Impact of Earth Science Integration on Student Learning in a High School Chemistry Course

Jonathon Grooms[1], Kevin Fleming[1], Alan R. Berkowitz[2], Mary Ellen Wolfinger[1], Bess Caplan[2], Chelsea McClure[2]
[1]*Graduate School of Education and Human Development, George Washington University*
[2]*Cary Institute of Ecosystem Studies*
(correspondence to Jonathon Grooms, jgrooms@gwu.edu)

Subject/Problem

In response to the adoption of NGSS, the focus school district has developed a new high school chemistry curriculum that integrates Earth science and chemistry content. The purpose of this present study is to investigate the implementation of the new integrated curriculum to identify changes in students' thinking related to integrated chemistry and Earth science content as well as specific science practices, such as modeling, and to contextualize students' learning based on teachers' and students' reflections on their curricular experiences during the school year.

Design and Procedures

Theoretical framework

Citizens are being asked to make pivotal decisions about the environment based on a plethora of arguments – in their schoolwork, in the media, and as they vote and make everyday choices. However, the arguments they encounter often have weak scientific reasoning, limited supporting evidence and many types of bias. Understanding how to make and evaluate claims based on the evidence provided is a key practice emphasized in the Next Generation Science Standards (NGSS Lead States, 2013) and Common Core Standards. Earth science is more relevant and needed than ever to understand complex global systems of today. The lack of trained Earth science teachers and the requirement of the three core science courses (biology, chemistry and physics) severely limit Earth science instruction at the high school level (Wyession, 2013; Horizon Research, 2013). NGSS requires adding meaningful Earth science experiences to high school science curricula, a shift from previous standards which suggested that students cover Earth science for the last time in middle school (Wysession, 2013). Developing a scientifically and environmentally literate citizenry, one that will be able to evaluate claims and make decisions, requires training our students to think critically and to engage in scientific inquiry practices (NRC, 2012; Berkowitz, Ford, & Brewer, 2005; NRC, 2012; AAAS, 1993; ACERE, 2009). Across the US, Earth science has traditionally been offered as an elective and is often perceived as an 'easier' science (Wysession, 2013). Only 48% of high schools across the country offer Earth science (Banilower et al., 2013). In 2013 the State Board of Education, where this district is located, adopted the NGSS which necessitated a shift whereby Earth science needed to be integrated across the district's high school biology, chemistry, and physics courses.

There are many risks and challenges in an integrated approach, e.g., the time it could take from teaching core chemistry concepts, isolation of Earth science content from its fuller context, and lack of familiarity among the teachers with Earth science content, practices and pedagogy. The Integrating Chemistry and Earth science (ICE) project addresses these challenges by using an interdisciplinary approach that more closely mirrors the actual work of environmental scientists to bring topics into the chemistry classroom that are relevant to students and address real world problems. Our approach builds on the following hypothesized elements of success: 1) careful identification of topics suited to interdisciplinary exploration across Earth science and chemistry; 2) three dimensional pedagogy that is engaging, place- and real-world based; 3) combining first hand inquiry with exploration of research datasets from the local environment; 4) seamless integration into the core chemistry curriculum; 5) an R&D process with researchers and educators working with classroom teachers trying things out with their students; and 6) research and teacher-driven professional development, educative curriculum, readily available and high quality teacher resources, and sustained support to foster broad implementation.

Research Questions

The data collected as part of this study address the following research questions:

1. To what extent do students develop content knowledge related to integrated chemistry and Earth science topics?
2. What evidence do students provide of their abilities to integrate their content knowledge with science practice, specifically modeling?
3. How do teachers' and students' reflections of their experiences contextualize any changes, or lack thereof, in students' achievement or abilities to integrate content knowledge and practices?

Methods and Data Sources

This paper is part of a larger, three-year study to develop and implement a new integrated chemistry and Earth science curriculum for the district's high school chemistry course. Although the curriculum is being implemented across the district, within the present student the research team worked with a group of six *development team teachers* (DTT) to try out specific aspects of the curriculum, give feedback during implementation, and engage in revisions based on classroom experiences. These development team teachers agreed to participate in periodic professional development, reflection, and revision sessions throughout the 2018-2019 school year. Additionally, these teachers received classroom equipment and materials in advance and during PD to ensure their students could engage in all ICE related curriculum experiences (all other district teachers have since received these materials).

