Title
Designing Critique Guidance for Revision of Science Ideas to Promote Self-Directed Learning

Author
Harrison, Emily Jean

Publication Date
2020

Peer reviewed|Thesis/dissertation

Designing Critique Guidance for Revision of Science Ideas to Promote Self-Directed Learning

By

Emily Jean Harrison

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Science and Mathematics Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Marcia Linn, Chair
Professor Michelle H. Wilkerson
Professor Rasmus Nielsen
Professor Sophia Rabe-Hesketh

Summer 2020

Designing Critique Guidance for Revision of Science Ideas to Promote Self-Directed Learning

by

Emily J. Harrison

Abstract

Designing Critique Guidance for Revision of Science Ideas to Promote Self-Directed Learning

by

Emily J. Harrison

Doctor of Philosophy in Science and Mathematics Education

University of California, Berkeley

Professor Marcia Linn, Chair

Scientific thinking involves continual critique and revision of ideas as an individual encounters new evidence and novel concepts. However, students often struggle to integrate new ideas with their prior knowledge, and as a result they continue to hold conflicting ideas about scientific phenomena. This is often reflected in students' revisions of scientific explanations, where they add unrelated information onto the end of their initial ideas. Graphing is another area where students struggle, and can benefit from continual revision. Engaging students in critique, of their own ideas and the ideas of others, has the potential to help students recognize inconsistencies in their scientific understanding. This dissertation research investigates iteratively designed critique guidance to clarify its impact on students' abilities to reevaluate their prior knowledge and successfully revise their science ideas and graphical representations. Emergent findings from these studies revealed the self-directed nature of revision and critique, and led to further investigation of the relationship between these practices and other aspects of self-directed learning, including generation and investigation of scientific questions.

Using the Knowledge Integration (KI) framework, I redesigned two web-based middle school curricular units, *Genetics & Simple Inheritance*, and *Graphing Stories*, in collaboration with teachers and a team of software designers. I conducted an iterative series of studies to clarify the impact of critique in helping students re-examine their prior knowledge, and help them distinguish ideas and improve their revisions of science explanations and graphs. The first set of studies investigated various types of critique guidance designed to help students improve their ability to revise explanations about mechanisms in genetics. The next set of studies includes similar critique guidance and applied it to revisions of student-generated position-time graphs, and looked at students' rationales for how they chose to revise their graphs. The last set of studies, motivated by the apparent self-directed nature of these practices, further investigated the relationship between revision, critique, and students' ability to generate and investigate their own questions.

These studies together help clarify the value of critique in promoting revision of science ideas and graphs. The findings show that success learning from critique depends heavily on prior knowledge of the content material. In the context of genetics and graphing, students attended most to aspects of the activity they already understood well rather than using critique to explore new concepts, and this was supported by students' revision rationales. However, engaging students in critique often encouraged them to revise their ideas more frequently, which is beneficial for improving their scientific understanding. This work reaffirms the value of revision, both for learning and asking further questions. Findings also support the notion that students can successfully critique, revise, and investigate their own questions with sufficient scaffolding. This work has implications for design of online curricula to promote revision, critique, and self-directed learning.

Table of Contents

Acknowledgements

I would like to thank my advisor, Marcia Linn, for providing me with support and mentorship throughout my time at Berkeley. I'm grateful for all the opportunities you've given me to conduct innovative classroom research, and to be a part of the network of researchers and teachers contributing to WISE. I would also like to thank my dissertation committee; I am grateful for Rasmus's scientific expertise, Sophia's guidance with analysis, and Michelle's amazing feedback and emotional support throughout this process.

Special thanks to Beth McBride for letting me follow in her footsteps every step of the way through the SESAME program. I wouldn't have made it without you guiding me through all the hardest parts of grad school, from bureaucratic frustration to picking hotels for conferences, as well as sharing everything you've ever written with me. Thank you to all of the members of the WISE research group for all the support and brainstorming sessions, particularly Jenn King Chen, Lauren Applebaum, Jonathan Vitale, Ady Kidron, Libby Gerard, Sarah Bichler, Korah Wiley, Allison Bradford, Jacquie Madhok, and Eliane Wiese. Of course, I would not have accomplished this without the WISE tech team; a million thanks to Jonathan Lim Breitbart, Hiroki Terashima, and Geoffrey Kwan for all the last-minute fixes to my graphing units, and for making my data tool design a reality. I also owe a lot to David Crowell for making everything at Berkeley so easy for me, and making me laugh every time I stopped by. I would also like to thank all of the teachers that welcomed me into their classroom and shared their enthusiasm for exploring new curricula and technology, especially Dan Gottheiner, Jen Coriell, Robin Stoneking, Ben Leech, and Erica De la Torre.

I could not have accomplished this without my Bay Area friends. I want to thank Laura Armstrong for hanging in there with me through every minute of the last 5 years. Our rooftop sushi and venting sessions were essential, and Tuesdays won't be the same without them. I also want to thank Hannah Gogel, my quaranteam during the covid-19 pandemic without whom I'd still be staring helplessly at a blank page. I appreciate you for indulging my desire for ridiculous adventures, and for introducing me to almost everyone I know in Berkeley. Thanks for reminding me that life is bigger than school, but I should probably still graduate. I am also thankful for my movie committee, Steve McNally, AJ Rahn, and Jeff Litt, who worked hard every week to keep my spirits up and educate me on films I should have seen years ago. I also write this in loving memory of my neighbor, Floyd Adams, and all the happiness and kindness he radiated throughout our neighborhood.

I want to thank my parents, George and Connie, for letting me move three thousand miles away to go to Berkeley and pursue this goal. And my brothers, Mat and George, for sending me ridiculous memes every day to help me process the academic and political climate, and for letting me rant excitedly about science education. And finally, to my fiancé Nabil, thank you for supporting me in every imaginable way. We knew moving to the west coast would be an adventure, and I'm so glad we did this together. I love your unique combination of ambition, silliness, and humility. This pandemic can't stop me from marrying you forever.

Chapter 1: Introduction

The research reported here explores an iterative process to design effective critique activities and clarify how they impact middle school students' ability to integrate new information through revision. Scientific thinking and critique should not only be school activities; too often, students develop a "school answer" to questions about scientific phenomena, while reverting back to their incomplete or inaccurate initial ideas once they leave the classroom (Berland & McNeill, 2010; Linn & Eylon, 2011). Here, I measure the impact of critique by studying how students revise explanations or graphs and generate new questions. I reframe the process of becoming adept at revision as an aspect of self-directed learning, and chose critique as an avenue for exploration due to its role in improving depth of scientific reasoning (Berland & McNeill, 2010), helping students distinguish ideas (Chang & Linn, 2013), and its general absence in middle school classrooms (Donnelly et al., 2020). Additionally, the Next Generation Science Standards (NGSS) build off of previous education reform policies (DeBoer, 1991; National Research Council, 2006; National Research Council, 2007) to further promote student participation in scientific practices, many of which require the ability to critique data, scientific evidence, and your own ideas as well as the ideas of others, in order to actively construct scientific knowledge (NGSS Lead States, 2013). The outcome measures explored here include students' revisions of their scientific explanations and graphical representations as well as student-generated scientific questions. This research, using the Knowledge Integration (KI) framework, emphasizes the importance of engaging prior knowledge in order to successfully build new understanding, and aims to emphasize to students the importance of the knowledge-construction process, through revision and critique, to promote self-directed learning and allow them to become lifelong learners in and outside of the classroom.

