6-19-2018

The Impact of the Use of Science Notebooks in Conjunction with a Learning Progression-based Science Unit in an Urban Middle School

Reagan Williams
rmwfm5@mail.umsl.edu

Recommended Citation

Williams, Reagan, "The Impact of the Use of Science Notebooks in Conjunction with a Learning Progression-based Science Unit in an Urban Middle School" (2018). *Dissertations*. 783.
https://irl.umsl.edu/dissertation/783

The Impact of the Use of Science Notebooks in Conjunction with a Learning Progression-based Science Unit in an Urban Middle School

Reagan M. Williams
M.Ed., University of Arkansas-Little Rock, 2004
B.S., University of Arkansas-Pine Bluff, 2003

A Dissertation Submitted to The Graduate School at the
University of Missouri—St. Louis
in partial fulfillment of the requirements for the degree Doctor of Philosophy in
Education with an emphasis in Teaching and Learning processes

August 2018
.

Advisory Committee

James A. Shymansky, Ph.D.
Chairperson

Natalie A. Bolton, Ph.D.

Charles R. Granger, Ph.D.

Gayle Wilkinson, Ph.D.

Abstract

Learning progressions are the latest tool to understand the ways science learning occurs and they underlie the structure and framework of the *Next Generation Science Standards*. Prior research indicated a variety of ways to develop and validate learning progressions and learning progression's general positive impact on students' science learning. However, no study has explicitly employed science notebooks as the cornerstone to the development and/or validation processes. Therefore, the research question is: what is the impact on students' science learning outcomes when a middle school science learning progression is developed and validated using science notebooks as part of an inquiry-based instructional intervention? A rock cycle learning progression based on the systems thinking hierarchy model was developed. Using a causal-comparative case study, the study validated the rock cycle learning progression by implementing a three-week instructional intervention with 22 rising 8th grade students in an urban charter school. Data were Rock Cycle Assessment pretest and posttest scores, symbolic media, and reflective conclusions. Three important results emerged: a) a statistically non-significant relationship existed between posttest scores of the On-campus and Learning Progression groups, but there was a statistically significant relationship between posttest scores of the Off-campus and Learning Progression groups; b) intervention participants were partially able or unable to describe their science learning; and c) there was moderate to strong association between each symbolic media categorical descriptor and the inquiry phase in which it was produced. The results suggest that the phase-placement of symbolic media in science notebooks influences science learning outcomes.

Table of Contents

LIST OF ABBREVIATIONS

1. AAAS—American Association for the Advancement of Science
2. AYP— adequate yearly progress
3. DCI—Disciplinary Core Idea
4. ESS—Earth and Space Sciences
5. NAEP— National Assessment of Educational Progress
6. NCLB—No Child Left Behind Act
7. NGSS— Next Generation Science Standards
8. NQT—NAEP Question Tool
9. NRC—National Research Council
10. NSF—National Science Foundation

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

A “learning progression” is a tool that focuses on understanding the ways science learning occurs (Duschl, Schweingruber, & Shouse, 2007). A learning progression is a systematic and well-organized description of’ thinking and/or understanding of a given science concept by students (Alonzo & Steedle, 2009). Many current researchers advocated learning progressions as a vehicle to transform science education (Duschl et al., 2007). As such, learning progressions were the premise of the Next Generation of Science Standards (NGSS). Although a learning progression is a model of cognition, it is not a single, linear, pathway. It is a probable idea—a conjectural model for learning core science ideas and practices (Alonzo & Steedle, 2009). Usually, a learning progression is arranged in hierarchal levels. Each level represented milestones along a trajectory from initial conceptual understanding to a scientific level of understanding (Plummer & Maynard, 2014).

Learning progressions have common features. Foremost, learning progressions were informed by research on student thinking and learning in a content domain and organized around the “big ideas” of that content domain (Duncan & Hmelo-Silver 2009; Duncan, Rogat, and Yarden, 2009; Duschl et al., 2007; Smith, Wiser, Anderson, & Krajcik, 2006). Secondly, learning progression development and validation occurred iteratively and/or concurrently through cycles of empirical testing and theoretical revising (Duncan & Hmelo-Silver, 2009). Lastly, all learning progressions inherently had three critical features: a) grade band, b) scope, and c) grain size.

Learning progressions have been developed and validated in a variety of ways. Researchers have employed case studies, cross-sectional studies, construct maps, instructional interventions, and a host of other techniques to develop, validate, and/or refine learning progressions; however, using learning progression with science notebooks is one method that has been exceptionally rare. According to Klentschy (2005), a science notebook is a living, working document. It is a central place where linguistics, data, and practice coalesce to construct meaning for the student (Klentschy, 2005). Science notebooks and their pedagogical function were highly researched. They were also advocated in many school districts and by many educational organizations, researchers, and practitioners. Specifically, research demonstrated the value of employing science notebooks in inquiry-based instruction (e.g. Aschbacher & Alonzo, 2006; Butler & Nesbit, 2008; Clidas, 2010).

1.0 Background of the Problem

Smith et al. (2006) reported the earliest learning progression research. They also defined the cardinal principles for much of the current learning progression research. Since their seminal work, several studies were published that demonstrated the contributions to the improvement of student outcomes by learning progression strategies. Overwhelmingly, learning progression research results indicated student improvements in various capacities. Songer, Kelcey, and Gotwals (2009) described a method to develop a learning progression on complex thinking about biodiversity. Their (2009) Hierarchical Linear Modeling (HLM) results showed noteworthy student success. Songer and Gotwals (2012) examined learning progressions with a similar sample as Songer et al. (2009). Their (2012) study results also indicated student improvement.

While learning progression research confirmed positive student outcomes, only Songer, Kelcey, & Gotwals (2009) employed science notebooks in the development and/or validation of a learning progression. Consequently, there is a gap in what is known about the role science notebooks play in inquiry-based science as it relates to the development and/or validation of a learning progression.

2.0 Purpose

The purpose was two-fold: a) to develop a middle school science learning progression validated in the context of inquiry by using science notebooks, and b) to study the impact of the notebook-based learning progression on middle school students' learning. The following research question was explored:

What is the impact on science learning outcomes when a middle school science learning progression is developed and validated using science notebooks as part of an inquiry-based instructional intervention?

Science notebooks served as the focal point for the instructional intervention. All student participants were administered a pretest of the targeted science content. The teacher-researcher utilized the learning progression and its associated materials in the Learning Progression group and did not utilize the learning progression, the instructional intervention, or science notebooks with the Computer-assisted group. Following the completion of the three-week intervention, all student participants were administered a posttest.

3.0 Significance

The study was important for several reasons. Foremost, it tested a different way to develop and validate a learning progression in a science content area while simultaneously addressing a research gap by using science notebooks. Many studies demonstrated the contribution of learning progressions to the improvement of student outcomes. For example, Songer et al. (2009) described a five-step process to develop a learning progression, and the results demonstrated significant student achievement. Schwarz et al. (2009) also presented a learning progression for scientific modeling, and results indicated the sample engaged in constructing and revising increasingly accurate models. Many studies have also shown science notebooks to be beneficial to student science achievement. For example, Huerta, Irby, Lara-Alecio, and Tong (2015) examined the relationship between language and concept science notebook scores of English language learners and/or economically disadvantaged students. The authors (2015) found positive, large, and significant correlations between students' language and concept scores; science notebook entries that had more academic language had the largest correlations. Klentschy and Molina (2004) illuminated the Valle Imperial Project in Science (VIPS) in their research of students' science notebooks and the inquiry process. Specifically, the VIPS project connected science and literacy through the use of science notebooks. It was found that there was a pattern of significant growth in student achievement in science achievement as well as reading and writing achievement for all students participating in the program.

Not only was the work of Schwarz et al. (2009) important to middle school students' science learning outcomes, but also it was also important to education stakeholders. Furtak (2009) noted learning progressions had the wherewithal to be used in teacher preparation and professional development. Learning progressions contain information about students' thinking and learning and therefore, were potentially a framework for developing coherent curricula and assessment in science (Shin, Stevens, Short, & Krajcik, 2009; e.g. Alonzo & Steedle, 2009; Berland & McNeill, 2010). Also, they could assist in preparing level-appropriate instruction (Shin et al., 2009). In some cases, they had suggestions for strategies and actions to help students learn (Furtak, 2009; e.g. Jin & Anderson, 2012; Lee & Liu, 2010). Additionally, learning progressions could also help teachers identify and judge collected artifacts as evidence of student thinking and learning. In turn, the artifacts could then be used to modify instruction and in some cases, revise the learning progression (Furtak, 2012). Consequently, as a teacher/professional development tool, learning progressions potentially increased teacher knowledge (Wilson, 2009). For policy makers, the associated assessments of learning progressions potentially provided (more accurate) diagnostic information about the level and nature of students' understanding (Steedle & Shavelson 2009; e.g. Neumann, Viering, Boone, & Fischer, 2013). For the researcher, the initial learning progression developed was important because it had the potential to bridge the gap between research and practice—between research on how students learn in a given content domain and the methods for teaching and assessing in science.

4.0 Delimitations

The problem was delimited to the role science notebooks play in inquiry-based science as it related to the development and validation of a learning progression. The problem was selected because, despite the established role of science notebooks in science education research, learning progression researchers rarely addressed them in research literature. The grade range was delimited to rising 8th grade students because of the summer enrichment program, ease of access to sample participants, and the availability of national science databases/resources. Grades five and 12 are other grade levels for potential selections because of their clear delineation in the *Next Generation Science Standards* (*NGSS*). However, they were not selected because of lack of access to a sample. Scope was delimited to earth and space science (ESS) disciplinary core ideas (DCI) of the NGSS. Specifically, the 5th and 8th grade DCI's bound the rock cycle learning progression.

Grain size was delimited to eight achievement levels. This size was selected based on the systems thinking hierarchal (STH) model in Earth and Space Science (ESS) and the grain size trend research literature. A grain size smaller than three achievement levels was not in research literature. Learning progression development was delimited to one of three of Duncan and Hmelo-Silver's (2009) recommended approaches: developing an "initial" learning progression based on existing research in student thinking and learning, and content domain analysis. The development approach was selected because of its feasibility and clarity for implementation.

5.0 Definition of Terms

- Construct map— a continuum that defined student understanding in addition to common errors at each performance level (i.e. achievement level) within the continuum. Wilson (2009) suggested a concept map be generated concurrently with the content domain analysis in order to guide the development of the intermediate levels of a learning progression.
- Mainstream students— students with social prestige, institutionalized privilege, and normative power; in the U.S., these students tended to be White, upper/middle class, and native English speakers (Duschl et al., 2007; Lee & Lukyx, 2007).
- Non-mainstream students— students who did not have access to the same prestige, privilege, and power as mainstream students; consequently they experienced social incongruency and were at an academic disadvantage (Lee & Lukyx, 2007).
- Symbolic media: according to Lehrer and Schauble (2012), drawings, diagrams, photos, and other similar models.
- Test blueprint—A guide that aided in test construction, it ensured the constructed test will sample important content areas and levels of cognitive complexity.

> According to Suskie (2009), a test blueprint is "an outline of the test that lists the learning goals that students are to demonstrate " (p.167). Test blueprinting linked the test to learning goals.

6.0 Summary

Education reform is traditionally a highly contentious topic in the United States (U.S.) and specifically, science education reform is no exception. The NGSS, released in April 2014, represented a major shift in science education reform, and learning progressions were foundational to that reform effort (Achieve Inc., 2013). In a learning progression, each level represented a significant milestone along the learning trajectories from initial conceptual understanding to a scientific level of understanding (Plummer & Maynard, 2014). They are systematic and well-organized descriptions of student thinking and understanding of a given science concept that are hierarchally arranged around the "big ideas (Alonzo & Steedle, 2009). Several studies demonstrated contributions of learning progressions to the improvement of student outcomes. While the learning progression research field is relatively new, and there was general consensus about many common features, there was also much ambiguity among researchers, across many dimensions of the research field, and consequently, gaps existed in the research literature. One gap was the role science notebooks played in inquiry-based science as it related to the development and validation of a learning progression.

The study was important for several reasons. Foremost, it examined a new way (using science notebooks) to develop and validate a learning progression that has yet to be established. Secondly, it bridged the gap between research and practice. Third, the results could act as teacher preparation and professional development tool, potentially be a framework for developing coherent science curricula and assessments, and assist in preparing level-appropriate instruction. They could also have suggestions for strategies/actions to help students learn, and, inform revisions for future iterations of the learning progression (Furtak, 2009; Shin et al., 2009).

Chapter 2: Review of Related Literature

1.0 Learning Progressions: A Promising Theme in Science Education Reform

There was never an intentional focus on children's thinking and learning in the historical context of U.S. science education (Kahle, 2007). Learning progressions aimed to remedy the neglect (Duschl et al., 2007). Consequently, learning progressions appeared to be the next theme as they systematically describe thinking and understanding by students of a given science topic, were informed by research on thinking and learning, and were foundational to the NGSS (Alonzo & Steedle, 2009; Duncan & Hmelo-Silver, 2009). Many researchers claimed learning progressions as potentially transformative for science education because of their capacity to better align curriculum, instruction, and assessment (e.g. Duncan & Hmelo-Silver, 2009; Mohan, Chen, & Anderson, 2009).

1.1 Historical Background of Learning Progressions

Smith et al. (2006) coined the phrase "learning progression" and designated it as a cognitive model that described the way students continuously and gradually refined ways of reasoning. Learning progression, as defined by Smith et al. (2006), was based on research synthesis and conceptual analysis. Their (2006) work initially stemmed from assessment systems development designed to track student progress (Kennedy, Brown, Drancy, & Wilson, 2005; Wilson, 2005). Part of the assessment system was construct maps. Generally, construct maps were considered the forerunner of learning progression research.

Smith et al. (2006) defined the cardinal principles for much of the current learning progression research. They (2006) recommended learning progressions be organized around big ideas—the central concepts and principles of a scientific discipline. Their

(2006) learning progression symbolized coalescence between not only theory and practice, but also between science learning researchers and measurement specialists. With their groundbreaking research, Smith et al. (2006) laid the foundation for NRC's 2007 policy, *Taking Science to School*, which in turn served as a guide for *NGSS*. Since Smith et al. publication, there was an ever-increasing amount of learning progressions research.

1.2 Description Of Learning Progressions

Several authors used synonymous terms for learning progressions such as learning pathways, conceptual pathways, and conceptual progressions. All of these phrases were used to describe a means of tracking student learning across time (Adadan et al., 2010; Liu & Lesniak, 2006). Several authors cited Duschl, et al. (2007) definition for a learning progression. (Duncan & Hmelo-Silver, 2009; Gunckel, Covitt, Salinas, & Anderson, 2012; Mohan et al., 2009; Plummer & Krajcik, 2010; Stevens et al., 2010). Among learning progression researchers, the general consensus was that the development of learning progressions must be informed by research on student thinking and learning in the content domain (Duncan & Hmelo-Silver, 2009; e.g. Plummer and Krajcik, 2010; Smith et al, 2006). Furthermore, there were a variety of ways to validate a learning progression in research literature. Learning progressions inherently aimed to develop and to deepen knowledge over time because they emphasized providing greater alignment among curriculum, instruction, and assessment as difficulty increased as grade levels increased.

Learning progressions were hierarchally organized in levels around big ideas (Duncan & Hmelo-Silver, 2009). At its upper end, a learning progression was anchored by what students should know and/or be able to do relative to societal expectations

(Duncan & Hmelo-Silver, 2009). At its lower end, it contained what students knew about the science ideas and practices upon entering school (Duschl et al., 2007). In between the upper and lower ends were the levels of achievement (i.e. performance levels). These levels articulated the understandings, alternative conceptions, and/or misconceptions characteristic to bridging the gap between its upper and lower ends (Duncan & Hmelo-Silver, 2009; Wilson, 2009). Several authors referred to the intermediate region as the "messy middle" (Furtak, 2012; Gotwals & Songer, 2013; Songer & Gotwals, 2012). Duncan and Hmelo-Silver (2009) provided a more formal description of learning progression levels by identifying four fundamental theoretical/structural components that unified all learning progressions:

- They focused on a few content ideas and inquiry practices.
- Upper and lower anchors bound learning progressions.
- Levels of achievement described the intermediate steps— a hypothesized order of the levels through which knowledge and skills progressed en route to the upper anchor.
- Targeted instruction and curriculum mediated learning progressions. Scaffolded curriculum and instruction reconciled the learning associated with the progression (p. 607).

2.0 Trends in Learning Progression Research

Despite the comparatively small amount of research literature, some salient trends and intriguing findings precipitated. Foremost, there were two extraordinarily broad classifications for learning progressions. One classification focused on curriculum and instruction. The other classification emphasized cognition and instruction. Across and within both classifications, there were current and characteristic practices researchers employed in developing and/or validating respective learning progressions. One example was the use of construct maps. Another trend was the virtual absence of poor and urban sample participants in learning progression research. Major findings included the contribution learning progressions to student outcomes, an emergent subdivision, the establishment of LeaPS conference, and U.S. stakeholder use of learning progression (e.g. Alonzo & Steedle, 2009; Lehrer & Schauble, 2012; Mohan, Chen, & Anderson, 2009; NSF, 2008; Schwarz et al., 2009).

2.1 Two Broad Classifications

Learning progression research concentrated on a variety of core ideas and practices (e.g. matter and the atomic-molecular theory, water in socio-ecological systems). Nevertheless, there were two broad classifications of learning progressions. One category focused on cognition and instruction; the other focused on curriculum and instruction (NSF, 2008).

2.11 Cognition and Instruction Learning Progressions

Cognition and instruction research typically began with a psychological analysis of the cognition, which was at the core of the content. For this category of research, the goal was fostering growth of the cognition as students moved from novice to expert in learning about a specific concept (Shavelson & Kurpius, 2012). Mark Wilson was an exemplar in this category for his work in developing the notion of construct maps. According to Wilson (2009), a construct map was less complicated than a learning progression. Wilson (2009) described a construct and its development in great detail. Alonzo and Steedle (2009) contended construct maps (which they conceptualize as smaller learning progressions) potentially provided the detail teachers needed so student thinking can be tracked over the course of instructional units.

Alonzo and Steedle's (2009) research demonstrated a cognition and instruction learning progression. The authors (2009) described the iterative process of developing a force and motion learning progression and its associated assessment items. They (2009) identified two areas of cognitive science research that were important to their learning progression development: a) the consistency of student responses, and b) language use. A learning progression was developed around these two areas. In efforts to foster cognition growth, the authors (2009) compared the use of ordered multiple-choice (OMC) to open-ended (OE) items in measuring comprehension of force and motion.

2.12 Curriculum and Instruction Learning Progressions

Typically, curriculum and instruction learning progressions began with a logical content analysis and were characterized by the development of an instructional unit (Shavelson & Kurpius, 2012). Plummer and Maynard (2014) exemplified the curriculum and instruction classification category. They explored how eighth grade students learned the seasons before and after an accompanying instructional intervention. The investigation began with developing a construct map by using the construct modeling methodology. Once developed, student-participants were given a pretest, a 10-day inquiry-based curriculum, and then a posttest. All students received identical instructional activities across each of the ten 50-minute class periods. Instead of developing instructional units, the authors utilized a curriculum based on lessons from *The Real Reasons for the Seasons* (Plummer & Maynard, 2014). Following the posttest, the authors revised the seasons construct map using a Rasch model analysis of pretest and posttests (Plummer & Maynard, 2014).

Leher and Schauble (2012) is another example of the curriculum and instruction focus. They described changes in representational and modeling practices for kindergarten through sixth grade students across three intertwined strands: a) change, b) variation, and c) ecosystems. Leher and Schauble (2012) identified potential milestones of conceptual progress in each of the three concept strands via one construct map. Once the construct map was complete, they illustrated examples of students' artifacts of models from classroom, developmental, and science learning empirical research. Instead of developing instructional units, they described instructional designs, which shaped classroom teaching and learning.

2.2 Construct Map in Developing Learning Progression

Several researchers also used construct maps in a variety of ways to develop their respective learning progressions. Both Plummer and Maynard (2014) and Lehrer and Schauble (2012) began their learning progression development by drafting construct maps. However, the authors used the construct maps differently. Plummer and Maynard (2014) used their construct map as a generated metric, transformed it into a learning progression, and iteratively refined the "initial" learning progression. Leher and Schauble (2012) used a construct map to represent the states of knowledge in representation and modeling across change, variation, and ecosystems for kindergarten through 6th grade students. Plummer and Maynard (2014) explicitly detailed construct map development, whereas Leher and Schauble (2012) implied the development of the construct map. Nevertheless, both construct maps helped delineate the content and/or skill, served as a precursor to learning progressions, and potentially guided instruction within the curriculum.

2.3 Learning Progression Development and Validation

The development and validation approach was the most crucial feature of learning progression research. Yet, there was no unified vision to accomplish this. The only general agreement was that the processes entwined (to varying degrees), took place through recurring cycles of empirical testing and theoretical revising, and was based on research of children's thinking and learning (Duncan & Hmelo-Silver, 2009, e.g. Mohan et al., 2009). Outside of those parameters, there were a variety of methods to develop and/or validate learning progressions. In rare cases, development and validation were not necessarily mutually inclusive (e.g. Leher & Schauble, 2012; Smith et al., 2006).