Data for this study are derived from three main sources: (a) student pre- and summative assessments for the ICE units, which provide evidence for changes in students' thinking related to integrated chemistry and Earth Science content; (b) student and teacher interviews about their curricular experiences during the school year, which provide data help contextualize student achievement data; and, (c) student learning artifacts collected from exemplar lessons. Student summative assessments were multimodal, comprised of single items and items clusters in multiple choice and constructed response formats. Two versions of the assessment were utilized, randomly assigned to students. A traditional version focused only on disciplinary core ideas, while an integrated version had items that included a disciplinary core idea and a science practice or cross-cutting concept. A total of 64 students completed a matching pre-assessment and post-unit assessment.

It was important to gain the perspectives of the DTTs as well as the students in their classroom. Each of the six DTTs participated in a reflective interview at the end of the school year to recount their experiences implementing the ICE curriculum materials. Each interview was conducted in the teacher's classroom during after school hours and followed a semi-structured protocol;

interviews were audio recorded with durations ranging from 35-96 minutes. Additionally, twelve students were randomly selected (from among those consenting to participate in interviews) from development team teachers' classes to participate in one-on-one interviews to share their perspective of experiencing the integrated curriculum during the 2018-19 school year. We limited our selection of students to the development team teachers' classrooms in part due to the consent process. Each interview was conducted during the school day in an empty classroom to avoid distractions or external influences. Students were interviewed during their regularly scheduled science class period. Each interview followed a semi-structured interview protocol that was audio recorded, lengths ranged from 16-34 minutes.

Analysis

Student learning data was analyzed from two sources: pre- and summative unit assessments and student learning artifacts. Students' unit assessments consisted of multiple-choice items and constructed response items. Multiple choice items were scored using a binary approach (correct/incorrect) and constructed response items were scored using researcher-generated rubrics. A 20% sample of the data was used to assess the validity of scoring criteria for the rubrics. Revisions were then made to the rubrics to alleviate inconsistencies or excluded criteria. Another random 20% sample of student assessments were used to assess interrater reliability. Interrater reliability was calculated using an intraclass correlation (ICC = 0.98), which indicated sufficient reliability (Landis & Koch, 1977) for the remaining assessments to be scored individually. Student assessment scores were then entered and analyzed using SPSS software to conduct paired- and independent-samples t-tests as relevant.

Student learning artifacts were scored using researcher-developed rubric using a model-based explanations framework (Zangori, Peel, Kinslow, Friedrichsen and Sadler, 2017). The rubric was divided into three sub-scores reflecting the three aspects of this framework: components, sequences, and explanations. Components are considered words, images or symbols used as pieces included in the model-based explanation. Sequences are cause and effect relationships demonstrating a linkage between two or more components. Explanations are multiple sequences linked together generating the model's explanatory power (Louca & Zacharia, 2012). Each sub-score (components, sequences, explanations) were criterion-reference scored along a three-point scale (see Figure 1 for rubric). A 20% sample of the collected learning artifacts was used to assess the validity of the rubric's scoring criteria. Revisions to the rubric were made to refine criteria and address inconsistencies. Another random 20% sample was used to assess interrater reliability. Interrater reliability was calculated using an intraclass correlation ($ICC_{components}$ = 0.920; $ICC_{sequences}$ = 0.955; $ICC_{explanation}$ = 1.00), which indicated strong reliability of the rubric and agreement between raters (Landis & Koch, 1977) sufficient for independent scoring of the remaining learning artifacts. Student scores on the learning artifacts were then entered and analyzed using SPSS software to generate descriptive statistics and comparative tests between sub-scores.

Each interview transcript of the six participating teachers and twelve participating students were open coded by members of the research team. Open codes were then analyzed using a qualitative approach of the constant comparative method (Strauss & Corbin, 2008). Emergent themes were then generated for the teacher and student interview sets. Each theme represents important experiences and views of the teachers and students from across the school year during the implementation of the integrated curriculum. The resulting themes were compared across teacher and student groups and refined by the research team until a core set of thematic patterns emerged.