With this research, I aim to help support teachers in designing curricula to meet the goals of NGSS and engage students in scientific practices; this work leverages the affordances of a novel web-based technological tool, Web-based Inquiry Science Environment (WISE). For these studies, I redesigned several WISE curricula, all of which were designed to promote knowledge integration for middle school students. Chapter 2 investigates ways to guide students to improve their capability to revise scientific explanations, including activities designed to help them recognize gaps in their knowledge or make the revision process more visible, a valuable skill for developing more sophisticated understandings of science concepts (Cavagnetto, 2010). Chapter 3 takes similar critique guidance and applies it to revisions of student-generated graphs, while looking at the rationales students have for how or whether they chose to revise their graphs. Evidence emerges from these two chapters that revision is a self-directed process; therefore, chapter 4 examines the relationship between revision and other components of self-directed learning, including students generating and investigating their own questions.

Research Questions

This dissertation research addresses the following research questions:

1) How can critique guidance encourage students to distinguish science ideas as reflected in revisions of their written explanations?

2) With insights regarding critique guidance gained from chapter 2:
   a) Can critique guidance help students improve their qualitative position-time graphs?
   b) What rationales do students have for their graph revisions?
   c) How do these compare to, or differ from, critique and revision of written science explanations?

3) What is the relationship between consistent revision of science ideas and student-generated questions?
   a) How can critique guidance support students to investigate their own questions in self-directed explorations of data?

Theoretical Framework

Knowledge Integration

The Knowledge Integration (KI) framework (Linn & Eylon, 2011) suggests the value of engaging students in critique, an essential aspect of scientific argumentation and general scientific literacy (Chang & Linn, 2013; Henderson et al., 2015). Critique is necessary for scientific thinking (Brownell et al., 2013) and can help students distinguish among different scientific ideas (Linn & Eylon, 2011; Sato, 2015). Promoting critique can help students revise their ideas, a neglected activity especially in science (Berland & Reiser, 2011). Students' revisions often involve surface-level changes rather than the reformulation of ideas necessary in science (Bridwell, 1980). The research presented here explores forms of critique guidance that could help students to revise more often and more productively.

The Knowledge Integration framework, along with NGSS, calls for developing coherent, linked understanding in science education (NGSS Lead States, 2013). Another aspect of revision involves constructing and interpreting data visualizations (Friel et al., 2001). Construction, and especially revision, of graphs is an area where students have few opportunities for practice in science (Chinn & Malhotra, 2002; McElhaney et al., 2014). As a result, graph construction is difficult for students. Research suggests that students benefit from the opportunity to sketch their initial ideas or predictions of trends and revise them after receiving guidance (Wu, & Krajcik, 2006; Vitale et al., 2015). In this research I explore the value of critique guidance to help students better distinguish aspects of graphs and what they represent, as well as helping students recognize misleading data representations in the real world.

Integration of new ideas with students' prior knowledge is a basic tenet of the Knowledge Integration (KI) theoretical framework and essential for revision to succeed. Students who are taught science through KI learn by building on their knowledge and exploring new content in the context of their initial ideas. In terms of learning affordances, the KI framework encourages making valid and coherent connections between scientific concepts, using evidence and reasoning (Linn & Eylon, 2011). This framework elicits students' prior knowledge in order to build on their ideas, and promotes discovering and distinguishing ideas, and finally reflecting on newly constructed knowledge (Linn & Eylon, 2006). These steps make this framework ideal for supporting students in revising their scientific ideas because supporting students through challenging practices, such as critique and revision, can be difficult in the context of new and demanding disciplinary content (Scheuer et al., 2015).

This work takes advantage of technology to deliver instruction, and track student and teacher actions. The WISE platform can engage students in graph construction, essay revision, and reflection. Technologically enhanced learning experiences are ideal for supporting revision in the context of knowledge integration (Wang & Hannefin, 2005).

Ultimately, students need to not only make revisions to their ideas when prompted, but to use these capabilities in self-directed learning (SDL). Recent research shows the value of supporting student engagement in the learning process with the goal of developing SDL (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Bannert et al., 2015). Building on my findings for guidance, in my final empirical chapter I explore the benefit of promoting more self-directed learning (SDL) through the use of the WISE online curricula to support student-generated questions. While recognizing gaps in understanding is a big part of revision (Flower & Hayes, 1981), strategies and motivation to revise by engaging more with the material is another major factor. Students who consistently revise, even if their revisions do not improve their responses initially, often have greater learning gains across experimental conditions (Tansomboon et al., 2017). Therefore, chapter 4 explores this relationship to try and isolate productive ways to promote self-directed learning.

Literature Review

Revision of Ideas in Science

This research involves students' construction and revision of explanations and arguments about complex scientific ideas. Writing in schools is historically designed for communication of ideas, rather than for articulating thoughts and ideas in an exploratory way, but writing to learn can enhance understanding of science ideas (Brownell et al., 2013); Rivard (1994) showed that both expository and expressive writing in science enhanced student learning. Additionally, writing and revising ideas can help students more effectively integrate ideas with their prior knowledge (Linn et al., 2014). Writing is taught to students as a linear process (Flower & Hayes, 1981). Science is often taught this way as well, rather than as a messy process involving failures and paradigm shifts (Campanile et al., 2013). In reality, writing involves a non-linear combination of planning, translating, reviewing, and metacognitive monitoring (Flower & Hayes, 1981). The nature of scientific thinking also involves continual revision of ideas as an individual encounters new evidence and experiences new phenomena (Berland et al., 2015).

Revision is stressed prominently by the Next Generation Science Standards (NGSS). The NGSS were developed around the idea that students continually build on and revise their knowledge (NGSS Lead States, 2013). Several NGSS science practices, including *constructing explanations, engaging in argument from evidence*, and *obtaining, evaluating, and communicating information*, describe an iterative process of incorporating new ideas and evidence into continually constructed scientific knowledge. Tansomboon (2017) found that students who revised during a science unit, even if their revisions did not directly improve their responses, gained more from pre to posttest; this suggests that revision helps students grapple with the material in new ways, resulting in better scientific understanding. Brownell et al. (2013) also found that revisions of science writing based on feedback resulted in clearer communication. Students need to be able to understand the epistemic nature of their arguments, citing evidence and reasoning rather than appealing to authority (Jimenez-Aleixandre et al., 2000); the process of revision can reinforce the epistemic nature of constructing scientific knowledge.

However, you cannot simply prescribe expert actions to novice students (Flower et al., 1986); it's more effective to build upon metacognitive skills such as diagnosing problems, recognizing inconsistencies in knowledge, and critiquing your own work by asking questions of yourself. When asked to revise, students often make surface-level changes rather than deeply evaluating their work and making substantial revisions to content (Bridwell, 1980). This is especially true when writing scientific explanations and arguments; surface level revisions do not involve integration of new ideas after encountering new information. Bridwell (1980) found that students who reread their drafts in their entirety had better, deeper-level revisions; while not surprising, this is an important beginner skill to help students practice. Peer feedback tends to be surface level as well (Zheng et al., 2015); expert guidance is sometimes needed to encourage conceptual revision.