There was no universal approach to developing and validating learning progressions. For example, Plummer and Maynard (2014) developed a construct map followed by administering a pretest, an instructional intervention, and a posttest. Students' pretest and posttest scores were analyzed with a Rasch model. The results were used to revise the construct map into a learning progression, which simultaneously validated the learning progression. Neuman et al. (2013) gleaned their initial learning progression from existing curriculum research on understanding and development of understanding in the domain of energy. These sources guided the development of the Energy Concept Assessment (ECA). To validate the learning progression, the ECA was administered to approximately 1800 6th through 10th grade students in German public and private schools. Revisions were made to the initial learning progression. Furtak (2012) modified another author's learning progression in order to investigate teacher engagement in the iterative development, enactment, and revision of formative assessments.

Duncan and Hmelo-Silver (2009) articulated three general development-validation approaches for learning progressions. Many researchers subscribed to one of the three methodologies, however, some did not adhere to Duncan and Hmelo-Silver (2009) recommendations. One approach was developing an initial learning progression from existing research on student learning and thinking in the content domain (Duncan & Hmelo-Silver, 2009). These learning progressions required validation studies, which involved the development and implementation of instructional interventions (Duncan & Hmelo-Silver, 2009). Plummer and Maynard (2014) demonstrated this approach. They presented the development of a learning progression for celestial motion and then explored how student learning of the seasons was supported by classroom instruction. The authors (2014) began their development process with an analysis of astronomy education and students' thinking and learning research in astronomy. Based on the analysis, they built a hypothetical construct map, pretested participants, and implemented an instructional intervention. The instruction supported students in building on, and changing, conceptions about incidents in the solar system. The lessons were based on Gould, Willard, and Pompea's (2000) *The Real Reasons for Seasons* and Coyle's (1993) *Project Star*. They also utilized teacher-created materials. The intervention called for students to examine important concepts for additional exploration of the seasons. Student-participants then used this information as they wrote reflections on how their understanding of Earth's orbit changed during the lessons. Based posttest performance, Plummer and Maynard (2014) revised the hypothetical seasons construct map as a means to validate the learning progression.

Duncan and Hmelo-Silver's (2009) second recommendation was for learning progressions to be built on carefully designed cross-sectional studies. The study chronicled knowledge development and reasoning of students on a specific topic across many grades (Duncan & Hmelo-Silver, 2009). This approach yielded an appraisal of the students' current learning trajectories (Duncan & Hmelo-Silver, 2009). Typically, the second approach did not involve instructional interventions. Mohan et al. (2009) utilized the second approach in developing a multi-year learning progression for carbon cycling in socio-ecological systems. Participants were 4th grade and 6th through 12th grade students in Michigan, Korea (on a U.S. military base), and California. The participant-teachers developed their own instruction between the pretest and posttest. The researchers offered instructional activities focusing on the principles of matter, energy, and scale during the carbon cycling processes. They also developed the initial learning progression and associated assessments, administered the assessments, and then used the assessment results to revise the initial learning progression. The revised learning progression led to new assessments for students each year of the study. Each iteration spanned one year (Mohan et al., 2009). Data were written assessments by students and clinical interviews, which informed the learning progression revisions (Mohan et al., 2009). The written assessments questions were iteratively developed during the three-year period, varied in length contingent on age level, and focused on what happened to matter during carbon transforming processes (Mohan et al., 2009). The clinical interviews used a set of cards, each showing a color picture and written description of a macroscopic event to stimulate students to develop ideas (Mohan et al., 2009). Students explained the underlying matter transformation and classification of the macroscopic events. Responses by students determined the interviewers probing questions (Mohan et al. 2009). The 30-minute student interviews were either video or audio recorded (Mohan et al. 2009).

Duncan and Hmelo-Silver's (2009) third recommendation involved developing a learning progression based on sequencing teaching experiments across multiple grades. Songer et al. (2009) employed this approach in describing an iterative, empirically driven process to develop a three-year learning progression for students in 4th to 6th grades that centered on complex thinking about biodiversity. Duncan and Hmelo-Silver (2009) presented and discussed a four-step process for its development and validation (p. 607):

1. Development of a preliminary content and a preliminary inquiry reasoning learning progression;
2. Development of eight weeks of curricular activities and the associated assessment items representative of both learning progressions;
3. Evaluation of learning that occurs with the curricular units using the initial assessment instruments; and
4. The revision and expansion of the initial learning progressions into a three-year content and three-year inquiry reasoning learning progression.

Step two detailed the careful sequencing of teaching experiments that were hallmarks of Duncan and Hmelo-Silver's (2009) third approach to development and validation. Songer et al. (2009) developed a preliminary content and a preliminary inquiry skill learning progression. They (2009) then translated the key points from both learning progressions into curricular activities and implemented the activities with students. The key points were then empirically tested. Drawing from cognitive scaffolding research, Songer et al. (2009) first worked with teachers to develop a scaffold format. The form served as a guide for developing evidence-based explanations. The authors (2009) and teachers then defined the essential components of an evidence-based explanation: "a scientific claim, two pieces of evidence (associated with a key scientific concept), reasoning that ties the claim to the evidence, and guidance in composing all of these pieces into one consistent whole (p. 613)." Secondly, the authors (2009) and teachers implemented the curricular activity that provided specific locations (such as boxes or lines) for the components of evidence-based explanations to be written (Songer et al., 2009). The curricular activity used the explanation-building format that was generated earlier in conjunction with symbolic media (e.g. drawings, diagrams) to help students develop their evidence-based explanations across each grade (Songer et al., 2009).

Despite Duncan and Hmelo-Silver's (2009) development and validation recommendations, there were some researchers who did not subscribe to the recommendations. (e.g. Neuman et al., 2013). Among those authors who did not adhere to the development-validation parameters, there were differences in how these researchers conceptualized and utilized their developed learning progressions.

2.4 Utility Of Learning Progressions

Another trend was the utility of a learning progression. There was an implied consensus as to how learning progressions could be definitively used—either as a diagnostic tool or as a tool to foster learning (NSF, 2008). Overwhelmingly, learning progressions were diagnostic. Diagnostic learning progressions identified precisely where, within the learning progression, a student's thinking was. Both Gunckel et al. (2009) and Jin and Anderson (2012) utilized a socio-ecological framework for their respective diagnostic learning progressions. Very few learning progressions were progressive, fostering conceptual change of students toward a scientific level of understanding.

2.41 Diagnosis of Student Thinking

Gunckel et al. (2009) demonstrated a diagnostic learning progression. They investigated explanation by students of water and substances in water moving through socio-ecological systems. Using a cross-section methodology and sampling from students in 5th through 12th grades, they employed an iterative design whereby each cycle moves through three phases: a) development of a model of cognition (i.e. learning progression); b) assessment; and c) interpretation. After each design cycle, the authors (2009) revised the learning progression based on results from the previous design cycle and in total conducted three cycles of assessments. After each cycle of assessment, items were then refined based on the results from the previous design cycle. During the interpretation phase, student explanations were analyzed and the results were used to inform revisions to the learning progression (Gunckel et al., 2009). The revisions enabled the authors to better articulate the intermediate levels and lower anchor. Gunckel et al. (2009) published the findings from the third cycle of assessment. The product was a four-level learning progression. The bulk of high school student-participants provided explanations between levels two and three. Very few students provided explanations at level four.

In the design and implementation of a diagnostic learning progression, Jin and Anderson (2012) focused on how K-12th grade students used energy-related concepts in their explanations of carbon-transforming processes (e.g. photosynthesis, cellular respiration, biosynthesis) in socio-ecological systems at multiple scales. The authors identified association and tracing as two hallmark practices, and they designed the learning progression around these two dimensions by analyzing explanations provided by the students. They conducted 48 clinical interviews and administered approximately

4,000 written tests to students. Jin and Anderson (2012) essentially used the same iterative process as Gunckel et al. (2009): observation (design/revise assessment); interpretation (data analysis); and model building (design/revise learning progression). Data were collected before and after instruction, whereas Gunckel et al. [2009] collected data after instruction for each assessment cycle. The Jin and Anderson’s (2012) learning progression was the product of five cycles. Based on the results of data, a four level learning progression was developed. Level four indicated students developed the sense that energy must be conserved and degraded in individual processes and in the system as a whole (Jin & Anderson, 2012). Level four was achieved by less than three percent of students sampled (Jin & Anderson, 2012).

2.42 Fostering Student Conceptual Understanding

In terms of fostering student progression, research was very limited. Schwarz et al. (2009) presented a learning progression that could potentially be used to foster progression. The evaporation and condensation learning progression developed by the authors centered on scientific modeling, combined metaknowledge, and elements of modeling practice. They described the progression of learning along two dimensions: a) scientific models as tools for predicting and explaining; and b) models change as understanding improved. The modeling process was operationalized to include four elements: a) constructing models; b) using models; c) evaluating the ability of different models; and d) revising models. Even though the authors (2009) endorsed an instructional modeling sequence based on an operationalization for the practice of modeling, they did not utilize an instructional modeling sequence in the 2009 research. In order to develop learning progression with empirical support, they presented samples of students' work, which demonstrated each dimension. The samples were drawn from various empirical investigations with 5th and 6th grade students. Data were written concerning assessments of reasoning with models, reflective interviews about modeling practice, and classroom discourse during modeling activities (Schwarz et al., 2009). The data, according to the authors (2009), helped demonstrate what kinds of student work could be achieved with good instructional support.

2.5 Neglect of Poor and Urban Research Participants

A disturbing trend was the neglect of non-mainstream, low-SES research participants in urban districts. There was an overwhelming focus on middle to upper class sample participants in rural and suburban areas (e.g. Alonzo & Steedle, 2008; Liu & Lesniak, 2006; Mohan et al., 2009; Plummer & Krajcik, 2010; Rivet & Kastens, 2012; Schwartz et al., 2009). Sample racial/ethnic demographic was either mainstream or not reported (e.g. Alonzo & Steedle, 2008; Liu & Lesniak, 2006; Mohan et al., 2009; Plummer & Krajcik, 2010; Rivet & Kastens, 2012). Schwartz et al. (2009) described their elementary sample participants as "ethnically and linguistically diverse." However, students' socioeconomic status was middle to upper class (Lehrer & Schauble, 2012; Mohan et al., 2009; Neumann et al., 2013; Plummer & Maynard, 2014). Songer et al. (2009), and the follow-up study, Gotwals & Songer (2013) were the only learning progression researchers, thus far, with explicit focus on non-mainstream students in a low SES urban district. Those sample participants were 4th through 6th grade and 6th grade students who attended Detroit Public Schools.

3.0 Major Findings in Learning Progression Research

The major findings that emerged from research literature were, by and large, promising. Four major findings materialized. First, learning progressions contributed to improvement in students' learning outcomes. Another major finding was models, modeling, and symbolic representations emerged as a subcategory within the research field. Two additional major findings were: a) up-and-coming discoveries from the Learning Progressions in Science (LeaPS) conference; and b) stakeholders' use of learning progressions across the U.S. (NSF, 2008; Missouri Learning Standards—ELA, 2015).

3.1 Contribution To Student Outcomes

A major finding was the learning progressions' general contribution to the improvement of student outcomes. Overwhelmingly, the results of many learning progression researchers indicated student improvement in various capacities. Songer et al. (2009) described a method to develop a learning progression on complex thinking about biodiversity. HLM results demonstrated noteworthy student success. Schwartz et al. (2009) presented a two-dimensional learning progression for scientific modeling. The results indicated 5th and 6th graders in the sample were building and modifying increasingly precise models. Songer and Gotwals (2012) investigated 4th through 6th grade students' learning outcomes in their experience with an eight-week scaffold-rich explanation formation intervention about biodiversity and ecology. The results (2012) demonstrated strong learning gains in all three grade-level cohorts. Nevertheless, there was research that demonstrated students not attaining the highest levels of achievement established by the learning progression (e.g. Gunckel et al., 2009; Jin & Anderson, 2012). However, those findings were anomalies.

3.2 An Emergent Subdivision

Within the total body of learning progression literature, a subcategory began to materialize. Several authors focused on scientific models, modeling, and/or symbolic media/representations. Schwarz et al. (2009) offered an operational definition of "a scientific model: a representation that abstracts and simplifies a system by focusing on key features to explain and predict scientific phenomena" (p. 633). Modeling was a fundamental scientific practice, skill, and a prominent facet of scientific literacy (Schwarz et al., 2009). However, models rarely appeared in science classrooms (Schwarz et al., 2009). When they did appear, they were restricted to drawings and were very rarely utilized as theory-building tools (Schwarz et al., 2009). Models were usually added to science curriculum at the high school or university-level and were either taken as obvious or for granted (Lehrer & Schauble, 2012).

Lehrer and Schauble (2012) classified models as a type of representational system. Symbolic media such as drawings, diagrams, photos, and other similar representations were also classified under representational systems. Although all models were a type of representation, all representations were not models (Schwarz et al, 2009). Rivet and Kasten (2012) distinguished three types of models: a) expressed; b) mental; and c) dynamic models. Expressed models were in the public domain (e.g. drawings, photographs). Dynamic models moved and/or changed in response to manipulation by the model user (Rivet & Kasten, 2012). Mental models were not explicitly addressed in learning progression literature, with the exception of Kasten and Rivets' (2012) very brief contrast of it to expressed models.

According to Lehrer and Schauble (2012), models were located anywhere along a representational continuum. The spectrum extended from models that served as examples to models that used symbolic media. They referred to models as "analogical structures" whose attributes changed between the base and the target system (Lehrer & Schauble, 2012). The base of the model was the objects/relations in the analogy; the target systems were the objects/relations that were explained for the phenomena (Lehrer & Schauble 2012). The analogical structures easiest to understand were those that kept the most likeness between the representing and represented worlds (Lehrer & Schauble, 2012).

Both Leher and Schauble (2012) and Schwarz et al. (2009) emphasized the importance of students participating in the practice of scientific modeling. Schwarz et al. (2009) emphasized students modeling the "elements of practice": constructing, using, evaluating, and revising their own models. Lehrer and Schauble (2012) had a similar sentiment: the backbone of science aimed toward building, modifying, using, and defending "natural world" models. According to Schwarz et al. (2009), modeling became accessible to learners when they engaged in the aforementioned practices. However, the skill developed over a long period of time because it was nuanced and had a complex epistemology (Schwarz et al., 2009). Lehrer and Schauble (2012) contended students needed to engage in the epistemic culture of modeling. This culture comprised of the goals, problems, representations, and forms of modeling (Lehrer & Schauble, 2012). However, science educational textbooks and curricula did not address these epistemological intricacies (Lehrer & Schauble, 2012). The authors recommended building modeling practices into school curricula. Schwarz et al. (2009) insisted that if scientific modeling was to be meaningful to learners, it had to be generative.

Learning progression research utilized modeling in various capacities. Plummer and Krajcik (2010) developed learning trajectories (synonymous with levels of achievement) for a full physical model ("dynamic" in Rivet & Kasten's [2012] terminology) of celestial motion. Emphasizing analogical reasoning, the authors (2010) examined ideas of students about celestial motion. Participants used a flashlight as they demonstrated their ideas about apparent celestial motion. Students performed the demonstration on the interior of a dome constructed of PVC pipe and dark canvas material. As students explained their reasoning, the researcher audio recorded students' responses. The interviewer drew visual information from demonstrations performed by their students on a two-dimensional dome template.

Rivet and Kasten (2012) emphasized analogical reasoning to a greater degree than Plummer and Krajcik (2010). Rivet and Kasten (2012) focused on "conceptualization, development, and testing the validity of an assessment of the ability to reason around physical dynamic models in Earth Science" (p. 713). Rivet and Kasten developed a two-dimensional construct map with three levels. The construct map exhibited the progressively refined forms of analogical reasoning between the model and the Earth System. After selecting moon phases as the topic, the authors developed assessment items. They then administered a pretest, the moon phase activities, and a posttest. Rivet and Kasten (2012) cited Getner's (1983) structure mapping analogy framework and employed it as the conceptual framework to guide their research. Getner (1983), as cited by Rivet and Kasten (2012), defined four levels of analogical reasoning. The reasoning occurred between a source and a target; it increased in complexity and abstractness as the levels progressed (Rivet and Kasten, 2012). Getner's (1983) analogy framework, according to Rivet and Kasten (2012), helped to orient the process of establishing alignment between a familiar source (i.e. base according to Lehrer & Schauble, 2012) and the unfamiliar target. The familiar source was the physical model in front of students; the unfamiliar target was the large-scale Earth process (i.e. lunar phases, causes of the seasons, and depositional processes; Rivet & Kasten, 2012). Among other things, Rivet and Kasten's (2012) conceptual framework articulated guidelines for mapping knowledge about the source onto the target.

Schwarz et al. (2009) developed a scientific modeling learning progression. They analyzed data, which helped show the types of knowledge and skills in modeling possible with 6th and 8th grade students. In addition to the learning progression, the authors generated two by-products: a) the potential components of metaknowledge (e.g. nature of models, purpose of models, and criteria for evaluating and revising models); and b) an instructional modeling sequence (Schwarz et al., 2009).

Lehrer and Schauble (2012) offered a modeling learning progression for elementary and middle school students with the goal of understanding the development of modeling "big ideas" with supportive forms of instruction. The big ideas eventually formed the foundation for reasoning about the theory of evolution (Lehrer & Schauble, 2012). Even though the premise of their work rested on the claim that modeling was best achieved by participating in the practice, the authors did not implement this in their research. Rather, they discussed changes in representations and modeling for K-6th grade. They also provided illustrations that exemplified the levels of the learning progression.

Adadan et al. (2010) identified and described conceptual pathways (i.e. learning progression) of 19 11th grade introductory chemistry students. The students voluntarily participated in multi-representational instruction of the particulate nature of matter (PNM). The study focused on stimulating PNM conceptual change; multi-representational instruction was the means to that end. The authors collected open-ended questionnaires and interviews, and then analyzed data with document analysis. The questionnaire, (NMDQ), contained tasks that included pictorial particulate representations coupled with open-ended questions. The open-ended questions required explanations of drawings for a given PNM phenomena (Adadan et al., 2010).

Although modeling became more notable in learning progression research, other forms of representational systems (e.g. microcosms, maps, globes) and/or use of symbolic media (e.g. drawings, diagrams) were missing in the context of learning progression. Adadan et al. (2010) was the only research to examine students' symbolic media and its role in learning progression development and validation. Gotwals and Songer (2013) did not research symbolic media use explicitly, yet they prudently used it as they researched the development of evidence-based explanation on ecology assessments. Outside of the aforementioned rare exceptions, the role of symbolic media in learning progression research was non-existent.

3.3 Learning Progression in Science (Leaps) Conference

The Learning Progressions in Science (LeaPS) conference was an NSF-sponsored conference founded by Amelia Gotwals and Alicia Alonzo (National Science Foundation, 2008). It provided a structured setting for facilitating discussions about challenges in the science learning progression field and it attempted to develop a consensus for possible solutions to these challenges (National Science Foundation, 2008). Nearly 100 science education and cognitive science researchers, measurement specialists, and practitioners gathered and had critical discussions about their work in and around various aspects of learning progressions (National Science Foundation, 2008). The conference was organized around four challenges identified within the research field:

- Defining learning progressions (the need for clearer definitions);
- Developing and validating assessments;
- Using statistical modeling to summarize students' level on learning progression; and
- Using learning progressions (implications of learning progression for curriculum, teacher education, and assessment).

Alonzo and Gotwals published the findings and conclusions from the conference proceedings in their 2012 text *Learning Progressions in Science: Current Challenges and Future Directions*.

3.4 Use of Learning Progressions by U.S. Stakeholders'

Learning progressions were not limited to science education. Several states implemented learning progressions (or derivatives thereof) in other content areas in respective school districts and/or state departments of education. Missouri developed vertical alignment charts for English Language Arts (ELA) for K-12th grade students. The charts were standards for reading literature, reading informational literature, reading foundations, writing, speaking/listening, and language. Missouri's ELA learning progressions were spirally developed, increased in rigor as grade levels increased, and sequentially built (Missouri Learning Standards—ELA, 2015). The Arizona Board of

Regents developed drafts for math learning progressions for students in K-12th grades. Illinois, Idaho, and Arkansas developed ELA learning progressions associated with Common Core ELA for elementary, middle, and/or high school.

The Pennsylvania Education Department, by far, had the most comprehensive use of learning progressions. There were documents for reading, writing, and math for the calendar years of 2013 and 2014. The science learning progression documents included 2010, 2013, and 2014. Pennsylvania also had a Voluntary Model Curriculum (VMC). The VMC was a series of units and lesson plans incorporating learning progressions and content resources aligned to the Pennsylvania standards within the curriculum frameworks. The VMC science unit plans included alignment (e.g. grade level, related academic standards), curriculum (e.g. big ideas, essential questions), and an assessment creator. The science VMC was available for kindergarten to 8th grades, biology, and chemistry.

4.0 Science Teaching, Learning, and Assessment With Science Notebooks

In research and teaching literature, science notebooks were referred to as journals, interactive journals, and learning logs (e.g. Audet, Hickman, & Dobriynina, 1996; Chesbro, 2006; Shepardson & Britsch, 1997). They had a multifaceted function. For example, they portrayed and reflected how science students practice inquiry in the classroom (Aschbacher & Alonzo, 2006). The entries provided a partial record of student instructional experiences in science class and contained students' interpretations of the goals and procedures of inquiry activities as presented by the teacher (Madden & Wiebe, 2013; Ruiz-Primo, Li & Shavelson, 2002). In concert, the multi-faceted role was considered to be "curricular evidence." Baxter, Bass, & Glasser (2000) noted curricular evidence was a critical aspect of science teaching and learning. Teaching and research literature demonstrated how science notebooks were used to monitor science instruction and assess students' learning. Teachers often used science notebooks as a tool for teaching, learning, and assessment within the confines of inquiry (Baxter et al., 2000). When this was the case, notebooks functioned as documentation of teacher instruction, provided differentiation and scaffolding opportunities, became a vehicle for tracking student progress over time, was a medium for student-teacher science dialogue, and served as a tool for formative assessment (Audet et al., 1996; Baxter et al., 2000; Madden & Wiebe, 2013; Ruiz-Primo, Li & Shavelson, 2002; Shepardson & Britsch, 1997).