Components

Level	Descriptor
0	No relevant components included
1	One component involving the occurrence of precipitation (e.g., rain, snow, ice) OR One component involving a human interaction with the environment (e.g. car, truck, plow) OR One component of a physical change occurring (e.g., cracks, holes, temp)
2	More than one component from any combination of categories (precipitation, human, physical change)
3	At least one component from each of the three categories (precipitation, human, physical change)

Sequences

Level	Descriptor
0	No sequences included
1	Includes one sequence (link between two components) from either weathering or erosion or human activity
2	Includes two sequences, from more than one category of weathering, erosion, or human activity
3	Includes one sequence from each category of weathering, erosion, and human-related activity

Explanations

Level	Descriptor
0	No links between sequences
1	Repeated sequences (e.g., ice wedging, car traffic) cause potholes to get larger or more potholes to form
2	Interaction between weathering and erosion cycles compounds the formation of more/bigger potholes
3	Interaction between human activity exacerbates natural weather/erosion processes

Figure 1. Pothole model-based explanation rubric

Findings

Analysis of student learning data, via paired-samples t-test, indicated significant, positive growth in student achievement from the pre-test (M = 3.16) to the post-test (M = 5.02), $t(63) = 6.78$, $p < .01$, $d = 0.85$, with a notably large effect size. Follow-up analyses indicate that students made significant growth on both versions of the assessments with traditional-style items (paired samples t-test: $t(25) = 5.01$, $p < .01$, $d = 0.98$) and integrated-style items (paired samples t-test: $t(37) = 5.22$, $p < .01$, $d = 0.86$). Despite no initial differences in scores between the two versions on the pre-assessment, perhaps not surprisingly, students made greater gains on the traditional-style assessment items compared to the integrated-style assessment items pre-to-post (independent samples t-test: $t(62) = 3.81$, $p < .01$, $d = 0.95$)

Reviewing classroom artifacts does not provide pre-post growth data, however these artifacts do suggest that students are garnering experiences integrating science practices and disciplinary core ideas within the curriculum (see Figure 2). Students had an average total score of 4.4 out of 9 points. On a three-point maximum scale, students had the highest average score for components (2.58) followed by sequences (1.42) and explanations (0.44).

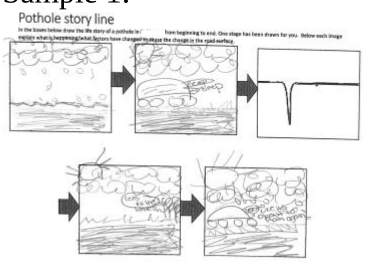
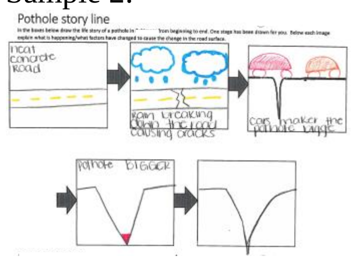
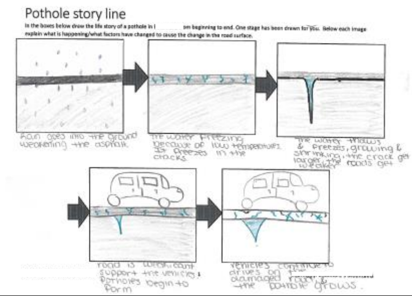
Student Learning	Sample 1: 	Sample 2: 	Sample 3: 
Interpretation	Student engages in explanation of the phenomena using a general narrative. Basic components are included (e.g. car traffic, precipitation, holes in the road) but cause and effect relationships or mechanistic explanations are not included. (Score: components = 3; sequence= 0; explanation =0)	Student engages in explanation of phenomena incorporating concepts of weathering and human impacts with an emergent use of descriptive model-based explanations. Components are linked into cause and effect relationships, but sequences are not incorporated into a mechanistic explanation. (Score: components =3; sequence = 2; explanation = 0)	Student engages in explanation of phenomena incorporating some concepts of weathering and human impacts; demonstrates emergent proficiency in model-based explanations where components are linked with cause and effect relationship integrated into an explanation. (Score: components = 3; sequence = 2; explanation = 1)

Figure 2. Classroom artifacts from student-experiences integrating core ideas and science practices indicate readiness for three-dimensional teaching and learning.