This work explores the value of critique in helping students distinguish new information to revise their scientific ideas into more sophisticated explanations. Critique is a useful tool for developing metacognitive awareness of your own understanding (Henderson et al., 2015). Critique also depends heavily on prior knowledge (Donnelly et al., 2015; Flower et al., 1986); students with low prior knowledge have a more difficult time critiquing their ideas or ideas of others, making it difficult to revise. However, explicitly supporting students in critique has been shown to increase the complexity of written explanations despite new and challenging content material (Berland & McNeill, 2010). In fact, generating science explanations without critique limits the depth of scientific reasoning, making the learning of science less effective (e.g. Chang & Linn, 2013). Technology can help scaffold the critique process, and can model the way science is done as well as make the revision process itself more visible (Tala & Vesterinen, 2015; Starbek et al., 2010; Breaky et al., 2008; Johnson & Stewart, 2001).

Graphing in Science

Due to the ubiquity of data availability via technology and the internet, graph literacy is an important skill, both in school science and for general scientific literacy. Beyond just interpreting data visualizations, the NGSS Science and Engineering Practices (SEP) promote going further: *analyzing and interpreting data, engaging in argument from evidence,* and *obtaining, evaluating, and communicating information* (NGSS Lead States, 2013). This involves construction of graphs beyond simple rote point-plotting, and critiquing your own and others' data visualizations (Vitale et al., 2015; Lai et al., 2016), "thinking with graphs" to interpret data in light of real-world scientific phenomena (Shah & Hoeffner, 2002; Donnelly-Hermosillo et al., 2020). Students need to be able to use graphs to communicate patterns, make predictions, elucidate mechanisms, and investigate relationships between variables.

Specific attention in literature is paid to position-time graphs because students have a great deal of difficulty with this type of graph (Boote, 2012; Brasell, 1987; McDermott et al., 1987), especially constructing qualitative position-time representations (Vitale et al., 2015). A common non-normative idea around position-time graphs include slope-height confusion, where students interpret line segments higher on the graph as representing a faster moving object, rather than attending to the slope of the line (Bell & Janvier, 1981; Clement, 1985). The abstract nature of this type of graph requires specific guidance to help students overcome these common misconceptions. Constructing graphs, especially qualitative representations, is a great way for students to learn how abstract aspects of data representations correspond to what is happening in the real world, such as slope and line direction (Vitale et al., 2015).

A major factor contributing to middle school students' difficulty is that they are rarely asked to interpret, let alone construct, graphs in science classes (Boote, 2012). Even less common is the opportunity for students to revise their own graphs. Students' graph construction can benefit from all the same processes involved in revision of written material, including improved ability to diagnose gaps in your own knowledge and other metacognitive skills, and improved communication of ideas. Opportunity for revision allows students to distinguish their ideas, reflect on their thinking, and improve their graphical representations to better communicate their ideas (Lai et al., 2016).

While constructing graphs is important, especially for learning about graphs themselves, use of large real-world datasets is a more practical skill; it is rare to plot individual points on a graph outside of school. More likely, students will use spreadsheet tools that auto-generate graphs from data, and they will need to be able to investigate real-world questions using the data through an inquiry process (Wolff et al., 2016; Schutt, 2013; Vahey et al., 2006). This skill is increasingly important as interactions with data become more common and people make judgements from data more frequently (Haddadi et al., 2015). Therefore, this work goes beyond graph construction to include data manipulation with generated graphs that students use as evidence to support conclusions in their own investigations.

Self-Directed Learning (SDL)

Critiquing and revising your own ideas are integral to self-directed learning (SDL). Historically popular in adult education, SDL is becoming recognized as a valuable goal for students of all ages (Gatewood, 2019; Hmelo, 2004). There are many definitions of SDL, but an aspect common to all involves giving students a broader role in the selection and evaluation of learning materials (Loyens et al., 2008; Fisher et al., 2001), giving the student, rather than the teacher, the power to define their own learning activities (Schmidt, 2000). Other components of SDL include personal autonomy, self-management, independent pursuit of learning, and learner control of instruction (Candy, 1991), as well as metacognitive awareness, ability to set learning goals, ability to identify gaps in knowledge that need to be investigated further, and ability to select appropriate learning strategies (Hmelo, 2004).

A prominent indication of self-directed learning is the ability to generate meaningful questions to both check understanding and pursue more personally meaningful learning goals (Biddulph et al., 1986). Generating, as opposed to simply answering, questions has, additionally, been recognized as a skill that promotes agency and engagement (King, 1992), particularly when science and engineering practices are involved (Barton & Tan, 2018). Students who generate questions about scientific concepts at higher levels have shown enhanced understanding of the content (Chin & Brown, 2002). Self-assessment is a crucial component of SDL as well (Candy, 1991; Blumberg, 2000). This skill is necessary for revision of students' ideas; revision involves the need to assess the gaps in your own knowledge (Flower et al., 1986) and investigate these further in order to revise by integrating newly learned ideas with prior knowledge (Hmelo, 2004; Liu et al., 2008).

Even within adult education, students still struggled with self-efficacy in recognizing gaps in their understanding and defining and determining their own learning issues to investigate (Dahlgren & Dahlgren, 2002; Loyens et al., 2008). Middle school students expectedly struggle with these issues as well, necessitating more guidance to achieve a higher level of self-directedness. Technology has shown to have a positive effect on self-directed learning and student engagement (Bannert et al., 2015; Muller & Seufert, 2018), and can provide a useful environment for appropriate scaffolding of various self-directed activities, including asking and investigating your own questions and revising scientific ideas (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Bannert et al., 2015). Self-directed learning is a learning goal as well as a process, and has a learning curve (Candy, 1991; Hmelo, 2004); a major goal of this work is to see what types of activities promote effective development of self-directed learning in middle school students.

Curricula and General Methods

Web-Based Inquiry Science Environment (WISE) Platform

All studies take place in the context of the Web-based Inquiry Science Environment (WISE) platform, which includes open-source online units covering various middle school content areas (Linn, Clark, & Slotta, 2003; Slotta & Linn, 2000; Linn & Eylon, 2011). These units are designed to meet national standards as well as specific teacher needs, and are continually iterated and improved upon with feedback from students, teachers, and new data after use in classrooms. Student data is logged within the WISE platform, including written responses, revisions, and graphs. This platform is an ideal setting for this research. This platform captures all student work, allows us to redirect students back to relevant material, create branching into various guidance conditions randomly by student ID, and import previous student work to allow revision of their ideas and graphs. The technological platform also supports embedded multimedia such as videos, interactive models, and various graphing tools designed to help students discover new scientific ideas and scaffold graph construction. All units are designed according to the KI framework, which elicits students' prior knowledge in order to make connections and integrate new science concepts.

Participants

All studies are conducted in Bay Area middle school classrooms with teachers that work intimately with our research group, many of whom attend professional development workshops in our group. Students work in teacher-assigned teams on the unit, and individually complete a pretest and posttest one day before and one day after completing the unit, respectively.

Assessments

For most items analyzed, student open-response essays are scored using a KI rubric, on a scale from 1-5 that rewards making links between normative science ideas and previously unlinked concepts generated from students' prior knowledge. For revision items, scores are given to initial and revised essays, allowing us to calculate gains due to revisions. For revision items, I also give various revision scores that note whether students revised by adding new ideas (as opposed to simply elaborating on ideas already present in their initial response), and whether students made connected revisions (as opposed to tacking on unrelated information to the end of their initial essay). All rubrics were co-developed within the WISE research group, and revised until an inter-rater reliability of 0.9 or above was reached. I also designed and utilized rubrics to capture the types of revisions students make on their own graphs, as well as the types of questions they generate about the material being studied.