Baxter et al. (2000) investigated monitoring instruction by examining the use of science notebooks during a unit on electricity with 5th grade students. Data (2000) were collected from 83 student notebooks in an urban school district. Baxter et al. (2000) found science notebooks consistently reflected what students did and what teachers focused on during the science class. Ruiz-Primo et al. (2002) focused on monitoring instruction and assessing learning. The authors (2002) examined 10 urban teachers and the teachers' classrooms in which two Full Option Science System® (FOSS) units were implemented. The study (2002) investigated the nature of activities encountered in science class, the nature of teacher feedback, and the interaction of those two dimensions. The authors (2002) analyzed the science notebooks of 60 5th grade participants. They concluded science notebooks permitted teachers to assess student understandings, and also gave the feedback students needed to improve performance. Aschbacher and Alonzo (2006) focused on monitoring student science notebooks as a means of formative assessment. Participants were 25 teachers and their 4th and 5th grade students. The inquiry unit focused on students' conceptual understanding of circuits. The teachers were divided into two groups. The first group engaged in a professional development workshop focusing on using science notebooks as a formative assessment tool. The second group did not receive the professional development. The classrooms whose teachers received the professional development were then compared to the classrooms of teachers who do not receive it. The authors (2006) concluded notebooks had tremendous "potential as a tool for formative assessment and they reveal student thinking (p. 200)."

4.1 Teacher Practices and Student Experiences as Reflected in Science Notebooks

Teacher practices and student experiences influenced student learning outcomes. Science notebooks represented student experiences with the science curriculum and were an abundant source for artifacts (Madden & Wiebe, 2013). Moreover, they were used to examine the impact on learning outcomes and the context of the science instruction (Klentschy et al., 2004).

Teacher identity—"what kind of person" a teacher was—was linked to science teachers' instructional practice (Madden & Wiebe, 2013). Baxter et al. (2000) found that the ways the teacher interpreted the unit was emulated in students' science notebooks. The content, organization, magnitude, and quality of students' science notebook entries were a reflection of the teacher's pedagogical methods (Baxter et al., 2000). Notebook entries revealed the type and duration of learning that transpired (Madden & Wiebe, 2013). They also gave teachers better insight into how their students understood their teaching because the notebooks were a window into students' thinking (Madden & Wiebe, 2013; Morrison, 2005).

Morrison (2005) presented a provocative study illustrating how science notebooks revealed the intersection of teacher practices and student experiences. Morrison (2005) explored how participants used their respective notebooks and how the notebooks influenced their understanding and predicted use of formative assessment involving science notebooks. Data were collected from notebooks. Participants were undergraduate and graduate pre-service teachers in a science methods course. They kept a science notebook during their matriculation. As evidence of growth in notebook use, one entry was photocopied from the beginning and from the end of the semester for each participant. During the semester, participants received informal and formal feedback from the instructor, wrote a formal paper about the use of science notebooks as a formative assessment tool, and wrote a reflection about their own use of science notebooks. Participants also completed an anonymous questionnaire at the end of the course, which explored participants' personal use of notebooks as an assessment tool, their future use for notebooks, and what they gained through assessment of their own notebooks. Morrison (2005) found pre-service teachers saw science notebooks as a way to continually gather information from students, and as an opportunity to provide students with consistent and constructive feedback**.**

5.0 Coalescencing Science Notebooks, Graphics, and Inquiry- Based Instruction

The research team of Wiebe et al. (2008; 2009a; 2009b) investigated student-generated graphic representations in science notebooks within the context of inquiry-based science instruction. These graphics were an integral part of the science notebook process (Wiebe et al., 2008). The research team concentrated on spatial intelligence and student-produced graphics with science-kit instruction in elementary education.

According to Wiebe et al. (2008), “there was a positive connection between student-generated graphics as part of science inquiry activities and conceptual learning of relevant science topics (p.1)”. Wiebe et al. (2008) aimed to determine how to enhance spatial intelligence as a learning tool for 2nd to 5th grade students. The study (2008) occurred in a single school in an urban/suburban district. Using classroom observations, assessments of utilized science kits, and student-generated graphics collected from science notebooks, there were four salient findings:

- Some teachers more than other were comfortable using graphics to further student thinking;
- Science kits and professional development did not position teachers to increase student-generated graphics;
- Graphic usage was not integrated across the inquiry process; and
- How different graphic types served and complemented parts of the inquiry cycle was not highlighted (Wiebe et al., 2008).

The findings served as a springboard for future research, which focused on the intersection of inquiry-based science and the role science notebooks played in the process.

Wiebe et al. (2009a) investigated the capacity of science notebooks to communicate evidence of inquiry practices in 2nd to 5th grade classrooms. They (2009a) focused on student-generated graphic representations in different stages of the inquiry process. Central to interpreting student-generated graphics was linking them to the classroom experiences that induced their creation. Science instruction was delivered through district adopted science kits and science notebooks were collected from two teachers per grade level. Each teacher selected between two and twelve notebooks per class to obtain a representative cross-section of student ability. Graphic representations were categorized according to the semiotic taxonomy: text-graphic, spatial organization, drawing’s scale representation, and drawing’s temporal representation. Each semiotic taxon represented a major analogical aspect of graphics (Wiebe et al., 2009a). Taxon arrangement was not hierarchal (Wiebe et al., 2009b).

Findings from Wiebe’s et al. (2009a) graphic analysis revealed very little pre-lab graphic activity and limited post-lab graphic activity. In the analysis of post-lab graphics of Wiebe et al. study (2009a), many of the student-generated entries were re-representations of text near it (e.g. Venn-diagrams). Many entries focused on during-lab activities (e.g. listing materials and procedures) and were heavily guided by the teacher. There was very little opportunity for student ownership/originality. Overall, the graphic analysis revealed strong teacher structuring of the content. Wiebe et al. (2009a) also demonstrated science notebook entries as evidence of in-class inquiry practices.

The findings of Wiebe et al. (2009a) informed Wiebe et al. (2009b). Wiebe et al. (2009b) investigated the capacity for science notebooks to efficiently inform a professional development aimed at guiding teachers in using student-generated graphics. A purposeful sample was analyzed for graphic content with an expanded semiotic taxonomy. The sample consisted of 32 science notebooks from a similar sample of students in Wiebe et al. (2008; 2009a). Wiebe et al. (2009b) found an uneven distribution of graphic production across the stages of inquiry, and teacher-driven entries dominated students’ notebooks. Furthermore, the analysis revealed students’ entries represented concrete, macro-scale, and real-time science phenomena.

Although Wiebe’s et al. (2009b) semiotic taxonomy gave insight into the intersection of teacher pedagogical content knowledge and skills, science kit-based curriculum, science notebooks, and student cognition, they (2009b) did not explicitly articulate how their analysis informed future professional development efforts for elementary teachers. Nevertheless, they (2009b) offered the suggestion to develop "graphical progressions"—master images that would be canonical representations of scientific phenomena. They (2009b) proposed both teachers and students could use graphical progressions throughout a kit and across grade levels.

6.0 Content Domain Analysis and Student Thinking

Multi-faceted systems were a significant emphasis of thinking and learning research. Most science textbooks, however, did not support learning science in this capacity. Rather, they supported learning science as a set of facts as opposed to big ideas to help foster integrated understanding and mediated behaviors of complex, interconnected systems (Liu & Hmelo-Silver, 2009). For the last two decades, science education research had been driven by the recognition of the importance of complex systems and the inadequacies of methods in helping students identify them (Kali, Orion, & Eylon, 2003). Furthermore, earth science education shifted towards a systems approach to teaching and curriculum development during the same time frame (Kali et al., 2003). The *Frameworks* (2014) recommended and emphasized the need for exposing students to the systems thinking approach and developing systems thinking skills among students beginning at the elementary level. It also delineated three dimensions in each of its content areas: a) scientific and engineering practices; b) cross-cutting concepts; and c) core ideas. "Systems and system models" was one of the cross-cutting concepts.

Coined by Barry Richmond in 1987, the definition of systems and systems thinking ranged from basic to broad (Arnold & Wade, 2015). Generally, researchers considered systems thinking as a vital skill set in a world in which systems were becoming increasingly complex (e.g. Arnold & Wade, 2015; Assaraf & Orion, 2005; Assaraf & Orion, 2010; Raved & Yarden, 2014). O'Conner and McDermott (1997) defined a system as an entity that preserved its existence and operated as a whole through the interaction of its parts. Kali et al. (2003) defined systems thinking as the type of thinking needed for understanding systems. Systems thinking was studied in medicine, and engineering, as well as other content domains outside of STEM fields (Orion & Libarken, 2014; Kali et al., 2003).

Systems thinking skill development was represented by several models (Orion & Libarken, 2014; Liu & Hmelo-Silver, 2009). Assaraf and Orion (2005) presented a model for systems thinking skills in earth science education. Their (2005) System Thinking Hierarchal (STH) model had eight developmental stages arranged in three hierarchal levels: a) analysis (stage 1); b) synthesis (stages 2-5); and c) implementation (stages 6-8). They (2005) utilized the STH model as part of their investigation of both 8th grade and high school students. The authors (2005) described hierarchal system thinking skills development as follows (p. 541):

1. The ability to identify the parts of a system and processes within the system;
2. The ability to identify simple relationships between or among the system's parts;
3. The ability to identify dynamic relationships within the system;
4. The ability to organize the systems' parts, processes, and interactions, within a framework of relationships;
5. The ability to identify the cyclic nature of matter and energy within the system;
6. The ability to recognize hidden dimensions of the system;
7. The ability to make generalizations—to solve problems based on understanding systems' mechanisms;
8. The ability to think temporally— retrospection and prediction.

Each of the eight facets of the systems thinking hierarchy (STH) model appeared independently in research literature, but they appeared in the context of different systems

(Assaraf & Orion; 2010). Traditionally, STH models were presented in the shape of a pyramid. The model's wide base represented the analytical skills. Moving toward the apex, the pyramid model narrowed and represented students possessing synthesis skills. The apex of the pyramid model represented students possessing implementation skills. Therefore, as systems thinking level increased (i.e. moves simultaneously through the eight hierarchal levels and three developmental stages), the amount of students possessing a particular systems thinking skill decreased. Consequently, a student reaching the implementation level (the highest systems thinking level) had to successfully complete the analysis and synthesis levels (see Figure 1). Although the STH model provided a system for delineating the development of systems thinking, it only highlighted the "touchstones" students passed through in their trajectory from lower to higher order systems thinking (Assaraf & Orion, 2010).

Kali et al. (2003) led a study describing the specific systems thinking required for understanding the rock cycle at the middle school level. The authors (2003) defined three general elements for systems thinking: a) understanding the parts of a system; b) understanding the connections between these parts; and c) understanding the system as a whole. Kali et al. (2003) found that most middle school students did not reach an understanding of both the dynamic and cyclic natures of the rock cycle even though they understood all the relevant geological processes and products.

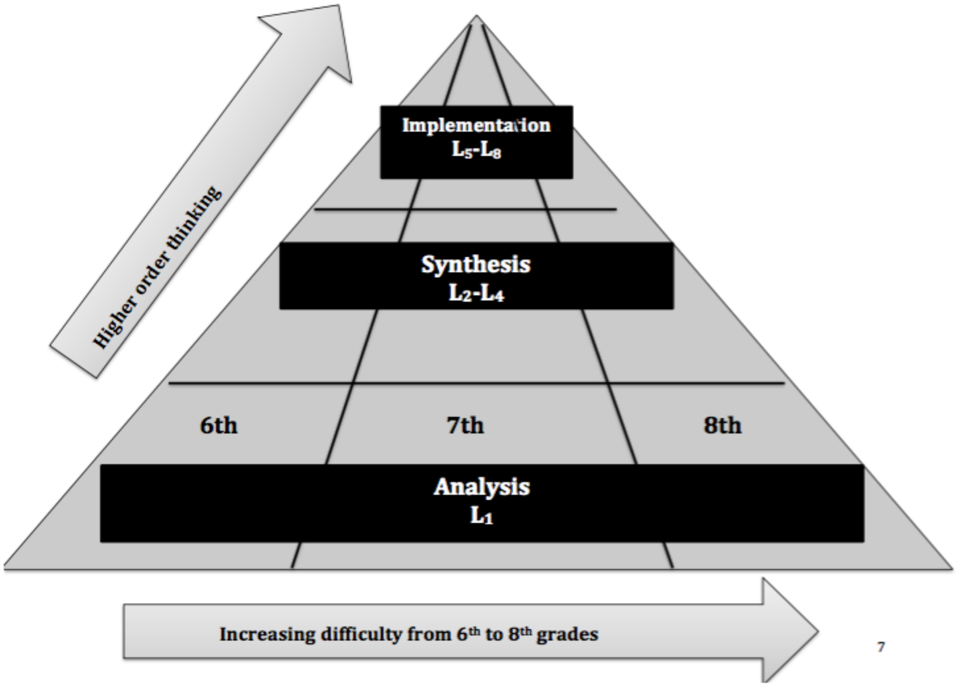

Figure 1. Rock cycle learning progression general structure (STH model). The pyramid illustrates the basic structure of the Rock Cycle learning progression. The lowest level, Analysis (L_1), is at the base of the model and represents the greatest population. As students transition to Synthesis (L_2-L_4) and Implementation (L_5-L_8), the thinking becomes more complex and fewer students inhabit those levels.

Assaraf and Orion (2005) examined an 8th grade earth system-based curriculum that focused on the water cycle in an inquiry context. There were three salient findings (Assaraf and Orion, 2005):

- Systems thinking development among middle school students was comprised of many stages arranged in hierarchical order.
- Even though students had marginal initial system thinking abilities, most achieved meaningful progress in system thinking.
- The factors that influenced the differential progress the most were students' initial system thinking cognitive abilities and their level of involvement in the inquiry-based activities.

Orion and Assaraf (2009) investigated the initial systems thinking levels of high school students who had not learned the middle school systems thinking unit. They (2009) found the initial STH levels of high school students did not differ significantly from 8th graders of the Assaraf and Orion 2005 study. They (2009) also found that students whose initial level of systems thinking was low developed much less than those with higher initial levels of systems thinking.

Assaraf and Orion (2010) examined elementary school students' complex systems thinking skills (based on their findings and recommendation of the 2005 study). Specifically, the authors (2010) studied 40 4th grade students in one school from a small town in Israel as the students studied the water cycle through an inquiry-based earth systems curriculum. The authors (2010) found that, despite students' minimal initial system thinking ability, most made significant progress with their ability to analyze the parts and processes of the water cycle. Some students even reached higher system thinking abilities. Assaraf and Orion (2010) also examined system thinking perception development. Specifically, four of the middle school student-participants from their 2005 study were observed via semi-structured interviews, observations, and a variety of "concept viewing" tools before, during, immediately after, and six years after completing the 2005 study. The authors (2010) concluded that students developed their systems mental models and remembered the learned material based on learning patterns that remained unchanged over time.

7.0 Curriculum Framework: Spiral Curriculum Design

Three features were indispensible to spiraling the curriculum of the Rock Cycle learning progression. First, students revisited the big ideas and the analysis level (L_1) of the Rock Cycle learning progression on several occasions during the intervention (Bruner, 1960; Harden & Stamper, 1999). Secondly, the Brunarian spiral curriculum design for the Rock Cycle learning progression had increasing levels of difficulty (Bruner, 1960; Harden & Stamper, 1999). The third feature was demonstrated as students participated in the inquiry-based labs. The situational context provided opportunities to discuss the learning connections across the STH model. New content and/or skills introduced at higher levels of the progression were related back and directly linked to learning in lower levels of the spiraled learning progression. Likewise, what was learned in the beginning of the learning progression was linked to what was learned at higher levels within the progression (Harden & Stamper, 1999). In terms of the organization and structure of the learning progression, the hierarchal and iterative nature of the STH model was an intrinsic feature and therefore facilitated the use of the spiral curriculum as its framework.

8.0 Theoretical Framework: Situated Cognition Theory

Brown et al. (1989) developed situated cognition theory (also referred to as situated learning theory, SitCog, situated action, and situativity). The theory contended that knowing was connected to doing. It was based on the supposition that knowledge should be presented in its authentic situation, which involved its application. Hence, situated cognition theory urged teachers to immerse students in a learning environment that imitates the real-world context. Students applied their new conceptions and skills in “real-world” learning environments (Brown et al., 1989). Brown et al. (1989) posited that the vital element of knowledge was positioned; it was anchored in the environment in which it was used. Furthermore, knowledge was partially created from the activity, context, and culture in which it was developed and was used (Brown et al., 1989). Activity, concept, and culture were interdependent, and learning must involve all three (Brown et al., 1989).

8.1 Theoretical Tenets

According to situated cognition theory, concepts were situated and increasingly developed through activities. Brown et al. (1989) suggested conceptual knowledge analogous to a set of tools in order to explain how learning takes place. Like tools, conceptual knowledge was understood through its use (Brown et al., 1989). Moreover, using tools/conceptual knowledge stimulated shifts in the user's perspectives and caused the user to adopt the belief systems of the culture (Brown et al., 1989). The situated cognition theory stated it was impossible to properly use tool/conceptual knowledge without understanding the community and culture in which it was used (Brown et al., 1989).

8.11 Enculturation

Enculturation was one tenet of situated cognition theory. According to the theory, enculturation process emphasized the socio-cultural context of the learning environment and ensured that learning and doing were not divorced from each other (Ho, 2015). The teacher's role was practitioner, and the teacher used the tools/conceptual knowledge in a way that called for students to wrestle with problems of the "real-world" (Brown et al., 1989).

8.12 Authentic Activity

Authentic activity was a second tenant of situated cognition theory. It addressed how practitioners orchestrate the "real-world" problems for their students. Brown et al. (1989) defined authentic activities as the prevalent practices of a culture that were coherent, meaningful, and purposeful. In an authentic activity, the teacher selected and implemented the situation. The teacher also provided the necessary scaffolding for learning within the situation (Ho, 2015).

8.13 Cognitive Apprenticeship

Cognitive apprenticeship was a third theoretical tenet. The pedagogic strategy aspired to contextualize learning and focused on skill acquisition (Brown et al., 1989). It was a process for teachers to impart their skill to students through a training process (Brown et al., 1989). The teacher intentionally elicited thinking to the surface and made it visible. Brown et al. (1989) listed three instructional procedures of cognitive apprenticeship:

- Identify the processes of the task and make them visible to students;
- Situate abstract tasks in authentic contexts so that students understand the relevance of the work; and
- Vary the diversity of situations and articulate the common aspects so that students can transfer what they learn.

Through cognitive apprenticeship, learning was fostered within the nexus of activity, tool, and culture because apprentices were encultured via activity and social interaction (Brown et al., 1989). Cognitive apprenticeship was not a compatible paradigm for all aspects of teaching, nor was it a “packaged formula for instruction” (Collins et al., 1991). It was a teaching paradigm to guide the pedagogical and theoretical issues that were associated with designing learning environments and experiences (Collins et al., 1991).

8.2 Empirical Evidence Supporting Situated Cognition Theory

Preece and Bond-Robinson (2003) used an ethnomethodological approach with three undergraduate novices who were selected for a NSF Research Experience. The authors (2003) examined cognition in “science-as-practice” based on situated learning with the novices. The interaction of individual, context, and activity was captured in 60 hours of video. The authors (2003) found that while apprenticeship was not efficient for their (2003) research, it was highly effective as a learning environment. There were two major findings: a) novices were cognitively and motivationally challenged; and b) novices exhibited difficulty transferring course knowledge to research. Sweeney and Paradis (2004) used the situated cognition theory framework to design and develop a laboratory-training course. The course provided two pre-service secondary science teachers with the opportunity to explore the pedagogical potential of the teaching laboratory and gaining hands-on experience running a general chemistry laboratory. A case study methodology was employed in the study, and the authors (2004) found the laboratory model of teacher preparation they developed positively influenced the pre-service teachers' abilities to design, organize, and manage chemistry laboratory experiments and activities. Sweeney and Paradis (2004) also found the model positively influenced the pre-service teachers' enculturation into the respective science subcultures of chemistry and science education. In 2005, Bond-Robinson and Preece-Stucky used ethnographic methods to explore the cognitive processes and the social environment in an organic synthesis laboratory. Specifically, the authors (2005) examined a graduate research group performing organic synthesis of molecules. The authors (2005) observed the daily work and problem solving in over 100 hours of video data as well as conducted informal and semi-structured interviews. Based on the findings, Bond-Robinson and Preece-Stucky (2005) concluded thinking and acting by the apprentice graduate researchers in the community of practice molded their everyday thinking into the scientific reasoning required to be a proficient organic research scientist. Brown et al.

(1989) examined two examples of mathematics instruction whereby children successfully solved math problems through authentic practices and activities.

Chapter 3: Methodology

The purpose of the study was two-fold: a) to develop and validate a middle school science learning progression in an inquiry context by using science notebooks; and b) to examine the impact of science notebooks' use with the learning progression on students' learning. The research question for the study was: What is the impact on students' science learning outcomes when a middle school science learning progression is developed and validated using science notebooks as part of an inquiry-based instructional intervention?

