groups must be identified and controlled for (WWC, 2021). To determine whether the Active Control and Treatment group were similar at the time of the pre-test, independent samples t-test were conducted to compare group means. Specifically, baseline equivalence at pre-test was established by analyzing group mean differences of gender, Topic Prior Knowledge, CIWS, Science Knowledge, and CER Scores across the two conditions. Hedges' *g* effect sizes were

computed to evaluate the magnitude of differences between the groups. According to WWC, baseline equivalence can be established if the effect size on identified covariates is no greater than 0.25 standard deviations (WWC, 2021, pp. 53). For indicators whose mean difference is small, as indicated by an effect size less than 0.05 *SD*, baseline equivalence is satisfied. For indicators with  $ES=|0.05- 0.25|$ , baseline equivalence standards can be met with a statistical adjustment whereby the identified covariate is entered into the main analysis.

### ***Comparison of Posttest Group Means***

To determine if there was a difference in the group means at posttest for students in the Treatment and Active Control condition, a One-way Analysis of Covariance (ANCOVA) was conducted. ANCOVA allows for the examination of the effect of the Independent Variable (condition) on the dependent variable (CER and Science Knowledge) while partialling out the effect of covariates that may otherwise obscure the results (Mertler & Reinhart, 2017). Bivariate correlations were computed to ensure that potential covariates that are correlated with the outcome measures. Additional assumptions of ANCOVA were checked including a) a linear relationship between the covariates and the dependent variable, b) reliability of the covariate and measuring it without error, and c) homogeneity of regression slopes (Mertler & Reinhart, 2017) Hedges' *g* effect sizes were also computed. Due to the small sample size of the current study, a post hoc power analysis was conducted to determine the likelihood of "correcting rejecting the null hypothesis when it is really false" (Lomax & Hahs-Vaughn, 2012).

## CHAPTER 4: RESULTS

Study data were collected and managed using REDCap electronic data capture tools hosted at Michigan State University (Harris et al., 2009, 2019). REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources.

### Missing Data

A summary of descriptive statistics for the raw data, including the identification of missing data, is available in Table 6. Missing data were largely due to absences. Three cases were missing for the NWEA, two due to absences and one due to not having yet been enrolled in the school. Two students were missing for the WA and Prior Knowledge Task administrations at pretest, one due to an absence and one student had not yet enrolled in the school. Two students' data were missing for the WA posttest. Field notes indicate that these students were present for the administration, especially as they had data for other measures collected on the day of the assessment, but further investigation found that a technical error had prevented their scores from being captured by the web-tool. The missing data accounted for 5% of the missing data for a given variable. 34 students, or 85% of cases, responded to all tasks assigned to them during the study. Little's chi square (1988) indicate that data were missing completely at random,  $\chi^2(61) = 38.06, p = .991$ . Because data were MCAR, multiple imputation (5 iterations) using the linear regression method was used to predict the missing values (Mertler & Reinhart, 2017; Rubin, 2018). The resulting imputations were pooled together for subsequent analysis.



**Table 6.***Descriptive Statistics of Students' Raw Performance on All Variables*

	Active Control (N=18)						Treatment (N=22)					
	pretest			posttest			pretest			posttest		
	n	M (SD)	Min-Max	n	M (SD)	Min-Max	n	M (SD)	Min-Max	n	M (SD)	Min-Max
Science Knowledge	18	12.11 (2.93)	8-19	18	14.28 (5.56)	4-24	22	11.32 (3.78)	4-18	22	15.72 (2.79)	11-22
CER Total Score	17	4.76 (4.71)	0-19	18	5.56 (3.82)	0-14	21	5.10(3.60)	0-13	20	6.65 (5.37)	1-27
Claims		3.24 (3.25)	0-11		1.72 (1.56)	0-5		2.38 (1.77)	0-6		3.45 (2.74)	0-12
Evidence		1.06 (1.52)	0-5		2.72 (2.08)	0-6		1.67 (2.22)	0-7		2.35 (2.62)	0-10
Reasoning		0.47 (0.94)	0-3		1.11 (1.18)	0-4		1.04 (1.11)	0-4		0.85 (1.42)	0-5
Copied Sentences		0.24 (0.56)	0-2		0.61 (1.50)	0-5		0.14 (0.48)	0-2		0.40 (1.23)	0-4
Topic Prior Knowledge	17	3.18 (2.83)	0-11	18	6.44 (5.46)	0-18	21	4.86 (4.20)	0-13	22	4.23 (3.64)	0-17
CIWS	17	20.65 (28.81)	-16-104	18	37.89 (47.96)	-20-174	21	32.14 (37.74)	-6-139	20	30.85 (45.73)	-9-205
Long Words	17	6.65 (6.67)	0-25	18	8.72 (10.43)	0-45	21	7.52 (6.53)	0-23	20	8.20 (10.67)	0-50
Sentence Length Diversity	17	2.43 (3.30)	0-10	18	6.84 (9.08)	0-34.50	21	3.62 (4.04)	0-12.50	20	3.87 (5.00)	0-20

## Baseline Equivalence

An independent samples t-test was conducted to determine if there were significant differences between the Treatment and Active Control for the following: gender and pretest scores for Topic Prior Knowledge. According to the Levene's test, the homogeneity of variance assumption was satisfied. Using a two-tailed test of significance, the independent samples t-test indicate that the gender ( $t=-1.34$ ,  $df=38$ ,  $p=.19$ ), and Topic Prior Knowledge ( $t=1.22$ ,  $df=38$ ,  $p=.23$ ) were not statistically significantly different for students in the Active Control and Treatment conditions. However, examination of Hedges'  $g$  (see Table 7) indicates that though these group differences are not statistically different, there is practical significance with respect to the Gender composition of each homeroom (Hedges'  $g=-0.42$ ; Treatment is comprised of 54.5% boys versus 33.3% boys in Active Control) and the number of pretest Topic Prior Knowledge units that students wrote (Hedges'  $g=0.38$ ). Because the mean differences of these potential covariates exceed WWC threshold of  $0.25 SD$ , I conclude that groups were not equivalent at baseline with respect to Gender and Topic Prior Knowledge. Due to the confirmed differences in gender composition across both groups, baseline equivalence of Gender on the outcome variables was further examined. Independent samples t-test yield insignificant group differences between boys and girls on CER Total scores ( $t=1.22$ ,  $df=38$ ,  $p=.23$ ,  $g=.38$ ), Science Knowledge ( $t=1.33$ ,  $df=38$ ,  $p=.19$ ,  $g=.42$ ), and CIWS ( $t=1.52$ ,  $df=38$ ,  $p=.14$ ,  $g=.47$ ). Yet, because these small differences were meaningful as indicated by large Hedges'  $g$  effect sizes, I concluded that boys' and girls' performance on each of these outcomes was not equivalent. Across conditions, girls pretest performance was greater than the boys on outcomes of CER Total ( $M_{\text{girls}}=5.61$ ,  $SD_{\text{girls}}=4.48$ ;  $M_{\text{boys}}=4.08$ ,  $SD_{\text{boys}}=3.17$ ), Science Knowledge ( $M_{\text{girls}}=12.32$ ,

$SD_{\text{girls}}=3.44$ ;  $M_{\text{boys}}=10.89$ ,  $SD_{\text{boys}}=3.28$ ), and CIWS ( $M_{\text{girls}}=35.34$   $SD_{\text{girls}}=34.08$ ;  $M_{\text{boys}}=19.03$ ,  $SD_{\text{boys}}=33.18$ ). A decision was made to include gender as a covariate in all ANCOVA.

**Table 7.***Descriptive Statistics of Imputed Student Performance at Pre- and Posttest*

	Pretest			Posttest	
	Active Control ( <i>n</i> =18) <i>M</i> ( <i>SD</i> )	Treatment ( <i>n</i> =22) <i>M</i> ( <i>SD</i> )	Hedges' <i>g</i>	Active Control ( <i>n</i> =18) <i>M</i> ( <i>SD</i> )	Treatment ( <i>n</i> =22) <i>M</i> ( <i>SD</i> )
Science Knowledge	12.11 (2.93)	11.32 (3.78)	-0.23	14.28 (5.56)	15.72 (2.79)
CER Total Score	4.75 (4.57)	5.06(3.52)	0.08	5.56 (3.82)	6.08 (2.94)
Claims Score	3.20 (3.15)	2.38 (1.73)	-0.33	1.72 (1.56)	3.19 (1.96)
Evidence Score	1.08 (1.48)	1.67 (2.16)	0.30	2.72 (2.08)	2.40 (2.50)
Reasoning Score	1.11 (1.18)	0.84 (1.35)	0.53	1.11 (1.18)	0.85 (1.35)
Copied Sentences	.24 (.55)	.20 (.52)	-0.09	.61 (1.5)	.46 (1.19)
Topic Prior Knowledge	3.41 (2.93)	4.82 (4.10)	0.38	6.44 (5.46)	4.23 (3.64)
CIWS	23.85 (31.09)	31.39 (36.99)	0.21	34.33 (38.76)	28.90 (30.22)
Long Words	7.08 (6.72)	7.46 (6.38)	0.06	7.22 (5.83)	6.67 (4.74)
Sentence Length Diversity	2.29 (3.25)	3.54 (3.96)	0.33	6.03 (6.84)	4.15 (5.06)

## Descriptive Statistics

Descriptive statistics of the pooled data with winsorization can be found in Table 7. Baseline equivalence data indicate that the mean differences in pretest scores for the Treatment versus Active Control group were insignificant for Topic Prior Knowledge, indicating that students had similar levels of knowledge about the science topics that they were going to write about at pretest. They also performed similarly on CIWS ( $t=0.69$ ,  $df=38$ ,  $p=.50$ ), Science Knowledge ( $t=-0.73$ ,  $df=38$ ,  $p=.47$ ), and CER Total Scores ( $t=0.25$ ,  $df=38$ ,  $p=.81$ ). With respect to students CER writing, differences in student's pretest Prior Knowledge were practically meaningful. Treatment group wrote an average of 4.82 ( $SD=4.10$ ) knowledge units, and though this difference was not statistically different from the Active Control ( $M=3.41$ ,  $SD=2.93$ ), the differences were enough to conclude that the groups baseline levels of prior knowledge was not equal (Hedges'  $g=0.38$ ) and favored the Treatment condition. To better understand student's response within the CER score, I broke down students' scores into their component parts: Claims, Evidence, and Reasons. At pretest, students in Active Control (Hedges'  $g=-0.33$ ) were perhaps more proficient in their use of Claims, and the Treatment group wrote more Evidence (Hedges'  $g=0.30$ ) and Reasoning (Hedges'  $g=0.53$ ) statements. Both groups at pretest did not have practically meaningful differences in their use of Long Words in their writing ( $t=0.18$ ,  $df=38$ ,  $p=.86$ , Hedges'  $g=.06$ ) but practical differences were observed for the varied length of their sentences ( $t=1.08$ ,  $df=38$ ,  $p=.29$ , Hedges'  $g=0.33$ ). Though these small differences are statistically insignificant and perhaps meaningful, as these items were not selected for inclusion as a covariate in the analysis as they are smaller components of a larger aggregate score (CER Total or CIWS) which is already included in the analysis.

Due to some concerns related to students copying the text, the number of copied sentences was also scored. It may also be that copying occurs in strategic ways to meet rhetorical goals. In this study, copied elements were included in the CER score because the primary aim of scoring was to assess student's abilities to answer the science question posed to them. Passages were scored using the rubric first, and then the number of copied elements were identified. A copied sentence was defined as a sentence for which there was at least 80% word-for-word correspondence to the source text (<https://osf.io/tfvx2>). It was noted that students had more copied elements in their posttest responses than at pretest but the overall presence of copy within the data set was minimal; only 13 student responses (16% of the entire dataset) had any instances of copied elements. For the 3 students at posttest whose text included 4 copied sentences, the arrangement of their sentences within their composition was unique and different from that of the source text. At pretest, there were no significant differences between the Treatment and Active Control with respect to the number of copied elements ( $t=-0.28$ ,  $df=38$ ,  $p=.78$ , Hedges'  $g=-0.09$ ), and as the influence of copy on student's CER writing is not a focus of the current study, it was not investigated further. Future iterations of this study would benefit from a codebook with a priori decision rules for scoring student responses with copied elements.

During the scoring process, coders made notes that the *Invasive Species* prompt may have been more difficult to respond to and so a decision was made to examine the differences in pretest scores for each prompt (see Table 8 for descriptives). Prompt administration was counterbalanced meaning that, at pretest, approximately one-half of students per condition were randomly assigned (via a random number generator) to responded to *Extinction* while the others responded to *Invasive Species*. Students responded to the other prompt at posttest.

**Table 8.***Descriptive Statistics of Imputed CER Writing Performance per Prompt at Pre-test*

	Extinctions			Invasive Species			Independent Sample's T-test		
	pretest			pretest			<i>t</i> ( <i>df</i> )	sig.	Hedges' <i>g</i>
	<i>n</i>	<i>M</i> ( <i>SD</i> )	Min-Max	<i>n</i>	<i>M</i> ( <i>SD</i> )	Min-Max			
Gender <sup>a</sup>	21	0.52 (.51)	0-1	19	0.58 (0.51)	0-1	-.34 (38)	.73	-0.11
Total Words Written	21	52.00 (35.07)	4-125	19	33.22 (30.82)	0-129	1.79 (38)	.08	0.55
Topic Prior Knowledge	21	3.95 (2.57)	0-11	19	4.44 (4.61)	0-13	-.42 (28) <sup>b</sup>	.68	-0.13
CIWS	21	31.95 (33.58)	-16-105	19	23.63 (35.35)	-6-139	.76 (38)	.45	0.24
CER Total Score	21	6.00 (4.27)	0-19	19	3.73 (3.32)	0-13	1.86 (38)	.07	0.58
Claims	21	3.47 (3.02)	0-11	19	1.94 (1.34)	0-5	2.10 (28) <sup>b</sup>	.05	0.63
Evidence	21	1.66 (2.10)	0-7	19	1.11 (1.62)	0-6	.92 (38)	.36	0.29
Reasons	21	0.85 (0.96)	0-3	19	0.67 (1.14)	0-4	.56 (38)	.58	0.17
Copied Elements	21	.19 (.51)	0-2	19	0.24 (0.56)	0-2	-.32 (38)	.75	-0.10
Long Words	21	9.28 (6.39)	1-25	19	5.08 (5.92)	0-23	2.15 (38)	.04	0.67
Sentence Length Diversity	21	3.51 (3.77)	0-12.5	19	2.38 (3.56)	-0.12-10.39	.98 (38)	.33	0.30

<sup>a</sup> Boys were coded as 0 and Girls were coded as 1.<sup>b</sup> Homogeneity of variances not assumed.

The homogeneity of variances assumption was satisfied for all variables except for Topic Prior Knowledge and Claims, a component of the CER Total score. Similar numbers of boys and girls were assigned to each prompt, despite more students ( $n=2$ ) responding to *Extinctions* at pretest. The number of words that student wrote per prompt was not statistically significant, ( $t=.1.79$ ,  $df=38$ ,  $p=.08$ ), though the practical importance of this difference was considerable as indicated by the large effect size (Hedges'  $g=0.55$ ). In suspecting that there were prompt effects, I hypothesized that differing amounts of Prior Knowledge with respect to the topic might influence student's responding. There was no statistical significance ( $t=-.42$ ,  $df=28$ ,  $p=.68$ ) or practical importance (Hedges'  $g=-.13$ ) in the pretest Topic Prior Knowledge scores for *Extinctions* ( $M=3.95$ ,  $SD=2.57$ ) and *Invasive Species* ( $M=4.44$ ,  $SD=4.61$ ). Students mean CER scores appear to be different between prompt, bordering on statistical significance ( $t=1.86$ ,  $df=38$ ,  $p=.07$ ). The magnitude of this difference is practically important as indicated by the large Hedges'  $g$  ( $ES=0.58$ ). An analysis of the breakdown in score shows that students' ability to make Claims is perhaps the most important. There was a significant difference [ $t(28)=2.10$ ,  $p<.05$ , Hedges'  $g=.63$ ] in the number of Claims made by students responding to *Extinctions* ( $M=3.47$ ,  $SD=1.94$ ) versus *Invasive Species* ( $M=1.94$ ,  $SD=1.34$ ). Students did not differ in the number of Evidence statements or Reasons that they provided. Students wrote considerably more Long Words when responding to *Extinctions* ( $t=2.15$ ,  $df=38$ ,  $p=.04$ , Hedges'  $g=0.67$ ). This pattern of responding suggest that, despite the two passages being similar in readability (See Table 4), perhaps students had a more difficult time articulating a response to the *Invasive Species* prompt. Given these significant differences, passage is included as a fixed effect in the ANCOVA.

It should be noted that some means were lower after the use of multiple imputation to fill in values for missing data (See Tables 6 and 7). To determine if student gains in CER



performance and Science Knowledge was because of the intervention, student posttest performance was assessed using ANCOVA.