Genetics and Simple Inheritance Curriculum

In this research, I employ one of the WISE units focused on teaching genetics and simple inheritance (for one version of the unit used in this research, see https://wise.berkeley.edu/project/25176) that I redesigned to address several high-level questions about inheritance and genomics, including what makes us look the way we do and what happens when the genetic code is altered. This unit covers content contained in the middle school NGSS science standards, while also delving into aspects of genetics pertinent to general scientific literacy; due to the growing presence of genetics in media and the public interest, developing and continually refining and evaluating one's knowledge of genetics and the mechanism of inheritance is of increasing importance. Additionally, I added several opportunities throughout the unit where students can record their initial ideas and revise them later in the unit. Two embedded essay questions include an activity which asks students to explain the genetic mechanism that makes siblings look similar but not the same, and another activity embedded within this WISE unit which asks students to explain how to determine probability of inheriting a trait using a Punnett square.

Graphing Stories Curriculum

This research also explores another unit in WISE, the Graphing Stories unit (for example see: https://wise.berkeley.edu/project/24729), which was designed as an introduction to interpreting and constructing data visualizations, while also addressing several NGSS science and engineering practices (NGSS Lead States, 2013). The curriculum focuses mainly on constructing and interpreting position-time graphs and includes animations that match up to student-constructed graphs to give visual feedback. Students construct graphs with various embedded graphing tools and have several opportunities to revise their graphs after receiving automated forms of feedback. I made several modifications to the unit to include different types of graphs, including scatter plots and bar graphs, in order to encourage students to recognize that different graphs all show a relationship between variables. In addition, I added several problematic graphs from the internet and newspapers for students to practice critique.

Overview

Together, these chapters explore revision of both essays and data visualizations as a way of tracing student thinking about topics middle school students historically find very difficult. Constructing knowledge, in writing or through graphing, is a very challenging process, and requires integrating new ideas with prior understanding to develop working scientific knowledge.

Revision of science explanations and arguments as well as data visualizations is an important skill for distinguishing, integrating, and communicating ideas as well as reflecting on your own learning. However, many students do not revise their ideas when prompted, maintaining their novice ideas in light of new information. This work aims to investigate what types of guidance help students revise more often and more productively. For example, is it more important to help students develop metacognitive monitoring processes to critique their own gaps in understanding, have students revisit relevant material to obtain new ideas, or to model how the revision process works? This work also investigates why students choose not to revise, and what rationales students have that do revise their work.

While the ability to critique ideas and recognize gaps in your own understanding helps with revision of ideas in science, revising also requires a degree of self-directedness. This prompted investigations into other self-directed learning processes to see the relationship between them and what type of activities can help improve students' self-directedness. A major goal of education in recent years is to have students develop and investigate their own questions about scientific phenomena. However, like revision, this is a skill that requires practice and guidance, especially for middle school students. Revision, as well as other elements of SDL, should be considered learning goals as well as learning processes, and students should be allowed ample opportunities to practice these skills.

Chapter 2: The Evolution of Guidance for Revision of Student Ideas about Genetics and Inheritance

This chapter investigates various forms of guidance designed to help students improve their ability to revise their scientific explanations about the mechanism of simple inheritance. The complex nature of this topic has proven difficult for middle school students to make sense of at more detailed levels. Revision can result in improved explanation and understanding but is practiced infrequently in middle school science classrooms. Three comparison studies are presented here, in three consecutive redesigns of an online genetics unit, to investigate which guidance activities help students revise their ideas successfully.

The first study investigates the benefit of critiquing common student misconceptions and non-normative statements about how genetic inheritance works; this explicit critique activity was designed for students to practice detailed analysis and assessment of a claim or theory, a skill necessary for the process of evaluating scientific information and evidence, as well as constructing and revising explanations. This was compared to a condition where students were directed back to relevant material from earlier in the curriculum, a strategy previously found to help students improve their scientific ideas.

The second study employs an annotator tool, designed to model the revision process for students unfamiliar with the general skill of revising their writing. This was compared to the benefit of revisiting relevant material in a more interactive way, having students label output of models they've used in the unit to reinforce information previously learned.

The third study compares the two most successful conditions from the first two studies, comparing explicit critique to the annotator tool, to see which is more beneficial in order to help us better understand what underlying skills are helping students revise their scientific explanations in each case.

A combined analysis of all three studies follows, in order to show a bigger picture of what types of guidance encouraged students to revise their ideas more frequently, and what helps them revise more productively.

Together, these studies reveal in more detail the difficulties that students have in parsing through the complex mechanism of inheritance and revising their scientific explanations. These studies also reveal promising methods of helping students improve their understanding and communication of ideas about genetics through revision, resulting in more coherent, connected, and/or mechanistically accurate scientific explanations.

Introduction

Revision of scientific explanations and arguments is an essential practice in learning and communicating science (Brownell et al., 2013). However, students have few opportunities to write in science class or modify their ideas throughout the learning process (Shepard, 2000). Therefore, students have a difficult time constructing coherent explanations. Writing in science can greatly enhance learning and understanding of difficult and abstract scientific concepts. Writing is one way that students can begin to take ownership of their science knowledge, and can help them integrate new complex ideas with their prior knowledge (O'Neill, 2010).

Students especially struggle with revising explanations in a way that incorporates new ideas or evidence (Berland & Reiser, 2009). Many students learn new science concepts separately from the context of their prior knowledge, and often return to their naive initial ideas when asked to construct explanations later (Mercier & Sperber, 2011). Revising ideas can help students integrate new concepts, especially those students with low prior knowledge of the content material. Revision is also stressed prominently by the Next Generation Science Standards (NGSS). The NGSS were developed around the idea that students continually build on and revise their knowledge (NGSS Lead States, 2013). The NGSS science and engineering practices (SEPs), particularly *constructing explanations, engaging in argument from evidence,* and *obtaining, evaluating, and communicating information*, include an iterative process of incorporating new ideas and evidence into continually constructed scientific knowledge. Although revision is a necessary skill for gaining the integrated understanding called for by the NGSS, there are many reasons students have limited opportunities to revise in a typical science classroom (Berland & Reiser, 2011).

Furthermore, students are challenged by NGSS not just to revise, but to make productive revisions. Many studies over the past 30 years have shown that when students revise their writing, if they revise at all, they often maintain conflicting ideas, tack ideas onto their initial response without integrating information, or they only make surface-level grammatical changes without incorporating new content (Rivard, 1994; Bridwell, 1980; Crawford et al., 2008). This suggests that prompting students to revise is not enough to support students in making meaningful or integrated revisions to their understanding. Students need more scaffolding to help them practice skills that lead to these types of revisions, including the ability to diagnose inconsistencies in their understanding and communication of their thoughts (Flower & Hayes, 1981). Additionally, literature is lacking on revision in science, which can include much more complex revision of ideas in addition to revision for clarity of communication.