A causal comparative case study was the research design for the study. Three groups were compared: a) a Computer-assisted instruction group that was on campus; b) a Computer-assisted instruction group that was off campus; and c) a Learning Progression group. The on-campus computer-assisted group received a computer-based rock cycle curriculum for 12 total hours while the off-campus group received five total hours. The Learning Progression group received the learning progression curriculum and participated in the instructional intervention. All participants took the pretest and posttest on the same respective day (with the exception of the Off-campus group).

1.0 Study Context, Population, and Sample

The study utilized both quantitative and qualitative research methods in data collection and analysis. The study took place in an urban public charter school during the summer of 2016 in conjunction with the school's summer enrichment program. The purpose of the summer program was to extend the school year, sustain students' overall achievement, and familiarize students with their teachers for the upcoming school year.

The program was volitional and grades were not issued to students for any classes. The researcher's role was that of the "8th science teacher." Three groups of students were compared: a) a learning progression-science notebook students; b) an on-campus computer-assisted instruction students; and c) an off-campus computer-assisted instruction students. The learning progression-science notebook students received the maximum instructional time in the inquiry-based learning progression. The on-campus computer-assisted instruction students also received maximum instructional time, but their curriculum consisted of science expository writing and comic strip production. The off-campus computer-assisted had the same curriculum, but students received substantially less instructional time.

Approximately 96% of the schools' students were non-mainstream and had a low-SES. The total enrollment for 2015 academic year was 572; the school had a 91% total attendance rate. Ninety-three percent of all enrolled student were black, while 7% were white. Ninety-four percent of students were eligible for free or reduced-price lunch.

The sample was four classes of rising 8th grade students whose ages ranged from 12 to 14 years old; three students did not report their age. Students' race and/or ethnicity were retrieved from the school's database (see Table 1). Two classes comprised the learning progression-science notebook students group, while the remaining two classes comprised the computer-assisted instruction group. Due to circumstances beyond the teacher-researcher's control, the computer-assisted instruction group was split into an on-campus and off-campus group. The learning progression-science notebook group consisted of 16 students. The on-campus group had six students while the off-campus group had 10 students. In general, the sample consisted of black, 13 year-old students. Females were predominately in the on campus group while males were predominately in the learning progression-science notebook group. Each class was an hour long and met Monday through Friday. The exception was the off-campus group; they received instruction five of the 12 days available.

Table 1

Racial/Ethnic, Gender, and Age Demographic of Students

Demographic	**Computer-Assisted Instruction**				**Learning Progression-Science Notebook**		**Grand Total**	
	On-Campus		**Off-campus**					
	n	Percentage	*n*	Percentage	*n*	Percentage	*n*	Percentage
Ethnicity/Race								
Black	3	50%	10	100%	15	93.7%	28	87.5%
Multi-racial	2	33.3%	0	0%	1	6.2%	3	9.4%
White	1	16.6%	0	0%	0	0%	1	3.1%
TOTAL	6	100%	10	100%	16	100%	32	100%
Gender								
Female	5	83.3%	5	50%	7	43.7%	17	53.5%
Male	1	16.6%	5	50%	9	56.2%	15	46.5%
TOTAL	6	100%	10	100%	16	100%	32	100%
Age								
12	0	0%	2	20%	1	6.2%	3	9.4%
13	5	83.3%	7	70%	10	62.5%	22	68.8%
14	1	16.6%	0	0%	3	18.7%	4	12.5%
Not Reported	0	0%	1	10%	2	12.5%	3	9.4%
TOTAL	6	100%	10	100%	16	100%	32	100%

On the first day of instruction, it was disclosed by the school's administration that the second period class was to participate in an off-campus program at a local university. Students were selected by administration based on their attendance, behavior, and grades during the academic school year. The off-campus program lasted two of the three weeks of the summer program. Consequently, the second period class only met on Fridays and three days of the last week.

1.1 Setting

Three groups were examined in efforts to describe the impact of the use of science notebooks in conjunction with the rock cycle learning progression: a) the learning progression-notebook group; b) the on-campus computer-assisted group; and c) the off-campus computer-assisted group. The researcher was the data-gathering instrument, participating in the study as the teacher and specifically collecting data from the Rock Cycle Assessment and science notebooks. Because of the qualitative approach, several sources of error persisted in the research design. Efforts were made to decrease error from the researcher, participating subjects, the social context, and during data collection and analysis and thereby increase the validity and reliability of the study. Foremost, the teacher-researcher made sure the student participants were very clear on the nature of the research. Secondly, a trust-relationship was built with the subjects as the teacher-researcher stayed in the setting (i.e. classroom) for the duration of the study. Third, informal interviews were conducted with many subjects on several occasions for the duration of the study.

Triangulation, multiple repetition, and thick description were used to address threats to internal and external validity. Three data sources were used in the study. The

Rock Cycle Assessment was analyzed with a *t*-test for independent means and the learning progression levels frequency distribution was extracted. Symbolic media and reflective conclusions were two data sources used from science notebooks. Symbolic media were analyzed using semiotic taxonomy while reflective conclusions were analyzed utilizing the constant comparative method. Multiple data sources and analysis approaches were employed to minimize researcher personal bias in addition to overcoming the inherent deficits to single-investigation, single-theory, and single-method studies. This increased the validity of the study.

There were other strategies employed to increase the consistency and trustworthiness of the results. There were two repetitions of the Rock Cycle Assessment, pre-lab, during lab, and post-lab during the study. Also, students were permitted to work in class outside of their scheduled class time (e.g. lunch or elective). Finally, a thick description was given for the development and validation of the middle school rock cycle learning progression such that the methodology was replicable. Nevertheless, threats to validity were present in the study. Descriptive validity was a threat because the researcher was unable to record while gathering data. Group composition effects were a concern because pre-existing differences among the groups could obscure the effects of the learning progression. Lastly, selective sample attrition was a tremendous threat as participants dropped out of the groups as the study progressed.

The research site was a charter school located in St. Louis, Missouri, in the Soulard neighborhood. The school facility was a miscellany of buildings in the process of coalescence. The main building was a fusion of four buildings and housed 7th to 12th grade students.

The middle school side of the main building had nine classrooms and the middle school administrative office. The study was facilitated in the 8th grade science classroom. There were six laboratory workstations located in the classroom and 25 desks in the classroom. The classroom was located in a high traffic area of the middle school, adjacent to the water fountain, restrooms, the shared principal and dean of students' office, and copy machines. Because of the high traffic, the teacher-researcher decided to keep the classroom door closed and locked during instructional time.

The study used one of Duncan and Hmelo-Silver (2009) recommended approach to develop the learning progression: developing an initial learning progression from existing research on student learning and thinking in the content domain. According to the authors (2009), developing an initial progression required a validation study that involves developing and implementing an instructional intervention. The treatment translated into the following methods:

1. Develop the learning progression;
2. Develop the instructional intervention;
3. Develop the assessment instrument;
4. Align the curriculum, instruction, and assessment of the learning progression;
5. Administer a pretest;
6. Implement the learning progression via instructional intervention;
7. Administer a posttest.

The methods employed to develop the rock cycle learning progression are described in this section: a) determining the upper and lower anchors; and b) constructing the intermediate levels. This section also describes the two-part validation process: a) developing the instructional intervention; and b) incorporating science notebooks in the intervention.

2.0 Rock Cycle Learning Progression Development

The first major task was to develop a learning progression. The content domain was Earth and space science (ESS), specifically, the rock cycle. To develop and inform the learning progression, the teacher-researcher surveyed student thinking and learning research in ESS relative to development of systems thinking skills and other relative domains. Kali et al. (2003) defined systems thinking as the type of thinking needed for understanding systems.

A two-dimensional pyramid diagram emerged. It was developed, deconstructed, and drafted into a middle school Rock Cycle learning progression. Assaraf and Orion (2005) presented a systems thinking skills model in earth science education. The model—the systems thinking hierarchal (STH) was a three-tiered model with an eight-level framework. It served as the draft for the study's concept map and consequently formed the intermediate region of the progression. Organized around the big ideas of ESS, NGSS core disciplinary ideas (CDI's) and the STH model, the learning progression had four theoretical tenets (Duncan & Hmelo-Silver, 2009, p. 67):

- It focused on a few content ideas and/or inquiry practices (i.e. rock cycle);
- An upper and lower anchor confined it (i.e. MS-ESS2-1 and 5-ESS2-1);
- Levels of achievement described the intermediate levels between the upper and lower anchors (i.e. STH model); and
- Targeted instruction and curricula mediated it (e.g. instruction and curriculum focused on parts of the rock cycle, understanding the connections between those parts, and understanding the system as a whole).

Other structural components included the grade band, scope, and grain size. The learning progression's grade band was 6th through 8th grades and each grade had eight levels of achievement.

2.1 General Anatomy of Rock Cycle Learning Progression

The rock cycle learning progression had three general parts: a) the upper anchor; b) the levels of achievement (i.e. the intermediate region); and c) the lower anchor. The upper and lower anchors were the boundary of the rock cycle learning progression while the STH model was in between those anchors. The modified STH model was the construct map, a draft of the intermediate region, and provided a means for delineating systems thinking skills development. It highlighted the "touchstones" of students' trajectory from lower to higher order systems thinking (Assaraf & Orion, 2010).

Fifth grade and middle school DCI performance expectations were the lower and upper anchors of the learning progression, respectively:

1. Lower Anchor: 5-ESS2-1—Develop a model using an example to describe ways the geosphere, biosphere, hydrosphere, and/or atmosphere interact;
2. Upper Anchor: MS-ESS2-1—Develop a model to describe the cycling of Earth's materials and the flow of energy that drives this process.

The lower anchor described what students should know (see Figure 2). The upper anchor described what students should know and/or be able to do relative to societal expectations (see Figure 3; Duncan & Hmelo-Silver, 2009; Smith et al., 2006).

2.12 Construct Maps: Drafting The Intermediate Levels

The construct map helped develop the intermediate levels of the learning progression and used the STH model as its organizational framework. The first step was identifying the big ideas within ESS systems thinking by analyzing research literature and documenting fundamental content skills. The two big ideas were: a) Earth was continuously changing; and b) Earth was a complex system of interacting rock, water, air, and life.

Lower Anchor: 5-ESS2-1 Develop a model using an example to describe ways the geosphere, biosphere, hydrosphere, and/or atmosphere interact (assessment is limited to the interactions of two systems at a time)			
ANALYSIS **(Touchstones/Levels of Achievement)**			
STH Model	**6th**	**7th**	**8th**
L_1: Ability to identify the components of a system and processes within the system	**Identify Earth Systems' Components** Hydrosphere, geosphere, atmosphere, and biosphere **Identify Earth Systems' Processes** The rock cycle, the food chain, the water cycle **Identify Rock Cycle Components** Sedimentary, igneous, & metamorphic rocks, rocks at the surface, soil **Identify Rock Cycle Processes** Weathering/erosion, high temps/pressure, melting, cooling	**Identify Earth Systems' Components** Hydrosphere, geosphere, atmosphere, and biosphere **Identify Earth Systems' Processes** The rock cycle, the food chain, the water cycle **Identify Rock Cycle Components** Sedimentary, igneous, & metamorphic rocks, rocks at the surface, soil, mobile sediments, sedimentary sequences **Identify Rock Cycle Processes** Weathering/erosion, high temps/pressure, melting, cooling, transportation, deposition, compaction/cementation, metamorphism,	**Identify Rock Cycle Components** Sedimentary, igneous, & metamorphic rocks, rocks at the surface, soil, mobile sediments, sedimentary sequences, sedimentary rocks, metamorphic rocks, magma (from below), intrusive igneous rocks, extrusive igneous rocks, layers of the earth, tectonics **Identify Rock Cycle Processes** Weathering, erosion/transportation, deposition, compaction/cementation, metamorphism, melting, extrusion, crystallization, uplift, convection

Figure 2. Lower anchor and analysis stage of the rock cycle learning progression. This figure illustrates the lower anchor boundary, the lowest level (L_1) of achievement, and the hallmark practices (in bold) of the Rock Cycle learning progression for 6th to 8th grades.

IMPLEMENTATION **(Touchstones/Levels of Achievement)**			
STH Model	**6th**	**7th**	**8th**
L_6: Ability to make generalizations	**Make Generalizations About** • The dynamic and cyclic nature of the rock cycle • Transformation of matter in the rock cycle • Energy in the rock cycle • Influence and/or interaction of either the atmosphere, hydrosphere, or biosphere • How the Earth changes as a consequence of the rock cycle • How the processes of the rock cycle affects the products of it		
L_7: Understanding the hidden dimensions of the system	**Recognize Patterns And Interrelationships Of The Rock Cycle Which Are Not Seen On The Surface** • Metamorphism of rocks • Melting of rocks (cooling of intrusive igneous rock) • Plate tectonic movements • Folding, faulting, and uplift		
L_8: Thinking temporally: retrospection and prediction	**Retrospective/Prediction Of The Temporal Component Of The Rock Cycle** • Changes in the rate that rocks are made and destroyed can have a profound affect on the planet. o E.g. As the rate of plate tectonic movements has changed over geologic time scales, the rock cycle has changed as well. ▪ at times when the rate of plate movements has been high, there is more volcanic activity, which releases more particles into the atmosphere. Faster plate tectonic movements also mean more mountains are built in areas where plates converge. As rocks are uplifted into mountains, they start to erode and dissolve, sending		

sediments and nutrients into waterways and impacting the ecosystems for living things.

Upper Anchor: MS-ESS2-1 Develop a model to describe the cycling earth's materials and the flow of energy that drives this process (emphasis is on the processes of melting, crystallization, weathering, deformation, and sedimentation, which act to together to form minerals and rocks through the cycling earth's materials)

Figure 3. Upper anchor and Implementation stage of rock cycle learning progression. This figure illustrates the upper anchor boundary, the highest levels (L_6- L_8) of achievement, and the hallmark practices (in bold) of the Rock Cycle learning progression for 6th to 8th grades.

The second step was diagramming the STH traditional pyramid model combined with the specific STH framework and its three hierarchal levels: a) analysis (stage one); b) synthesis (stages two through five); and c) implementation (stages six through eight). The diagram was deconstructed such that each level at every stage could be clearly and fully articulated. The eight stages were listed and respective touchstones/achievement levels were expressed at every grade level. The third step involved arranging the construct map according to the tenets of a spiral curriculum. Recommendations and the teacher-researcher's professional judgment was used to hierarchically arrange the touchstones/levels of achievement so that a spiraled continuum was achieved. The construct map was reviewed and edited. The final edit served as the intermediate levels for the learning progression. The edited construct map was fused with the upper and lower anchors to form the completed rock cycle learning progression.

3.0 Learning Progression Validation

Once developed, the learning progression had to be validated. According to Duncan and Hmelo-Silver (2010), validating a learning progression required the development and implementation of an instructional intervention and an assessment instrument.

3.1 Instructional Intervention Development and Implementation

Several inquiry activities were harvested from several sources. After careful review, the activities were aligned to the levels of the learning progression. Multimedia presentations, curricular materials, and instructional resources were secured, prepared, and organized for each lab according to Shepardson and Britsch's (1997) instructional outline. The rock cycle learning progression, its supplement materials, situated cognition instructional principles, and the instructional framework were incorporated into the teacher-researcher's pedagogical repertoire and facilitated daily. To ensure fidelity of the intervention's execution, five faculty-participants observed the classes at least three times weekly. Each class met for one hour daily. With the exception of field trip attendance, the Learning Progression group received at most 11 hours of instructional intervention time.

3.2 Science Notebook-Based Instructional Intervention

The intervention used Shepardson and Britsch (1997) instructional outline for children's science journals. The authors provided definitive parameters for the three phases of inquiry: a) pre-lab, b) during lab, and c) post-lab. Symbolic media were generated in each phase. In pre-lab, students explained existing ideas/understandings, described the purpose of the investigation, stated questions to be answered, made predictions/hypotheses, and explained procedures. During the lab, students recorded observations and created drawings, charts, and tables for organizing data. In post-lab, students used data and other resources to explain the results, reflected on existing ideas and predictions in light of findings, and identified ways of conducting the investigation differently or improving the investigation.

The instructional intervention centered on using science notebooks in inquiry-based activities. Symbolic media and reflective conclusions were produced in science notebooks. Symbolic media included drawings, diagrams, photos, and organized text written and drawn in students' notebooks; they provided a record of students' development of science concepts (Wiebe et al., 2009). There were six categorical qualities of symbolic media: a) text-graphic relationship; b) spatial organization; c) scale representation; d) temporal representation; e) re-representation; and f) driving force of notebooks (Wiebe et al., 2009). Symbolic media were generated at each phase of inquiry while reflective conclusions were generated only in post-lab.

The first class day was for introduction and organization. On the second day of instruction, the study's permission forms were collected. Students who were absent on the first day received permission forms, completed the applicable forms, and were instructed to return parental consent forms on the next (third) class day. The pretest was also administered to all students on the second instructional day. On the third day, the instructional protocol and culminating project (a diorama) were introduced. Students were shown a video of how to build a diorama, pictures of various themed dioramas, discussed materials, scale and creativity.

Students participated in two labs during the intervention. The first lab (Lab #1) focused on components and processes of the rock cycle (L_1; see Fig. 2). Lab #2 focused on L_2 of the learning progression: how the processes of the rock cycle affect the parts of the rock cycle.

3.21 Level One—The Parts And Processes Of The Rock Cycle

Level one lesson began with a two-minute video of the rock cycle and the path a rock might take through the cycle as it is transformed. Students were introduced to the major principles of the lessons, informed to look for these principles as they studied the rock cycle, and were shown a very simple concept map of the rock cycle. Students were also shown the recycling symbol and asked what the symbol indicated and where they had seen the symbol. The recycling symbol was the springboard for discussion about Earth systems' overlapping cycles and how matter is constantly recycled. Familiar cycles were discussed like day and night as well as seasons. The rock cycle was connected to the discussion about cycles.

Lab #1 focused on three types of rocks and emphasized the differences among them. Students determined rock types based on the rocks' physical characteristics. In a multimedia presentation, students saw various examples of each type of rock. Some sketched examples in their notes; others wrote the definitive characteristics. As the examples were discussed, students were given background about why the rock looked the way it did. Similar rocks were compared and contrasted to each other (e.g. sandstone and conglomerate).

At the end of the multimedia presentation, students were given a testable question: "What type of rock do I have?" Students observed unknown rocks, generated predictions based on their observations and pre-lab notes, and explained their thinking for their predictions. After approval of their predictions, students were given materials to collect qualitative data. Data included drawings of their assigned, unknown rocks, the type of rock they believed it was, and citing evidence from notes. At the end of data collection, the reflective conclusion writing frames were introduced. The teacher-researcher discussed each part of the reflective conclusion with students while students completed the writing frame.

3.22 Level Two—The Effect Of Processes On The Rock Cycle

Level two lesson focused on how rocks change. It started with a do-now/quick write question, “Where do rocks come from? Provide evidence to support your answer.” After the do-now question, students wrote background/pre-lab notes in their science notebooks. Students copied a flow chart that illustrated how the processes of the rock cycle changed. The five ways rocks changed were identified and listed with their representative picture from a more elaborate rock cycle concept map than in Lab #1. Each process was discussed in detail.

In Lab #2, students created and modeled sedimentary, metamorphic, and igneous rocks from crayon and other common materials. The testable question was given and students were to select three of the five processes discussed in the pre-lab. Students made predictions about how the processes would affect the crayons, explained their thinking with the prediction, and used pre-lab notes as the basis for predictions. Students were given the procedure in a handout. After pre-lab approval, students collected qualitative data, which consisted of before and after pictures for each “rock” type modeled from crayon and identifying the process that caused the rock changes. Students obtained the teacher-researcher’s signature to ensure all parts of the data were accurately recorded after each drawing. Once data collection was complete, students wrote reflective conclusions.

4.0 Rock Cycle Assessment: Instrument Development and Administration

The American Association for the Advancement of Science's (AAAS) Science Assessment and the National Center for Education Statistics' (NCES) National Assessment of Educational Progress (NAEP) Question Tool (NQT) released items were used to adapt a Rock Cycle Assessment instrument. Each test question data bank was harvested independently for potential items. Next, the questions were preliminary screened and organized separately.

4.1 AAAS and NQT Preliminary Item Screening

The teacher-researcher created a free account with AAAS science assessment to establish an item bank. The released items were relative to earth science and placed into one digital document. Three AAAS released-item topics were selected from the item banks: a) Plate Tectonics; b) Weathering, Erosion and Deposition (WED); and c) Weather and Climate II: Seasonal Differences (WCII). Questions were eliminated if they were outside the scope of the rock cycle. Remaining questions were re-numbered with the original AAAS released-item number. Approximately 50 questions qualified for inclusion on the Rock Cycle Assessment instrument from the AAAS item bank. The item code and performance details were recorded, organized into a table and analyzed for apparent trends.

Of the released items available from NQT, 43 were selected from the ESS domain and were put into a NQT item list on NAEP's website. The ESS released items were categorized into five topics:

- Using Science Principles
- Identifying Science Principles
- Using Scientific Inquiry

- Scientific Investigation
- Conceptual Understanding

Irrelevant questions were eliminated from the NQT potential item pool and eleven questions remained. NQT Question Identification Numbers, content classification, question type, and subject were recorded for each question. The difficulty levels—described as easy, medium, or hard—were given a quantitative equivocal rating (i.e. easy = 1; hard = 3). NQT released items were then sorted by grade level (4th or 8th) and then by topic in a table for comparative purposes.