### **Analysis of Covariance**

Results of baseline equivalence testing indicate that covariate mean differences with effect sizes falling within WWC's range of  $|.05-.25|$  standard deviations require statistical adjustment by being entered into ANCOVA. Topic Prior Knowledge and Gender were above this threshold. However, due to the potential girl advantage observed at baseline on the outcomes, Gender was entered as a covariate. Bivariate correlations were computed to confirm whether to exclude Topic Prior Knowledge from being entered as covariate (See Table 9). As Topic Prior Knowledge was not significantly correlated with any outcome measure, the decision to not include it as a covariate was upheld.

The assumption that covariates must have high reliability is satisfied for all potential covariates, as duly described in the Measures and Interscorer Reliability sections. Reliability metrics and interrater reliability estimates range from high to excellent reliability.

**Table 9.***Pearson Correlations Between Pre-test and Posttest Variables for All Instructional Conditions*

	1	2	3	4	5	6	7	8	9
1. Gender	1	.20	.07	.21	.21	.24	.20	.02	-.01
2 . CER Total, pretest		1	.20	.07	.21	.21	.24	.20	.07
3. CER Total, posttest			1	.18	.38*	.55**	.67**	.17	.58**
4. SK, pretest				1	.44**	.18	.25	-.13	.02
5. SK, posttest					1	.49**	.48**	.12	.33*
6. CIWS, pretest						1	.82**	.1	.69**
7. CIWS, posttest							1	.18	.66**
8. TPK, pretest								1	.12
9. TPK, posttest									1

**Note:** SK=science knowledge; TPK=Topic Prior Knowledge; SLD=Sentence Length Diversity.

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed)

### ***RQ1. Intervention Effects on CER Writing***

No serious curvilinearity was observed on the within-cells residual scatter plots of the DVs with the covariates, satisfying the linearity assumption. Visual inspection of DV-covariate slopes (for CER Total at pretest) yielded intersecting regression lines, indicating a violation of the homogeneity of regression assumption. The regression slopes used to predict a group's CER post test score due to the covariate are not equal. As result of this assumption, "the error terms are not reduced as fully as they could be and group means are incompletely adjusted...[which could] result in errors in statistical decision making" (Mertler & Reinhart, 2017). The homogeneity of regression assumption was met as indicated by the F-test produced by the factor-covariate interaction, [ $F(1,39)=.48, p=.48$ ]. A violation of this assumption would call into question the results of ANCOVA. The small sample size skews the visual inspection of regression slopes as the best fitting line is constructed from few data points. With caution, the decision to continue the ANCOVA was made. The homogeneity of variance assumption was violated, [ $F(1,38)=9.47, p=.01$ ] indicating that the error variance of CER Total scores was not equal across condition. There is no adjustment for this in ANCOVA. Interpretations of the ANCOVA to assess mean differences on students CER Total scores because of membership in the Treatment group or Active Control should be suspect.

An analysis of covariance was conducted to determine the effect of instructional condition on student's posttest CER Total scores after controlling for the effects of covariates (gender and writing prompt) and student's pretest CER performance. Prompt at posttest and Gender was entered into the model to control for potential prompt effects or girl advantage. ANCOVA (Table 10a) results indicate no significant main effect of gender [ $F(1,35)=0.07, p=.79, \text{partial } \eta^2=.00$ ], or instructional condition [ $F(1,35)=.04, p=.84, \text{partial } \eta^2=.24$ ] on student's

posttest CER Total scores. A significant main effect of prompt [ $F(1,39)=6.91$   $p=.01$ , partial  $\eta^2=.16$ ] and students pretest CER performance [ $F(1,35)=10.90$ ,  $p=.00$ , partial  $\eta^2=.24$ ] was observed. Table 10b presents the adjusted and unadjusted group means for CER Total scores per condition. After adjusting for students gender, post-test prompt, and their CER performance at pretest, the average CER Total score for the Treatment group was ( $M=5.93$ ) not significantly different from the average score of the Active Control ( $M=5.74$ ). To determine the practical significance of the intervention on student outcomes, treatment effect sizes were determined by computing Hedges'  $g$  according WWC guidelines for assessing the effect size of quasi-experimental studies using the posttest adjusted means and standard deviation (WWC, 2022). The practical significance of the pre-test to posttest treatment differences due to condition on student's CER Total scores was small, Hedges'  $g=0.07$ , such that the intervention group performed 0.07 standard deviations higher than the active control.

Statistical power (denoted as  $1-\beta$ ) is the likelihood of rejecting the null hypothesis and finding an effect if one is present (Lomax & Hahs-Vaugh, 2012; Mertler & Reinhart, 2017) If the statistical power of an inferential statistic is limited, it is unlikely to detect an intervention effect if one is present resulting in a false-negative. For example, the observed power of the inferential statistic in using the regression slope of condition to predict CER posttest scores was .04. Given the current sample and effect size, only 4% of the time would I correctly conclude that instructional condition had no influence on students' CER scores and observe an effect of instruction on CER scores if such an effect were present. Increasing the statistical power by way of a larger sample size would be beneficial for reducing Type II error and increasing the confidence of results. Generally, the observed statistical power of the inferential statistics is high due to the limited number of covariates included in the analysis given the current sample.

**Table 10a.***Analysis of Co-Variance Summary Table Examining Group Differences at Posttest for CER Total*

Source	SS	df	MS	F	p	partial n2	Observed Power
Corrected Model	126.63	4	31.66	3.61	.01	.29	.82
Gender	0.64	1	0.64	0.07	.79	.00	.06
Prompt	60.54	1	60.54	6.91	.01	.16	.72
CER pretest	95.49	1	95.49	10.9	.00	.24	.89
Condition	0.34	1	0.34	0.04	.84	.00	.05
Error	306.52	35	8.76				
Total	1799.7	40					

**Table 10b.***Treatment Effects and Adjusted and Unadjusted Group Means for Posttest CER Total*

	Adjusted <i>M</i>	Unadjusted <i>M</i>
Active Control	5.70	5.56
Treatment	5.93	6.08
Treatment Effect	0.07	

**Note:** Treatment Effect calculated with Hedges' *g*.

## ***RQ2. Intervention Effects on Science Knowledge***

Homogeneity of regression slopes was violated according to a significant interaction between Condition and Science Knowledge pretest, [ $F(2,35)=5.29, p=.01$ ]. This indicates that student's performance on the pretest Science Knowledge measure is not equivalent across groups. Similarly to the ANCOVA for RQ1, visual inspection of line plots shows an interaction between Science Knowledge at posttest and the covariate. The homogeneity of variance assumption was not satisfied, [ $F(1,38)=8.17, p=.01$ ] indicating that the error variance of Science Knowledge posttest scores was not equal across condition. For the current study, ANCOVA is an inappropriate analysis for assessing student learning condition on students' science learning outcomes. This may be due to bias resulting from the violation of the heterogeneity assumption. Interpretation of results should be treated with caution regardless.

An analysis of covariance was conducted to determine the effect of instruction condition on student's posttest Science Knowledge scores after controlling for the effects of gender and student's pretest CER performance. ANCOVA (Table 11a) results indicate no significant main effect of condition [ $F(1,36)=3.12, p=.09, \text{partial } \eta^2=.08$ ] on student's posttest Science Knowledge scores. A larger sample size may have different effects. The observed power of this estimate was moderate ( $1-\beta=.41$ ), indicating there is a 41% likelihood of correctly rejecting the null hypothesis (Lomax & Hahs-Vaughn, 2012) and observing an effect of instructional condition on Science Knowledge. Gender likewise was not a significant covariate, [ $F(1,36)=1.31, p=.26, \text{partial } \eta^2=.04$ ]. Pretest Science Knowledge was a significant [ $F(1,36)=8.88, p=.01, \text{partial } \eta^2=.20$ ] predictor as performance was different between the groups. Accordingly, adjustments in means were conducted due to covariates. Table 11b presents the adjusted and unadjusted means for condition on student's posttest science scores. The treatment

effect size for the pretest to posttest differences of the Science Knowledge outcome due to condition was practically meaningful, Hedges'  $g=0.51$ .

**Table 11a.***Analysis of Co-Variance Summary Table Examining Group Differences at Posttest for Science Knowledge*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial $\eta^2$	Observed Power
Corrected Model	192.57	3	64.19	4.46	.01	.27	.84
Gender	18.81	1	18.81	1.31	.26	.04	.20
SK pretest	127.78	1	127.78	8.88	.01	.20	.83
Condition	44.94	1	44.94	3.12	.09	.08	.41
Error	518.2	36	14.39				
Total	9801	40					

**Table 11b.***Adjusted and Unadjusted Group Means for Posttest Science Knowledge*

	Adjusted <i>M</i>	Unadjusted <i>M</i>
Active Control	13.87	14.27
Treatment	16.06	15.72
Treatment Effect	0.51	

**Note:** Treatment Effect calculated with Hedges' *g*.



### ***RQ3. Intervention Effects on CIWS***

Scatterplots show no curvilinear pattern in the standardized residuals for the DV and covariates of Gender and Prompt, indicating that the linearity assumption was satisfied. A visual inspection of the regression lines shows parallel slopes, indicating that the homogeneity of regression slopes was satisfied. The interaction between CIWS pre-test scores and condition was insignificant ( $F(1,34)=3.59, p=.07$ ), confirming that homogeneity of regression was satisfied. The Levene's test also indicate that the homogeneity of variances assumption was met, [ $F(1,38)=3.69, p=.06$ ]. With statistical assumptions satisfied, the results of the following ANCOVA can be more reliably interpreted.

An analysis of covariance was conducted to determine the effect of instruction condition on student's posttest CIWS scores after controlling for the effects of Gender, Prompt, and pretest CIWS performance. Students' pre-test CIWS scores were significantly different between conditions, [ $F(1, 35)=77.42, p<.001, \text{partial } \eta^2=.69$ ]. This suggests that students pretest spelling and grammar/syntax knowledge (i.e., CIWS) above their gender, writing prompt, or instructional condition. ANCOVA (Table 12a) results indicate no significant main effect of condition [ $F(1,35)=3.89, p=.06, \text{partial } \eta^2=.10$ ] or prompt [ $F(1, 35)=0.40, p=.53, \text{partial } \eta^2=.01$ ] on student's posttest CIWS scores. It is notable that the treatment effect of Condition on CIWS posttest scores had a practically significant negative effect with a Hedges'  $g$  effect size = -0.37. This indicates that the active control instruction was superior to the instruction happening in the treatment condition for addressing students' science writing needs. The observed statistical power for most of the computed statistics were low, including the power of testing instructional condition which was less powerful than a coin flip ( $1-\beta=.48$ ). . The adjusted and unadjusted means are presented in Table 12b.

**Table 12a.***Analysis of Co-Variance Summary Table Examining Group Differences at Posttest for CIWS*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial $\eta^2$	Observed Power
Corrected Model	31645.25	4	7911.31	20.70	.001	.70	1.00
Gender	100.32	1	100.32	0.26	.61	.01	.08
Prompt	151.54	1	151.54	0.40	.53	.01	.09
CIWS pretest	29587.64	1	29587.64	77.42	.001	.69	1.00
Condition	1487.32	1	1487.32	3.89	.06	.10	.48
Error	13375.58	35	382.16				
Total	84324.36	40					

**Table 12b.***Adjusted and Unadjusted Group Means for Posttest CIWS*

	Adjusted <i>M</i>	Unadjusted <i>M</i>
Active Control	38.39	34.33
Treatment	25.59	28.9
Treatment Effect	-.37	

**Note:** Treatment effect calculated with Hedges' *g*.

### **Post-Hoc Intervention Fidelity Measures**

Following similar procedures as the primary analysis, I conducted a post-hoc analysis of group differences in student's incorporation of Long Words and Diverse Sentences to assess whether instructional condition had any relation on proximal metrics of instruction. For example, the vocabulary instruction was intended to directly impact the number of long words and increase in long words is associated with more distal writing outcomes. As reported in the descriptive statistics (Table 4), there were no statistically significant group differences in these variables but there was a practical difference (Hedges'  $g=0.33$ ) with respect to student's use of diverse sentences, favoring the Treatment. Topic Prior Knowledge was not included as a covariate due to the insignificant correlations with any outcome of interest (Table 13).

**Table 13.**

*Pearson Correlations Between Pre-test and Posttest Post Hoc Variables for All Instructional Conditions*

	1	2	3	4	5	6	7
1. Gender	1	.02	.12	.28	.21	.30	.35*
2. TPK, pretest		1	.28	.17	.21	-.20	.26
3. TPK, posttest			1	.27	.05	-.12	.21
4. Long Words, pretest				1	.17	.21	-.20
5. Long Words, posttest					1	.20	.64**
6. SLD, pretest						1	.02
7. SLD, posttest							1

**Note:** SK=science knowledge; TPK=Topic Prior Knowledge; SLD=Sentence Length Diversity.

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed)

An analysis of covariance was conducted to determine the effect of instruction condition and prompt on student's posttest Long Word scores after controlling for the effects of covariates (Gender and Prompt) and student's pretest Long Word performance. The linearity assumption was satisfied as no curvilinear pattern was detected on residual scatter plots. Visual inspection of line graphs showed parallel slopes for the pretest Long Word covariate, and the interaction term was insignificant [ $F(1,34)=0.58, p=.45$ ] indicating that the regression slope for pre-test Long Word can be used to predict groups posttest Long Word performance (Mertler & Reinhart, 2017). The homogeneity of variances assumption was violated [ $F(1,38)=16.59, p=.001$ ]. ANCOVA (Table 14a) results indicate no significant main effect of condition [ $F(1,35)=0.95, p=.34, \text{partial } \eta^2=.03$ ] but there was a significant main effect of prompt [ $F(1,35)=24.65, p<.001, \text{partial } \eta^2=.41$ ] on student's posttest Long Word scores. Students' pre-test Long Word scores was significant in predicting students' posttest Long Word score [ $F(1,35)=34.94, p=.001, \text{partial } \eta^2=.50$ ]. The adjusted mean scores (Table 14b) indicate that students in the Active Control included more long words in their writing ( $M=7.60$ ) compared to the Treatment group ( $M=6.46$ ), and these differences were statistically significant, but it is not clear that such a difference was attributable to instructional condition. Whether condition had an influence is inconclusive given the low observed statistical power ( $1-\beta=.16$ ). Yet, Hedges'  $g$  effect size for the practical significance of the pretest to posttest differences of Long Words due to condition was meaningful (Hedges'  $g = -0.21$ ), though in the opposite direction as was anticipated. This may indicate that the vocabulary instruction in the treatment condition was limited, or instruction was not intensive enough to compensate for effects of the prompt on students writing. .

**Table 14a.***Analysis of Co-Variance Summary Table Examining Group Differences at Posttest for Long Words*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial $\eta^2$	Observed Power
Corrected Model	605.85	4	151.46	11.86	.001	.58	1.0
Gender	2.87	1	2.87	0.22	.64	.01	0.07
Prompt	314.72	1	314.72	24.65	.001	.41	1.0
Long Words, pretest	446.12	1	446.12	34.94	.001	.50	1.0
Condition	12.12	1	12.12	0.95	.34	.03	.16
Error	446.93	35	12.77				
Total	2996.72	40					

**Table 14b.***Adjusted and Unadjusted Group Means for Posttest Long Words*

	Adjusted <i>M</i>	Unadjusted <i>M</i>
Active Control	7.60	7.22
Treatment	6.46	6.77
Treatment Effect	-.22	

**Note:** Treatment effects calculated with Hedges' *g*.

ANCOVA was conducted to determine the effect of instructional condition and prompt on student's inclusion of diverse sentences after controlling for covariates and pretest performance. The linearity assumption was satisfied. The interaction of Sentence Length Diversity at pretest and condition was insignificant suggesting that the homogeneity of regression slopes was satisfied, [ $F(1,34)=.20, p=.66$ ]. The Levene's test was also insignificant, [ $F(1,38)=1.43, p=.24$ ]. ANCOVA results (Table 15a) yielded no significant main effects of condition on student's posttest Sentence Length Diversity scores. Likewise, student's pretest performance also did not yield a significant effect [ $F(1,35)=.008, p=.78$ ]. Only Gender had a significant effect on students' Sentence Length Diversity scores, [ $F(1,35)=4.11, p=.05$ , partial  $\eta^2=.11$ ], with an advantage for girls. This difference can't be attributed to the prompt as the effects of prompt on the dependent variable were insignificant, [ $F(1,35)=.42, p=.52$ , partial  $\eta^2=.01$ ]. The posttest differences had a negligible practical significance (Hedges'  $g=0.02$ ), indicating that treatment likely did not include intense enough sentence instruction to make an impact on sentence variety (see Table 15b).