Among other guidance tools, these studies explore the value of critique in helping students distinguish new information and revise their scientific ideas into more sophisticated explanations of the content. Indeed, critique is not simply an exercise, it is fundamental for epistemic vigilance and critical thinking (Henderson et al., 2015). Construction of explanations without critique limits the depth of scientific reasoning making the learning of science less effective (e.g. Chang & Linn, 2013). Written arguments and explanations depend heavily on content knowledge, but explicitly supporting students in skills such as critique has been shown to increase the complexity of written explanations even in the context of new and challenging content material (Berland & McNeill, 2010). The value of critique has been demonstrated repeatedly in the context of genetics (Jimenez-Aleixandre & Duschl, 2000; Zohar & Nemet, 2001), making the WISE Genetics unit an ideal context for this study.

Theoretical Framework: *Knowledge Integration*

The following studies draw on and build upon research using the Knowledge Integration (KI) framework. Knowledge integration serves as an overarching framework for designing all curricula, interventions, and guidance for all studies presented here. This framework is built upon constructivist foundations that support the idea that students do not learn ideas in a vacuum, but construct knowledge by building on their prior ideas (Linn et al., 2014; Liu et al., 2008). Many students learn new science concepts separately from the context of their prior knowledge, and often return to their naive initial ideas when asked to construct explanations later (Mercier & Sperber, 2011). In order to support student learning, especially in complex topics such as genetics, it is important to take into account the ideas that students bring to a topic. The KI framework does this by *eliciting* student ideas, *adding* new content, encouraging *distinguishing* of ideas, and prompting *connection* of new ideas and understanding to prior knowledge (Linn & Eylon, 2006). These steps make the KI framework ideal for supporting students in revising their scientific ideas, where students build and improve upon their initial understandings. Throughout the following studies, students' ideas are elicited through embedded essay questions, encouraging students to explicate their existing ideas regarding DNA and inheritance of traits. The various guidance conditions are built to help students either *add* ideas or *distinguish* between new information and their prior knowledge in various ways, particularly our critique activities. Students then revise their original essays to help them *reflect* and *connect* all of their ideas in an integrated way.

Curriculum

In the three studies presented below, all instructional materials were part of an online unit in the Web-Based Inquiry Science Environment (WISE) platform. This is an open-source platform ideal for designing inquiry activities, and leveraging the use of interactive models, dynamic visualizations, and assessments (Linn, Clark, & Slotta, 2003). The WISE unit used in all three studies is on the topic of genetics and simple inheritance, and covers several NGSS disciplinary core ideas (DCIs) at the middle school level on this topic. Throughout the unit, content learning goals include independent assortment of alleles, dominance and recessivity, probability of inheriting certain traits depending on parental genotypes, tracking of alleles through several generations using a pedigree, the relationship of DNA to phenotypic expression, and how the environment and human influence lead to changes in genetic expression and inheritance over time. This unit was designed for the first study in accordance with the Knowledge Integration (KI) theoretical framework, which aims to support connecting new scientific content with students' prior knowledge. The unit was subsequently modified in small ways to meet the research questions of the following studies.

I redesigned the WISE Genetics and Simple Inheritance unit to address 4 essential questions related to genetics (Figure 2.1): (a) Why do we look the way we do? (b) How can we predict disease? (c) How do mutations affect DNA? (d) How do we control our world with genetics?

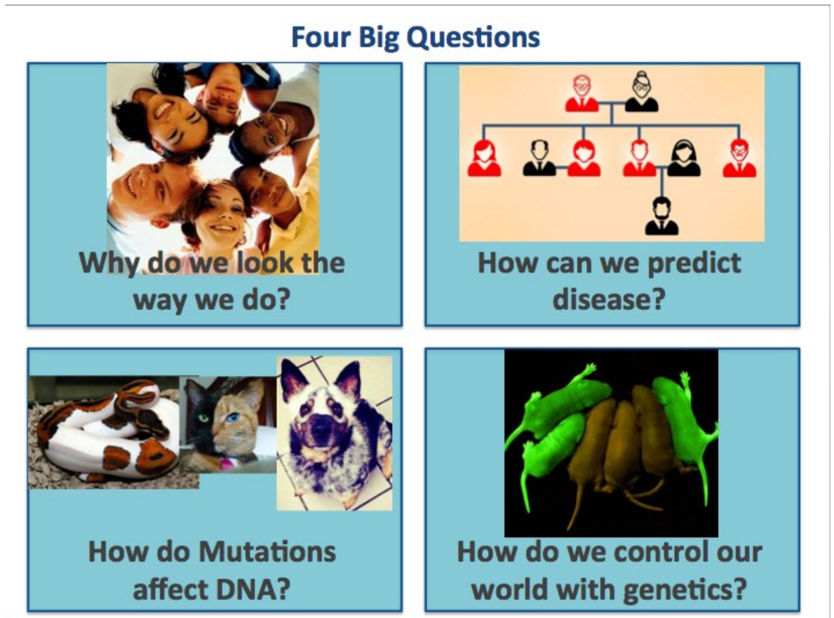


Figure 2.1: Introduction to the essential questions covered in the WISE Genetics and Inheritance unit (Link https://wise.berkeley.edu/project/20893#!/project/20893/node3)

To address these essential questions, it is important to understand both the mechanism of simple inheritance and that inheritance is rarely simple. This unit does this by having students discuss and explore genetics at varying levels; for example, students learn about DNA base pairs, how DNA is organized in a cell, two-generation and several-generation pedigrees with Mendelian traits, and traits that are controlled by more than one gene. The WISE platform scaffolds this by employing manipulatable computer models and various interactive question types to help students understand the complexity of genetics while simplifying situations into solvable problems. Students explore several of their own traits through visuals that allow them to explore different combinations of alleles (Figure 2.2).

Drag the letters into the allele boxes to make different combinations of alleles (genotypes) and see what kind of dimples *phenotype* results:

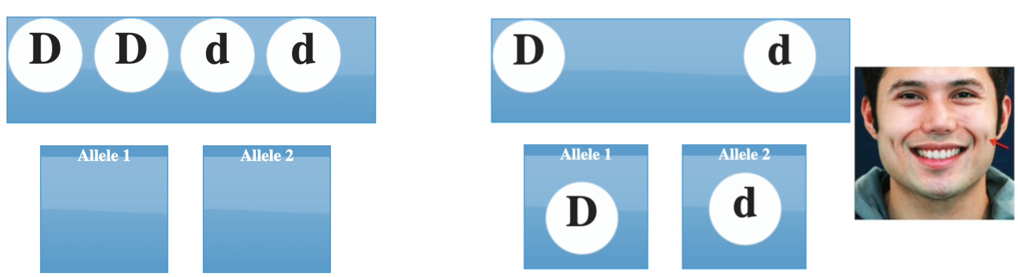


Figure 2.2: Interactive model allowing students to test different combinations of alleles and see phenotypic outcomes.

Supporting students through challenging practices, such as critique and revision, can be difficult in the context of new and demanding disciplinary content (Scheuer et al., 2009);

technology in the WISE platform is ideal for supporting these two tasks simultaneously. Multiple choice or matching items are used throughout to give students immediate feedback to check their understanding before moving on. These items also provide visible formative feedback for teachers circulating the room as students work at their own pace. In addition, this platform is useful for recording student behaviors; the WISE platform is able to track and save all student work from various assessments embedded throughout the unit, including revisions made on essays. Students' original writing can be imported to pages later in the unit, allowing them to revise with their new understanding in the context of their prior knowledge. Additionally, this technology enables random branching to different guidance activities, allowing for random and unbiased assignment of comparison conditions by WISE ID.