4.2 Item Analyses and Instrument Construction

After preliminary screening, each released item was placed into a table for evaluation. In the first evaluation, each item code /question ID, questions with their respective answers, items' source, topic/description, percentage responding correctly (when applicable), and difficulty level (when applicable) were listed. Several items were eliminated during the first evaluation and an explanation was given for every item eliminated. Sixty questions were evaluated, 23 questions were eliminated, and 37 questions went on to a second evaluation.

During the second evaluation, the same parameters were listed as were in the first evaluation. Items and their definitive parameters were examined much closer. Inappropriate and repetitive items were eliminated. An explanation was given for each eliminated released item. Seven items were eliminated from the second evaluation. The remaining 30 items were then arranged according to their learning progression alignment, and this determined the items' Rock Cycle Assessment assigned number. The items were listed sequentially by learning progression level: a) L_1; b) L_2; c) L_3; and d) L_{2-3}. A test blueprint was constructed to ensure a balanced instrument (see Appendix A). Finally, each item was copied from its respective website and pasted onto a Word document for the final draft of the assessment instrument.

4.3 Instrument Validation And Reliability

The number of items selected for the instrument (and consequent percentage of test items) was contingent upon the number of levels of achievement, the learning progression's grain size, and the amount of relevant released items available. Test items were approximately equally distributed across three 8th grade levels of the learning progression in light of the limiting factors. The item source, percentage of students responding correctly, and difficulty level varied across the e Rock Cycle Assessment instrument.

Both national databases validated their respective released question items. The AAAS research team carefully validated the *Science Assessment* test questions. The released items measured students' conceptual understanding, tested for misconceptions/alternative ideas, and aligned the science ideas (AAAS, 2015). NAEP *NQT* item-development process used many steps to validate the test items including internal (i.e. NAEP) and external test specialists reviewing and revising the items, editorial and fairness reviews, a pilot test, and selection based on pilot test analysis (NAEP, 2009). Cronbach's alpha was used to determine internal consistency of the Rock Cycle Assessment instrument (30 items; α=. 315) and the instrument had low reliability.

4.4 Rock Cycle Assessment Pretest-Posttest Administration

The Rock Cycle Assessment pretest was administered the second day of instruction to students in first through third periods. Fourth period students were administered the pretest the first Friday of instruction in a likewise manner. The pretest was a 30-question, multiple-choice exam and it was not timed. Students had approximately 40 minutes to complete the pretest. Students were informed that there were no adverse consequences for their test score. They were instructed to ask for clarity if they did not know a word, to do the easy questions first, and to make intelligent guesses. The identical Rock Cycle Assessment posttest was administered the day before the last day of instruction to all students in a similar manner.

5.0 Data Collection, Management and Analysis

Rock Cycle Assessment pretest-posttest scores, student-generated symbolic media, and reflective conclusions were collected from students. Pretest-posttest data were collected from all students at the completion of each exam. In addition to descriptive statistics, the posttests were analyzed with a *t*-test for independent means. Student-generated symbolic media and reflective conclusions were collected from the Learning Progression group's science notebooks at the end of the study. They were analyzed with Wiebe et at. (2009) semiotic taxonomy and the constant comparative method, respectively.

5.1 Rock Cycle Assessment Pretest-Posttest Collection, Management and Analysis

The study utilized the Rock Cycle Assessment to gauge the learning progression's impact on students' science outcomes. The pretest-posttest data were collected after each administration of the instrument. Scores were first recorded manually and then entered into a spreadsheet. After all pretest-posttest exams were administered, posttest scores were analyzed with *t*-test for independent means.

5.2 Data Collection from Science Notebook

Symbolic media and reflective conclusions were collected from science notebooks. All science notebooks were passed out at the beginning of class and collected at the end of class. Entries were made in notebooks daily. At the end of class, students placed science notebooks in respective milk crates and were dismissed if their science notebooks were in the milk crate. The notebooks were stored in a restricted area of the classroom and remained in a locked classroom when the teacher-researcher was not in the classroom. They remained intact as a complete unit for the duration of the study.

5.21 Data Collection and Analysis of Symbolic Media

The study also tracked the distribution of student-generated symbolic media across, and within, the phases of the inquiry process. Symbolic media were harvested from the Learning Progression group's science notebooks and they were analyzed using Wiebe et al. (2009a; 2009b) semiotic taxonomy. After the study was completed, each page of every science notebook was labeled with students' corresponding identification number. The pages were given an entry number and a taxonomy analysis form was stapled to every page, labeled with the corresponding identification number and entry number. All notebook entries were kept in a data log: a large three ring binder.

There were six qualities for categorizing symbolic media and the context in which the symbolic media were generated. Each taxon's categorical descriptors were identified on the taxonomy form by marking an "X" for the respective investigative phase in which it was produced (i.e. pre, during, after, unknown). A tally sheet was generated; totals were summed for the six taxa and investigation phases. The frequency was determined for each investigative stage of the inquiry process. Descriptive statistics were calculated across, and between, the phases of the inquiry process. Frequency distribution was analyzed using a chi-square test.

5.22 Reflective Conclusions Collection, Management, and Analysis

Reflective conclusions were isolated from other notebook entries following semiotic analysis and were analyzed using the constant comparative method. The data unit was the reflective conclusion because “it is heuristic and the smallest piece of information interpretable in the absence of any additional information” (Merriam, 2009, p. 345).

Several measures were utilized to ensure validity and reliability as the constant comparative method was employed. Internal validity was established because of the multiple sources of raw data (i.e. the reflective conclusions). Students wrote their reflective conclusions during instructional time; however, some students required additional instructional time either during lunch, elective time, or an additional class day. Many reflective conclusions received an in-situ assessment. A student was given written or oral feedback while other students were working on their reflective conclusions. Others were given written feedback in their notebooks after instructional time had ended. As the informal assessment took place, the teacher-researcher observed and compared reflective conclusions to each other and recorded the observations in the teacher log. Raw data were constantly compared among, and between, both lab activities. After visual examination of all reflective conclusions, three codes emerged to describe students’ reflective conclusions:

- Satisfactory (S): correctly completed the writing frame;
- Needs improvement (NI): attempted to complete the writing frame;

- Incomplete (INC): only wrote the stem of the frame or did not write the stem of the frame.

Each frame was coded for both labs in a handwritten table, which included students' ID codes, data log entry number, and the teacher-researcher's random "self-notes." The handwritten chart was converted into a digital document. The code totals and percentages were calculated for each of the five frames for both labs. Codes were double checked against the original handwritten copy for accuracy. After corrections were made and double-checked, data were extrapolated according to codes and the percentage of each code was calculated. The data were then bar-graphed.

External validity was ensured through the use of rich, thick descriptions of the setting and participants of the study, the findings and sufficient evidence from notebook entries, the teacher-researcher's log, and the faculty-participants' classroom observations. The classroom, the school, and the teacher-researcher's observations were described in the teacher log. Students' personalities, struggles, limitations, and other characteristics were also described in the teacher-log. Student anonymity was, however, maintained. As a means of reliability, the teacher-participant generated an audit trail. The trail was recorded in the teacher log and in memos as data were examined.

5.23 Data Collection and Analysis of Reflective Conclusions

Each day after school, the teacher-researcher reviewed a few students' notebooks and gave written feedback. Memos were made of the most common trends in the notebooks and the teacher-researcher addressed those trends in subsequent instruction. Students' reflective conclusions were retrieved and managed the same way as the symbolic media (i.e. given identification number, given data log entry number, and placed in the data log). The reflective conclusions were isolated from the other symbolic media. The constant comparative method was employed to analyze reflective conclusions.

There were multiple sources of raw data examined at different times. Students wrote their reflective conclusions in class, during lunch, during elective time, and/or during an additional class day. Some reflective conclusions were assessed in-situ; others were given written feedback after instructional time had ended. The teacher-researcher observed and compared written reflections to each other and recorded those observations in the teacher log as well as generated “self-notes”/memos of patterns that emerged.

Visual examination was done to obtain axial codes. Each writing frame was a representative code because it highlighted a significant understanding in the inquiry process. The five axial codes were a) purpose of investigation; b) process of investigation; c) results of investigation; d) accuracy of investigation; and e) further investigation. After sorting students’ reflective conclusions into the five categories, it was observed that many writing frames were either complete or incomplete. Closer examination revealed that some of the “completed” writing frames were partially to completely erroneous. Nevertheless, students attempted to describe a particular portion of the inquiry experience. Each writing frame was coded for both labs in a handwritten table, and then converted into a digital document. Afterward, a detailed word-by-word content analysis was done for each category of the 16 reflective conclusions, and the result emerged: students were not fully able to define their science experience based on the overwhelming amount of writing frames that needed improvement or were incomplete.

6.0 Computer-Assisted Group

The computer-based rock cycle curriculum was found online and modified such that it only included content around the rock cycle. It called for students to research the three types of rocks and the rock cycle. Students wrote letters at the end of their individual research and used writing frames (different writing frames from the Learning Progression group) to help scaffold their writing process. Students had two options for completion: complete a hardcopy or a digital portfolio. Digital portfolios were overwhelmingly selected.

The Computer-assisted group researched how rocks were made. They investigated four aspects: a) rocks' composition; b) ways rocks were made; c) rocks' different physical characteristics; and d) ways rocks transform. The curriculum consisted of five partnered projects and one individual project. Students' products were stored in folders in a locked classroom and in their Google drives. Instructions and requirements were given at the start of each project. Students selected with whom they partnered and each student was responsible for submitting the requirements of the project.

Five projects required a scientific letter addressed to a fellow colleague. Students used the Internet and other sources to conduct their research. At the start of each project, students were given a list of websites to assist them; they also had the option to explore other relevant websites. Students were given a set of questions to answer; the answers helped students write their scientific letters. The grading rubric was also given for each letter. The rough draft letters were peer-edited, typed, and shared (via Google docs) with the teacher-researcher. Students used the “Eight Sentence Paragraph Structure”—a school-wide template that assisted students at every grade level to write consistent paragraphs. The sixth project had two parts: a) each student wrote about the journey through the rock cycle from the perspective of a rock; and b) each student created a comic strip about the experience.

Chapter 4: Results

The study reported examined students' science learning outcomes when a middle school science learning progression was implemented. The research question for the study was: What is the impact on students' science learning outcomes when a middle school science learning progression is developed and validated using science notebooks as part of an inquiry-based instructional intervention? In general, the study tested the hypothesis that students' learning outcomes would be greatly impacted when a learning progression instructional strategy was utilized.

Shepardson and Britsch (1997) instructional outline for using science journals provided the instructional framework for the intervention as well as the definitive parameters for delineating the instructional phases: a) pre-lab, b) during lab, and c) post-lab. In pre-lab, students explained existing ideas and understandings, described the purpose of the investigation, stated the testable question, and made hypotheses/predictions. During the lab, students recorded observations, organized data, and created drawings and diagrams. Post-lab, students answered the testable question using observations and data collected during the lab, explained their results using information from their notes, and reflected on their existing ideas in light of their findings.

Collected data helped to describe the impact. A *t*-test of independent means was performed to compare the means of the On/Off-campus and Learning progression groups' posttest in order to determine if the means were significantly different. The Rock Cycle Assessment scores were separated by the three groups (i.e. learning progression, on-campus, and off-campus) and put into a frequency table (see Appendix B). Constant comparative method was used to generate grounded theory of students' reflective conclusions. A Pearson's Chi-square was used to test how likely it was that the frequency distribution of graphic representations in students' science notebooks was due to chance.

1.0 Rock Cycle Posttest Results

As can be seen in Table 2, the Rock Cycle Assessment scores were an acceptable, normal distribution for the purpose of conducting a *t*-test. The On-campus group (*n*=6) had a mean posttest score of *M* = 11.5 (*SD* = 3.39) on a scale of 0-30 while the Off-campus group (*n*=10) had a mean posttest score of *M* = 13.8 (*SD* = 3.88). By comparison, the learning progression group (*n*=16) had a numerically smaller mean posttest score of *M* = 10.43 (*SD* = 3.52). To test the hypothesis that the On-campus/Off-campus groups and the learning progression group had statistically significant different mean posttest scores, an independent samples t-test was performed.

Table 2

Rock Cycle Assessment Scores' Descriptive Statistics

	n	*M* Pretest	*SD* Pretest	*M* Posttest	SD Posttest	Skew	Skew *SE*	Kurtosis	Kurtosis *SE*
On-Campus Group	6	10.67	2.42	11.5	3.39	-.462	.845	-2.07	1.741
Off-Campus Group	10	10.1	3.48	13.8	3.88	1.04	.77	0.71	1.6
Learning Progression Group	16	7.25	3.60	10.43	3.52	-.699	.564	.316	1.09

The On-campus group did not have a statistically significant different mean posttest score than that of the learning progression group, *t*(20)=. 636, *p* = .532. However, the Off-campus group did have a statistically significant different mean posttest score than the learning progression group, *t*(24) = 2.27, *p* = .0319.

2.0 Notebook Analyses

The notebooks analyses enabled a comprehensive examination of the changes that occurred in the notebook entries. Specifically, the reflective conclusions examined students' capacity to explain and make meaning of what they learned about the rock cycle in the context of inquiry. The semiotic taxonomy partially revealed how students experienced the learning progression curriculum while categorizing symbolic media and the context in which the symbolic were generated.

2.1 Science Notebook Results: Reflective Conclusions

Reflective conclusions were written in post-lab. Three concerns led to the research study: a) establishment of learning progression utility; b) phenomenological perspective of learning progression research; and c) facilitating learning progression research in low SES and non-mainstream learning environments. Consequently, in their reflective conclusions, students explored their inquiry experience by clarifying what they learned and they made meaning out of what they studied. Randall (1999) sentence starters were used as writing frames. This research study aimed to address the gap between the role science notebooks play in inquiry-based science and the development and validation of learning progressions. The study also aimed to establish a pattern for using a learning progression with science notebooks by a science classroom teacher.

Comparative and content analyses yielded three ratings to describe students' reflective conclusions: a) satisfactory; b) needs improvement; and c) incomplete.

2.11 Rated Satisfactory

There were 21 satisfactory codes across both labs. Writing Frame One focused on the purpose of the investigation and it dominated the satisfactory category (see Figure 4). Of the total 21 possible responses, 10 were categorized as being satisfactorily completed for identifying the purpose of the investigation (i.e. writing frame one). More students in the satisfactory category completed the purpose of the investigation than in any other rating. Students were overwhelmingly able to articulate the purpose of Lab #1 (e.g. entries #10, 61, 120; see Table 3).

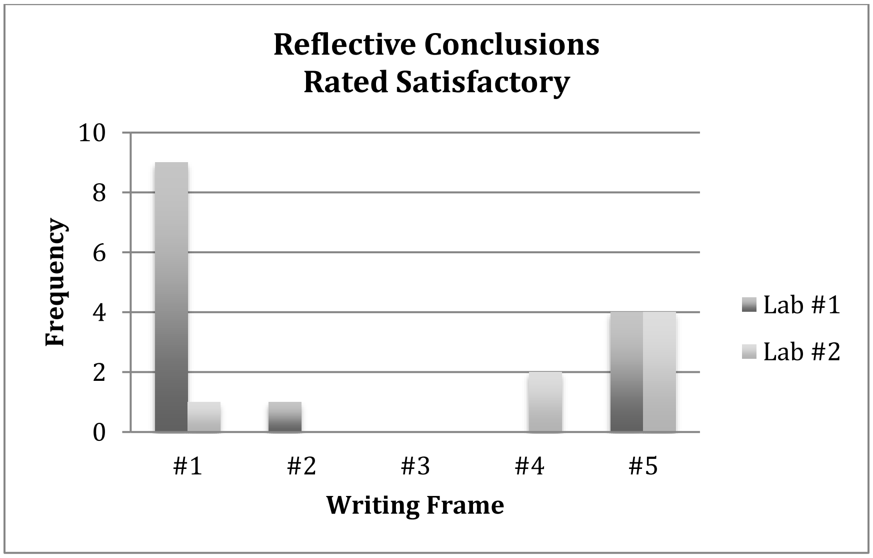

Figure 4. Frequency distribution of satisfactory rating for each writing frames of reflective conclusions. The bar graph compares the frequency distribution of satisfactory ratings for Labs #1 and #2.

However, it was found that more students did not use their own words to express the purpose of the investigation compared to the number of students who did. For example,

Table 3

Content Analysis Results of Writing Frames with Satisfactory Rating

Entry Number	Writing Frame Number and Content Observations
10	1. Didn’t use own words
31	1. Didn’t use own words 5. Expressed how the study could be extended; didn't explain why more time was needed.
61	1. Didn’t use own words
104	4. Stated results were accurate because they were reviewed by "expert" (teacher) for accuracy of results; selected if it was accurate. 5. Use different kind of material to get different results (change/manipulate variable)
120	1. Used own words 5. Identified how to extend the investigation; didn't explain why it was
160	1. Used own words
206	1. Didn’t use own words 2. Discussed the procedure(s) used to identify rocks 5. Used own words; proposed to change variables for further investigation
217	1. Mentioned processes used to change rocks a.k.a. crayons 5. Used own words; proposed to change variables for further investigation (change of wax)
256	1. Used own words; expressed the focus/topic of lab
277	5. Wording a little off; proposed to change variables for further investigation (change of wax)
288	1. Didn’t use own words
313	1. Used own words 5. Proposed to change variables for further investigation (different numbered rock[s])
319	4. Indicated accuracy of results; identified the specific methods taken to ensure accuracy of results 5. Proposed to change variables for further investigation (change of crayon for easier melting)

of all satisfactory ratings for Lab #1, entries #120, 160, 256, and 313 used their own words, which was approximately 45% of the ratings. No student used his or her own words for Lab #2. Of all satisfactory ratings for Lab #1, entries #10, 31, 61, 206, and 288 had the exact wording from the pre-lab notes to complete writing frame one. For Lab #2, entry #217—the only satisfactory rating for Writing Frame One—also used the exact wording from pre-lab notes. Writing Frame Two articulated the methods utilized to investigate the topic of study. One entry had a satisfactory rating for Writing Frame Two in Lab #1, entry #206. No entry had a satisfactory rating in Lab #2.

Writing Frame Three articulated the results of the investigation. It included using claims, evidence, and reasoning (C-E-R) to explain the results. No entry had a satisfactory rating for Writing Frame Three. Writing Frame Four articulated the accuracy of results. For Writing Frame Four to be rated satisfactory, students had to select "accurate or inaccurate" and explain why their results were either accurate or inaccurate. Writing Frame Four had no satisfactory rating in Lab #1. Entries #104 and 319 had a satisfactory rating in Lab #2. Writing Frame Five articulated further investigation of the studied phenomenon. It had eight of 21 (approximately 40%) possible satisfactory ratings. All students used their own words for Writing Frame Five. Students' responses articulated what would be done differently, or what would be changed, upon further investigation. For example, entry #31 specified a need for more time to carry out the investigation. Entry #104 expressed utilizing a different type of material to get different results. Entry #313 specified investigating a different set of unidentified rocks. One entry, #277, had "off wording" (e.g. grammatical errors, did not proof read prior to submitting the assignment).

2.12 Rated Needs Improvement

There were 47 total needs improvement ratings across both labs. Lab #1 had 26 entries while Lab #2 had 21 entries. Writing Frame One had eight needs improvement ratings. In Lab #1, there was one entry; Lab #2 had seven entries (see Figure 5). Wording was "off" (e.g. grammatical errors, missing words, incomplete thoughts) for many of the entries (e.g. #21, 64, 104, 133, and 319). It was found that students used their own words, partially explained the salient concepts, and neglected to mention the nuanced concepts that were essential for a fuller understanding (see Table 4). For example, entry #64 only identified the processes; no mention was made of how those processes transform the rocks, much less mentioned the rocks specifically.

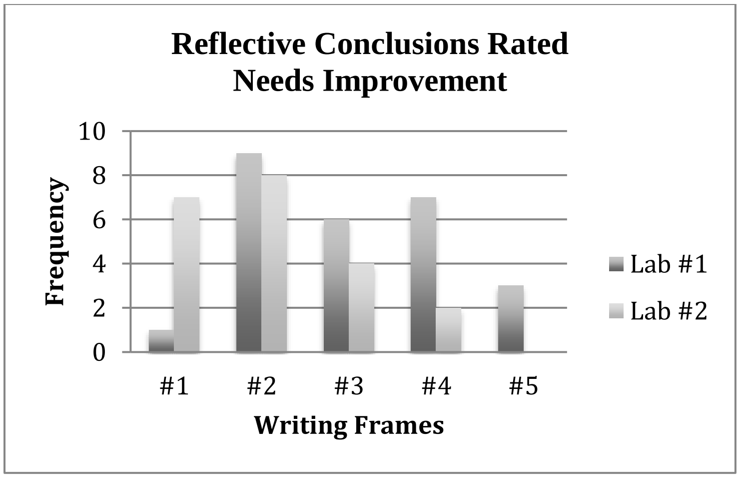

Figure 5. Frequency distribution of needs improvement rating for each writing frames of reflective conclusions. The bar graph compares the frequency distribution of needs improvement ratings for Labs #1 and #2.