**Table 15a.***Analysis of Co-Variance Summary Table Examining Group Differences at Posttest for Sentence Length Diversity*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial $\eta^2$	Observed Power
Corrected Model	7.09	4	1.77	1.48	.23	.14	.41
Gender	4.94	1	4.94	4.11	.05	.11	.50
Prompt	0.5	1	0.5	0.42	.52	.01	.10
SLD, pretest	0.09	1	0.09	0.08	.78	.00	.06
Condition	0.15	1	0.15	0.12	.73	.00	.06
Error	42.06	35	1.2				
Total	239.88	40					

**Table 15b.***Adjusted and Unadjusted Group Means for Posttest Sentence Length Diversity*

	Adjusted <i>M</i>	Unadjusted <i>M</i>
Active Control	2.18	2.35
Treatment	2.07	2.04
Treatment Effect	-.02	

**Note:** Treatment effect calculated with Hedges' *g*.



## **Social Validity**

In this study, students across conditions participated in additional writing time during science class while only the Treatment students were recipients of the science writing scaffolds. In general, students across conditions indicated that they felt like they were better at science writing and were better at writing overall. 71% of students receiving the Active Control and 95% of students receiving the Treatment condition indicated that they were better at writing “like a scientist”; similar percentages were reported for writing in general. Students in the Active Control condition indicated that they like brainstorming solutions to the problem-based writing prompt at the end of the unit and found planning, including discussing the problem as a class, to be the most helpful. In contrast, the Treatment group overwhelmingly reported that they benefitted from drawing and diagramming their thinking to each topic outlined in the prompt and that this method of planning helped them to write their paragraphs. Both classes reported that the week they spent responding to the prompt was challenging. Students reported that the end of unit writing task involved “too much writing” and “not enough time to write”. Some students reported that writing was difficult because they didn’t know how to start writing, when to stop writing, and found some elements of the prompt conceptually more difficult to respond to than others. Overwhelmingly, students in both conditions reported that writing was an important skill for them to have especially because they would need it in their work as adults. Though not a target of instruction, students across conditions provided valuable insights about the how the nature of scientific writing compares to other disciplines. In addition to the general statement that scientists write to inform, writing is used to help scientists remember technical information, solve problems, and communicate the same information to scientists from different cultures. Finally, of the students that answered question 5 of the open-ended part of the survey, two

students indicated that it would be helpful to know why students need to know science and what is the relative importance of writing about it.

Mr. Frizzle and Mrs. Honey both reported that they believed the science writing scaffolds were helpful supporting their students writing in science and intend to use them in the future. Both teachers reported that they felt students benefitted from the modified science block in that it allowed more time for students to process the material, and they intend to incorporate more writing into the science block. Mr. Frizzle noted that due to the lack of science curriculum at his school, having any writing scaffold available to him was beneficial. Meanwhile, Mrs. Honey noted that using the science scaffolds was more difficult because there was not a curriculum infrastructure available to her in the way that the English curriculum supports her teaching during any given writing block. She notes that these supports, embedded within a curriculum with examples, non-examples, and pedagogical tips would be helpful. However, since she often supplements her English curriculum with additional resources and student graphic organizers to address the specific needs of her students, having the science writing scaffolds available to her in her teaching toolkit will benefit her future instruction, especially because it tends to be rich in history and science content. The teachers note that the modification of the science block, which required the two classrooms to separate so that contamination during the writing sessions did not occur, was the most difficult component of the study. For Mrs. Honey, it required additional prep and study so that she could answer student's science-related questions. Both teachers indicate that they would welcome a year-long scope and sequence for the sciences that integrated the science scaffolds across reading, writing, experimentation, and class discussion. To echo a point made by some students, such a scope and sequence would also include some emphasis on how science writing is similar and different from other types of writing.

## CHAPTER 5: DISCUSSION

The current study sought to explore the effectiveness of a multi-component intervention teaching Grade 4 students to “write like scientists” through the incorporation of science-specific writing scaffolds intended to support students’ morphology development, sentence writing, and both comprehension of texts/experiments and writing a CER text using a graphic organizer. Due to the unique teaching and learning context of a small convenience sample of students, the intervention package was contrasted against an Active Control condition. What resulted was all students, regardless of treatment, received 2-3 additional hours of writing time per week (as reported by the teachers) in the science classroom, with or without the science writing supports. As the teachers in this study co-teach, a unique opportunity was presented to understand the role that the researcher-generated writing scaffolds could play across instructional settings (ELA or Science) during the learning of science-related material. This process illuminated instructional and logistical barriers that make it difficult to assess the effectiveness of the intervention’s effect on students’ science argument writing. With these difficulties in mind, I was able to identify factors that facilitate ongoing efforts to integrate science writing in late elementary classrooms. This study has implications for future research exploring the teaching and learning of science in real classrooms as well as the development of future science and content-rich ELA curriculum. The results of this study demonstrate that the award-winning science teacher (Mr. Frizzle) had a significant impact on science outcomes (i.e., science knowledge) and that the master ELA teacher (Mrs. Honey) had a significant impact on writing outcomes (CIWS, sentence diversity). Neither teacher had an impact on science writing (CER). This is not surprising given that neither teacher was able to effectively teach CER given the multiple barriers of implementation. As noted from the interviews, these barriers include usability of the CER materials, lack of

integration with science and ELA content, competing materials in the ELA curriculum (i.e., the RACES strategy and lack of science information content), and, most importantly, lack of scheduled time to teach writing.

### **Inconclusive Findings Still Hold Practical Significance**

Writing is one of the most complex communicative acts that a person can engage in as they attempt to articulate their ephemeral thoughts into tangible products governed by morphosyntax, structure, and rhetorical expectations (Kim & Park, 2019; Nagy & Townsend, 2012; Norris & Phillips, 2003). While writing can serve as an important tool to promote learning (Graham et al., 2020; Phillips Galloway, 2020), it is easy to set aside during instruction in favor of instructional areas that are equally as important (e.g., reading) and less effortful for both students and teachers (Troia & Maddox., 2010). Despite writing instruction having meaningful effect sizes on learning for elementary students (Hedges'  $g=.29$ ; Graham et al., 2020) and their science learning outcomes (Hedges'  $g=.31$ ; Graham et al., 2020), middle school teachers tend to spend fewer than 32 minutes a week on writing instruction (Graham et al., 2014). That middle school students write an average of 2-5-minutes a day in science class calls into question whether the benefits of writing to learn is being fully realized (Graham et al., 2014). What is more concerning about the paucity of writing instruction in classrooms is that while students benefit from writing instruction, a substantial proportion of students require explicit writing instruction and are educated alongside their typically achieving peers (NCES, 2023; NCW, 2003). If elementary students are expected to start thinking and writing like scientists, as laid out by science standards and indicated by state assessments of science achievement (e.g., MSTEP), it then becomes essential that students a) write in science and b) are supported while doing it. Yet, a systematic review conducted by Lee and De La Paz (2021a) note that few ( $N=14$ ) science

writing interventions include students who have language learning difficulties and even fewer ( $n=3$ ) focus their study on students in Grade 4. The supports described in each of these studies included cognitive, linguistic, and general learning supports. This dissertation was an attempt to incorporate linguistic and cognitive scaffolds into a Grade 4 classroom for the purposes of supporting student's science arguments.

This study of group mean differences offers some insights and future directions for supporting elementary children's science argument writing despite the small effect sizes and indeterminable results. Small treatment effect sizes and indeterminate effects should not immediately discount an intervention from future study (Vaughn et al., 2010). Graham and colleagues (2020) meta-analyzed 56 writing-to-learn studies to understand the types of study features. 41 of the studies contrasted the Treatment with a Control condition in which no additional writing occurred. 15 of the studies compared the Treatment with a condition that received less writing. Neither description fits this study as both Treatment and Control, experienced a school-level barrier of no additional time for writing. As effect sizes are generated by the contrast between Treatment and Control, smaller effect sizes are observed when the Control condition is also receiving additional instruction (Vaughn et al., 2010). The additional instruction in this study was particularly strong given the skills of the participating teachers.

The intensity and duration of an intervention are also likely to impact the magnitude of an intervention. As the intervention in this study is a Tier 1 intervention and occurred over 6 weeks, it is unlikely that these factors alone would garner large effects. Intervention studies with larger sample sizes, more intense instruction, and duration of intervention than the current study have found disappointing results (Vaughn et al. 2010). With respect to the Next Generation Science Standards, science content knowledge is built gradually from grade-to-grade in a series

of learning progressions and so it would be unreasonable to expect large gains in knowledge over a short period of time (Gotwals, 2018). Rather than discard an intervention all together because of its perceived ineffectiveness, it is possible that the lack of an effect indicates a need for ongoing interventions administered to students on a systems-level rather than be the responsibility of a single teacher during a singular point in time (Vaughn et al., 2010).

Most importantly, the science writing intervention described in the current study was not feasible to implement as originally conceived and so key components were not implemented, which is like many other large-scale studies of academic language instruction (Corrin et al., 2014, 2022). After considering all that can be improved upon in the current study, I will discuss my findings, implications for research and practice, and next steps.

## **Multicomponent Intervention’s Effect on Learning Outcomes**

### ***Science-Specific Writing Outcomes***

The primary writing outcome of this study was the quality of student’s written science arguments (CER). To do this, teachers were provided training on how they might implement the science specific writing scaffolds within their specific teaching contexts. For Mr. Frizzle, the writing supports provided some much-needed structure for navigating discussions and modeling during the science blocks for which students responded to end-of-lab questions. And for Mrs. Honey, the science instructional scaffolds helped to incorporate discipline-specific language during the reading and writing about science content while in English. In contrast, student’s not receiving these instructional scaffolds received the same English Language Arts instruction minus the scaffolds and wrote in science for about 2 hours every week. Results from this study indicated that there is no conclusive evidence that this was due to the intervention. According to the results of the ANCOVA, the pre- to posttest differences in student’s CER writing cannot be

attributed to the instructional condition that they were placed in. The small effect size that was observed (Hedges'  $g=0.07$ ) indicate that the Treatment Group outperformed the Active Control group at posttest on their Total CER scores. This effect is smaller than the effect size needed to be deemed practically significant (WWC, 2021). An investigation of what did not go to plan during this study can inform future studies to assess the role of these instructional scaffolds on elementary students' science argument writing.

Previous studies have been successful in implementing graphic organizers as the primary tool of support for students when learning to write science arguments (Bulgren et al., 2009; Mason et al., 2006) and graphic organizers tend to be the most implemented instructional scaffold used to support students with the cognitive demands of writing in science (Lee & De La Paz, 2021a). Specifically, as students become more proficient at using a graphic organizer, the number of conceptual ideas and complex vocabulary included in their written responses increase (Mason et al., 2006, Lee & De La Paz, 2021a). Unfortunately, the graphic organizer that was designed for this study went unutilized due to time constraints. This may help to explain the disappointing practical effects observed in the current study with respect to writing quality (Hedges'  $g=0.07$ ), whereas studies that incorporated a graphic organizer found a large practical effect ( $d=1.44$ ) favoring the Treatment group (Bulgren et al., 2009). In the current study, teacher participants felt they lacked the time to teach students how to use another graphic organizer and still get through content, which is a point of contention noted by previous studies evaluating educators' beliefs about writing's role in the classroom (Troia & Maddox, 2010). Because of its perceived ease of implementation, teachers implemented the CER sentence starters during their writing instruction with the Treatment group 2-3 times a week. This scaffold was also observed being used with the Treatment group during every observation. The use of the CER sentence

starters is an implicit scaffold as it does not teach students what these three structural components are, but particularly advanced students may be able to intuit these structural elements.

Another reason for the ineffectiveness of the intervention on student's CER writing was because Active Control received a second exposure to a previously learned text structure . Furthermore, the RACES (**R**estate and **A**nswer the question, **C**ite your evidence and **E**xplain it, **S**ummarize your answer) strategy that was taught in the Active Control seems to share some overlapping language with CER and may generalize to CER (and possibly other content areas) when student's use the structure alongside science-rich ELA curriculum.

Without explicit text structure support (Lee & De La Paz, 2021a, 2021b), instruction on the specialized language of science alone is not powerful enough to make noticeable gains in discourse-level measures in the short span of 6 weeks. Using teacher feedback, the CER Organizer should be modified to better meet the needs of elementary classrooms (it was modeled after Bulgren et al.'s (2009) work with high school students) and perhaps, in response to the worries teachers expressed regarding implementation, be more aligned with the informational text structure used in the ELA class.

### ***General Writing Outcomes***

This dissertation extends the work conducted by previous researchers by including a general writing outcome measure (CIWS) which captures student's word and syntax fluency (Kim, et al., 2019; Tortorelli & Truckenmiller, 2024). While many studies reviewed for this dissertation evaluated students' writing for CER quality (Klein & Rose, 2010; Sampson & Clark, 2009), structure (Benedek-Wood et al., 2014; Herbert et al., 2018), or science content (Bulgren et al., 2009), few studies also examined student's general writing abilities (Bulgren et al., 2009; Lee



& De La Paz, 2021b). Using a 6-Traits rubric, Bulgren et al., (2009) found that a small sample of high school with disabilities taught how to use a graphic organizer to summarize lesson information and respond to a science question grew from pre-test to posttest on all Traits except Conventions. Meanwhile Lee & De La Paz (2021b) examined three middle school students writing for grammatically and lexically sophisticated sentence following an instruction on language parts that scientists use to write more clearly and accurately. In both instances, these researchers examined these writing outcomes using rubrics, which at larger sample sizes, can become time consuming (Troia et al., 2019) and subject to a rater's discretion as time goes on (Leckie & Baird, 2011). The general writing outcome in this study was measured within the context of a curriculum-based measurement (CBM) framework using the Writing Architect (Truckenmiller et al., 2019) which allows assessors to glean valuable information with respect to student's writing expression (WE) abilities in a short duration of time (Romig et al., 2017). Past research has called for CBM work to examine student performance on other types of writing tasks (Truckenmiller et al. 2021). Most science CBMs assess students via vocabulary matching or statement verification (Conoyer et al., 2018). To the best of my knowledge, this is the first study that has attempted to use CBM-WE to directly assess student's science writing abilities. Despite this, any gain in student's CIWS scores and any posttest differences between groups could not be attributed to instructional condition. CBM-WEs role for measuring students' progress in the content areas other than reading and math (Conoyer et al., 2019; Romig et al., 2017) is a worthwhile area for future research.

The two homerooms in this study differed with respect to their initial CIWS levels, with the students in Mrs. Honey's Homeroom (who received the Active Control) having higher mean levels of CIWS at pretest compared to Mr. Frizzle's Homeroom. The magnitude of these

differences was practically meaningful, Hedges'  $g = -.31$ . As teachers reported drastically different levels of explicit instruction in writing, it is reasonable to assume that students in the Treatment group only got this type of instruction during ELA when they were with Mrs. Honey but not in the science classroom. Meanwhile, AC students also likely received this type of instruction during the 2 extra hours of writing instruction that they spent with Mrs. Honey. These differences in instruction, in part an artifact of the instructional design where classes had to stay separate during science writing periods, may be the reason that AC outperformed the Treatment condition at posttest. At posttest, AC's adjusted mean scores were 12.8 points higher than the treatment group; the two classes differed by 8.32 points at pretest. As correlations indicate that CIWS and the quality of student's CER writing at posttest (but not at pre-test) are related ( $r = .55$ ,  $p < .01$ ), supporting students' automaticity in writing should not be neglected during science writing instruction. In fact, students' automaticity in writing shares overlapping variance with their reading speed (Truckenmiller & Tortorelli, 2024); low CIWS scores may indicate that students are spending valuable writing time and cognitive energies simply processing the text rather than responding to the science question posed to them. For these reasons, supporting students learning via linguistic supports, alongside strategies that promote comprehension, are reasonable to include in an intervention package.

Yet, despite measuring general and content-specific writing outcomes, results of this dissertation indicate no effect due to condition. The effects of prompt on student's writing outcomes are a worthwhile area of exploration and there is some evidence in the current study to suggest that the writing prompts influenced student's writing outcomes. At pretest, the mean differences in student responding (regardless of the condition they were placed in) on CIWS (Hedges'  $g = .24$ ), CER Total (Hedges'  $g = .58$ ), Long Words (Hedges'  $g = .67$ ), and Sentence

Length Diversity (Hedges'  $g=.30$ ) were practically meaningful. Specifically, the students who responded to the *Invasive Species* prompt at posttest had lower CER ( $M_{adj}=5.03$ ) and CIWS ( $M_{adj}=30.46$ ) scores, incorporated fewer Long Words ( $M_{adj}=4.03$ ), and wrote shorter sentences ( $M_{adj}=1.99$ ) compared to students who responded to *Extinctions* despite having comparable readability scores. Student's prior knowledge might have played a role in their differential performance on these topics as previous experiences and content knowledge influence both reading comprehension (Cabell & Hwang, 2020) and writing (Brown et al., 2010; Kim & Park, 2019). After following recommendations to measure and control for Topic Prior Knowledge (Benedek-Wood et al., 2014; Brown et al., 2010; Lee & De La Paz, 2021), Topic Prior Knowledge did not differ in any meaningful way between students responding to the prompt. That Prompt had a significant main effect on the number of Long Words wrote in their CER essays suggesting that students had more content vocabulary knowledge to discuss *Extinctions*.