The three studies discussed below aimed to test guidance with the goal of encouraging revision to help students move beyond rote skills and towards purposeful and usable construction of knowledge. Each study compares two guidance conditions that target specific skills hypothesized to be important for productive revisions. Previous studies in this group have shown that using the technological environment to direct students back to relevant material in the unit helped students revise their ideas more frequently and more successfully. I build on these findings by directing students back to material already explored, but with new questions or ways of interacting with the material. This type of guidance is referred to as *revisiting* in the following studies, but involves students answering new open-response questions or labeling output of models they already explored earlier in the unit. The goal of this guidance is to help students *add* new ideas to their essays that they might have overlooked the first time through the material. In the first study, this is compared to *critique* guidance, which provides students with several common incorrect student explanations about the mechanism of inheritance, and asks students to explicitly write what is incorrect about the statement and how they would improve it, and are then directed to their own initial essay to revise. This is designed to help students *distinguish* more explicitly between commonly held incorrect ideas and new knowledge that they have learned at this point in the unit. The third guidance activity involves an *annotator* tool, that presents a vague or incorrect essay by a fictional student, and students place labels on the essay to suggest where and how to improve the explanation. This tool is meant to help students not only distinguish vague or incorrect ideas to make them more precise or mechanistically accurate, but also to make visible how to revise in a more integrated way. Students add labels at various points throughout the fictional essay to show that changes should be made throughout. This is to address the issues that students often tack inconsistent (if correct) statements onto the end of their original essays rather than changing and integrating new ideas with their prior knowledge.

Study 1: Encouraging revision of scientific explanations with critique in genetics

This study investigates how engaging students in critique can promote productive revision of student written explanations. I conducted this study in the context of the WISE Genetics and Simple Inheritance unit with 8th grade students. I compared critique activities to having students revisit relevant material and interactive models to gain new ideas. Both strategies promote knowledge integration by encouraging students to look for concepts that are missing from their explanations and incorporate them into their existing ideas, and making valid and coherent connections between scientific concepts as well as using evidence and reasoning (Linn and Eylon, 2011).

Students in the *critique* guidance condition explicitly explained what was wrong or missing from several common non-normative student ideas regarding difficult topics in genetics. This condition was designed to support the "distinguishing" step in the knowledge integration process, helping students take vague notions about inheritance and parse out a more nuanced understanding. The non-normative statements that students critiqued were developed from ideas that students generally found to be the most challenging. The goal of this activity was to encourage students to consider flaws in their own reasoning, explicate them, and then incorporate these newly distinguished ideas into their own revisions. The activity was designed to help students unpack the complexity of constructing and revising scientific ideas by focusing explicitly on the practice of critique. This was compared to a method used in the past to encourage students to add new ideas to their essays; students in the *revisit* condition were directed back to relevant information and interactive models rather than practicing critique, supporting the "adding ideas" step of knowledge integration. Students were prompted to attend to specific aspects of the revisited material with new questions. This study was designed to investigate whether critique, as a skill, encourages more nuanced understanding of complex scientific phenomena as well as whether practicing this skill encourages more frequent and productive revision of student ideas.

Methods

Participants and procedures

Three teachers from two middle schools participated in this study, with a total of 13 classes of 8^{th} grade students (277 students, 197 student teams). Teacher 1 taught 4 classes at the first school (49% non-white, 32% free/reduced lunch, 7% ELL). Teachers 2 and 3 taught at the second school (62% non-white, 22% free/reduced lunch, 12% ELL). Students completed the 8-day Genetics WISE unit during 50 minute class periods. Students worked in collaborative workgroups assigned by their teachers, mostly pairs with a few students working individually or in groups of 3. Students completed the pretest one day before beginning the unit, and the posttest one day after completing the unit. Both pre and posttest were completed individually. Workgroups were randomly assigned to one of the two conditions (*critique* or *revisit*) by the computer based on their WISE Workgroup ID. All teachers implemented the unit as planned, during 50 minute class periods over the course of two weeks. Teachers intermittently reminded students of guidelines for productive collaboration with their partners throughout the project.

Curricular materials

Throughout the two-week-long unit, students encountered their assigned activity, *critique* or *revisit* (see Table 2.1 for outline of conditions), while the rest of the unit was identical for all students.

Table 2.1: Outline of a sample sequence of *critique* and *revisit* conditions.

Critique	**Revisit/ Experiment**
Essay Prompt: Explain how you would use a Punnett square to figure out the probability of getting a certain genotype.	
Critique **Example Critique:** **Student 1**: "Their 4th child will have attached earlobes." Explain how this statement is incorrect or too vague, and how to make it more accurate.	**Revisit** Use the Punnett square model again to answer the questions below. (Model is embedded on the next page, along with the questions) **Example Question:** When both parents have the genotype EE, what is the probability of having a child with attached earlobes? Explain.
Revise: Now that you've learned a bit more about Punnett squares and probability, take some time to *revise* or *improve* your answer to this question from earlier. What is a method you would use to calculate probability of getting a certain genotype using a Punnett square? *(Students' original responses are imported automatically)*	

The unit features an interactive Punnett square model to help students see the effects of allele combinations on specific phenotypes (Figure 2.3). It also includes drag and drop questions to help students sort evidence and receive immediate feedback, and interactive graphing tools to allow students to construct visualizations of data illustrating connections between different genotype crosses. Activities were added that focus on genetic modification, both through artificial selection and engineering, as well as common mutations and their effects.

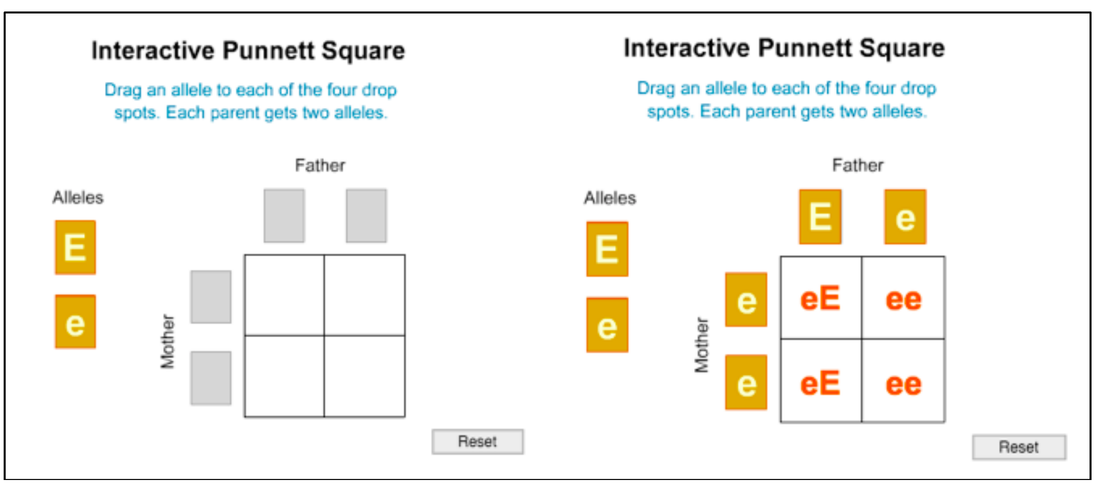


Figure 2.3: Screenshots of the interactive Punnett Square model; students drag different alleles to populate the boxes and test different combinations

Assessments

The pretest and posttest were used as measures of student knowledge integration, looking at students' prior knowledge and relative improvement. These items asked students for a written explanation that required synthesis of several genetics concepts. Two relevant open-response items were scored and these scores were averaged for students' overall pre/post scores. For example, one pretest question (*SiblingsPrePost)* prompted: "Siblings look similar, but not exactly the same unless they are identical twins. If they inherited their DNA from the same parents, why don't siblings look exactly the same? Explain." A critique-style question was also included for all students on the pre and posttest; Students were given a completed Punnett square showing two parents heterozygous for brown/blue eyes and were asked to predict the probability of having a child with blue eyes. The follow up critique question (*CritPrePost)* prompted: "Another student said: 'From the Punnett square, you can tell that the couple's 4th child will have blue eyes.' Do you agree or disagree with this statement? Explain."