Table 4

Content Analysis Results of Writing Frames with Needs Improvement Rating

Entry Number	Writing Frame Number and Content Observations
10	2. No summary/explanation of procedure
21	1. Wording in frame is "off"; components of rock cycle but not how they interact to form "new" rocks 2. Didn't describe the procedure (or summarize) tools used to investigate the problem/topic of the lab
31	2. Explained what they were supposed to do; did not summarize/ explain how to do it 3. Had claim; evidence & reasoning were missing; expressed that hypothesis was correct 4. Didn't articulate if results were accurate; attempted to explain that they were accurate (implied)
61	2. Attempted to complete the frame but did so incorrectly; the wording is "off"; didn't summarize/explain the procedure used to investigate the physical appearance of rocks 3. Claim present; evidence & reasoning missing; stated sedimentary rock, didn't mention metamorphic or igneous rocks; stated hypothesis was correct 4. Described methods of investigation to compare hardness to other rocks; stated results were accurate because of comparison to other rocks; no explicit indication of accuracy 5. Described checking additional sources but not clearly articulated
64	1. Wording is "off"; components of rock cycle but not how they interact to form "new" rocks 2. Attempted to complete the frame; ideas are not expressed clearly; circled word can't read
94	2. "It" = rocks; discusses one way the rocks were examined (texture); compared textures; didn't identify which rocks were compared 3. Wording is not clear; expression/articulation of ideas aren't clear; has claim but no evidence or reasoning; indicated the hypothesis was correct 4. Expressions not clear; how were data "checked"? (e.g. compared to the notes or to a neighbor's results); stated results were accurate because it was self-verified. 5. Very vague response; didn't describe what would be done procedurally to achieve different results; described what was not done; perhaps student is explaining what could be done overall to make the investigation easier?

104	1. Wording is "off"; components of rock cycle but not how they interact to form "new" rocks 2. Used vocab words; some unclear wording; describes only one part of the procedure; maybe didn't finish because of attendance or just didn't address the other processes of the rock cycle or maybe this is what was most memorable; doesn't summarize the entire process. 3. Used vocab & describes procedure; no C-E-R; wording unclear; discussed #2 succinctly as #3 (this is what also makes it incorrect); hypothesis as correct/incorrect not indicated
120	2. Didn’t summarize/explain the procedure used 3. Didn’t discuss all the results; didn't indicate if hypothesis was correct; claim is a sentence fragment, no E-R.
127	1. Identified two processes of the rock cycle; no mention of other processes or components of rock cycle. 2. Stated how to complete the frame, but didn't follow the explanation; directions to complete the frame correctly
133	1. Incomplete idea expression (types of rocks); but attempted to complete frame 2. Attempted to complete (but incorrect); no explanation of process to complete procedure 3. Response unclear, can't tell if the rocks were numbered #1-3 or if the rocks are just listed; didn't express if hypothesis was correct; no CER 4. Indicated results are accurate; didn't indicate the measures taken to make sure they were good; attempted to complete the frame.
160	2. Didn’t express the procedure; the response is a restatement of #1 4. No indication of accuracy of results; attempted to explain measures taken to ensure good data
192	1. Articulated the process but not its effects; sentence fragment 3. Indicated the hypothesis was correct; no CER
206	3. Claim, but no evidence or reasoning; expressed the hypothesis as correct 4. Didn’t indicate if results are accurate; did discuss measures taken to ensure accurate data
217	2. Attempted to summarize the procedure; discussed/summarized the igneous formation parts of weathering /erosions 4. No explicit indication of accurate result—implied results are accurate because there's an explanation of what was done to make sure results were accurate
256	2. Discussed procedure for one type of rock; not the other two

277	1. Mentions two process but not the effect on rocks; didn't use own words 2. Wording isn't clear; attempted to summarize 2/3 processes with the model 3. No CER; stated hypothesis was correct; restated process, didn't articulate the effect of the processes in the type of rock generated 4. No explicit articulation of accurate results; it's implied because student states what was done to make sure good data was obtained; no explicit mention of what was checked in the notes.
288	2. Stated what was done, but not how it was done—partial explanation 5. Completed with erroneous & unintelligible info
299	2. Stated what should have been done, but didn't summarize how it was done; mixed model; sedimentary, igneous, & metamorphic rock (not crayons)
313	4. Did not indicate if results were accurate or inaccurate; it is implied (states' s/he tried his/her best)
319	1. Incomplete thought; attempted to complete frame with 2 of 5 processes 2. Summarized 2 of 5 processes (erosion/weathering & melting) for rock formation 3. Stated correct hypothesis; no CER

Entries #21, 104, 192, 277, and 319 followed suit with that of entry #64. Entry #127 identified two of the processes (the requirement was three of the five processes) but did not describe the processes' effects on the components of the rocks (i.e. crayons). Also, students did not explicitly articulate that crayons were representative of the three types of rocks. Writing Frame Two had a total 17 entries. Students' responses to Writing Frame Two reflected a diversity of misunderstanding. It was found that many did not describe and/or summarize the process used to complete the investigation. For example, entries #21, 133, 160, and 192 provided completely irrelevant responses to Writing Frame Two. Entry #133 was unintelligible. Entries #104, 217, 256, and 319 explained one part of the procedure, as opposed to the entire process. Entries #127 and 299 explained how to complete Writing Frame Two, but did not follow the explanation written to satisfactorily complete Writing Frame Two. Entry #277 had the most comprehensive response to Writing Frame Two. The response described a majority of the procedure, but the wording and grammar were so poor that it was difficult to properly interpret. Writing Frame Three had a total of 10 entries rated needs improvement. A diversity of misunderstanding in fundamental concepts persisted; claims were made but omitted evidence and reasoning, and some entries omitted claims, evidence, and reasoning altogether. It was also found that many students indicated correct hypotheses or omitted reference to the hypothesis. Furthermore, entries #104 and 207 had irrelevant responses for Writing Frame Three. Writing Frame Four had nine entries. The entries had very vague explanations of the accuracy of results. Students' explanations did not indicate (by circling) if the results were accurate; entries #160, 206, 217, 277, and 313 implied accuracy. Some responses included following directions, checking notes, or trying their best. Entries #94 and #133 explicitly expressed accuracy, but the measures taken to do so were not indicated. Writing Frame Five had the least number of entries. Across both labs, three entries were rated needs improvement. There was no general description of what could be done, procedurally, to achieve different results/outcomes of the investigation. Entry #61 responded, "look at more definitions." Entry #94 had an equally vague response. To further investigate the problem, the student responded "to listen." Entry #288 completed the frame, however, the student's response was unintelligible and erroneous.

2.13 Rated Incomplete

There were 36 total incomplete ratings across both labs (see Figure 6). Writing Frames One and Two each had three total entries. Writing Frame Three had the most incomplete ratings with 11 total entries. Writing Frames Four and Five had 10 total entries each.

Based on the overwhelming amount of writing frames needing improvement or that were incomplete, it appears that students were not able to fully explain their science experience. Based on verbal class feedback, it was found that many students were unsure and not confident participating in inquiry; they were uncomfortable with using the language of inquiry and became frustrated while writing reflective conclusions because they had little to no command of inquiry language or process. At best, students were able to partially explain their inquiry experience.

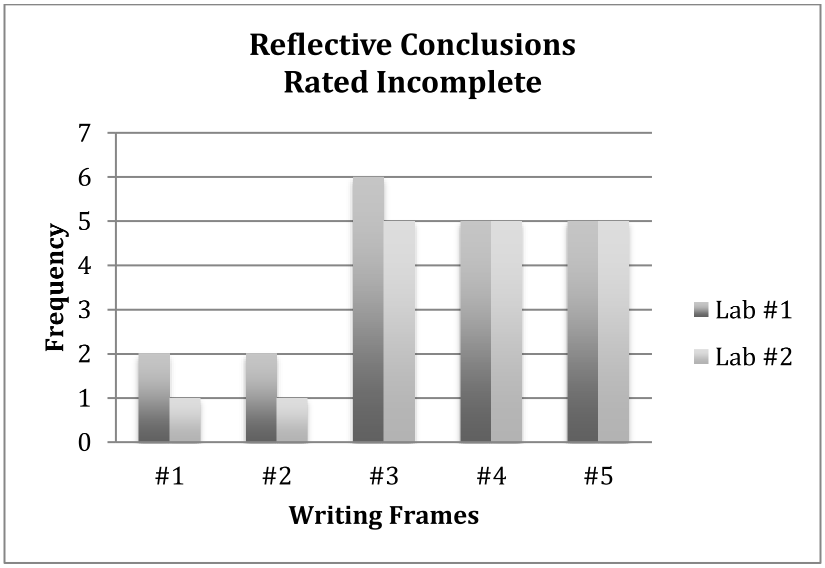

Figure 6. Frequency distribution of incomplete ratings for each writing frames of reflective conclusions. The bar graph compares the frequency distribution of incomplete ratings for Labs #1 and #2.

3.0 Science Notebook Results: Semiotic Taxonomy Analysis

Sixteen notebooks were disassembled. There were 326 pages retrieved from notebooks. The contents were given an entry number, a student identification code, and analyzed according to Wiebe et al. (2009a; 2009b) semiotic taxonomy. A chi-square test was calculated to compare the frequency distributions of the six categorical qualities of graphic representations in notebooks across three phases of inquiry.

3.1 Text Graphic Relationship

Table 5 shows the frequency distribution and percentages of the text-graphic representation across the phases of inquiry. Across the phases, sub-categorical entries were “balanced” and “drawing driven” during lab. “Text-driven” entries were mostly distributed in the pre-lab phase. The majority of entries across the phases were text-driven.

Within respective phases, there was relatively little distribution of text-graphic entries post-lab (see Table 6). However, the pre-lab had the most text-graphic entries and 98.5% of the text-graphic relationships were text-driven; the remaining entries were drawing-driven. During lab, this trend essentially reversed. The majority of text-graphic entries were drawing-driven (70%) while small portions of text-graphic entries were text-driven or unknown (see Table 6). A significantly strong association was found between the three phases of inquiry and the text-graphic relationship ($\chi^2_{(6)} = 140.68$, $\phi = 0.682$, $p < 0.05$).

Table 5

Frequency Distribution and Percentage of Semiotic Notebook Codes: Text-Graphic Across Phases

Text Graphic Relationship	Pre-Lab	During-Lab	Post-Lab	Total
Drawing-driven	4% (1)	96% (28)	0% (0)	100% (29)
Text-driven	73% (65)	1% (1)	26% (23)	100% (89)
Balanced	0% (0)	100% (10)	0% (0)	100% (10)
Unknown	0% (0)	100% (1)	0% (0)	100% (1)

Table 6

Frequency Distribution and Percentage of Semiotic Notebook Codes: Text-Graphic Within Phases

Text Graphic Relationship	Pre-Lab	During-Lab	Post-Lab
Drawing-driven	1.5% (1)	70% (28)	0% (0)
Text-driven	98.5% (65)	2.5% (1)	100% (23)
Balanced	0% (0)	25% (10)	0% (0)
Unknown	0% (0)	2.5% (1)	0% (0)
Total	100% (66)	100% (40)	100% (23)

3.2 Spatial Organization

Table 7 displays the frequency distribution and percentages of the spatial organization relationship across the inquiry phases. Across the phases, there was equal distribution in the “1-dimentional” subcategory while all “2 or more” and most “unknown” were distributed in pre-lab (see Table 7). The spatial organization was unknown for most entries, while very few entries displayed “1-dimension” spatial organization.

Table 7

Frequency Distribution and Percentage of Semiotic Notebook Codes: Spatial Organization Across Phases

Spatial Organization	Pre-Lab	During-Lab	Post-Lab	Total
1-dimension	33.3% (1)	33.3% (1)	33.3% (1)	100% (3)
2 or more dimensions	100% (16)	0% (0)	0% (0)	100% (16)
Unknown	60% (59)	20% (20)	20% (20)	100% (99)

Within each respective phase, the spatial organization was largely unknown (see Table 8). Pre-lab had the majority of spatial organization entries. These entries were mostly two or more dimensions. “One-dimensional” spatial organization accounted for only 5% of during and post-lab entries. A significantly moderate association was found between the three phases of inquiry and spatial organization ($\chi^2_{(4)} = 16.25$, $\phi = 0.217$, $p < 0.05$).

Table 8

Frequency Distribution and Percentage of Semiotic Notebook Codes: Spatial Organization Within Phases

Spatial Organization	Pre-Lab	During-Lab	Post-Lab
1-dimension	1.4% (1)	5% (1)	5% (1)
2 or more dimension	21% (16)	0% (0)	0% (0)
Unknown	77.6% (59)	95% (20)	95% (20)
Total	100% (76)	100% (21)	100% (21)

3.3 Scale Representation

Table 9 displays the frequency distribution and percentages of the scale representation relationship. Across the phases, all macro, macro-micro, and super macro frequencies were only distributed during lab. No entries displayed macro-molecular, micro, or molecular level scale. For most entries, the scale was unknown and there was uneven distribution with the most unknown scale occurring in the pre-lab (≅ 74%; see Table 9). For the entries that could be categorized by their scale, most were at the macro or macro-micro level.

Table 9

Frequency Distribution and Percentage of Semiotic Notebook Codes: Scale Representation Across Phases

Scale Representation	Pre-Lab	During-Lab	Post-Lab	Total
Macro	0% (0)	100% (36)	0% (0)	100% (36)
Macro-micro	0% (0)	100% (28)	0% (0)	100% (28)
Macro-molecular	0% (0)	0% (0)	0% (0)	0% (0)
Micro	0% (0)	0% (0)	0% (0)	0% (0)
Molecular	0% (0)	0% (0)	0% (0)	0% (0)
Super-macro	0% (0)	100% (1)	0% (0)	100% (1)
Unknown	73.5% (61)	1.2 % (1)	25.3% (21)	100% (83)

Within the phases, nearly all “unknown” scale frequency was in the pre-lab and post-lab (see Table 10). The majority of scale representation entries were in pre-lab and during lab. However, a particular scale could only be identified during lab. A significantly strong association was found between the three phases of inquiry and scale representation ($\chi^2_{(6)} = 192.61$, $\phi = 0.697$, $p < 0.05$).

Table 10

Frequency Distribution and Percentage of Semiotic Notebook Codes: Scale Representation Within Phases

Scale Representation	Pre-Lab	During-Lab	Post-Lab
Macro	0% (0)	54.5% (36)	0% (0)
Macro-micro	0% (0)	42.4% (28)	0% (0)
Macro-molecular	0% (0)	0% (0)	0% (0)

Micro	0%	0%	0%
	(0)	(0)	(0)
Molecular	0%	0%	0%
	(0)	(0)	(0)
Super-macro	0%	1.5%	0%
	(0)	(1)	(0)
Unknown	100%	1.5%	100%
	(61)	(1)	(21)
Total	100%	100%	100%
	(61)	(66)	(21)

3.4 Drawing's Temporal Representation

Table 11 displays the frequency distribution and percentages of drawing's temporal representation. Across all phases, there was uneven distribution demonstrated on the temporal scale (see Table 11). Temporal representation was mostly not applicable. Comparatively, real-time, slower than real time, faster than real time, and unknown were all under-represented across the phases.

Table 11

Frequency Distribution and Percentage of Semiotic Notebook Codes: Drawing's Temporal Representation Across Phases

Drawing's Temporal Representation	Pre-Lab	During-Lab	Post-Lab	Total
Real-time	0%	100%	0%	100%
	(0)	(7)	(0)	(7)
Slower than real time	0%	100%	0%	100%
	(0)	(3)	(0)	(3)
Faster than real time	0%	100%	0%	100%
	(0)	(2)	(0)	(2)
Not applicable	56.5%	24%	19.5%	100%
	(61)	(26)	(21)	(108)
Unknown	50%	50%	0%	100%
	(1)	(1)	(0)	(2)

Within the phases, all “real-time”, “slower than real time,” and “faster than real time,” frequencies were distributed during lab; these same sub-categories had no distribution in pre-lab and post-lab (see Table 12). Pre-lab had the most entries while the post-lab had the least amount. Examination of each phase revealed the majority or all frequencies were distributed in the “not applicable” sub-category (i.e. 98%, 66%, and 100% for each respective investigation phase). A significantly moderate association was found between the three phases of inquiry and drawings’ temporal representation ($\chi^2_{(8)} = 31.49$, $\phi = 0.346$, $p < 0.05$).

Table 12

Frequency Distribution and Percentage of Semiotic Notebook Codes: Drawing’s Temporal Representation Within Phases

Drawing’s Temporal Representation	Pre-Lab	During-Lab	Post-Lab
Real-time	0% (0)	18% (7)	0% (0)
Slower than real time	0% (0)	7.7% (3)	0% (0)
Faster than real time	0% (0)	5% (2)	0% (0)
Not applicable	98% (61)	66% (26)	100% (21)
Unknown	2% (1)	3.3% (1)	0% (0)
Total	100% (62)	100% (39)	100% (21)

3.5 Re-representation

Table 13 displays the frequency distribution and percentages of re-representations across the inquiry phases. There was uneven frequency distribution across all the phases for each sub-category. The “no” sub-category had the most entries across the phases while the “unknown” sub-category had the least (see Table 13). “Yes” and “no” subcategories had their greatest distribution in the pre-lab phase and their least distribution in the post-lab.

Table 13

Frequency Distribution and Percentage of Semiotic Notebook Codes: Re-Representation Across Phases

Re-Representation	Pre-Lab	During-Lab	Post-Lab	Total
Yes	72.4% (55)	0% (0)	27.6% (21)	100% (76)
No	45.3% (48)	33.2% (37)	21.5% (21)	100% (106)
Unknown	50% (1)	50% (1)	0% (0)	100% (2)

There was uneven distribution within the phases (see Table 14). Most re-representations were generated in pre-lab; virtually no re-representation was unknown. During-lab, "no" was 97% of the frequencies distributed, indicating that the drawings constructed during lab were almost all student-generated. During-lab also had the least frequencies distributed. Post-lab, none of the re-representations were unknown. A significantly moderate association was found between the three phases of inquiry and re-representation ($\chi^2_{(4)} = 48.34$, $\phi = 0.306$, $p< 0.05$).

Table 14

Frequency Distribution and Percentage of Semiotic Notebook Codes: Re-Representation Within Phases

Re-Representation	Pre-Lab	During-Lab	Post-Lab
Yes	53% (55)	0% (0)	50% (21)
No	46% (48)	97% (37)	50% (21)
Unknown	1% (1)	3% (1)	0% (0)
Total	100% (104)	100% (38)	100% (42)

3.6 Driving Force of Notebook Entries

Table 15 displays the frequency distribution and percentages of the driving force of notebook entries. Across the investigation phases, notebook entries were mostly student-driven or teacher-student driven (see Table 15). Teacher-driven entries were concentrated in pre-lab (90%) while student-driven entries were primarily during lab (70%). Furthermore, there were no teacher-driven, teacher-student driven, or unknown entries during lab.

Table 15

Frequency Distribution and Percentage of Semiotic Notebook Codes: Driving Force of Notebook Entries Across Phases

Driving Force of Notebook Entries	Pre-Lab	During-Lab	Post-Lab	Total
Teacher-driven	90% (18)	0% (0)	10% (2)	100% (20)
Student-driven	9.2% (5)	70% (37)	20.8% (11)	100% (53)
Teacher-Student driven	84.6% (44)	0% (0)	15.4% (8)	100% (52)
Unknown	0% (0)	0% (0)	0% (0)	100% (0)

Within each phase, pre-lab had the greatest distribution while post-lab had the least. Pre-lab had 65.6% and post-lab had 38% of the frequency distributed in "teacher-student driven" sub-category. During-lab had 100% and post-lab had 52% of the frequencies distributed as student-driven (see Table 16). A significantly strong association was found between the three phases of inquiry and the driving force of notebook entries ($\chi^2_{(4)} = 84.93$, $\phi = 0.583$, $p < 0.05$).

Table 16

Frequency Distribution and Percentage of Semiotic Notebook Codes: Driving Force of Notebook Entries Within Phases

Driving Force of Notebook Entries	Pre-Lab	During-Lab	Post-Lab
Teacher-driven	26.8% (18)	0% (0)	10% (2)
Student-driven	7.6% (5)	100% (37)	52% (11)
Teacher-Student driven	65.6% (44)	0% (0)	38% (8)
Unknown	0% (0)	0% (0)	0% (0)
Total	100% (67)	100% (37)	100% (21)

Chapter 5: Conclusions and Recommendations

This chapter presents a summary of the study and important conclusions drawn from the data presented in Chapter Four. It provides a discussion of major findings, implications for practice, limitations, and conclusions.

Learning progressions are systematic and conjectural models that are research-based descriptions of students' thinking and/or learning of a scientific concept and/or skill. There was a general consensus in the research field around a few fundamental features and the role learning progressions play as a means to transform science education. At the same time, there was much ambiguity in several arenas. Nevertheless, research demonstrated that learning progressions generally improved students' science learning outcomes.

The writings and drawings in students' science notebooks portrayed and reflected how they practice inquiry within the science classroom (Aschbacher & Alonzo, 2006). Science notebooks contained "curricular evidence" that was a critical aspect of science teaching and learning (Baxter, Bass, & Glasser, 2000). Teachers used science notebooks as a tool for teaching, learning, and assessment within the confines of inquiry (Baxter et al., 2000). Many studies also demonstrated science notebooks to be beneficial to student science achievement (Huerta, Irby, Lara-Alecio, & Tong , 2015; Klentschy & de la Torre, 2004).