These descriptive statistics suggest that students succeeded in identifying components of the phenomenon within their model and struggled the most with articulating explanations of interactions between sequences. Students seem to have an emerging ability to identify cause and effect relationships as sequences related to the phenomenon. Additional pairwise Wilcoxon tests were performed between each of these aspects of the model-based explanation framework. The tests indicated that that students indeed performed significantly better at identifying components of the phenomena when compared to identifying cause and effect sequences within the phenomena ($z = -6.11$, $p < .001$) or including explanations of interactions between sequences ($z = -6.01$, $p < .001$). Similarly, students were better able to identify sequences of cause and effect relationships when compared to their ability to include explanations of the connections between sequences ($z = -5.37$, $p < .001$). While students tended to score higher with the more basic aspect of their models (components) than in the more rigorous aspects (explanations), these results taken together with the pre/post assessment data, suggest there is a degree of readiness or emerging proficiency among the major shifts in NGSS-based teaching and learning. Learning artifacts also provide supportive evidence that students are receptive toward experiencing pedagogy that is moving toward the facets of three-dimensional learning.

An analysis of student and teacher interviews reflect a similar pattern of receptivity and recognition of challenges in teaching and learning at this stage of implementation. Themes generated from the qualitative analysis are grouped into the broad categories of *synergies* (see Table 1) and *tensions* (see Table 2) to capture aspects of alignment and misalignment between the teachers' and students' experiences.

Table 1. Synergies between Teachers and Students.

Theme	Teacher	Student
Alignment of topics	T3: Yeah I think the space, like there are some students who are really into like why aren't we doing more space stuff.	S2: I learned a lot it was from the periodic table and then it connected to the stars and …the Big Bang and how Earth was made and the different elements that stem from that…everything was sort of like connected to each other… I probably wouldn't have known as much as I did
The importance of metrics of "doing well"	T2: What do you need for the final? …some of the stuff like I thought this was really cool…I think it's cool but again I'm thinking it wouldn't be helpful for the final.	S4: I mean I try for the grade 'cause apparently it's a requirement and that matters to get into colleges and stuff

Two synergistic themes that emerged are related to the *coherence of specific topics* (i.e., integration of Earth science and chemistry) and an emphasis on *metrics and success measures* (e.g., final exams or state tests). Related to these two themes, both teachers and students were able to articulate connections between various topics within the curriculum where chemistry and Earth science integration was attempted. With regard to the 'metrics' theme, both teacher and students placed high emphasis on accountability in terms of learning material for the purpose of testing or getting good grades. So, in this way, teaching and learning were each influenced by this extrinsic factor.

The two tension-related themes that emerged point to different views on the role of *local phenomena or data* and *teachers' classroom instruction*. One goal of the integrated curriculum was to leverage students' local context (e.g. their city or data collected from nearby locations) as an engaging feature of lessons. Teachers' and students' perception of this feature of the curriculum differed, with teachers' interpretation that students' lacked interest in this feature, but students indicating that the local connections did hold some interest. As may be expected, there was also a tension in how classroom instruction was perceived. Teachers' described classroom instruction that was interactive and included investigations (in alignment with the curriculum and target lesson), whereas students described the classroom as centering on notetaking and drill-based activities.

Table 2. Tensions between Teachers and Students.

Theme	Teacher	Student
Teacher's approach to instruction	T3: like a lot of the times, there has been some sort of investigation…it's not like just a straight teacher demo, it's just certain stations that just have one set of things.	S5: we always do drill and then we go right into notes...sometimes labs yeah.
Role of local phenomena or local data	T2: to me they're more attracted to sort of umm (pauses) what's the word I'm looking grandiosity than familiarity	S2: I think it's always good to learn about the area where you're from…I think that would be interesting.

Conclusions

Curriculum development is an important aspect of science teaching and learning as well as important for supporting efforts to move classrooms into alignment with new standards efforts. Therefore, understanding curriculum development and implementation from a variety of perspectives is a crucial aspect for successful reform efforts. Evaluating students' assessment data and classroom artifacts contextualized by teachers' and students' experiences highlight that some design features of a curriculum may not have the intended effects. For example, teachers felt that including local data might not induce further engagement in learning, thinking that large-scale or grand phenomena might be more effective. However, student responses suggested a preference for a more experiential connection to the local phenomena. This desire for more experiential learning is also seen in a tension between teacher approaches in implementing the curriculum and student learning preferences. Teachers described varied and somewhat limited use of hands-on instructional approaches while students expressed a desire to do more hands-on learning in place of the perceived more commonplace drill and lecture notes. This tension indicates that while curricular changes have been enacted, the adaptability toward the necessary pedagogical shifts to more student-centered, three-dimensional learning remains an area of focus.