Patricia Alexander (1988) asserts that as children are asked to engage in more discipline-specific practices in schools, such as producing arguments in a manner resembling a scientist, they require higher levels of conceptual knowledge to efficiently enact discipline specific strategies. A scientist in training is simultaneously learning about science concepts at the same time they are learning about scientific language, genres of discourse, and the unique mechanics of style. A young scientist's expression of their declarative knowledge (content knowledge) via acceptable procedures (argument writing) is hypothesized to be dependent upon the condition onto which they need to access that knowledge. In the Writing Architect task, students procedural and conditional knowledge at posttest should have been greater than that at pre-test as they had prior experience with this task. What's left is their declarative knowledge. Grade 4 students, at least at Aster Elementary, may have more nascent conceptions of what an invasive

species is compared to their knowledge of extinctions. According to the NGSS, as part of students' mastery of science standard Life Science 2A: Interdependent Relationships in Ecosystems, students should know that "newly introduced species can damage the balance of an ecosystem" by the end of Grade 5 (NRC, 2012 pp.152). In contrast, students start to develop an understanding about "some kinds of plants and animals that once lived on Earth [but] are no longer found anywhere" as early as Grade 2 (NRC, 2012, pp. 162). Despite attempting to make the invasive species prompt more relatable by making the context of the story perhaps personally familiar, students' limited experiences with the invasive species concept impaired their ability to perform as well as they might have if they responded to a prompt whose concepts were more well developed. As a result, students' CIWS scores may have been negatively affected and reflect their attempts at processing the text more so than demonstrating their procedural knowledge.

CIWS may be more sensitive to student's growth over the course of the 6-week unit due to the increased writing time rather than instructional condition. However, this sensitivity may also make the outcome more vulnerable to prompt effects. One means of getting ahead of this issue is to include a variety of science-specific CMB-WE prompts that span the grade bands and include topics that are most extensively covered by the science standards. Alternatively, the questions that are asked during the CBM-WE administration could be phrased in a way that is less abstract and more accessible for diverse learners to respond to. If CBM-WE is to be included in the currently limited list of CBMs for science content, it will be important that CER elements are included so that students' argumentation can be monitored over time. Currently, it is not clear if 6 weeks was enough instructional time to make an impact on students learning. Given that the rater reliability for scoring CER (ICC=.91) and CIWS (ICC=.99) was high given this small

sample, developing the codebooks and scoring systems for such a CBM is a reasonable line of future research, particularly given the writing demands of end-of-year science assessments. Ensuring students success on these types of high-stakes assessments necessitates the monitoring of students writing so that instruction and timely intervention can be implemented.

### ***Science Knowledge***

Few studies reviewed for this dissertation examined the effects of a science writing intervention on both student science writing outcomes and conceptual knowledge (Lee et al., 2009; Rouse et al., 2017). While some research aligns the writing outcome with the unit of instruction (Benedek-Wood et al., 2014; Rouse et al., 2014), others call for assessing students' pretest knowledge of taught science concepts so that researchers can better understand if science learning occurred (Lee & De La Paz, 2021). In this study, the pre- and posttest science measure was directly aligned with the unit of instruction and was content validated by the teachers who planned the unit of instruction. Learning did occur over the duration of the study as indicated by the change in Cronbach's alpha from pre- ( $r=.42$ ) to posttest ( $r=.69$ ). While practice effects might have been contributed to the change in reliabilities, it is more likely that the low Cronbach's alpha at pre-test was due to students guessing (Taber, 2018) especially as these Grade 4 students rarely take multiple-choice science assessments. In this study, there was no significant main effect of learning condition on student's posttest science knowledge, but the differences for the adjusted posttest scores between AC ( $M_{adj}=13.87$ ) and Treatment ( $M_{adj}=16.06$ ) was practically meaningful (Hedges'  $g=0.51$ ). To better understand if the intervention played a role in student's Science Knowledge, it would be reasonable for students to respond to a writing task that was directly related to the unit and the unit assessment.

It should be noted that measuring a student’s conceptual knowledge is more nuanced than simply examining the pre- to posttest differences on assessments like the one administered for this study. When measuring a child’s science knowledge, their reading comprehension is likewise being measured. Previous studies have indicated that reading comprehension accounts for more than 70% of the variance in state science achievement scores for Grades 5 through 8 (Reed et al., 2107). Because content experts engage with text in ways that are unique to their discipline (Shanahan & Shanahan, 2008), it is worthwhile to integrate discipline-specific reading comprehension practices when teaching science content. Previous studies have been successful at incorporating reading comprehension strategies within existing instructional units for social studies (Vaughn & Wanzek, 2024). For the current study a graphic organizer was developed to support students’ comprehension of science texts and experiments, but it went unutilized in favor of the CER Sentence Starters. It then remains unclear how the intervention, beyond additional writing time, contributed to students’ abilities to comprehend science text, including the questions posed to them on the test.

### **The Role of Writing Instruction in Learning Progressions**

Learning progressions (LP) can be conceived as curricular maps whereby “the knowledge, skills, and understandings of a learning area are sequenced” and the performance expectations of a given grade level are defined (Master & Forster, 2013; Shepard et al., 2013). The NGSS make use of LPs. As students advance in their schooling, their knowledge of scientific subjects deepens, and students’ prior science learning becomes essential to the development of more sophisticated knowledge. At any given grade level, there are two “anchors”: the baseline level of competencies coming into a grade band and an upper limit (Gotwals, 2018). As described earlier, Grade 4 students have had years of exposure to topics

relating to extinctions as the lower limit of competency was first established in Grade 2.

Conversely, knowledge of invasive species is an upper limit anchor where students should understand this concept by the end of Grade 5.

Scientists develop new knowledge via experimentation and utilize empirical data to craft their scientific arguments. In an ongoing effort to acclimatize students to the science discipline, seven of the eight science practices described in the NGSS Framework describe the type of work scientists engage in over the course of their inquiry (e.g., asking questions, planning and carrying out experiments, constructing explanations; NRC, 2012). How to incorporate LPs into a system of formative assessment (Gotwals, 2024) is an ongoing conversation in the science educational space, including the use of multiple modes of assessment (e.g., performance tasks, video assessments, oral arguments), with a primary objective being to assess how well students have internalized the *doing of science*. Yet, the language and structure of scientific talk is an important instructional element to attend to over the course of any learning progression, particularly because teachers are teaching *students*-who are not scientists engaging in primary data collection- who must use textual means at all stages of a learning cycle to engage with the content (Norris & Philips, 2003). For example, students who are designing a model of how plants use carbon dioxide, water, and light to grow have had to examine mentor texts of other science models to construct their own.

Reading and writing instruction are not separable from scientific practice. While Practice 8 explicitly addresses reading and writing in the sciences (Obtaining, Evaluating, and Communicating Information), Practice 6 and 7 ask students to construct explanations and engage in argument from evidence. Students are routinely asked to do this via writing using informational text passages, including on end-of-year state assessments (e.g., MSTEP), and so it

is instructionally relevant to provide explicit language instruction when students write in response to text. Discipline-specific language knowledge is built upon foundational and domain-general academic language skills (Schleppegrell, 2001; Uccelli et al., 2015) but there is evidence to suggest that explicit language instruction is rare in the disciplines (Lee & De La Paz, 2021a; Drew & Thomas, 2018). While the current study evaluated a language-focused intervention for elementary science, it may be worthwhile for researchers to explore how language instruction in the disciplines might fit within current LPs. If LPs are to be used as an assessment framework to accommodate all language learners' success in science, they must also consider the role that language plays in supporting students' full access to the content.

### **Implications**

The results of this study have implications for future curriculum development with respect to content-rich ELA and science curriculum. A national survey found that middle school ELA and content teachers shared the responsibility to teach their students but lacked the time, resources, and/or knowledge to implement this instruction in their classrooms (Graham et al., 2014). Other studies have suggested that supporting children's background knowledge through the incorporation of content-rich ELA instruction could support children's later reading comprehension (Cabell & Hwang, 2020) which has lasting effects on children's academic performance in the content areas (Reed et al., 2017). Mrs. Honey and Mr. Frizzle have built a strong working relationship that leverages their backgrounds in literacy and science instruction to increase the number of opportunities that their students must read, discuss, and explore issues that integrate all the academic content areas. That they rely on each other's skill areas to improve their instruction speaks to the fact that they view the responsibility to their students learning as something that is shared. And still, both teachers express difficulty in implementing more writing



instruction. This dissertation study assessed the feasibility in incorporating science writing scaffolds into their typical instruction and in doing so, it revealed some areas of improvement that can be applied to existing ELA and Science curricula.

For students to become better science writers, they need to be opportunities to express their disciplinary knowledge in disciplinary ways. While general writing strategies can be facilitative of this endeavor, “talking like a scientist” involves the interaction of knowledge and ways of thinking that is emblematic of that discipline (Lemke,1990). Content-rich ELA curriculums, like the one used in this study, attempt to support children’s content knowledge by providing opportunities to read, write, and discuss texts that help to answer a driving question. The ELA unit in this study asked students to write an informational text that answers the science question, “How do animal’s bodies and behavior help them survive?”

Mrs. Honey expressed that the CER scaffolds, but especially the CER sentence starters, helped students to write responses that were more science like and gave them some structure for making the moves common in an argument. As a supplementary scaffold, the CER scaffolds were helpful but without a fully realized curriculum to accompany them, she would struggle to support her students learning if she did not have her English curriculum. Mr. Frizzle was the most enthusiastic about all the scaffolds, even the CER graphic organizer that he did not use. For him, the Mystery Science curriculum is sorely lacking in supports for how to teach writing at all and the CER scaffolds provide him with some type of structure for building his own instructional practice. Both teachers would like to see a fully realized, year-long science curriculum that not only includes a phased approach to introducing these scaffolds before they become routinized, but suggestions for differentiation, tips for implementation, instructional texts, and exemplars. Amplify Science is a curriculum on the market as of this writing that purports to be a literacy

rich science curriculum. It would be beneficial for districts to have access to more options. Feedback from teachers also suggests that content rich ELA curriculum could also incorporate discipline-specific resources that help students to adopt the language and literacy practices that are common in that content. The sentence starters used in this dissertation appear to have been easy to adopt and interfered very little with the ELA curriculum.

According to the Council for Exceptional Children, there are four areas of instructional practice that educators can implement to support struggling learners, including those with reading and writing difficulty, in schools. Among those practices is collaboration.

“Collaboration with individuals or teams requires the use of effective collaboration behaviors (e.g., sharing ideas, active listening, questioning, planning, problem solving, negotiating) to develop and adjust instructional or behavioral plans based on student data, and the coordination of expectations, responsibilities, and resources to maximize student learning. (CEC, 2023)”

Collaborative relationships should exist between literacy and content teachers as their skill sets are complimentary to one another. Yet, a focus group conducted by Troia and Maddox (2004) suggest a disconnect between special education teachers (who are experts in literacy instruction) and general education teachers that hinders their abilities to maximize student learning. For example, despite their extensive knowledge for implementing and differentiating writing instruction for struggling writers, Special Education teachers felt their general education counterparts did not see a role for them with respect to instruction. While general education teachers worried about how to implement reading and writing during content instruction, it was difficult to implement individualized instruction for every learner. Mrs. Honey and Mr. Frizzle represent the possibility that something different can exist in schools. Though this study forced

the teachers to separate during science writing instruction for experimental purpose, these teachers jointly support their students during science. Future studies should explore the effects of and feasibility of team teaching (through ELA and content-area teacher or Special Education and content-area teacher partnerships) on content area writing further.

### **Limitations**

There are several limitations to consider when interpreting any results of this study. First, the sample size of this convenient sample is small (>100 cases), and so it is likely that there is not enough statistical power to make an accurate statistical inference regarding the effects of the intervention on student writing and knowledge outcomes (Lomax & Hahs-Vaughn, 2012). Small sample sizes can result could be an increase in Type II error in which the researcher incorrectly fails to reject the null hypothesis, thus resulting in a “false negative” (Lomax & Haha-Vaughn, 2012; Mertler & Reinhard, 2017). If an effect due to the intervention is present it is unlikely to be detected. I conducted a post hoc power analysis to assess the statistical power of the ANCOVA results and result of that analysis do indicate that this study is underpowered and the likelihood of drawing a correct inference from statistical tests is low. Future studies should incorporate larger sample sizes and consider a power analysis during the conceptualization of the method and plan for recruitment.

Additionally, there was a violation of the homogeneity of regression slopes for ANCOVAs, usually through visual inspection of the regression slopes (Mertler & Reinhart, 2017). A violation of this assumption could indicate that student performance on the outcome differed according to student achievement levels and if so, ANCOVA therefore becomes an inappropriate analysis to conduct (Mertler & Reinhart, 2017). It is also possible that the observed floor effects for CER at pretest reduced the predictive validity of measures at posttest (Catts et

al., 2008). Quantile regression is an alternative analysis that could address these limitations and understand how students at varying performance levels grow because of instruction (Catts et al., 2008).

Another potential issue affecting the interpretation of results is the decision to use multiple imputation to contend with missing data. Though it is the preferred option for dealing with missing data (Cokluk & Kayri, 2011; Rubin, 2018), there are tradeoffs (Mertler & Reinhart, 2017). In the case of this study, multiple imputation resulted in lower post-test means for the Treatment group than was observed but did allow for more degrees of freedom during statistical tests. Greater degrees of freedom are important for making it easier to reject the null hypothesis (Lomax & Hahs-Vaughn, 2012). With a small sample already hindering the statistical power of the study, a decision to increase the degrees of freedom at the expense of a small change in any pre-posttest differences was reasonable. As Hedges' *g* indicate that small insignificant differences between groups at posttest were still practically meaningful, it may be the case that, had some other treatment for missing data been applied, significance between groups may have been observed.

Finally, this study was not truly randomized and so it is feasible unobserved spillover effects occurred between conditions, especially because teachers taught both the Treatment and Active Control conditions. Future studies should consider recruiting a greater number of teacher pairs, perhaps from different schools to account for this. The underpowered nature of this research study limited the number of covariates that could be included in the analysis and detect an effect. A general rule for selecting covariates to include in ANCOVA is one covariate for every 15 participants (Mertler & Reinhart, 2017). Past studies have recommended that writing interventions control for variables such as typing fluency (Graham et al., 2020; Tortorelli &

Truckenmiller, 2024) and often include measures of reading comprehension (Hebert et al., 2018; Mason et al., 2006; Tortorelli & Truckenmiller, 2024). While these variables were collected during data collection, a decision was made to exclude these variables as potential covariates to align with best practices in variable selection. Since CIWS has associations with both constructs, such an exclusion was thought to be permissible. A more strongly powered iteration of this study should incorporate these controls.

### **Conclusion and Future Directions**

Due to worries about adequately covering the required content in a limited time frame, it was not feasible for teachers to implement the CER graphic organizer at all. Past studies (Lee & De La Paz, 2021a) identified graphic organizers as the most implemented instructional scaffold used in studies assessing science argument writing. Future studies might evaluate the effectiveness of instruction using each of the scaffolds individually and their effect on student writing outcomes. To draw clearer conclusions about the role of explicit instruction using these routines on student outcomes, it would be beneficial for future research to have a clearly defined more feasible protocol for instruction and utilize intervention checklists to monitor the implementation of key instructional components. Leaving teachers to figure out how to use the scaffolds with minimal direction is perhaps overly. An analysis of teacher/student talk and its mediating role in learning could be useful for understanding elements of the instructional routines that most support the practices of the disciplines using them. As these instructional routines were used by an English teacher when reading science text versus a science teacher engaging in experimentation, understanding such differences in teacher usage could be facilitated by such a close examination. The group design of this study did not assess how the different components of CER grew over time in a manner like past studies (Klein & Samuels, 2010) but

an individual differences design could consider such a detailed analysis. Furthermore, there was some evidence to suggest that students' performance levels on pre-test measures (CER Total, CIWS) influenced students posttest outcomes. If intervention research is to help inform teacher decision making about instruction and intervention response, a more nuanced analysis of student groups, such as through a quantile regression, may be productive. Future researchers should consider implementation issues with their teacher partners when designing interventions that introduce new instructional scaffolds and even consider how those very scaffolds could complement existing ones so that teachers are not overwhelmed by "just one more thing". To bridge the research to practice divide, instructional resources and practice guides should be produced for consumption by the individuals who will ultimately implement the routines-teachers.

This dissertation sought to investigate whether a multicomponent intervention using instructional scaffolds supporting students' morphology, sentence construction, and comprehension of texts/experiments supported Grade 4 students' science argument writing. Though this study is inconclusive at best regarding the role that the intervention played on student writing outcomes, it is reasonable to conclude that introducing more writing time into elementary classrooms is possible via cooperative teaching arrangements between content and literacy teacher experts.