1.0 Summary of the Study

Learning progressions were developed and validated in a variety of ways. However, using learning progression with science notebooks was one method that had not been researched explicitly. The pedagogical importance and uses of science notebooks have been heavily researched (e.g. Klentschy, 2005). While learning progression research results, overall, demonstrated positive student outcomes, no learning progression to date explicitly employed science notebooks as the cornerstone to the development and/or validation process. Many learning progressions were diagnostic; very few learning progressions were progressive—fostering students' conceptual change toward a scientific level of understanding. As such, the phenomenological perspective was rarely examined. Furthermore, learning progression research was rarely conducted in urban schools and, therefore, the various complicated and fragile nuances that strain urban students, teachers, and schools have largely been ignored in learning progression research. Consequently, three concerns led to the research study: a) establishment of learning progression utility; b) phenomenological perspective of learning progression research; and c) facilitating learning progression research in low SES and non-mainstream learning environments.

The purpose of the study was twofold: a) to develop a middle school science learning progression validated in the context of inquiry by using science notebooks; and b) to study the impact of the notebook-based learning progression on middle school students' learning. The study sought to answer the question: what is the impact on students' science learning outcomes when a middle school science learning progression is developed and validated using science notebooks as part of an inquiry-based instructional intervention?

Situated cognition was the theoretical framework, Bruner's spiral curriculum was the curricular framework, and Shepardson's and Britsch's (1997) instructional outline provided the definitive parameters for the intervention. The study utilized a causal-comparative approach. First, the learning progression was developed using the Systems Thinking Hierarchy (STH) model. A three-week intervention was administered to 22 rising 8th grade students in which the computer-assisted curriculum or the learning progression was facilitated. The Rock Cycle Assessment was administered as a pretest and posttest. Data collected consisted of: a) Rock Cycle Assessment pretest-posttest scores; b) symbolic media from notebooks; and c) reflective conclusions from notebooks. Data were analyzed with a *t*-test for independent means, semiotic taxonomy, and constant comparative analysis.

2.0 Impact of Learning Progression on Students' Rock Cycle Learning

The learning progression group had a smaller average pretest score (*n*=16, *M*= 7.25, SD =3.6) and posttest score (*n*=16, *M*= 10.43, *SD* = 3.52) than that of the On-campus group (*n*=6, *M*=10.67, SD=2.42; *M*=11.5, *SD*=3.39). The trend was the same for the Off-campus group average pretest score (*n*= 10, *M* = 10.1, *SD* = 3.48) and average posttest score (*n*= 10, *M* = 13.8, *SD* = 3.88). To determine if the average difference between the On-campus/Off-campus and learning progression groups was statistically significant, an independent sample *t*-test was performed. There was no statistically significant difference between the On-campus group and learning progression group. However, there was a statistically significant difference between the Off-campus group and the learning progression group. Therefore because of unforeseeable selection bias, the researcher failed to reject the null hypothesis.

The results run counter to the general trend in learning progression research literature. For example, Plummer and Maynard's (2014) study explored how student learning of the seasons was impacted by classroom instruction that incorporated a learning progression. Thirty-eight 8th grade students participated in a 10-day curriculum. The authors administered a 13-question pretest three weeks prior to instruction, administered the posttest one week after instruction, and incorporated open-ended items in the data collection. Three findings materialized: a) 29 students improved, six stayed at the same level, and three regressed; b) posttest mean score was much higher than pretest mean score; and c) the difference between the mean pretest-posttest scores was significantly different.

Several possible explanations exist as to why the results ran counter to Plummer and Maynard (2014). Primarily, the ethnic/racial and economic demographics and methodology differed. Demographically, the Plummer and Maynard (2014) sample was 94% white and approximately 50% of the school population was low SES. This is in contrast to the author's study where the sample was approximately 90% black had low SES. The sociological challenges associated with urban schools, teachers, and students are well documented (e.g. Barton, 2007; Lee and Luykx, 2007). In the author's study, challenges such as attendance and transiency influenced the amount of data able to be collected. The off-campus program also impacted the results of the study. Effective urban schools build relationships with resources. In the case of the author's study, the resource was the off-campus program at a local university. Students were selected to participate based on their academics, attendance, behavior, and citizenship. Essentially, these students were ambassadors for the school. So while the school made strides to be a high achieving urban school, the other side of the school's efforts contributed to predisposing the sample to students who were not necessarily ambassadors.

Methodologically, Plummer's and Maynard's data collection, instrument design, and data analysis were much more sophisticated and in-depth. The methodology used by Plummer and Maynard differed from that of the author's study, which had a longer assessment instrument, lacked variety in the types of assessment questions, and used a different source for assessment items. In contrast to methodology of the author's study, Plummer and Maynard had six multiple-choice and seven open-ended questions on their assessment. The selected assessment items were from *Reason for Seasons*, the SCALE-uP project (a previously developed in-depth assessment of the seasons) and teacher generated. Plummer and Maynard's learning progression development was grounded in a metric approach, whereas the author's learning progression was grounded in a theoretical approach. For example, Plummer and Maynard developed and revised a seasons construct map using the construct modeling approach for learning progression development. Construct modeling precipitated from assessment-based learning progression research. The author developed and validated her learning progression from science education learning progression research and systems-thinking research. The Construct modeling used by the Plummer and Maynard was a four-step cycle of measurement that began with the researchers making observations to determine the subjects' understanding of the construct, inferring the respondents' level of the construct by categorizing, and scoring the responses to rank student responses according to their scientific accuracy. Finally, an interpretational model (a Rasch analysis and Wright map) was applied—this was a process by which the researcher compared results from the assessment to the hypothetical construct map. The author did not utilize metric-based methods to develop her construct map.

The cross-sectional study by Songer et al. (2009) aligned with the author's study in terms of sample ethnic/racial and economic demographic, plagues of urban school districts, and experimental design. However, Songer et al. findings differed from the author's findings. Songer et al. described an empirically driven, five-step process to develop a three-year learning progression focused on complex thinking about biodiversity. They sampled approximately 1800 Detroit Public School 4th through 6th graders. The control group engaged in the district-approved textbook-based curricular program while the experimental group received the learning progression intervention. Both groups participated for eight weeks, and the pretest and the posttest were administered. Songer et al. used multiple imputations for missing data and four findings emerged: a) posttest scores were descriptively higher than pretest scores; b) empirically, target domain achievement was substantially higher for intervention students; c) standardized measures were significantly better for intervention students; and d) intervention students gained 0.34 *SD* more on average.

Several explanations accounted for the author's results being counter to those of Songer et al. (2009). Foremost, the author conflated content and skill in the development-validation process, whereas Songer et al. distinguished content and skill. The author produced one content learning progression developed in conjunction with its validation process; validation included students' using the skill of explanation in the form of reflective conclusions. Songer et al. developed two preliminary learning progressions. One learning progression emphasized content (i.e. biodiversity) and the other emphasized skill (i.e. complex reasoning, specifically written explanations). Validation consisted of an identical pretest-posttest, which had a total of 23 items. This fundamental difference between the two studies could help explain why students' reflective conclusions—particularly Writing Frame Three—was very challenging. Secondly, the validation instruments were very different for both studies. The author used a simple 30-question multiple-choice adapted instrument from national databases. Songer et al. used an instrument with 16 multiple choice/fill-in-the-blank items and seven open ended explanation items. Six items on the instrument were from released standardized tests (two multiple choice items) from the Michigan Educational Assessment Program and four items (two multiple choice, two open-ended explanations items) from the National Assessment of Educational Progress, and the remaining 17 items written and pilot tested for the curriculum by the research team. This variety in Songer et al. instrument permitted greater variety in their analyses (e.g. HLM, growth model) and, therefore, a more in-depth explanation of students' progression and pretest-posttest scores. Third, Songer et al. development-validation process was complex, time-consuming, and outside the scope of this researcher's capacity and ability. For example, the authors communicated closely with expert scientists' in determining the focal points for learning progression development. For seven years, the research project worked with zoologists to transform scientific resources (e.g. Animal Diversity Web) designed for an adult audience into resources (e.g. Critter Catalog) that support inquiry questioning and explanation-building for elementary students. First drafts arose from these conversations. The author did not have access to such a resource. Last, the author's study limitations were not circumstantial constraints for Songer et al. For example, Songer et al. had the capacity to implement a bias-free quasi-experimental design in a much larger district: a) with fewer time constraints; b) a much larger sample and therefore more teachers in more schools to implement the progressions; and d) access to sophisticated statistical methods. The aforementioned were limitations to the author's study.

2.1 Contribution of Rock Cycle Learning Progression and Intervention

While results were unexpected, they are no less substantial. On the Rock Cycle Assessment, the learning progression group's difference between the means (*MD*=3.18) was higher than that of the On-campus group (*MD* = 0.83), even though—numerically—the On-campus group mean posttest score (*M*=11.5) was higher than that of the learning progression group's (*M*=10.43). The Off-campus group's difference between the means (*MD*=3.7), pretest average score (M = 10.1), and average posttest score (M =13.8) were all higher than that of the learning progression group. And, although the Rock Cycle Assessment scores did not yield statistically significant results for the On-campus and Learning Progression groups, the notebook analysis revealed two things: a) there was a statistical relationship between every type of graphic representation in students' notebooks and the phase in which the graphic representation was generated and b) students' explanatory skills needed to be explicitly developed.

Tangentially, the Rock Cycle learning progression was a product of this study. No prior learning progression existed which focused on rock cycle learning. Furthermore, none of the systems-based learning progressions utilized the STH model. Using the STH model provided a cohesive and systematic framework hierarchally arranged to clearly delineate the learning goals that foster systems-based thinking. A learning progression for 6th, 7th, and 8th grades was developed an eight-level rock cycle during the study.

As part of the intervention, notebook function was multi-faceted, which is in harmony with science teaching, learning, and assessment with notebooks research (e.g. Aschbacher & Alonzo, 2006; Madden & Wiebe, 2013; Ruiz-Primo et al., 2002, Baxter et al., 2000). The notebooks helped support the learning progression intervention, tracked and organized students' progress over the course of the intervention, and revealed how the experimental group practiced inquiry. They also provided a fractional account of students' experience in the learning progression and contained students' understanding of inquiry activities.

Semiotic taxonomy was used to analyze students' notebooks and it gave insight into students' learning progression experience. Symbolic media's categorical qualities were examined across the prelab, during lab, and post lab, in addition to between each group (e.g. prelab only). Students were the driving force for notebook entries during lab. In contrast, prelab was primarily teacher-driven. By post lab, most entries were either student or teacher-student driven.

A possible explanation is that students needed more assistance navigating their understanding in the beginning of the lab, because the teacher created the "situation;" this phenomenon was a function of the study's theoretical framework. As students transitioned to data collection (i.e. during lab), they needed little assistance from the teacher because they were able to refer back to their prelab notes and drawings. Post lab, activities/entries were writing-based and completely independent, although some students required assistance. The text-graphic relationship results further support this explanation. Prelab entries were overwhelming text-driven (73% of all text-driven entries were in this phase). During lab, entries were balanced (i.e. text and drawing) or drawing-driven. In post lab, the text-graphic relationships were all text-driven.

Wiebe et al. (2009b) define spatial organization as the dimensionality of a drawing. When text was organized in the prelab, it was overwhelmingly in two or more dimensions (e.g. concept map, table). Wiebe et al. also define re-representation as distinguishing if entries were copies of symbolic media (i.e. “yes”) or original generations of such (i.e. “no”). In prelab, re-representation was an approximate balance of yes and no. In other words, there was a balance of student-copied and student-generated symbolic media. Post lab displayed the same trend. However, during lab, re-representations were nearly non-existent, and, therefore, entries were almost all student-generated. This indicates data collection was primarily the only place in the investigative process where students generated drawings. Specifically, students generated symbolic media of various temporal and scale representations during lab. Outside of this, students copied symbolic media in their notebooks.

Some categorical descriptors were semiotically specific. For example, a drawing’s temporal representation referred to the amount of time change occurs. Wiebe et al. (2009) qualifies real time as change seen with the unaided eye in less than one hour. Temporally, “change” occurred during lab as students collected data; it was reflected in collected data. However, temporal representation was not applicable to any text entries, as alphanumeric characters do not display “change.” Scale representation was another semiotically specific categorical descriptor and it referred to a representative drawing that can be seen with the unaided eye in a single view. It was the baseline for macro scale drawings. Across phases, all entries were either at the super-macro, macro, or macro-micro scale and they were generated during lab.

Prelab entries were overwhelmingly text driven, re-representations and teacher-student driven. This indicated an overwhelming amount of symbolic media that was not student-generated in the prelab; rather, symbolic media were student-copied. For this result, one possible explanation is that students were encultrated, which was a function of the study's theoretical framework. The teacher gave students the necessary conceptual knowledge to navigate the upcoming authentic activity to occur during lab. Also, writing frames were used, for example, to assists students' formation of predictions/hypotheses. Spatial organization, scale representation, and temporal representation were not applicable as these categorical qualities were attributed to drawings and not to text.

During lab, notebook entries were largely student-driven with a smaller portion of entries being a balance of text and drawing. Entries were student-generated (i.e. not re-representations) and drawing-driven at the macro-scale. The spatial organization was unable to be determined while the temporal representation was not applicable for the majority of symbolic media. When applicable, the drawing's temporal representation was in real time (i.e. under one hour). This suggests students observed a change in variables during lab. Student-generated drawings were concrete, macro scale items—things that could be seen with the unaided eye in a single view.

Symbolic media were produced the least post lab. Text-graphic, scale representation, temporal representation, and notebooks' driving force all had the smallest frequencies distributed in the post lab; spatial organization and scale representation were unknown. Low frequency suggests many students did not make notebook entries post lab.

It also suggests that many reflective conclusions were incomplete, there were relatively few drawings generated and/or copied, and that drawings were not emphasized post lab. Waning attendance/attrition also could have contributed to the relatively few symbolic media post lab. This result aligned with the higher frequency of text-driven symbolic media post lab. Temporal representations were not applicable after investigation. This was also consistent with text-driven entries.

Songer and Gotwals (2012) was the only study to incorporate the use of notebooks. The 2012 study was conducted in an urban district that had high levels of student mobility and absenteeism plaguing the district. Songer and Gotwals focused on student explanations and integrated science notebooks to collect these explanations—specifically claim, evidence, and reasoning— as part of the data analyses. Of particular interest was the use of scaffolding to support students' explanations. No such scaffolding was employed with students' explanatory reflective conclusion data in the author's study. Also, Songer and Gotwals' study had attendance/attrition issues very similar to the author's study. In both studies, some students were missing either the pretest or posttest. However, missing pretest-posttest scores were removed in the author's study, whereas Songer and Gotwals opted to impute data for missing 4th through 6th grade pretest-posttest scores. The imputed scores were not very different (in terms of achievement) from students who had all data. But, the authors were not able to fully empirically test the data set. Nevertheless, pretest-posttest mean scores of both Songer and Gotwals and the author's studies demonstrated learning progressions that contribute to students' improvement.

2.3 Students' Perspective of their Science Experience

Students wrote reflective conclusions to articulate their science experience. Reflective conclusions provided an opportunity for students to explain what they learned and to make meaning out of what they studied. Randall (1999) sentence starters were used as writing frames. The five writing frames each addressed a critical portion of the inquiry process. Students wrote reflective conclusions post lab.

Writing Frame One expressed the purpose of the lab. Results suggest the majority of students had satisfactory understanding for Lab #1, but partial understanding in Lab #2. A possible explanation is that the attempt to scaffold Writing Frame One in lab two, in conjunction with the steady decrease in student attendance, combined to cause the shift in students' understanding of lab two's purpose. Writing Frame Two articulated the investigative methods. Nearly all students had a partial understanding of the methods for both labs or Writing Frame Two was incomplete. In Writing Frame Three, students were to explain the results of their investigation by stating a claim, supporting it with evidence from collected data, and justifying the reason the evidence supported the claim. This proved to be the most challenging writing frame for students. For both labs, all students had either partial understanding of their results or they did not complete the writing frame. Writing Frame Four focused on elucidating the accuracy of the results. Every student had a partial understanding of how to maintain accuracy of the results or they did not complete the writing frame for lab one. This trend continued in lab two, with the exception of two students. Writing Frame Five communicated the means to further investigate the topic of study. Many students had partial understanding of how to extend an investigation of the lab's problem in lab one; some also did not complete the writing frame. For lab two, approximately half of the students had satisfactory understanding, while the other half had incomplete writing frames.

Based on the results of students' reflective conclusions, it was concluded that students were able to partially define, or unable to define, their scientific world as evidenced by the relatively few writing frames satisfactorily completed for both labs. The reflective conclusions revealed a breakdown of students' ability and capacity to explain their science (and consequently their learning progression experience) to a wider audience. While notebooks used in conjunction with the learning progression helped students' science outcomes, students were not able to fully explain their science experience. Data analysis revealed many students correctly expressed the experiments' purposes. However, very few students articulated the procedure, results (i.e. claim, evidence, and reasoning), or accuracy of the investigations. Some students discussed further investigation of both labs. Overall, an overwhelming majority either needed improvement in explaining their science experience or they had incomplete writing frames.

3.0 Limitations

The study was limited by three major factors. Length of the study was the most substantial factor. The full Rock Cycle learning progression was eight-levels at each grade level and it required more time to implement than what was available. Originally scheduled for an eight-week period during the academic year in a large school district, the intervention presented persisted for three weeks, occurred during the summer and took place at a small charter school. The allotted time was not sufficient for students to make adequate progress. Furthermore, Off-campus group participants were selected by the school's administration to participate in an off-campus enrichment program for eight of the 15 days of the intervention. School-wide field trips were also scheduled (prior to the researcher's knowledge) during the school's summer program. Consequently, all participants did not receive the potential maximum amount of time for the already truncated intervention.

Methodologically, sample size was a second substantial limitation. Original enrollment was approximately 85 students. However, absenteeism and attrition increased as the summer progressed. After data were cleaned, a sample of 32 remained. The computer-assisted and learning progression groups each had 16 students. However, 63% of the computer-assisted group participated in the summer off-campus program. Students were selected based on their academic performance, attendance, citizenship and behavior. Essentially, these students were school ambassadors; the remaining students were not considered as such. Because of the reduced sample size and selection bias, the availability of data was also reduced. This impacted the reliability of the results.

The third and most surprising limitation was the lack of released-items available for adapting the measurement instrument. Over 70 released-items were initially selected for the Rock Cycle Assessment instrument. After reviewing and eliminating irrelevant and repetitive items, 30 test items remained. However, the available test items only addressed levels one, two, and two-three (i.e. L_1, L_2, and $L_{2\text{-}3}$) of the Rock Cycle learning progression. Therefore, the synthesis and implication levels (i.e. the upper levels of the learning progression) were not addressed by the Rock Cycle Assessment because there were no released-items available. This limitation reduced the validity and reliability of the Rock Cycle Assessment. The internal consistency of the Rock Cycle Assessment indicated the instrument did not accurately measure students' rock cycle learning outcomes (30 items; $\alpha = .315$).

4.0 Implication for Practice

The study's results have implications for learning progression researchers, educators, and educational stakeholders. The results give more insight into a different way of incorporating a learning progression with science notebooks, particularly with graphic representations. For example, teachers can identify and evaluate collected artifacts as evidence of students' thinking and leaning. Teachers can then, in turn, use the results to modify instruction, revise the learning progression, and/or help students' learning advance. Researchers can investigate students' science understanding through collecting and surveying notebook artifacts, determining where students encounter symbolic media, and examine how symbolic media influences science learning outcomes based on where encountered. The results can inform professional development, as well as inform and assist in curriculum and/or instruction modification.

Symbolic media results imply a need to utilize student-generated graphics more effectively. Specifically, the focus should be on examining the use of student-generated graphic representations in the prelab and post lab stages of inquiry. Results also suggest a need for a science graphic representation canon. In other words, there needs to be standard graphic representations that all students should utilize as they progress in their science education. The graphic representations canon should become increasingly complex as learning progresses, particularly for systems-based science graphic representations. Researchers in the emerging learning progression sub-division of models, modeling, and symbolic representations can best meet the need. The symbolic media results can also support researchers and educational stakeholders in precisely identifying where symbolic media are student-generated in the inquiry process, and the nature of the graphic representation. That information can in turn be used to inform researchers of revisions and/or modifications for additional iterations of a learning progression and its curricular and instructional products.

Reflective conclusions results can inform researchers concerned with students' capacity to communicate their learning in science. In the author's study, the results demonstrated a need for repeated practice in using and writing of science language. Repeated practice in using language and writing (and by implication reading) of science language can help students gain the knowledge and skills needed to handle the intellectual expanse of science classes. By implication, writing in science and teaching students how scientist use writing is indicative for generating a model. Moreover, it implies relating the nature of science to middle school science writing and writing-to-learn progressions. Songer et al. (2009) is the only learning progression to offer a model-based mechanism for communicating science.

Independent t-test results imply a need to re-examine some current sampling practices in the research field. Learning progression validated in the absence of ethnically, racially, and economically diverse populations have limited reliability, generalizability, and validity. The unique challenges urban schools pose have been well documented (e.g. Barton, 2007; Lee and Luykx, 2007). However, this demographic was virtually ignored by the learning progression research community. The methodological challenges that arose while conducting research were surprisingly difficult to foresee, accommodate and they potentially derailed the study. Additionally, learning progression research occurred during the academic year when data was collected from students, leaving the assumption that students' grades were impacted by academic performance involving the learning progression research. Further iterations of the Rock Cycle Learning Progression must have revisions to the Rock Cycle Assessment. Mainly, it should be conducted during the academic year when attendance rates are more stable, students are accountable for learning, there is less influence of extra-curricular programs during instructional time, and class assignment is less biased.