With the help of the WISE group, I scored students' initial and revised explanations using KI rubrics to measure the value of critique versus revisit guidance. Specifically, I looked at the types of revisions students made on two embedded KI synthesis essay questions. The first question was identical to the first pre/post question regarding the mechanism of inheritance: Why do siblings from the same parents look similar but not exactly the same? (*Siblings*). The second question focused on the use of a Punnett square to determine the probability of getting a certain genotype in various cases (*PunnettSquare*): "What is a method you would use to calculate probability of getting a certain genotype using a Punnett square?"

Student essay responses on the pre and posttest, as well as embedded essay items, were scored using a 5-point Knowledge Integration (KI) scale (See Table 2.2 for sample rubric). KI scoring is designed to reward students for making connections between ideas, thereby integrating new information with their prior knowledge (Linn & Eylon, 2011; Liu et al., 2008). The rubric for the *Siblings* question demonstrates how links are scored (Table 2.2).

Table 2.2: KI Rubric Example for the item: "Why do siblings from the same parents look similar but not exactly the same?"

KI Score	**Description**	**Examples**
1	No Answer	"I don't know"
2	Non-normative/irrelevant: Token mechanism only ("skips a generation") with no elaboration. Incorrect ideas: "you get different amounts of DNA from each parent"	Because you and your sibling have close genes but they are not the same genes. You inherit similar amounts of the same traits from the same parents at slightly different amounts. Because it's not exactly the same, you look a little different.
3	Partial link (one correct statement, but not connected to other scientific ideas, or student does not elaborate)	They get different parts of dna from their parents. You get a different set of genes then your sibling.

4	One full link between normative scientific ideas	Because you get half of your parents DNA but it does not specify which half you will inherit from them. This means that the half that you might get will not be the same that your sibling will get.
5	At least two full links	Siblings do not look exactly the same because they have slightly different alleles. Each child has a chance of receiving a different allele from its parents than its sibling because of probability.

I also analyzed the *way* students revised on our two embedded essay questions: *Siblings* and *Punnett Square*. Specifically, I was interested in whether students made revisions that were connected to their prior knowledge, rather than simply tacking on to the end of their original explanation. Additionally, I was interested in whether students added new scientific ideas to their explanations, or if they simply elaborated on or clarified their original ideas. Therefore, qualitative revision codes were given to the embedded essay revisions based on how a student revised. A code was given for whether students made connected (C) or disconnected (D) revisions. Another code was given for whether students added new (N) ideas in their revision or expanded existing (E) ideas that were already present in their initial response (see table 2.3 for examples of each code combination).

Table 2.3: Rubric and examples for embedded essay revisions (Student revisions underlined)

Initial Response	**Revision**	**Score (C/D) (N/E)**	
You would use the method of counting by 25's. Each square is a 25% chance.	You would use the method of counting <u>by quarters. The two letters from each parent would represent a quarter of the genotype and all the quarters combined would show what phenotype would be dominant over the other.</u>	C	N
A method I would use is that I would choose the number of squares that have a certain genotype and find out what percent of the squares have that genotype.	A method I would use <u>to calculate the probability of getting a certain genotype</u> is that I would choose the number of squares that have a certain genotype and find out what percent of the squares have that genotype.	C	E
You would put the alleles of each parent on the outside of the square and the possible alleles for their children would be in the square.	You <u>can find the dominant and recessive traits to calculate the probability out of four.</u>	D	N

The figure shows how many possible genotypes that children can have.	The figure shows how many possible genotypes that children can have. <u>D means dimples and d means no dimples.</u>	D	E

Results

Pre/posttest analysis (Individual Students)

<u>Learning gains</u>

Students began the unit with moderately low genetics knowledge, with an average pretest KI score of 2.49. Students in the *critique* and *revisit* conditions were not statistically significantly different at the beginning of the unit [*Critique*: Pre mean=2.51; *Revisit*: Pre mean=2.46, t(276)=-0.565, p=0.714].

An average of two pre/posttest items were used to calculate learning gains for students from the Genetics unit. These items were chosen for analysis because they require the synthesis of various genetics concepts. The item *SiblingPrePost* involves independent assortment of alleles, leading to offspring from the same parents receiving different combinations of alleles, and different genotypes resulting in different phenotypes. The item *CritPrePost* asks students to explain how the probability of a certain trait can be predicted using a Punnett square, and involves a critique component. A t-test for all students (in both conditions) showed that students achieved learning gains in genetics by the posttest (average of the 2 pre/post item scores: *SiblingsPrePost* and *CritPrePost*) [Posttest mean: 3.19, pretest mean: 2.49; $p<0.001$]. A t-test on students' posttest scores by condition showed that there was no statistically significant difference in posttest score between the *critique* and *revisit* conditions [t(276)=1.358; p=0.089]. Students in the *critique* condition had a mean posttest score of 3.13, while students in the *revisit* condition had a mean posttest score of 3.27.

<u>Influence of prior knowledge</u>

Students were then separated into high or low prior knowledge groups based on their pretest score on the average of the two items to see if one condition was more beneficial for students with lower content knowledge. Students that scored below a 3 were considered low prior knowledge (LPK), and those that scored 3-5 were considered high prior knowledge (HPK). This cutoff was chosen because students must include at least one normative scientific idea to achieve a score of 3. With this cutoff, 214 students were categorized as LPK, and 63 students were categorized as HPK.

Regression analysis on only those students with LPK (N=214) revealed that LPK students scored an estimated average of 0.36 points lower in posttest gain in the *critique* condition compared to LPK students in the *revisit* condition [t(213)=-2.06; $p<0.05$]. This may be because success at critique depends on understanding the content material. There was no statistically significant difference in posttest gain score between condition for HPK students.

Revision on Embedded Assessments (Student Teams)

<u>Nature of student revisions and learning gains</u>

Overall, I found that critique motivated adding more new ideas to essay revisions than did revisiting relevant material.

For revisions on the *Siblings* embedded essay question, logistic regression revealed that the odds of a student team in the *critique* condition adding new ideas to their revisions were 1.84 times as great compared to student teams in the *revisit* condition [$z(196)=2.05$; $p<0.05$]. Condition had no statistically significant effect on KI score gain from initial to revised response for this item. However, student teams that did revise on this item scored, on average, 0.41 points higher on their revised essay than students who kept their original answer (did not revise) [$t(196)=3.40$, $p<0.001$]. This suggests that students that revised made productive changes to their essays on this item.