## REFERENCES

- Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education*, 29(14), 1745–1765. <https://doi.org/10.1080/09500690601075629>
- Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14. <https://doi.org/10.3102/0013189X032008010>
- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58(4), 375–404. <https://doi-org.proxy1.cl.msu.edu/10.2307/1170279>
- Andreev, L., & Uccelli, P. (2023). The secret life of connectives: A taxonomy to study individual differences in mid-adolescents' use of connectives in writing to persuade. *Reading and Writing*. <https://doi.org/10.1007/s11145-023-10425-3>
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). Guilford Press.
- Benedek-Wood, E., Mason, L. H., Wood, P. H., Hoffman, K. E., & McGuire, A. (2014). An experimental examination of quick writing in the middle school science classroom. *Learning Disabilities: A Contemporary Journal*, 12(1), 69–92. [EJ1039825.pdf \(ed.gov\)](#)
- Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93(1), 26–55. <https://doi-org.proxy2.cl.msu.edu/10.1002/sce.20286>
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, 43(2), 79–120. [https://doi-org.proxy1.cl.msu.edu/10.1207/s15326950dp4302\\_1](https://doi-org.proxy1.cl.msu.edu/10.1207/s15326950dp4302_1)
- Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities*, 35(1), 39–56. <https://doi.org/10.1177/002221940203500104>
- Block, N. C. (2019). Evaluating the efficacy of using sentence frames for learning new vocabulary in science. *Journal of Research in Science Teaching*, 57(3), 1–25. <https://doi.org/10.1002/tea.21602>
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22(1), 1–14. [https://doi.org/10.1016/S0022-5371\(83\)80002-4](https://doi.org/10.1016/S0022-5371(83)80002-4)

- Brown, B. A., Donovan, B., & Wild, A. (2019). Language and cognitive interference: How using complex scientific language limits cognitive performance. *Science Education*, 103(4), 750–769. <https://doi.org/10.1002/sce.21509>
- Brown, B. A., Ryoo, K., & Rodriguez, J. (2010). Pathway towards fluency: Using ‘disaggregate instruction’ to promote science literacy. *International Journal of Science Education*, 32(11), 1465–1493. <https://doi.org/10.1080/09500690903117921>
- Bulgren, J. A., Marquis, J. G., Lenz, B. K., Schumaker, J. B., & Deshler, D. D. (2009). Effectiveness of question exploration to enhance students’ written expression of content knowledge and comprehension. *Reading & Writing Quarterly*, 25(4), 271–289. <https://doi.org/10.1080/10573560903120813>
- Cabell, S. Q., & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly*, 55(S1), S99–S107. <https://doi.org/10.1002/rrq.338>
- Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers’ processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429–441. <https://doi.org/10.1037/a0022824>
- Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, 45(4), 464–487. <https://doi.org/10.1598/RRQ.45.4.5>
- Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of learning disabilities*, 42(2), 163-176. <https://doi.org/10.1177/0022219408326219>
- Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In E. B. Moje, P. P. Afflerbach, P. Enciso, & N. K. Lesaux, *Handbook of Reading Research, Volume V* (1st ed., pp. 237–260). Routledge. <https://doi.org/10.4324/9781315676302-13>
- Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing*, 29(4), 761–779. <https://doi.org/10.1007/s11145-016-9628-x>
- Coker, D., & Lewis, W. E. (2008). Beyond writing next: A discussion of writing research and instructional uncertainty. *Harvard Educational Review*, 78(1), 231–250. <https://doi.org/10.17763/haer.78.1.275qt3622200317h>
- Çokluk, Ö., & Kayri, M. (2011). The effects of methods of imputation for missing values on the validity and reliability of scales. *Educational Sciences*.



- Collins, G., Wolter, J. A., Meaux, A. B., & Alonzo, C. N. (2020). Integrating morphological awareness in a multilingual structured literacy approach to improve literacy in adolescents with reading and/or language disorders. *Language, Speech & Hearing Services in Schools, 51*(3), 531–543. [https://doi.org/10.1044/2020\\_LSHSS-19-00053](https://doi.org/10.1044/2020_LSHSS-19-00053)
- Common Core of Data. (2023). *Public School Data*. National Center for Education Statistics (NCES). [https://nces.ed.gov/ccd/schoolsearch/school\\_detail.asp](https://nces.ed.gov/ccd/schoolsearch/school_detail.asp) Date Accessed: May 01, 2024.
- Conoyer, S. J., Ford, J. W., Smith, R. A., Mason, E. N., Lembke, E. S., & Hosp, J. L. (2019). Examining curriculum-based measurement screening tools in middle school science: A scaled replication study. *Journal of Psychoeducational Assessment, 37*(7), 887–898. <https://doi.org/10.1177/0734282918803493>
- Corrin, W., Sepanik, S., Gray, A., Fernandez, F., Briggs, A., & Wang, K. K. (2014). Laying tracks to graduation: The first year of implementing Diplomas Now. *MDRC*.
- Corrin, W., Zhu, P., Shih, M., Brown Jr, K. T., Teres, J., Darrow, C., ... & Lack, K. (2022). The effects of an academic language program on student reading outcomes. Appendix. NCEE 2022-007a. *National Center for Education Evaluation and Regional Assistance*.
- Council of Chief State School Officers. (2010) *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Math* (pp. 7). [https://learning.ccsso.org/wp-content/uploads/2022/11/ELA\\_Standards1.pdf](https://learning.ccsso.org/wp-content/uploads/2022/11/ELA_Standards1.pdf)
- Council for Exceptional Children. (2023). *About the HLPs*. <https://highleveragepractices.org/about-hlps>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213. <https://doi.org/10.2307/3587951>
- Crosson, A. C., & Lesaux, N. K. (2013). Connectives: Fitting another piece of the vocabulary instruction puzzle. *The Reading Teacher, 67*(3), 193–200. <https://doi.org/10.1002/TRTR.1197>
- Crosson, A. C., Lesaux, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics, 29*(4), 603–625. <https://doi.org/10.1017/S0142716408080260>
- Davidson, M., & Berninger, V. (2016). Informative, compare and contrast, and persuasive essay composing of fifth and seventh graders: Not all essay writing is the same. *Journal of Psychoeducational Assessment, 34*(4), 311–321. <https://doi.org/10.1177/0734282915604977>
- De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94*(4), 687–698. <https://doi-org.proxy1.cl.msu.edu/10.1037/0022-0663.94.4.687>

- Dixon, W. J., & Yuen, K. K. (1974). Trimming and winsorization: A review. *Statistische Hefte*, 15(2), 157–170. <https://doi.org/10.1007/BF02922904>
- Drew, S. V., & Thomas, J. (2018). Secondary science teachers' implementation of CCSS and NGSS literacy practices: A survey study. *Reading and Writing*, 31(2), 267–291. <https://doi.org/10.1007/s11145-017-9784-7>
- EL Education. (2024). Curriculum: Grade 4. [ELA 4 | EL Education Curriculum](#)
- Ferretti, R. P., Andrews-Weckerly, S., & Lewis, W. E. (2007). Improving the argumentative writing of students with learning disabilities: Descriptive and normative considerations. *Reading & Writing Quarterly*, 23(3), 267–285. <https://doi.org/10.1080/10573560701277740>
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practice*, 7(1), 25–35. <https://doi.org/10.1111/j.1745-3992.1988.tb00422.x>
- Gillespie-Rouse, A., Graham, S., & Compton, D. (2017). Writing to learn in science: Effects on Grade 4 students' understanding of balance. *The Journal of Educational Research*, 110(4), 366–379. <https://doi.org/10.1080/00220671.2015.1103688>
- Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, 17(4), 257–285. <https://doi.org/10.1080/10888438.2012.689791>
- Gotwals, A. W. (2018). Where are we now? Learning progressions and formative assessment. *Applied Measurement in Education*, 31(2), 157–164. <https://doi.org/10.1080/08957347.2017.1408626>
- Gotwals, A., Songer, N., & Bullard, L. (2012). *Assessing Students' Progressing Abilities to Construct Scientific Explanations* (pp. 183–210). [https://doi.org/10.1007/978-94-6091-824-7\\_9](https://doi.org/10.1007/978-94-6091-824-7_9)
- Graham, S. (2020). The sciences of reading and writing must become more fully integrated. *Reading Research Quarterly*, 55(S1). <https://doi.org/10.1002/rrq.332>
- Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing*, 27(6), 1015–1042. <http://dx.doi.org.proxy1.cl.msu.edu/10.1007/s11145-013-9495-7>
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710–744, 784–785. <https://doi.org/10.17763/haer.81.4.t2k0m13756113566>

- Graham, S., Kiuahara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, 90(2), 179–226. <https://doi.org/10.3102/0034654320914744>
- Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research*, 88(2), 243–284. <https://doi.org/10.3102/0034654317746927>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Halliday, M. A. K., & Hassan, R. (1976). *Cohesion in English*. Routledge. <https://doi.org/10.4324/9781315836010>
- Halliday, M.A.K., & Martin, J.R. (1993). *Writing Science: Literacy and Discursive Power* (1st ed.). Routledge. <https://doi.org/10.4324/9780203209936>
- Hammill, D. D., & Larsen, S. C. (2009). *Test of written language: TOWL4*. Pro-ed.
- Hand, B., Chen, Y.-C., & Suh, J. K. (2021). Does a knowledge generation approach to learning benefit students? A systematic review of research on the science writing heuristic approach. *Educational Psychology Review*, 33(2), 535–577. <https://doi.org/10.1007/s10648-020-09550-0>
- Hand, B., Wallace, C. W., & Yang, E. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. *International Journal of Science Education*, 26(2), 131–149. <https://doi.org/10.1080/0950069032000070252>
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., ... & REDCap Consortium. (2019). The REDCap consortium: building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. <https://doi.org/10.1016/j.jbi.2019.103208>.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Hebert, M., Bohaty, J. J., Nelson, J. R., & Roehling, J. V. (2018). Writing informational text using provided information and text structures: An intervention for upper elementary struggling writers. *Reading and Writing*, 31(9), 2165–2190. <https://doi.org/10.1007/s11145-018-9841-x>

- Kamberelis, G. (1999). Genre development and learning: Children writing stories, science reports, and poems. *Research in the Teaching of English*, 33.
- Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science (Version 1). Deakin University. <https://hdl.handle.net/10536/DRO/DU:30099108>
- Kim, Y.S. G., & Park, S.H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing*, 32(5), 1319–1343. <https://doi.org/10.1007/s11145-018-9913-y>
- Klein, P. D., & Rose, M. A. (2010). Teaching argument and explanation to prepare junior students for writing to learn. *Reading Research Quarterly*, 45(4), 433–461. <https://doi.org/10.1598/RRQ.45.4.4>
- Klein, P. D., & Samuels, B. (2010). Learning about plate tectonics through argument-writing. *Alberta Journal of Educational Research*, 56(2), Article 2. <https://doi.org/10.11575/ajer.v56i2.55398>
- Leckie, G. & Baird, J. A. (2011). Rater effects on essay scoring: A multilevel analysis of severity drift, central tendency, and rater experience. *Journal of Educational Measurement*, 48, 399-418. <https://doi.org/10.1111/j.1745-3984.2011.00152.x>
- Lee, O., Mahotiere, M., Salinas, A., Penfield, R. D., & Maerten-Rivera, J. (2009). Science writing achievement among English language learners: Results of three-year intervention in urban elementary schools. *Bilingual Research Journal*, 32(2), 153–167. <https://doi.org/10.1080/15235880903170009>
- Lee, Y., & De La Paz, S. (2021a). Science writing intervention research for students with and at risk for learning disabilities, and English learners: A systematic review. *Learning Disability Quarterly*, 44(4), 261–274. <https://doi.org/10.1177/07319487211018213>
- Lee, Y., & De La Paz, S. (2021b). Writing scientific explanations: Effects of a cognitive apprenticeship for students with LD and English learners. *Exceptional Children*, 87(4), 458–475. <https://doi.org/10.1177/0014402921999310>
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Lomax, R.G. & Hahs-Vaughn, D.L., (2012). *An Introduction to Statistical Concepts* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203137819>

- Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA + PLANS strategies for expository reading and writing: Effects for nine fourth-grade students. *Exceptional Children*, 73(1), 20. <https://doi-org.proxy1.cl.msu.edu/10.1177/001440290607300104>
- McCann, T. M. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English*, 23(1), 62–76.
- McNamara, D., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Matrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- McNeill, K. L., & Krajcik, J. (2006). *Supporting students' construction of scientific explanation through generic versus context specific written scaffolds*. American Education Research Association, San Francisco. [http://websites.umich.edu/~hiceweb/papers/2006/McNeill&Krajcik\\_AERA2006.pdf](http://websites.umich.edu/~hiceweb/papers/2006/McNeill&Krajcik_AERA2006.pdf)
- McNeill, K. L., & Krajcik, J. (2007). Middle school students use of appropriate and inappropriate evidence in writing scientific explanations. In M. C. Lovett & P. Shah (Eds.), *Thinking with Data* (pp. 233–265). Lawrence Erlbaum Associates, Inc.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153–191. [https://doi.org/10.1207/s15327809jls1502\\_1](https://doi.org/10.1207/s15327809jls1502_1)
- McNeill, K. L., & Martin, D. M. (2011). Claims, evidence, and reasoning. *Science and Children*, 48(8), 52–56. <http://searkscience.pbworks.com/w/file/attach/70117336/2-Claimsevidence.pdf>
- Mertler, C. A., Reinhart, R. V. (2017). *Advanced and multivariate statistical methods: Practical application and interpretation*. Routledge.
- Miller, A.C., Keenan, J.M. (2009). How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia*, 59, 99–113. <https://doi.org/10.1007/s11881-009-0025-x>
- Moher, D., Schulz, K. F., Altman, D., & Consort Group. (2005). The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. *Explore*, 1(1), 40-45. <https://doi.org/10.1016/j.explore.2004.11.001>
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>

- National Center for Education Statistics. (2023). Students with disabilities. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved [date], from <https://nces.ed.gov/programs/coe/indicator/cgg>.
- National Commission on Writing in America's Schools and Colleges. (2003). *The Neglected R: The Need for A Writing Revolution*. The report of the National Commission on Writing in America's Schools and Colleges.
- National Research Council (U.S.) (Ed.). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.
- Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Report submitted to the Gates Foundation.
- Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224–240. <https://doi.org/10.1002/sce.10066>
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1), 45–65. <https://doi.org/10.1007/s11145-012-9392-5>
- Phillips Galloway, E., Qin, W., Uccelli, P., & Barr, C. D. (2020). The role of cross-disciplinary academic language skills in disciplinary, source-based writing: Investigating the role of core academic language skills in science summarization for middle grade writers. *Reading and Writing*, 33(1), 13–44. <https://doi.org/10.1007/s11145-019-09942-x>
- Pear Deck Learning. (2024). Why Pear Deck Learning? Efficacy data. <https://www.peardeck.com/efficacy>
- Reed, D. K., Petscher, Y., & Truckenmiller, A. J. (2017). The contribution of general reading ability to science achievement. *Reading Research Quarterly*, 52(2), 253–266. <https://doi.org/10.1002/rrq.158>
- Regents of the University of California. (2022). Smarter Balanced: Sample items. [https://sampleitems.smarterbalanced.org/Item/200-182667?&isaap=TDS\\_SCNotepad](https://sampleitems.smarterbalanced.org/Item/200-182667?&isaap=TDS_SCNotepad)
- Reynolds, G. A., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology*, 30(3), 265–300. <https://doi.org/10.1080/02702710802411547>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, 74(4), 445–458. <https://doi.org/10.1037/amp0000356>
- Rivard, L. P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education*, 88(3), 420–442. <https://doi.org/10.1002/sce.10114>

- Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education, 51*(2), 72–82. <https://doi.org/10.1177/0022466916670637>
- Rubin, D. B. (2018). Multiple imputation. In *Flexible imputation of missing data, second edition* (2nd ed.). Chapman and Hall/CRC.
- Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448–484. <https://doi.org/10.1002/sce.20306>
- Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. *Science Education, 97*(5), 643–670. <https://doi.org/10.1002/sce.21069>
- Sarmiento, C.M., Hennenfent, Lauren G., Truckenmiller, A.J. (in prep). Qualitative analysis of 7+ letters in middle school children’s academic writing.
- Sarmiento, C. M., & Truckenmiller, A. J. (2024). What is important to measure in sentence-level language comprehension? *Assessment for Effective Intervention, 0*(0). <https://doi-org.proxy1.cl.msu.edu/10.1177/15345084241265620>
- Sarmiento, C. M., Truckenmiller, A., Cho, E., & Wang, H. (2022, June 29). Academic language use in middle school informational writing. <https://doi.org/10.31234/osf.io/umt7w>
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*(4), 431–459. [https://doi.org/10.1016/S0898-5898\(01\)00073-0](https://doi.org/10.1016/S0898-5898(01)00073-0)
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content- area literacy. *Harvard Educational Review, 78*(1), 40–59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Shepard, L., Daro, P., & Stancavage, F. B. (2013). *The Relevance of Learning Progressions for NAEP* (ED-04-CO-0025/0012). American Institute for Research. <https://files.eric.ed.gov/fulltext/ED545240.pdf>
- Snow, C. (2010). Academic language and the challenge for learning about science. *Science, 328*(5977), 450–452. <http://www.jstor.org/stable/40655773>
- Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325–344. <https://doi.org/10.1080/19345740903167042>
- Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching, 49*(2), 141–165. <https://doi.org/10.1002/tea.20454>

- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tortorelli, L. S., & Truckenmiller, A. J. (2024). Automaticity in writing in response to reading: Relations between oral reading fluency and compositional writing fluency in grades 3–5. *Reading & Writing Quarterly*, 40(2), 103–117. <https://doi.org/10.1080/10573569.2023.2172757>
- Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional levels of language writing measures in grades four to six. *Written Communication*, 36(2), 231–266. <https://doi.org/10.1177/0741088318819473>
- Troia, G. A., & Maddox, M. E. (2004). Writing instruction in middle schools: Special and general education teachers share their views and voice their concerns. *Exceptionality*, 12(1), 19–37. [https://doi.org/10.1207/s15327035ex1201\\_3](https://doi.org/10.1207/s15327035ex1201_3)
- Truckenmiller, A.J. & Bowles, R., (July 2019). Diagnostic profiles of written expression in middle grades. In H. Gerde (Chair), *Writing development: Predictors, profiles, and intervention*. Symposium presented to the Society for the Scientific Study of Reading, Toronto, Canada.
- Truckenmiller, A. J., Cho, E., & Troia, G. A. (2022). Expanding assessment to instructionally relevant writing components in middle school. *Journal of School Psychology*, 94, 28–48. <https://doi.org/10.1016/j.jsp.2022.07.002>
- Truckenmiller, A. J., McKindles, J. V., Petscher, Y., Eckert, T. L., & Tock, J. (2019). Expanding curriculum-based measurement in written expression for middle school. *The Journal of Special Education*, 002246691988715. <https://doi.org/10.1177/0022466919887150>
- Truckenmiller, A. J., & Petscher, Y. (2020). The role of academic language in written composition in elementary and middle school. *Reading and Writing*, 33(1), 45–66. <https://doi.org/10.1007/s11145-019-09938-7>
- Truckenmiller, A., Shen, M., & Sweet, L. E. (2021). The role of vocabulary and syntax in informational written composition in middle school. *Reading and Writing*, 34(4), 911–943. <https://doi.org/10.1007/s11145-020-10099-1>
- Truckenmiller, A.J., Valentine, K.A., & Sarmiento, C.M. (2023). Essential Practices in Writing Checklists – Writing Architect. [OSF link and doi to be determined].
- Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sánchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 36(5), 1077–1109. <https://doi.org/10.1017/S014271641400006X>



- Valentine, K. A., Truckenmiller, A. J., Troia, G. A., & Aldridge, S. (2021). What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change? *Assessing Writing*, 50, 100567. <https://doi.org/10.1016/j.asw.2021.100567>
- Vaughn, S., Wanzek, J., Wexler, J. et al (2010). The relative effects of group size on reading progress of older students with reading difficulties. *Reading and Writing*, 23, 931–956. <https://doi-org.proxy1.cl.msu.edu/10.1007/s11145-009-9183->
- Vaughn, S., & Wanzek, J. (2024). Promoting Adolescents’ Comprehension of Text: Efficacy and Effectiveness. *Remedial and Special Education*, 45(1), 58-67. <https://doi-org.proxy1.cl.msu.edu/10.1177/07419325231190805>
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://doi.org/10.1037/a0036620>
- What Works Clearinghouse. (2021). Reporting guide for study authors: Group designs. *Institute for Educational Sciences*, 4. [https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final\\_WWC-HandbookVer5\\_0-0-508.pdf](https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf)
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). <https://ies.ed.gov/ncee/wwc/Handbooks>.
- Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34, 16–36. <https://doi.org/10.1016/j.asw.2017.08.002>
- Wright, K. L., Hodges, T. S., Zimmer, W. K., & McTigue, E. M. (2019). Writing-to-learn in secondary science classes: For whom is it effective? *Reading & Writing Quarterly*, 35(4), 289–304. <https://doi.org/10.1080/10573569.2018.1541769>

## APPENDIX A: WRITING ARCHITECT SCIENCE PASSAGES

### How to Speed up Extinctions

It seems like every day there is a story about a polar bear struggling to find food in a melting Arctic. Polar bears are just the tip of the iceberg when it comes to animals on the brink of extinction. According to an analysis of 15,000 studies, 1 in 8 plant or animal species may disappear. That means almost 1 million species could join the dinosaurs as creatures of the past.

What's behind this surprising finding? People. The number of people on the planet has doubled in the last 50 years. In 1970, there were about 3.7 billion people on the planet, but that number is 7.6 billion today! The activities that have helped humans thrive, have also caused plants and animals to go extinct. The report states that 40% of amphibians, 33% of ocean mammals, 33% of sharks, and 10% of insects could soon go extinct.

If humans are not more mindful, extinctions will continue to speed up. Therefore, it is helpful to know how people negatively impact the planet. Here are four ways that people are speeding up extinction.

#### 1. Fewer Places to Live

The top threat to species on land is habitat loss. About 75% of land on Earth has been changed to build cities and increase farmland. Since 1992, cities have grown by more than 100%. To feed more people, healthy habitats have been turned into farmland. 85% of wetlands and 32% of rainforest that were around in 1760 are gone. For example, rainforests are being replaced with cattle ranches. All that land development makes it hard for animals to find a good place to live.

#### 2. Overfishing the oceans

Overfishing is the greatest human danger that ocean creatures face. People love seafood- over 3 billion people rely on seafood for protein! 55% of the ocean's surface is fished, and about 33% of the ocean's fish are overfished. This means there is not a lot of these fish left behind after the fishing is done. Tuna is one of the most overfished species in the world and their numbers are shrinking in the wild. Another downside of overfishing is that other animals, like turtles and dolphins, also get trapped in fishing nets. These unwanted catches are called bycatch.

### **3. Dirtying the environment**

Humans have not done enough to cut down on pollution. One of the biggest problems is our love of plastic. It's everywhere! Ocean plastic has increased by ten times since 1980. It has harmed 267 species of ocean animals. They mistake the plastic for food or get trapped in it. On land, tiny pieces of plastic end up in the soil or in drinking water. Other sources of pollution, like oil spills or dirty drinking water, is also a problem. Pollution can harm an animal's health and make a place unlivable.

### **4. Paving the way for invaders**

Humans bring invasive species to new areas as they travel the world. Invasive species not only compete with native species for food but can wipe them out. Across 21 countries, the number of invasive species has grown by 70 % since 1970, the report finds.

### **But there's hope ...**

Humans can slow extinctions, the researchers note. Conservation efforts have lowered the risk that many plants and animals will go extinct. To save more species, people need to rethink their behavior including how they use land, grow food, and what they throw out.

## **A Diet to Fuel an Invasive Carp**

84% of the freshwater in the United States is in the Great Lakes. People use the lakes for fishing, boating, wildlife viewing, and recreation. That business brings in \$8.5 billion a year. Keeping the lakes enjoyable depends on a healthy ecosystem. But keeping Lake Michigan healthy and free of invading critters is a stressful job. Two species of invasive fish, the bighead and silver carp, have been wreaking havoc on the Mississippi River. At first, scientists were not worried about the fish getting into the Great Lakes. They thought there was not enough for the fish to eat. However, a new study found that if invasive carp reach Lake Michigan, they might be able to survive on a buffet of mussel poop. Now, scientists are both worried and disgusted.

### **Not Such a Great Place to Live**

Scientists once thought that the Great Lakes was a food desert for the carp. A food desert is any place that an animal has a hard time finding food. Animals tend to look somewhere else for food. Sandra Cooke studies freshwater ecosystems, like the Great Lakes. Cooke says the carp prefer to eat phytoplankton, which is a type of algae. These tiny plants live in warm water. Because the Great Lakes get so cold, scientists thought there wasn't enough food to satisfy the carp's big appetite. A warm lakeshore might have enough algae for the fish to eat in the summer. In the winter though, the fish would starve.

Peter Alsip, a scientist from the University of Michigan, says the carp are not picky eaters. When there are no good food options available, the fish will live off detritus. Detritus, or animal waste, includes fish and mussel poop and dead organisms. Detritus is like junk food for fish. Alsip and his team studied Lake Michigan, the fishes' appetite — including for junk food — and their energy needs to predict where the carp could live. Invasive zebra and quagga mussels cover much of the lake floor. Mussel poop could make the deeper, colder parts of the

lake livable for the invasive fish. The carp could eat mussel poop to survive a Lake Michigan winter.

### **A Costly Mistake**

Bighead and silver carp were brought to the United States in the 1970s to control the growth of algae, which thrived in polluted rivers. But during floods, the fish escaped into the wild. Soon, the fish found new homes in the Mississippi River. They have been traveling north ever since and are knocking on the Great Lakes' doorstep. If the carp can eat mussel poop and algae, 75% of Lake Michigan could be a comfortable place for the fish to live. The fish would likely use the lake as a highway, traveling from lake to lake in search of warmer waters and better food.

The last invasive species to sneak into the Great Lakes was the quagga mussel. It now covers the bottom of Lake Michigan. When the mussels first invaded, they changed the chemistry of the water. This change killed many native whitefish and caused waterbirds to get sick. Businesses spend \$100,000 a year to remove the mussels from water that comes out of the lake. Once an invasive species travels someplace new, the damage is expensive and can't be undone.

If carp gain a foothold in Lake Michigan, their populations could take off. "We should be doing everything we can to keep bighead and silver carps out of the Great Lakes," says Sandra Cooke. "Time and again, what we actually observe is worse than what we predicted in the first place," she says.

## **CO<sub>2</sub>-loving Plants**

There are holidays for everything. We celebrate our love for our dogs, for donuts, and so much more during unofficial holidays. On the last Friday in April, the nation celebrates Arbor Day. Arbor Day is all about appreciating and celebrating trees. Volunteers across the country plant trees to celebrate nature and the environment. A new study may give the country more of a reason to celebrate trees. According to the study, during 2002 to 2014 plants slowed the rate of CO<sub>2</sub> collecting in the air.

### **Plants Make Their Own Energy**

From the tallest trees to the smallest flower, plants are amazing. Plants use the energy in sunlight to make their own food. Plants collect carbon dioxide gas (also called CO<sub>2</sub>) from the air and water from the soil and turn it into sugar. This chemical process is called photosynthesis.

When there is more CO<sub>2</sub> in the air, plants can make more sugars through photosynthesis. The plants can grow quickly! The sugar that plants make can be stored for safekeeping until a later time. Plants may store their sugars in their roots as bulbs (like tulips) and tubers (like potatoes), or in their stems (like celery). This stored sugar becomes a food source that plants can use to grow, to flower, and to make seeds. Plant growth is slow. This process is the opposite of photosynthesis. It is called respiration, and it releases CO<sub>2</sub> into the air.

### **Slowing the Rise of CO<sub>2</sub>**

81% of the energy that the United States uses comes from fossil fuels. A fossil fuel is energy that comes from burning fossilized plants and animal. Fossil fuels include coal, natural gas, and oil. These fuels release a lot of energy when burned. They also release CO<sub>2</sub> into the air. CO<sub>2</sub> is very good at trapping heat. This can fuel climate change. 75% of the carbon dioxide in the air is from fossil fuels. What can be done to lower the amount of carbon dioxide in the air?

According to the study, from 2002 to 2014 the amount of CO<sub>2</sub> people released in the air rose from 372 parts per million to 397 parts per million. During that time, people burned a lot of fossil fuels. Though CO<sub>2</sub> increased, it wasn't as quick as scientists were expecting. After lots of study, they concluded that plants slowed the rate that CO<sub>2</sub> collected in the air. Each year, land plants and the oceans remove about 45% of the CO<sub>2</sub> emitted from human activities. The amount of CO<sub>2</sub> they have absorbed has doubled over the last 50 years.

### **Plants Can't Do It All**

Today, carbon dioxide enters the air more quickly than plants can absorb it. As CO<sub>2</sub> collects, the climate warms. When plants are too warm, they are less effective at photosynthesis. Instead, plants respire. All that plant breathing releases CO<sub>2</sub>.

In the early 2000's, the rate of rising CO<sub>2</sub> concentrations outpaced the rate of global warming. This caused plants to absorb more CO<sub>2</sub> during photosynthesis than they released during respiration. That imbalance slowed the buildup of atmospheric CO<sub>2</sub>. While the amount of CO<sub>2</sub> increased at a rate of 0.75 parts per million in 1959, it rose to a rate of 1.86 parts per million thirty years later. But between 2002 and 2014, the rate held at around 1.9 parts per million.

Trevor Keenan wrote the study. "If we keep emitting as much as we are, and what we emit keeps going up, then it won't matter very much what the plants do," he warns. While plants keep CO<sub>2</sub> out of the air, planting more trees every Arbor Day is not the solution to slowing climate change.

## **Fertilizer for Rooftop Gardens**

Living in a city can be exciting. Cities provide many chances for work and play. New York is one of the largest and densest cities in the United States. 8.8 million people live in 468 square miles of space! According to the United Nations, 55% of the entire world's population lives in cities. Providing homes to so many people in a small area often means building upwards. The most famous cities in the world are known for their impressive skyscrapers. By 2050, over 68% of the world's population will live in cities. Scientists and urban planners are working together to find ways to make city living even better. They want to lower the amount of pollution in cities by increasing the amount of green space. One way to do this is to turn unused rooftop space into gardens.

## **Lots of Energy to Keep Cool**

Living in a city can be tough. Cities have few green spaces, such as parks or forests, compared to rural areas. Land must be cleared to make room for buildings and factories. Cities need to build hundreds of miles of roads and sidewalks so people can move around. Keeping a city running requires a lot of energy, which produces a lot of gases that linger in the air. These gases are sometimes called greenhouse gases. As a result, cities have a great deal of air pollution, or smog. Smog can make people ill. In fact, cities are responsible for 70% of the greenhouse gases that are released into the air.

Since cities have fewer shade trees, cities are warmer than nearby areas. This is called the heat island effect. A lot of energy used by cities is used to keep buildings cool. Bare city rooftops soak up heat from the sun, which warms the building. In response, air conditioners turn on to bring the temperature back down. Air conditioners use energy that often comes from coal-fired power plants. When coal burns, it releases CO<sub>2</sub> into the air. CO<sub>2</sub> is a greenhouse gas.



## Turning Carbon into Fertilizer

Many scientists are studying how well rooftop gardens work in lowering city pollution. Plants “breathe in” CO<sub>2</sub> from the air during photosynthesis. This process allows plants to make their own energy and grow. Most rooftop garden plants are smaller and less healthy than plants in regular gardens. This may be because rooftop gardens get more solar radiation and wind. The water in the soil may evaporate faster. These conditions may limit how fast young, fragile plants can grow.

Dr. Sarah Buckley works at Boston University. She came up with a genius way of helping rooftop plants out. She created a new device that connects to exhaust fans at the tops of buildings. It funnels any CO<sub>2</sub> gas from the building onto a garden bed. The gas acts like a fertilizer to help the plants grow. Her team grew spinach and corn plants on a roof to test how well the device works. Some plants grew next to the building exhaust vents, which was rich in CO<sub>2</sub>. The control plants grew next to a regular fan. They did not get extra CO<sub>2</sub> from the building. The plants that grew next to the building’s vents grew 4 times larger than the control plants!

The CO<sub>2</sub> inside buildings can help rooftop plants grow larger. The new devices are part of a plan to make rooftop gardening easier. Buckley says rooftop gardens have many benefits, “...such as energy savings for the building, urban heat reduction, local food production, community building, and aesthetic and mental health benefits.” Windy rooftops are still a challenge for rooftop gardens since wind stunts a plant’s growth. However, her invention is just the first step in making gardens on every roof possible.

## **Multi-tasking Windmills**

Americans have been dealing with their share of natural disasters, such as hurricanes and heat waves. Many are worried that these disasters are fueled by climate change. After all, climate change raises the odds that bad weather will happen. Climate change is a shift in the normal patterns of temperature and weather of an area. Scientists agree that human activities, like burning coal, cause climate to change more quickly today than in the past. Coal releases a lot of carbon dioxide into the air when it is burned. CO<sub>2</sub> in the air acts like a blanket: it traps heat and warms the planet. It is also a blanket that is hard to take off. It is easier to put CO<sub>2</sub> into the air than it is to take it out.

The United States gets most of its electricity from coal. Scientists are looking for ways to get power from other energy sources. Renewable energies, like wind power, release almost zero CO<sub>2</sub> into the air. They also never disappear. Now, some engineers are trying to build a windmill that not only makes energy but also removes CO<sub>2</sub> from the air.

### **Windmills Pull in Dirty Air**

Windmills are also called wind turbines. They are tall structures with blades that reach into the sky. The blades are attached at an angle so that they spin when the wind blows. This spinning turns a set of gears, which then starts a generator. The generator turns the energy from the spinning blades into electricity! The faster a windmill spins, the more energy it makes.

Luciano Castillo is an engineer. He says windmills also pull air down behind them when they spin. Castillo's team uses computers to test the carbon removing power of windmills. According to their data, windmills may pull down CO<sub>2</sub> from the air. CO<sub>2</sub> could be removed from the air if it can make it to the windmills.

### **Pros and Cons**

Having carbon removing windmills in cities could be useful. Windmills may lower the amount of pollution in and around cities by pulling it out of the air. This would be helpful for cities since they seem to always be clouded in dirty air. The dirty air comes from cars and factories. Windmills could also lower the cost of electricity. Windmills could make the cost of taking CO<sub>2</sub> out of the air cheaper too.

Some people doubt that Castillo's idea will work. They say that the CO<sub>2</sub> made by power plants is too high in the air. The windmills would not reach it to pull it down. Others worry that a windmill farm could damage the environment. According to a Harvard study, the number of windmills required to meet America's energy needs would heat the country up by 0.43 degrees! That warming would cancel out the climate benefit of the windmills for at least 100 years. The study also says that the US is unlikely to use wind power as its only energy source. Perhaps the pros of using wind power outweighs the cons.

### **Next Steps**

Castillo's windmills are not in use, yet. He hopes to scale up his study to test if windmills really can capture CO<sub>2</sub>. He would like his next study to take place in Chicago. Chicago is called The Windy City. "The beauty is that around Chicago, you have one of the best wind resources in the region, so you can use the windmills to take some of the dirty air in the city and capture it,"

Castillo says.