The study informs the practice of professionals in the research and practical science education fields. Complex systems were emphasized in thinking and learning research, particularly in earth science. Systems thinking promoted an integrated understanding of complex, interconnected systems. The *Frameworks* (2014) recommends and emphasizes students' exposure to the systems thinking approach. Yet, systems thinking has been ignored in learning progression development and validation, despite the prominence of systems-based topics in learning progression research. However, if STH model was ignored in learning progression research, there is a likelihood the model will be ignored by educational practitioners. Implementing STH model at the elementary and middle school levels can support students' understanding of the interconnectedness of earth's systems, develop systems-thinking skills, and encourage their awareness of the dynamic and cyclic nature of the world. STH was not limited to geology, but to content and skills that are systems based. Therefore, a need exists for future research in the learning progression field.

5.0 Conclusions

As states give consideration to or are currently adopting and/or implementing NGSS, it is critical to take a serious look at the historical context of science education reform in light of the current and relevant research literature. Historically, children's thinking and learning did not have an intentional focus (Kahle, 2007). Furthermore, after two decades of emphasizing standards-based reform, U.S. science curriculum and instruction have not yielded the type of science learning that resulted in student conceptual understanding and meaningful engagement (Duschl et al., 2007). The 2009 science NAEP results were evidence of this phenomenon. The results showed that less than 50% of students performed at or above the proficient level in science at all three grade levels (NCES, 2011). Many researchers asserted and demonstrated learning progressions' potential to transform science education because of their capacity to better align curriculum, instruction, and assessment (e.g. Duncan and Hmelo-Silver, 2009; Mohan, Chen, & Anderson, 2009).

Consistent involvement in inquiry-based activities should be the cornerstone of science instruction and curriculum development at every school level. For the past 25 years, there was an intense instruction and curricular focus on literacy and numeracy. Consequently, elementary and middle school science was—by and large— neglected. Compounded with the tremendous breadth and shallow depth of NSES, science education focused on content mastery and not inquiry-based activities. Learning progressions and their associated products has the potential to turn the contemporary tide of science education from its two-decade course. Even though there was no credible evidence to suggest the learning progression had a significant impact on students' science learning outcomes, and the data provided little evidence that the null hypothesis was false, there is one invaluable imperative: researchers, stakeholders, and educators should critically examine practices associated with learning progression development and validation and move forward with caution.

References

AAAS. (1993). Benchmarks for Scientific literacy. Retrieved from http://www.project2061.org/publications/bsl/online/index.php?intro=true.

AAAS. Science Assessment item. Retrieved from http://assessment.aaas.org/items/WC023002#/0.

AAAS. Science Assessment item. Retrieved from http://assessment.aaas.org/items/WC036002#/0.

AAAS. (2015). Frequently asked questions. Retrieved from http://assessment.aaas.org/pages/faq#testing.

Adadan, E., Trundle, K.C., & Irving, K.E. (2010). Exploring grade 11 students' Conceptual pathways of the particulate nature of matter in the context of multi-representational instruction. *Journal of Research in Science Teaching,* 47, 1004-1035.

Alonzo, A.C. & Steedle, J.T. (2009). Developing and assessing a force and motion Learning progression. *Science Education*, 93, 389-421.

Anderman, E. M. (1998). The middle school experience: Effects on the math and science achievement of adolescents with LD. *Journal Of Learning Disabilities*, 31, 128.

Anderson, C.W. (2007). Perspectives on science learning. In S.K. Abell & N.G. Lederman (Eds.). *Handbook of Research on Science Education* (3-30). New York, NY: Routledge.

Arnold, R.D. & Wade, J.P. (2015). A definition of systems thinking: a systems approach. *Procedia Computer Science*, 44, 669-678.

Aschbacher, P. and Alonzo, A. (2006). Examining the utility of elementary science Notebooks for formative assessment purposes. *Educational Assessment, 11* (3&4), 179-203.

Assaraf, O.B. & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, 42, 518–560.

Assaraf, O.B. & Orion, N. (2005). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47, 540–563.

Assaraf, O.B. & Orion, N. (2009). A design based research of an earth systems based environmental curriculum. Eurasia Journal of Mathematics, Science, and Technology Education, 5, 47–62.

Bedward, J., Wiebe, E. N., Madden, L., Carter, M., & Minogue, J. (2009). Presented at the ASEE Annual Conference: *Graphic Literacy in Elementary Science Education: Enhancing Inquiry, Engineering Problem-solving and Reasoning Skills*. Austin, TX.

Development of system thinking skills in the context of earth system education

Berland, L.K., & McNeill, K.L. (2010). A learning progression for scientific Argumentation: understanding student work and designing supportive instructional context. *Science Education*, *94,* 765-793.

Bond-Robinson, J. & Stucky, A.P. (2005). Grounding scientific inquiry and knowledge in situated cognition, Retrieved from http://csjarchive.cogsci.rpi.edu/Proceedings/2005/docs/p310.pdf

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 15, 32-42.

Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Butler, M.B. and Nesbit, C. (2008). Using science notebooks to improve writing skills and conceptual understanding. *Science Activities*, 44, 137-145.

Chokski, N. (2014). Washington becomes first state to lose its waiver from NCLB. Retrieved from http://www.washingtonpost.com/blogs/govbeat/wp/2014/04/24/washington-becomes-first-state-to-lose-its-waiver-from-no-child-left-behind/.

Clidas, J. (2010). A laboratory of words: Using science notebook entries as pre-assessment creates opportunities to adapt teaching. Science and Children, *48*, 60-63.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: making thinking visible. *American Educator*, 6, 38–46.

Duncan, R.G. & Hmelo-Silver, C.E. (2009). Learning progressions: aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*. 46, 606-609.

Duncan, R.G., Rogat, A.D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. *Journal of Research in Science Teaching*. 46, 655–674.

Duschl, A., Schweingruber, H., & Shouse, A. (2007). Taking science to school: learning and teaching science in grades K-8. Washington, DC: National Academies Press.

Earle, J. & Wan, J.C. (1995). The National Science Foundation and systemic reform. In *On Common Ground: Partnerships in Science and Technology*. Retrieved from http://www.yale.edu/ynhti/pubs/A17/earle.html.

Emma, C. (2014). Duncan yanks NCLB waiver from Washington state. Retrieved from http://www.politico.com/story/2014/04/washington-state-nclb-waiver-arne-duncan-105997.html.

Furtak, E.M. (2009) Proceedings from Learning Progressions in Science (LeaPs) Conference. *Toward Learning Progressions as Teacher Development Tools*: Iowa City, IA.

Furtak, E.M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, 49,1181-1210.

Furtak, E. M. & Heredia, S.C. (2014). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching*, 51,982-1020.

Gall, M.D., Gall, J.P., & Borg, W.R. (2007). Educational research: An introduction. Boston, MA: Allyn and Bacon.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10, 486–489.

Gibbs, B.C. (2014). Reconfiguring Bruner: compressing the spiral curriculum. *Phi Delta Kappan*, 95, 41-47.

Gotwals, A.W. & Songer, N.B. (2013). Validity evidence for learning progression-based assessment items that fuse core disciplinary ideas and science practices. *Journal of Research in Science Teaching*, 46, 597-626.

Gunckel, K.L, Covitt, B.A., Salinas, I., & Anderson, C.W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching*, 49, 843-868.

Hamza, K.M. & Wickman, P. (2013). Supporting students' progression in science: continuity between the particular, contingent, and the general. *Science Education*. *97*, 113-138.

Harden, R. M.; Stamper, N. (1999). What is a spiral curriculum? *Medical Teacher.* 21, 141-143.

Hewson, P.E. (1992). Conceptual change in science teaching and teacher education. Retrieved from http://learner.org/workshops/lala2/support/hewson.pdf

Ho, W. (2015). Situated learning. Retrieved from http://www.personal.psu.edu/wxh139/Situated.htm

Huerta, M., Irby, B.J., Lara-Alecio, R., & Tong, F. (2005). Relationship between language and concept science notebook scores of English language learners and/or economically disadvantaged students. *International Journal of Science and Mathematics Educat*ion. *14*, 269-285.

Huerta, M. & Jackson, J. (2010). Connecting literacy and science to increase achievement for English language learners. *Early Childhood Education Journal.* *38*, 205–211.

Jin, H. & Anderson, C.W. (2012). Learning progression for energy in socio-ecological systems. *Journal of Research in Science Teaching*, 49, 1149-1180.

Johnson, B. R., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112-133.

Johnson, P. & Tymms, P. (2011). The emergence of a learning progression in middle school chemistry. *Journal of Research in Science Teaching*, 48, 849-877.

Kahle, J.B. (2007). Systemic reform: research, vision, and politics. In S.K. Abell & N.G. Lederman (Eds.). *Handbook of Research on Science Education* (911-942). New York, NY: Routledge.

Kali, Y., Orion, N., and Eylon, B. (2003). The effect of knowledge integration activities on students' perception of the Earth's crust as a cyclic system. *Journal of Research in Science Teaching*, 40, 545-565.

Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation,* Retrieved October 25, 2015 from http://PAREonline.net/getvn.asp?v=4&n=10 .

Kennedy, C. A., Brown, N. J., Draney, K., & Wilson, M. (2005). Paper presented at the annual meeting of the American Education Research Association: Using Progress Variables and Embedded Assessment To Improve Teaching And Learning: Montreal, Canada.

Kesidou, S. & Roseman, J.E. (2002). How well do middle school science programs measure up? Findings from project 2061's curriculum review. *Journal of Research in Science Teaching*, 39, 522–549.

Klentschy, M. (2005). Science notebook essentials: A guide to effective notebook components. *Science and Children*, *Nov/Dec*, 24-27.

Klentschy, M. & Molina-de-la Torre, E. (2004). Students' science notebooks and the inquiry process. In E.W. Saul (Ed.). *Crossing Borders In Literacy And Science Instruction: Perspectives On Theory And Practice*. VA: NSTA Press

Klew, A. (2015). Could Washington state get its NCLB waiver back? Retrieved from http://blogs.edweek.org/edweek/campaign-k 12/2015/03/could_washington_state_get_its.html

Lee, H. & Liu, O.L. (2010). Assessing learning of energy concepts across middle school grades. *Science Education*, *94*, 665-688.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, 31, 3-12.

Lee, O. and Luykx, A. (2007). Science education and student diversity: Race/ethnicity, language, culture, and socioeconomic status. In *Handbook of Research on Science Education*: S.K. Abell & N.G. Lederman (Eds.). New York: Routledge.

Lehrer, R. & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, 96, 701-724.

Liu, L. & Hmelo-Silver, C.E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching,* 49, 1023–1040.

Liu, X. & Lesniak, K. (2006). Progression in children's understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, 43, 320-347.

Lofgren, K. (n.d.) YouTube. Normality test using spss: how to check whether data are normally distributed. Retrieved from https://youtu.be/IiedOyglLn0.

Lunetta, V.N., Hofstein, A., & Clough, M.P. (2007). Learning and Teaching in the

School Science Laboratory: An Analysis of Research, Theory, and Practice. In S.K. Abell & N.G. Lederman (Eds.). *Handbook of Research on Science Education* (393-442): New York, NY: Routledge.

Masters, K. & Gibbs, T. (2007). The spiral curriculum: implication for online learning. *BMC Medical Education*, 7, 52-62.

Mathematical Association of America (MAA). (2015). Critique of situated cognition. Retrieved from http://www.maa.org/research-sampler-21-critique-of-situated-cognition.

Merriam, S.B. (2009). Qualitative research. San Francisco, CA: Jossey-Bass.

Miller, J.A. (1991). Report questioning 'crisis' in education triggers an uproar. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/1991/10/09/06crisis.h11.html

Missouri Learning Standards—ELA. (2015). Retrieved from http://www.missourilearningstandards.com/ela/

Mohan, L., Chen, J., & Anderson, C.W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, 46, 675-698.

Morrison, J.A. (2005). Using science notebooks to promote pre-service teachers' understanding of formative assessment. *Issues in Teacher Education*, 14, 5-21.

Muthukrishna, N., Carnine, D., Grossen, B. & Miller, S. (1993). Children's alternative frameworks: should they be directly addressed in science instruction? *Journal of Research in Science Teaching*, 30, 233-248.

NAEP. (2009). Technical documentation on the web. Retrieved from http://nces.ed.gov/nationsreportcard/tdw/sitemap.aspx

National Center for Education Statistics. (2011). The nation's report card: science 2009. Institute of Education Sciences, U.S. Department of Education: Washington, DC.

National Commission on Excellence in Education (1983). A Nation at Risk. Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html.

National Research Council. (2000). How people learn: brain, mind, experience, and school. National Academy Press: Washington, DC.

National Research Council. (2005). *How students learn: history, mathematics, and science in the classroom*. Retrieved from http://www.nap.edu/openbook.php?record_id=10126.

National Research Council. (2014). Next Generation Science Standards (NGSS). Retrieved from from http://www.nextgenscience.org/search-standards-dci.

National Science Foundation. (2008). Award Abstract #0824432 Learning Progressions in Science (LeaPS) Conference. Retrieved from http://www.nsf.gov/awardsearch/showAward?AWD_ID=0824432.

Neressian, N.J. (2007). Mental modeling in conceptual change. Retrieved from http://www.cc.gatech.edu/aimosaic/faculty/nersessian/papers/Nersessian-ModelingConceptualChange.pdf.

Neumann, K., Viering, T., Boone, W.J., & Fischer, H.E. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching*, 50, 162-188.

NGSS. (2014a). 5-ESS2 Earth's system. Retrieved from http://www.nextgenscience.org/5ess2-earth-systems

NGSS. (2014b). MS-ESS2 Earth's system. Retrieved from http://www.nextgenscience.org/msess2-earth-systems

O'Connor, J. & McDermott, I. (1997). The art of systems thinking :essential skills for creativity and problem solving London : Thorsons.

Orion, N. & Libarkin, J. (2014). Earth systems science education. In N.G. Lederman & S.K. Abell (Eds.), *Handbook of research on science education Volume II.* New York: Routledge.

Preece, A.B. and Bond-Robinson, J. (2003). *Learning by doing: cognitive apprenticeship in chemical research*. Retrieved from http://csjarchive.cogsci.rpi.edu/Proceedings/2003/pdfs/327.pdf.

Porter-Magee, K. (2004). Teacher quality, controversy, and NCLB. *The Clearing House*, 78, 26-29.

Plummer, J.D. & Krajcik, J. (2010). Building a learning progression for celestial motion: elementary levels from an earth-based perspective. *Journal of Research in Science Teaching,* 47, 768-787.

Plummer, J.D. & Maynard, L. (2014). Building a learning progression for celestial motion: An exploration of students' reasoning about the seasons. *Journal of Research in Science Teaching.* 51, 902–929.

Randall, C. (1999). Lab report terms and sample lab report. Retrieved from http://donnayoung.org/apologia/labhow-cr.htm.

Raved, L., & Yarden, A. (2014). Developing Seventh Grade Students’ Systems Thinking Skills in the Context of the Human Circulatory System. *Frontiers in Public Health*, 2, 260

Rivet, A. E. & Kastens, K. A. (2012). Developing a construct-based assessment to examine students' analogical reasoning around physical models in earth science. *Journal of Research in Science Teaching*, 49,713–743.

Rivet, A.E. & Krajcik, J.S. (2004). Project-based science curricula: Achieving standards in urban systemic reform. *Journal of Research in Science Teaching*. *41*, 669-692.

Sáez. L., Lai, C. F., Tindal, G. (2013). *Learning progressions: Tools for assessment and instruction for all learners* (Technical Report No. 1307). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Salkind, N.J. (2008). Statistics for people who (think they) hate statistics. Los Angeles, CA: Sage Publications

Schwarz, C.V., Reiser, B.J., Davis, E.A., Kenyon, L., Acher, A., Fortus, D., Scwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46, 632-654.

Smith, C.L., Wiser, M., Anderson, C.W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: a proposed learning progression for matter and the atomic molecular theory. *Measurement: Interdisciplinary Research and Perspective*, 14, 1-98

Schroeder, K. (1995). Urban achievement gap. *Education Digest*, 61, 74.

Schmidt, D.L., Saigo, B.W., & Stepans, J.I. (2006). Conceptual Change Model: The CC Handbook. St. Cloud, MN. Saiwood Publications.

Shavelson, R. & Kurpius, A. (2012). Reflections on learning progressions. In

Learning Progressions in Science: Current Challenges and Future Directions. A.C. Alonzo & A.W. Gotwals (Eds). The Netherlands: Sense Publishers

Shepardson, D.P. & Britsch, S.J. (1997). Children's Science Journals: Tools for Teaching, Learning, and Assessing. *Children & Science*, NSTA.

Shin, N., Stevens, S.Y., Short, H., and Krajcik, J. (2009). Paper presented at the Learning Progressions in Science (LeaPS) Conference: Learning Progressions To Support Coherence Curricula In Instructional Material, Instruction, and Assessment Design, Iowa City, IA.

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.* 49,141-165.

Songer, N.B., Kelcey, B. Gotwals, A.W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching.* 46,610-631.

Spector, B. (1984). Qualitative research: Data analysis framework generating grounded theory applicable to the crisis in science education. *Journal of Research in Science Teaching.* 21,459-467.

Stahly, L.L., Krockover, G.H., & Shepardson, D.P. (1999). Third grade students' ideas about the lunar phases. *Journal of Research in Science Teaching,* 3, 159-177.

Steedle, J.T. & Shavelson, R.J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching. 46,* 699-715.

Stevens, S, Delgado, C, & Krajcik, J.S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, 47, 687-715.

Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.) San Francisco: Jossey-Bass.

Sweeney, A.E. & Paradis, J.A. (2004). Developing a laboratory model for the professional preparation of future science teachers: a situated cognition perspective. *Research in Science Education*, 34, 195-219.

Vosniadou, S. & Brewer, W.F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papadimitriou (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11, 381-419.

Wallace, C. (2004). An illumination of the roles of hands-on activities, discussion, text reading, and writing in constructing biology knowledge in 7th grade. *School Science and Mathematics*. 104, 70-79.

Wells, C.S. & Wollack, J.A. (2003). An Instructor's Guide to Understanding Test Reliability. Retrieved from http://testing.wisc.edu/Reliability.pdf

Wiebe, E.N., Madden, L.P., Bedward, J.C., Carter, M., & Minogue, J. (2008). Presented at the Conference on Research and Training in Spatial Intelligence: *Improving Early Spatial Intelligence Through Science Notebook Graphic Production: Effective Elementary Classroom Practices*. Evanston, IL

Wiebe, E.N., Bedward, J.C., Madden, L.P., Minogue, J., & Carter, M. (2009a).

Presented at AERA: *Graphic Representations in Science Notebooks: A Vehicle for Understanding Science Inquiry in the Elementary Classroom*. San Diego, CA.

Wiebe, E.N., Madden, L.P., Bedward, J.C., Minogue, J., & Carter, M. (2009b). Presented at NARST: *Examining Science Inquiry Practices in the Elementary Classroom through Science Notebooks*. Garden Grove, CA.

Wilson, M. (2005). Constructing measures: An item response theory approach. Mahwah, NJ: Erlbaum.

Wilson, M. (2009) Measuring Progressions: Assessment structure underlying a learning progression. *Journal of Research in Science Teaching*, 46, 716-730

Appendix A: Test Blueprint

Achievement Level	Big Idea	Test Item (Item Code/Question ID)	Test Percentage
8-L1	1. Earth is continuously changing. 2. Earth is a complex system of interacting rock, water, air & life.	1. 011-8S11 #1 K119401 5. 2009-8S10 #2 K111701 9. 2005-4S13 #7 K036001 13. PT007001 17. PT019001 21. PT017001 25. PT029001 27. PT016001 29. CL014002 30. PT018001	33%
8-L2		2. 2009-4S11 #10 K106601 6. 2009-4S11 #12 K106604 10. WE037004 14. WE021002 18. WE042002 22. CL021002	20%
8-L3		3. 2005-8S14 #9 K037801 7. PT025001 11. WE059001 15. WE039002 19. WE011002 23. WE056001	20%
8-L2 & 3		**4. 2009-4S11 #2K154301** **8. WE064001** **12. WE032003** **16. WE014004** **20. WE018003** **24. WE012003** **26. WE015003** **28. WE053001**	**27%**
8-L4 to 8-L8		N/A	0%

Appendix B: The Frequency[1] of Correct Responses on the Rock Cycle Assessment by Learning Progression Achievement Level[2]

Achievement Level	Comparison Group	Pretest	Posttest	Total
8-L1[3]	Learning Progression	33	48	81
	On-Campus Computer Assisted	14	18	21
	Off-Campus Computer Assisted	22	37	59
8-L2[4]	Learning Progression	32	36	68
	On-Campus Computer Assisted	16	14	30
	Off-Campus Computer Assisted	25	31	56
8-L3[5]	Learning Progression	23	42	65
	On-Campus Computer Assisted	12	11	23
	Off-Campus Computer Assisted	20	27	47
8-L2 & 3[6]	Learning Progression	27	42	69
	On-Campus Computer Assisted	14	21	35
	Off-Campus Computer Assisted	20	29	49

---

[1] Frequency refers to the number of times the event occurred. In this case, the numbers of correct responses for each of the three groups were tallied for the pretest and posttest; they were then totaled.

[2] The Achievement Levels are the levels of the learning progression. They represent what a student should be able to know and do at a particular point in the learning progression.

[3] On Achievement Level 8-L1, students should be able to identify rock cycle components and processes.

[4] On Achievement Level 8-L2, students should be able to identify relationships among rock cycle components.

[5] On Achievement Level 8-L3, students should be able to identify dynamic relationships within the rock cycle.

[6] On Achievement Level 8-L2&3, students should be able to identify rock cycle components, processes, and dynamic relationships.