This is further supported by the student data that indicate students' abilities to better respond to unidimensional items as opposed to multidimensional items. However, when given the opportunity to engage in a specific science practice, e.g. modeling a local phenomenon, students are able to demonstrate an emerging proficiency in integrating multiple dimensions of science learning. Collected learning artifact data demonstrate a degree of readiness to engaging in multi-dimensional tasks (e.g. model-based explanations). Students were more successful in incorporating the components aspect of their model-based explanations compared to sequences and explanations. This suggests students may be struggling with the higher learning demand those aspects impose or that there is a need for greater curricular interventions to improve their proficiency.

One synergism among students and teachers was a positive sentiment toward including Earth science within chemistry. Both valued examining the origin of the universe and the elements in an integrated fashion promoting meaningful conceptual connections. Teachers reported a perception that students responded more positively toward the space science aspects than the geologic science across the units. Some students recalled prior learning of geologic science in middle school and enjoyed learning more about certain topics and building upon that knowledge. While most teachers commented positively toward integration of the two science disciplines, for some a tension remained in expressing consternation of such a major change in the course, favoring a single-disciplinary approach. This mixed reaction suggests that ongoing efforts in promoting integration and refining its enactment within the curriculum is need to support both teachers and students' confidence and agency with implementation efforts.

Implications

Supporting and leading curriculum development, implementation, and research is an important aspect in this era of systemic reform. Therefore, elucidating the intricacies within and between teachers and students at early stages of implementation can further inform curriculum revision and targeted professional learning support. The synergies between teachers and students can be used to build capacity, supporting common ground amid uncertainty in early years of implementation and counter-balance perceived tensions to enhance teacher practice and student learning. This work illustrates the utility of a curriculum integration and implementation effort that has the potential to serve as a catalyst for teachers to examine the various conceptual, pedagogical, cultural, and political dilemmas (Windschitl, 2002) that impact their practice and classroom learning environment. Assessment of student learning in this study suggests that students are displaying a degree of readiness with aspects of three-dimensional learning driven by NGSS-related reform initiatives. Identifying successes and challenges in student learning data can provide clarity in adapting instructional decision making to support student learning. This work also has the potential to locate critical areas to develop research-based learning progressions in students' conceptual understanding, particularly in the context of an integrated approach to learning science (e.g. students' model-based explanations of erosion and weathering in the context of pothole formation).

Acknowledgement

This work was supported by funding through the National Science Foundation, award number 1721163. The opinions expressed are those of the authors and do not represent the views of the Foundation.

References

AAAS. 1993. *Benchmarks for Science Literacy*. Project 2061, American Association for the Advancement of Science. Oxford UP, Oxford.

Advisory Committee for Environmental Research and Education (ACERE) (2009). *Transitions and Tipping Points in Complex Environmental Systems*. A Report by the NSF Advisory Committee for Environmental Research and Education.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education.* Chapel Hill, NC: Horizon Research, Inc

Berkowitz, A., Ford, M., & Brewer C. 2005. A framework for integrating ecological literacy, civics literacy, and environmental citizenship in environmental education. *Environmental*

*Education and Advocacy. Changing Perspectives of Ecology and Education.*11, 227-266. U.K. Cambridge University Press.

Horizon Research, Inc. (2013). *2012 National survey of science and mathematics education: Highlights Report.* Chapel Hill, NC.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Louca, L. & Zacharia, Z. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review,* 64(*4*), 471-492.

National Research Council (NRC). 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

NGSS Lead States (2013). *The next generation science standards: for states, by states.* Achieve, Inc., on behalf of the twenty–six states and partners that collaborated on the NGSS.

Strauss, A., & Corbin, J. (2008). *Basics of Qualitative Research* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72(2), 131-175.

Wysession, M.E. (2013). The Next Generation Science Standards and the Earth and Space Sciences. *The Science Teacher*. April/May, 31-37.

Zangori, L., Peel, A., Kinslow, A., Friedrichsen, P., & Sadler, T. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching*, 54(*10*), 1249-1273.