## APPENDIX B: MORPHOLOGY ROUTINE

### The Wonderful World of Words

The word that I am learning today is:

#### Spelling by Sound

If I cover the word up, I can sound it out to help me spell it. (Try it out! Cover the word and try sounding it out).

#### Morphology

Words are made of word parts, called morphemes. This can help me learn the meaning of the word.

Does this word have any prefixes? yes no

The prefix of the word \_\_\_\_\_ is \_\_\_\_\_.

The prefix means:

Does this word have a root word? yes no

The root word of \_\_\_\_\_ is \_\_\_\_\_.

The root word means:

Does this word have any suffixes? yes no

The suffix of the word \_\_\_\_\_ is \_\_\_\_\_.

The suffix means:

Based on these word parts, I predict that the word \_\_\_\_\_ means:

#### Dictionary Definitions

A dictionary or glossary of terms can help me figure out the meaning of words. According to the dictionary, the word \_\_\_\_\_ means:

#### Words in the Wild

I may see my new word used in many ways when reading. Because my word can be used in different ways, the meaning of my word might change.

<p>Example 1:</p>	<p>My word is being used as a:</p> <p><input type="checkbox"/> Noun- a person, place, thing, or idea.</p> <p><input type="checkbox"/> Adjective- a word that describes something.</p> <p><input type="checkbox"/> Verb- an action</p> <p><input type="checkbox"/> A process- describing how something happens.</p> <p>In this text, my word means:</p>
-------------------	--

<p>Example 2:</p>	<p>My word is being used as a:</p> <p><input type="checkbox"/> Noun- a person, place, thing, or idea.</p> <p><input type="checkbox"/> Adjective- a word that describes something.</p> <p><input type="checkbox"/> Verb- an action</p> <p><input type="checkbox"/> A process- describing how something happens.</p> <p>In this text, my word means:</p>
-------------------	--

**Synonyms and Other Word Relatives**

<p>Synonyms are words that mean the same as my word. Some synonyms for the word structure are:</p>	<p>When I think about my word _____ I also think about other words and ideas like:</p>
--	--

**My Own Definition**

<p>After lots of study, I can make my own definition for the word _____. It means:</p>	<p>If I could draw what my word means, it would look like this:</p>
--	---

## APPENDIX C: CER GRAPHIC ORGANIZER

### Let's Argue- For Science!

Name:	Date:
Class:	
This information source is (check one): <input type="checkbox"/> an experiment <input type="checkbox"/> an article.	
Title:	
What is the main science question being asked?	
Key terms and definitions that will help answer the question:	Materials (if an experiment)
1.	
2.	
3.	
<b>Claim #1:</b> What idea helps to answer the main science question?	<b>Evidence:</b> What evidence is there to support this claim?  <b>Source:</b> Who or what gave you this information?
<b>Reasoning:</b> How does the evidence support the claim? What does the evidence mean?	

<b>Claim #2:</b>	<b>Evidence:</b>  <b>Source:</b>
<b>Reasoning:</b>	
<b>Claim #3:</b>	<b>Evidence:</b>  <b>Source:</b>
<b>Reasoning:</b>	
Write a science argument, using science terms, that answers the main science question. Remember to include claims, evidence, and reasons.	

## APPENDIX D: CER SENTENCE STARTERS

Scientists use arguments to convince others of their answers to research questions. They make these arguments in three parts: claims, evidence, and reasons. You can make them too!

CLAIM	
A claim is the answer to a scientific question. It is often written as a statement of fact that helps to draw a conclusion. Claims can be made to answer questions during scientific experiments or from reading scientific texts.	
Sentence Starter	Example of You Might Use It
“I think/predict...”	“ <b>I think</b> seeds sprout faster when the soil is warm.”
“I conclude that...”	“ <b>I conclude that</b> sunflowers grow best in full sunlight.”
“I believe...”	“ <b>I believe that</b> recycling is good for the planet.”
“I know that...”	“ <b>I know that</b> bees are important plant pollinators...”
“It is true that...”	“ <b>It is true that</b> cold water is more dense than hot water.”
“It is false that...”	“ <b>It is false that</b> all bears hibernate in the winter.”

EVIDENCE	
Evidence is proof that the answer to a science question is correct. Evidence can come from texts, lab observations, data tables, charts, or graphs. Good scientists mention where they got their evidence from.	
Sentence Starter	Example of You Might Use It
“I know this because...”	“ <b>I know this because</b> seeds planted in warm soil sprouted 3 days faster than those in cold soil.”
“During the experiment, I observed...”	“ <b>During the experiment, I observed</b> that plants grown in sunlight were 2 inches taller than those grown in the shade.”
“According to the data table/chart/graph...”	“ <b>According to the data in Table 2,</b> countries who recycle more have lower rates of air pollution and asthma than those who don’t.”
“Research on/from...”	“ <b>Research on</b> bees, shows that they pollinate most of the plants that we harvest for food.”
“My research ...”	“ <b>My research</b> testing how fast color diffused in a cup of water showed that cold water was the slowest to completely change color.”
“According to the article/text...”	“ <b>According to the article</b> Do Bears Hibernate? bears that live in cold climates may hibernate but bears that live in warmer climates may not. ”
“For example...”	“ <b>For example,</b> brown bears in lower latitudes often leave their dens in the winter.”



## REASONING

Reasoning is a process of persuading the reader that the answer to the question is correct and can be believed. Reasons make use of science ideas and terms to help the reader understand what the evidence means and how it helps to answer the question.

Sentence Starter	Example of You Might Use It
“Since...”	“ <b>Since</b> plants need sunlight to grow, warm soil is likely to indicate favorable growing conditions.”
“Therefore...”	“ <b>Therefore</b> , cold soil likely indicates that conditions are not favorable for plant growth and germination may take longer.”
“In other words,”	“ <b>In other words</b> , sunflowers are not shade-loving plants and grow best when their light needs are met.”
“As a result,”	“ <b>As a result</b> of lowering the amount of food waste in the city, the local landfill produced less greenhouse gases.”
“So, this means...”	“ <b>So, this means</b> that without bees we would not enjoy food like almonds, apples, or tomatoes.”
“This demonstrates...”	“ <b>This demonstrates</b> that cold-water particles were more tightly packed together and prevented the dye from diffusing as quickly as in the hot water.”
“Based on the evidence...”	“ <b>Based on the evidence</b> , how tightly water molecules are packed together, also called density, depends on temperature.”
“According to [credible person/institution’s name] ...”	“ <b>According to Dr. Brown</b> , most bears enter a voluntary sleep called torpor and can wake up on their own.”
“My hypothesis is correct/incorrect because...”	“ <b>My hypothesis was correct</b> because my data shows seeds planted in warm soil sprouted faster than those planted in cold soil.”

### Example

**Do snails prefer strawberries or bananas? How do you know?**

**I predicted** that snails like strawberries more than bananas. **In our experiment**, we placed a snail on a plate with two fruits and then observed which fruit it picked first. We did this 10 times. 6 out of 10 times the snail picked the strawberry. **Based on the evidence, my hypothesis is correct. I think** the snail picked the strawberry because it has more sugar than the banana. **Since** sugar is energy, the snail would get more energy by eating the strawberry rather than the banana.

## APPENDIX E: TOPIC PRIOR KNOWLEDGE TEST ADMINISTRATIVE SCRIPT

**Time of administration:** Day before Writing Architect administration

**Location of administration:** English Language Arts class, beginning of period, whole class.

**Duration of task:** 5 min

1. At the beginning of the period, give students a face-down bulleted piece of paper with the writing task written on top.
2. Project the writing task on the class white board.
3. After students have received their writing paper, say the following: “Today you will have five minutes to tell me everything you know about a topic. You may have a topic that is different from your neighbor. You will write one thing that you know next to one bullet point. You will try to write as many facts as possible that you know in five minutes. Do not worry about spelling. Does anyone have any questions?”
4. Answer any clarifying questions for the students.
5. Read the writing task that is presented on the screen out loud, then say, “You can begin writing.”
6. Students have 5 minutes to complete the task. When they are done, thank them and collect the responses.

### Writing Task

Please tell me everything you know about **animal extinctions**. You may write your answer as bullet points.

## APPENDIX F: SCIENCE KNOWLEDGE TEST

1. What is the loosening and movement of weathered bits of rock or soil from one place to another?
  - a. Weathering
  - b. Erosion
  - c. Deposition
2. What can cause water erosion? **Choose two.**
  - a. Wind
  - b. Sun
  - c. Rain
  - d. Snow/ice
3. Select **two** sentences that describe examples of erosion.
  - a. Ice melts on a lake.
  - b. Waves rise and fall in the ocean.
  - c. Rainwater moves soil down a hill.
  - d. Rock forms at the bottom of the ocean
  - e. Wind blows sand on a beach to a different area.
4. The Amite River washes the riverbanks away because of its fast current during storms. This is an example of:
  - a. Weathering
  - b. Runoff
  - c. Erosion
5. Which of the following best describes how most soil forms?
  - a. Through the growth of trees in a forest.
  - b. Through the buildup of snow on an iceberg.

- c. Through the weathering of rock by wind and water.
  - d. Through the cooling of lava from a volcanic eruption.
6. The drawings below show the same rock at two different times. [image depicting ice wedging].
- Which set of information **best** identifies and describes the process shown in the drawings?
- a. **Process:** Erosion.  
**Description:** Water froze in a crack in the rock and caused the rock to break apart
  - b. **Process:** Weathering.  
**Description:** Water froze in a crack in the rock and caused the rock to break apart
  - c. **Process:** Erosion.  
**Description:** Water moved through a crack in the rock and carried small pieces of the rock to a new area
  - d. **Process:** Weathering.  
**Description:** Water moved through a crack in the rock and carried small pieces of the rock to a new area
7. What effect does weathering have on rocks? **Choose two.**
- a. It makes them larger.
  - b. It changes their color.
  - c. It breaks them down.
  - d. It makes them smoother.
8. What is the main difference between weathering and erosion?
- a. Weathering breaks down rocks at the same place, while erosion carries the broken parts away.
  - b. Weathering and erosion are the same process.
  - c. Weathering moves rocks, while erosion breaks them up.
  - d. Erosion only happens with the help of wind, while weathering happens even without wind.
9. A fault is a:

- a. Break in the earth's crust.
  - b. A large, flat piece of land.
  - c. A special type of volcano.
  - d. None of the above.
10. What is the melted rock that was once under Earth's surface called once it reaches Earth's surface?
- a. Core
  - b. Magma
  - c. Lava
  - d. Mantle
11. Volcanoes erupt when
- a. Plateaus are formed.
  - b. Liquid rock breaks through the crust.
  - c. Water and wind wear away at the volcano.
  - d. Glaciers and ice change the surface of the earth.
12. Students observe two hills, Hill A and Hill B. After a heavy rainstorm, the students observe that Hill A has more soil at the bottom than Hill B. Which sentence **best** describes what **most likely** caused the loose soil?
- a. Hill A has fewer plants than Hill B.
  - b. Hill A erodes more slowly than Hill B.
  - c. Hill A has a slope not as steep as Hill B.
  - d. Hill A receives less average rainfall than Hill B.
13. Which of the following is a solution to stop water erosion? **Choose 3.**
- a. Using mulch
  - b. Removing plants
  - c. Building walls
  - d. Planting trees on slopes

14. Use your knowledge of weathering and erosion to sort the following items into the correct category:

weathering, erosion, or weathering AND erosion. Be careful and choose wisely.

- a. Rocks fall down the side of a mountain
- b. Wind moving sand at the beach
- c. Plant roots breaking rock
- d. A process that moves rock that has been broken into smaller pieces
- e. A process that breaks rock down into smaller pieces
- f. Can be caused by ice, wind, and water
- g. Rocks being washed down a river and hitting other rocks that break them

15. The students were asked to look at locations that will be best to construct a new building. They study

a chart that shows factors that affect rates of weathering. [chart] Next, the students study a chart showing characteristics of four locations in Wisconsin. [chart] Which location most likely has the slowest rate of rock weathering?

- a. Location 1
- b. Location 2
- c. Location 3
- d. Location 4

## APPENDIX G: TEACHER CONSENT FORM

### For participation in a research study

**Study title:** A Pilot Study of Explicit Instruction in Science Argument Writing

**Researcher:**

Department of Counseling, Educational Psychology and Special Education,  
Michigan State University

**Address:**

**Phone:**

**Email:**

Dear Teacher,

My name is Cherish Sarmiento, and I am a doctoral candidate at Michigan State University. I am asking you to participate in a research study to better understand and improve writing instruction within the context of science education. Your participation is voluntary. This consent form will describe the project, explain the risks and benefits of participation, and empower you to make an informed decision. Please feel free to call me or email me if you have any questions.

Purpose of research

You are being asked to participate in this project because of your valuable perspective on implementation of writing instruction and formative assessment. The purpose of this study is to explore the role of explicit instruction in the language components of writing on student’s science writing outcomes and its subsequent impact on student achievement.

What you will do

Your homerooms will be randomly assigned to a condition, group 1 or group 2. We are asking for your participation in the following activities:

Date for Group 1	Date for Group 2	Activity
Sept.- Oct.	Jan.	Participation in a 1-day workshop about evidence-based writing instruction practices; Researcher and teachers co-develop instructional unit for study.
Oct.		50-minute assessment of your students and additional observation of typical writing instruction, conducted by the MSU research team.
Nov.- Dec.	Feb.	Implementation of instructional routines and co-developed instructional unit; Observation of implementation, conducted by the MSU research team.
Dec.	Mar.	50-minute assessment of your students.
Jan.	Apr.	Review of student performance

Potential benefits of participation

The instructional materials used in this study have had positive impacts on student writing achievement and we intend to enhance the impact on student achievement by integrating with science instruction and ELA.

At the workshop you will receive writing instruction materials. You will receive a total of \$300 for your participation in project activities throughout the year.

#### Potential risks of participation

There are no foreseeable risks associated with participation in this study.

The observation is being conducted to evaluate the feasibility of the instructional routines in real classrooms and is NOT designed to evaluate your performance. The observation will NOT be shared with your supervisor.

If you consent to participate in this study, you agree to implement the new practices in one section in the fall and wait to implement those same practices with the other condition.

#### Privacy and confidentiality

The information collected from the surveys and observations will be kept confidential to the maximum extent allowable by law. The survey will be transmitted through MSU's license with RedCap, which follows industry standard privacy and security guidelines. Your survey form will have a randomly assigned identifying number on it and no identifiable information. The link between your name and the identifier used in this study will be kept in a separate password-protected file on my computer and will not be shared with anyone else.

Professional learning community meetings may be audio recorded solely for the purpose of checking reliability of our summarization.

The electronic files containing your survey responses, the observation data, and audio recordings will be kept in a restricted folder on a secure cloud (MSU's OneDrive), which is HIPAA and FERPA compliant. The restricted folder can only be accessed by me, my research assistants, and the Human Research Protections Program personnel.

At the completion of this study, we will be writing a report about the results. This report will not include any identifiable information about you or your school.

We also plan to make the results of the surveys and observations available to other researchers who may be able to enhance what is known about student writing development. Any information that can link you to this study will be removed prior to any data being made publicly available or shared with other researchers who request the data. There will be no information that can link your participation to the study, and as such no one outside of those directly involved with the research will know that you took part in this study. We cannot guarantee that reidentification is impossible but will take several steps to ensure that your data is as safe and private as is currently possible.

#### Your rights to participate, say no, or withdraw.

Participation is voluntary. Refusal to participate will involve no penalty and will not affect your relationship with your school or Michigan State University. You may discontinue participation at any time without penalty. You have the right to say no. You may change your mind at any time and withdraw. You may also choose not to answer specific questions.

#### Contact Information

If you have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report a complaint, please contact Dr. Adrea.

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program.



**Your signature below means that you voluntarily agree to participate in this research study. Please retain a copy of this document for your records.**

I, \_\_\_\_\_ give my consent to participate in the study.  
(Please print your name)

I agree to be audio recorded  YES  NO

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

## APPENDIX H: PARENT CONSENT FORM

### For participation in a research study

**Study title:** A Pilot Study of Explicit Instruction in Science Argument Writing

**Researcher:**

**Address:**

**Phone:**

**Email:**

Dear Parent or Guardian,

My name is Cherish Sarmiento, and I am a doctoral candidate at Michigan State University. I am asking you to participate in a research study to better understand and improve writing instruction within the context of science education. I am asking for your permission for your child to participate in this research study at school. Participation is voluntary. This consent form will describe the project, explain the risks and benefits of your child's participation, and empower you to make an informed decision. Please feel free to call me or email me if you have any questions. You may also contact my dissertation director, Dr. Truckenmiller by phone or email

#### Purpose of research

You have been selected as a potential participant in this study because the language arts and science teachers at your child's school volunteered to participate and are interested in understanding instructional routines that may improve students' science writing. The purpose of this study is to evaluate writing instructional routines that will support students' learning as demonstrated in a classroom written composition assessment. The goal is to improve elementary students' science learning and written composition.

#### What you and your child will do

First, if you agree to allow your child to participate, we ask that you sign this form. If you choose not to have your child participate in the study, please indicate that on this form. My research team will be administering brief writing assessments (no more than four assessments over the course of the year). During the assessments, your child will be asked to participate in grade-appropriate writing activities with pencil/paper and on the computer. All activities will be very similar to the writing practices your child's teacher uses. Your child's responses will be transmitted via an internet application to secure servers at Michigan State University. I will also provide a copy of your child's written responses, features of your child's response, and instructional recommendations to your child's language arts teacher. If you would like a copy of your child's responses from any of the assessment activities, please contact me and I will provide them to you. Finally, I will request that the school provide me with your child's score on the school-administered reading screener at the beginning of the school, and demographic information about your child, including gender, age, race/ethnicity, special education status, qualification for free and reduced-price lunch, and English Learner status.

#### What will we ask you to do

As part of this research study, we are interested in the role of instructional routines on your child's writing development. Beyond providing your consent for your student to participate in this study, there are no additional tasks required of you.

#### Potential benefits of participation

The potential benefit to your child participating in this study is extra practice with writing. Your child's teacher may use your child's responses to identify strengths and weaknesses and adjust instruction for your child accordingly.

#### Potential risks of participation

The risks of participating in this study are minimal. We will be transmitting your child's name with their written compositions and their scores on the written composition via the web. We are employing industry-standard data security protocols to ensure that only your child's teacher and the researchers see your child's scores. However, we wanted to make you aware that there is always a small risk of a data breach. No other sensitive data will be transmitted other than your child's name. We have also confirmed with your child's school that the scores from the written compositions will not affect your child's grade or any high stakes decision (e.g., retention). If you do not want your child's name transmitted in the online written composition, please contact me.

#### Privacy and confidentiality

Information collected during this study will be kept confidential to the maximum extent allowable by law. That is, the work that your child produces will not be shared with anyone outside the research team, your child's teacher, and personnel from the Human Research Protections (HRPP) program. HRPP personnel may have access to all research records. After the study is concluded, all of your child's work will be assigned a random identifier so that your child's work cannot be linked to your child. The link between your child's name and the identifier used in this study will be kept in a separate password-protected file on my computer and will not be shared with anyone else. IP addresses will not be collected. The rest of the de-identified information that I collect, including your child's education records will be kept in a file that can only be accessed by me, my research assistants, and the Human Research Protections Program personnel. Your child's participation in the study will NOT affect your child's grades or your relationship with your child's teacher.

At the completion of this study, we will be writing a report about the results. This report will not include any identifiable information about your child or your child's school.

We also plan to make the final dataset available to other researchers who may be able to enhance what is known about student writing development. Any information that can link you/your child to this study will be removed prior to any data being made publicly available or shared with other researchers who request the data. There will be no information that can link your participation to the study, and as such no one outside of those directly involved with the research will know that you took part in this study. We cannot guarantee that reidentification is impossible but will take all of the steps we can to ensure that your data is as safe and private as is currently possible.

#### Your rights to participate, say no, or withdraw.

Your child's participation in this study is voluntary. You are free to choose not to have your child's work included in this study. You may also withdraw your child from the study at any time, for whatever reason, without risk to your child's school grades or relationship with the school. In the event that you do not give consent or withdraw consent, your child's work will be kept in a confidential manner.

#### Costs and compensation

Participation in this study does not involve any cost to you or your child. Your child's teacher will receive a gift card to purchase classroom supplies. The gift cards will be distributed during

the final session. We will not administer the study or the gift cards to classrooms with fewer than 10 students providing affirmative consent.

Contact Information

If you have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report a complaint, please contact my dissertation chair.

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University’s Human Research Protection Program.

**Please sign yes or no and return this to the school. After signing this form, please keep a photograph copy of it for your records.**

**Yes**, I give my consent for \_\_\_\_\_ to participate in the study,  
(Name of child)

“A Pilot Study of Explicit Instruction in Science Argument Writing.”

\_\_\_\_\_  
(Parent/ caregiver signature)

\_\_\_\_\_  
(Date)

We also ask for the child’s assent to participate in the study. Please ask your child to sign their name here to indicate that they assent to participate in the study.

\_\_\_\_\_  
(Child signature)

**No**, I do not give my consent for \_\_\_\_\_ to participate in the study,  
(child’s name)

“A Pilot Study of Explicit Instruction in Science Argument Writing”.

\_\_\_\_\_  
(Parent/ caregiver signature)

\_\_\_\_\_  
(Date)

## Demographic Information

We are asking for demographic information about your child so that we can describe the population for which the study was conducted. You may choose not to answer.

1. Child's birthdate (mm/dd/yyyy): \_\_\_\_\_
2. Gender identity of your child: \_\_\_\_\_
3. Does your child have a disability? If yes, please list the disability:  
\_\_\_\_\_
4. Is English your child's native language?  
 Yes  
 No
5. Select all categories that apply to your child's identity:  
 American Indian or Alaskan Native  
 Asian  
 Black or African American  
 Hispanic  
 Indigenous group other than Native American  
 Latino/a/x  
 Middle Eastern origin  
 Native Hawaiian or Other Pacific Islander  
 White  
 Choose not to answer.
6. Does your child receive free or reduced-price lunch?  Yes  No

## APPENDIX I: STUDENT SOCIAL VALIDITY SURVEY

### Science Writing Project Student Survey

**Directions:** Circle how much you agree or disagree with each statement.

1. I enjoy writing about things that I have learned in science class.  
**Never!                  Not really.                  Sometimes.                  Always!**
2. Writing helps me make sense of new information.  
**Never!                  Not really.                  Sometimes.                  Always!**
3. Writing helps me figure out what I learned from a lab experiment.  
**Never!                  Not really.                  Sometimes.                  Always!**
4. Writing helps me figure out what I learned from reading informational books and passages.  
**Never!                  Not really.                  Sometimes.                  Always!**
5. Writing helps me figure out how to use my science knowledge to find solutions to difficult problems.  
**Never!                  Not really.                  Sometimes.                  Always!**
6. Compared to the beginning of the year, I feel like I am better at writing “like a scientist”.  
**No!                  Not really.                  A little bit.                  Yes!**
7. Compared to the beginning of the year, I feel like I am better at writing overall.  
**No!                  Not really.                  A little bit.                  Yes!**
8. I will use what I know about science arguments to help my friends and family learn about science things *in the real world*.  
**Never!                  Not really.                  Sometimes.                  Always!**
9. I will use what I know about writing to help my friends and family learn about new things *in the real world*.  
**Never!                  Not really.                  Sometimes.                  Always!**
10. Whenever I must write something for class, I feel \_\_\_\_\_. **(circle all that apply)**  
**Happy                  Angry                  Excited                  Frustrated                  Determined**  
**Confused                  Knowledgeable                  Lost                  Pride                  Unsure                  Enjoyment**  
**Stressed                  Accomplished                  Bored                  Motivated                  Annoyed**

**These next few questions are open-ended questions where you get to tell Miss Cherish your thoughts about writing. You can write in complete sentences or in bullet points.**

1. When writing about the *Trouble at Hill Crest*, what were your favorite and least favorite parts of the activity?

**Favorite**

**Least Favorite**



2. **Why do you think writing is important for scientists to do their job?**

3. **How is “writing like a scientist” different from other types of writing that you do?**

4. **Do you think writing is an important skill for you. Explain your answer.**

5. **What other thoughts do you have about science writing that you want Miss Cherish to know?**

## APPENDIX J: TEACHER ACCEPTABILITY SURVEY

Page 2

### CER Acceptability Survey

Thank you for your work in the Science Writing Intervention. Your responses to the following questions will be helpful in future iterations of research. In this survey, I will ask you some questions about the strengths of the study, perceptions on how it may have helped your students learning, and your inputs on how it might be improved. You have the ability to save and return to the survey at any time.

**This survey is designed to understand what components about the science writing intervention and co-teaching arrangement helped to contribute to feelings of success about the science writing intervention on your students science learning and writing. This survey will also solicit your thoughts on the aspects of the study that helped or hindered feelings of success. Finally, this survey will ask you for recommendations for improving future iterations of this study with respect to how researchers work with co-teaching teachers.**

Morphology Routine: The following questions will ask you about your use of the Morphology Routine, including any alternative/similar versions that you may have used to support your student's organization of science information.

[Attachment: "Morphology Routine (1).docx"]

- 1) About how many times did you use the Morphology Routine (or something similar) over the 7-week instructional period from Jan 6 to March 1? \_\_\_\_\_
- 2) If you modified this scaffold so that it better suits your/your student's needs, how did you modify it? \_\_\_\_\_
- 3) What are aspects of this organizer that you like? \_\_\_\_\_
- 4) What are aspects of this organizer that you do not like? \_\_\_\_\_
- 5) Describe an instance in which you would use this graphic organizer if you got the chance to. What might that lesson look like? \_\_\_\_\_

Graphic Organizer: The following questions will ask you about your use of the CER graphic organizer, including any alternative/similar versions that you may have used to support your student's organization of science information.

[Attachment: "CER Graphic Organizer.docx"]

- 6) About how many times did you use the CER graphic organizer (or something similar) over the 7-week instructional period from Jan 6 to March 1? \_\_\_\_\_
- 7) If you modified this scaffold so that it better suits your/your student's needs, how did you modify it? \_\_\_\_\_
- 8) What are aspects of this organizer that you like? \_\_\_\_\_
- 9) What are aspects of this organizer that you do not like? \_\_\_\_\_



- 
- 10) Describe an instance in which you would use this graphic organizer if you got the chance to. What might that lesson look like? \_\_\_\_\_

---

CER Sentence Starters: The following questions will ask you about your use of the CER sentence starters, including any alternative/similar versions that you may have used to support your student's organization of science information.

[Attachment: "cer sentence starters.pdf"]

- 
- 11) About how many times did you use the CER sentence starters (or something similar) over the 7-week instructional period from Jan 6 to March 1? \_\_\_\_\_

- 
- 12) If you modified this scaffold so that it better suits your/your student's needs, how did you modify it? \_\_\_\_\_

- 
- 13) What are aspects of this organizer that you like? \_\_\_\_\_

- 
- 14) What are aspects of this organizer that you do not like? \_\_\_\_\_

- 
- 15) Describe an instance in which you would use this graphic organizer if you got the chance to. What might that lesson look like? \_\_\_\_\_

---

### Instruction

**There were three active ingredients in this study: (1) science specific writing supports, (2) additional time spent writing, and (3) teacher collaboration via co-teaching. This section will ask you about each of these ingredients and your perceptions of their usefulness in your class with respect to students science learning and writing.**

- 
- 16) Science Writing Supports (for Knowles HR specifically)- Having science specific writing supports available helped me to organize my thoughts about teaching writing during Explore and Do.  Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 17) Science Writing Supports (for Knowles HR specifically)- Having science specific writing supports available helped my students to organize their thoughts about science writing during Explore and Do.  Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 18) Science Writing Supports (for Knowles HR specifically)- The science writing supports helped to structure class conversations around science topics during Explore and Do.  Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 19) Science Writing Supports (for Knowles HR specifically)- The science writing supports helped to structure class conversations around about non-science topics.  Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.

- |   |  |
|---|--|
| 20) Science Writing Supports (for Knowles HR specifically)- I think my students have made improvements with respect to their science writing over the period of the intervention (Jan 6-March1).  | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 21) Science Writing Supports (for Knowles HR specifically)- I think my students made improvements with respect to their overall writing during the period of the intervention (Jan 6-March1).   | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 22) Science Writing Supports- Having science specific writing supports available to me will make it more likely that I incorporate writing during Explore and Do.   | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 23) Science Writing Supports- I will use one or more of these science writing supports in the future.   | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 24) Science Writing Supports- I will recommend these writing supports to fellow teachers who express interest in incorporating more writing in their science instruction.   | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 25) Additional Writing Time- Modifying Explore and Do to include more writing time (e.g., One lab period+One writing period) was beneficial for most students.  | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 26) Additional Writing Time (for both homerooms)- Modifying Explore and Do to include more writing time (e.g., One lab period+One writing period) improved students science understanding.  | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 27) Additional Writing Time (for both homerooms)- Modifying Explore and Do to include more writing time (e.g., One lab period+One writing period) improved students writing.  | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 28) Additional Writing Time (for both homerooms)- I found that dedicating some Explore and Do periods to discuss and respond (via writing) to lab based questions helped students conceptual understanding of the material that was covered in the lab experiment.                            | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 29) Additional Writing Time (for both homerooms)- In the future, I will dedicate some Explore and Do periods to discuss and respond (via writing) to lab based questions.   | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |
| 30) Additional Writing Time (for both homerooms)- I found that dedicating some Explore and Do periods to discuss and respond (via writing) to problem-based critical questions (e.g., The Trouble at Hill Crest) at the end of an inquiry unit was useful in assessing what students learned. | <input type="radio"/> Strongly Disagree.<br><input type="radio"/> Disagree.<br><input type="radio"/> Agree.<br><input type="radio"/> Strongly Agree. |

- 
- 31) Additional Writing Time (for both homerooms)- In the future, I will dedicate some Explore and Do periods to discuss and respond(via writing) to problem-based critical questions (e.g., The Trouble at Hill Crest) at the end of an inquiry unit.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 32) Additional Writing Time (for both homerooms)- In the future, I will dedicate some Explore and Do periods to discuss and respond(via writing) to problem-based critical questions (e.g., The Trouble at Hill Crest) at the end of an inquiry unit.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 33) Co-teaching- When it comes to teaching my students how to "write like a scientist", I bounce ideas off of my teacher partner.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 34) Co-teaching- When it comes to teaching my students how to write generally, I bounce ideas off of my teacher partner.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 35) Co-teaching- My teaching partner and I check in with each other regularly to assess our students learning.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 36) Co-teaching- My teaching partner and I check in with each other regularly to assess specific students learning.
- Strongly Disagree.  
 Disagree.  
 Agree.  
 Strongly Agree.
- 
- 37) Co-teaching- My teacher partner and I spend \_\_\_\_\_ minutes/week to discuss our instructional plans.
- 
- 38) Co-teaching- My teacher partner and I tend to discuss these things at our co-teaching meetings: \_\_\_\_\_
- 
- 39) Co-teaching- Without these specific resources in place at my school, it would be impossible to have the same co-teaching dynamic that I have with my teaching partner. \_\_\_\_\_
- 
- 40) Co-teaching- The most valued professional skills and/or experiences that my teacher partner brings to our co-teaching partnership are: \_\_\_\_\_
- 
- 41) Co-teaching- These resources/skills/learning opportunities would enhance our co-teaching partnership and improve student outcomes: \_\_\_\_\_

### Future Research

**This study was considered a pilot study. Pilot studies help members of the research team identify elements of research that worked well and not so well for the purposes of improving future studies. As this research occurs in schools with real teachers and real students, I am interested components of the study that impacted teaching and learning, specifically with respect to writing and science knowledge outcomes.**

- 42) Logistical difficulties- What were some of the difficulties that participating in this project placed on your planning? How did you respond to those difficulties? \_\_\_\_\_
- 
- 43) Pedagogical difficulties- What were some of the difficulties that participating in this project placed on you with respect to how to teach? How did you respond to those difficulties? \_\_\_\_\_
- 
- 44) Curricular difficulties- What were some of the difficulties that participating in this project placed on what/how deep you taught? How did you respond to those difficulties? \_\_\_\_\_
- 
- 45) What were your favorite parts about participating in this project? \_\_\_\_\_
- 
- 46) What were your least favorite parts about participating in this project? How would you change those parts? \_\_\_\_\_
- 
- 47) Knowing what you know now, if you were given an opportunity to participate in a year-long version of this study, how would you go about using these instructional scaffolds to teach students how to "write like a scientist"? \_\_\_\_\_
- 
- 48) After participating in this study, what elements of the study are you more likely to include as part of your "instructional toolbox"? \_\_\_\_\_
- 
- 49) What concluding remarks/thoughts/recommendations do you have for improving science writing research so that it makes the most impact on students/teachers? \_\_\_\_\_

## APPENDIX K: PROFESSIONAL LEARNING AGENDA

### Science Writing Study Planning Meeting

#### Meeting Purpose

Morning Session: To understand the research purpose and corresponding study design. Introduce the scaffolds and discuss how they could be implemented in the classroom.

Afternoon Session: Work through the logistics of implementation. Plan the unit of instruction.

**Agenda-**December 2<sup>nd</sup>, 2023

*Coffee and snacks provided. Teachers will need their laptops and curricula materials.*

**8:00-9:00 am:** Study Background

Rationale behind the study- Writing in the Content Areas and Explicit Instruction

**9:00-10:00 am:** CER Graphic Organizer

Introduction to each part of the organizer

Discussion of how it might be used

Practice

**10:00-10:10 am:** Break

**10:15-11:00 am:** Morphology Routine

Introduction to each part of the organizer

Discussion of how it might be used

Practice

**11:00-12:00 pm:** CER Sentence Starters

Introduction to each part

Discussion of how it might be used

Discussion of additional sentence frames

Practice

**12:00-1:00 pm:** Lunch

**1:00-2:30 pm:** Logistics of Implementation

**2:30-2:45 pm:** Break

**2:45-4:30 pm:** Curricular planning

**4:30-5:00 pm:** Questions

Pass out scaffolds.