For the *PunnettSquare* embedded essay, logistic regression revealed that the odds of a student team in the *critique* condition revising by adding new ideas were 2.74 times as great compared to student teams in the *revisit* condition [$z(196)=2.72$; $p<0.01$]. In addition, student teams in the *critique* condition scored an estimated 0.20 points higher on this item than student teams in the *revisit* condition from initial to revised response [$t(196)=2.54$, $p<0.05$]. This is likely due to the addition of new relevant science ideas in their revisions.

For both embedded essay questions (*Siblings* and *PunnettSquare*), there was no statistically significant difference between the two guidance conditions in terms of students making connected revisions.

Frequency of revision

Logistic regression showed that the odds of a student team in the *critique* condition revising their explanation on at least one of the two embedded essay assessments were 2.75 times as great as student teams in the *revisit* condition [$z(196)=3.25$; $p<0.001$].

In order to examine the effect of prior knowledge on students' revisions, students were again separated into high and low prior knowledge groups. 161 students were classified as HPK and 36 students were classified as LPK. Logistic regression of only HPK students showed that the odds of a student team in the *critique* condition revising at least one of the two embedded essay items were 2.22 times as great as students in the *revisit* condition [$z(160)=2.26$; $p<0.05$]. Logistic regression of just LPK students showed that the odds of a student team in the *critique* condition revising at least one of the embedded essays were 5.2 times as great compared to students in the *revisit* condition [$z(35)=2.27$; $p<0.05$]. Therefore, not only did the *critique* condition promote revision more often, it especially encouraged students with low prior knowledge to revise more often.

Qualitative Critique Examples

These results suggest that both conditions were effective in promoting student understanding of genetics and simple inheritance from pre to posttest. While low prior knowledge students in the critique condition gained slightly less than high prior knowledge students, the critique activities still encouraged them to revise their responses more often than the revisit condition. For revisions, critique was helpful for adding new scientific ideas, but these ideas were not necessarily connected to their previous responses. Below are examples of revisions that students in the critique condition made in order to see how ideas were being added (see table 2.4).

Table 2.4: Revisions from students in the critique condition (revisions in **bold**) on the question: "How did Eric inherit cystic fibrosis if neither of his parents had the disease?"

Initial Response	**Revision after Critique**
His grandparents passed it to his parents, but the disease skipped a generation	His grandparents passed **down one recessive and one dominant gene** to his parents **and then they both gave him their recessive genes so he would have cystic fibrosis.**
Eric inherited cystic fibrosis because his grandparent had it and it could very possibly skip a generation. Eric's mother most likely recieved the dominant gene which was not getting cystic fibrosis.	Eric inherited cystic fibrosis because his grandparent had it and **[ ]** it skipped their generation. Eric's mother most likely recieved the dominant gene **from her mother and a recessive gene from her father which was hidden and allowed her** not to get cystic fibrosis.
Eric might have inherited cystic fibrosis by his grandpa's genes skipping over his parents and going straight to him.	Eric might have inherited cystic fibrosis by his grandpa's genes skipping over his parents and going straight to him. **His parents could have a heterozygous genotype which would give Eric a 25% chance of developing the disease.**
We think that siblings look similar to each other but not exactly the same because the traits of the parents are different and each child gets different traits from each parent.	We think that siblings look similar to each other but not exactly the same because **they have different combinations of alleles from their parents.**

These examples illustrate how students in the critique condition distinguished their ideas. Rather than keeping their vague answers (ex. "each child gets different traits"), they revised to include more specific scientific vocabulary and more nuanced definitions of phenomena (such as "a heterozygous genotype which could give Eric a 25% chance"). Distinguishing vague ideas to give more detailed scientific explanations is an important part of the KI framework. In this study, the critique condition showed promise in helping students achieve this difficult task.

Conclusions and Implications

Revising ideas after encountering new information is an essential scientific literacy skill. This study shows the benefit of critique activities in promoting revision of scientific ideas. Both conditions in our study were effective in helping all students achieve learning goals in genetics and simple inheritance. Further analysis revealed advantages for critique regarding motivating revision more frequently. This is likely because these students were exposed to flaws in their own thinking by analyzing common incorrect ideas, motivating them to rethink and clarify their original responses. Students in the *critique* condition also regularly added more new ideas to their essay revisions. This is likely because students had to consider their logic more carefully while critiquing, encouraging them to distinguish between ideas, whereas the models revisited in the other condition did not explicitly encourage students to think about the mechanisms of inheritance, such as that of allele movement. While I hope to create guidance and activities that encourage students to make more integrated revisions, studies have found that even attempts at revision have been shown to result in greater learning gains (Tansomboon et al., 2017). Our critique activity was successful at motivating students to at least attempt to revise their ideas more often, especially those students with low prior knowledge in the content area.

Overall, revisions that students made in this unit were highly relevant, and attempted to add value to their responses in the form of new or better-clarified ideas. This is in distinct contrast to studies, including Bridwell (1980) and Crawford et al. (2008), that found most student revisions were occurring at the word or surface level. This again promotes the practice of critique in encouraging students to revise their scientific ideas rather than just their grammar.

While critique was effective in some ways, the next study investigates further ways to encourage students to revise more often, and ways to encourage students to make more connected, integrated revisions of their scientific ideas. The next study introduces a new guidance tool in the WISE Genetics unit that models how to revise in order to help students practice this specific skill.

Study 2: Guiding students to revise scientific essays by modeling the revision process in genetics

Revising scientific ideas is difficult for students, due partly to lack of practice and guidance. Study 1 investigated the effects of writing critiques of incorrect science ideas on student revisions of their own ideas in genetics. This study examines the effects of an essay annotator activity that explicitly models the revision process (*annotator*) compared to an activity where students revisit interactive models from the unit (*revisit*), designed to support interpreting evidence. The essay *annotator* activity was designed to more explicitly model the revision process, prompting students to add labels to a fictional student's essay to suggest where and how to improve their explanation before revising their own ideas. This was designed based on a previous finding that annotating a fictional students' essay helped students with low prior knowledge improve their own revisions (Harrison et al., 2018). Previous work from our group also found that directing students back to relevant models helped students gain more new ideas; the *revisit* activity was designed to scaffold gaining more new ideas from the evidence in the models. There was evidence that some students in the *revisit* condition in study 1 may have overlooked the material meant to help them add ideas, due to the fact that they had seen that information previously in the genetics unit. The *revisit* activity in this study was designed to improve upon the *revisit* condition from study 1. Instead of simply being redirected back to material relevant to revise their explanations, students were directed to label screenshots of output from models they interacted with throughout the unit. For example, students labeled output of a Punnett square model with labels showing what was being represented by the various parts. While students were still looking back at material they had previously explored, this activity involved more explicit interpretation of evidence from interactive models.

Additionally, this iteration included a pre/post revision item to show whether students improved in their ability to revise at the end of the unit (see Figure 2.6). This item consisted of an open response item regarding the interpretation of a pedigree. Students wrote their answer, and on the next page were given a new piece of information about the pedigree, and were prompted to revise their (imported) original essay. This was designed to show students' abilities to integrate new information into their science explanations before learning the content and practicing our revision guidance activities, and show how students in different conditions improved by the end of the unit. For this revision pre/post item, I looked at how students revised as well as looking at overall gains and gains specifically from their revisions.

Methods

This study compares the effect of two guidance activities on the types of revisions students made to their written explanations (Table 2.5). Student workgroups (2-3 students) were randomly assigned to one of the two conditions (*annotator* or *revisit*). This study examines the types of revisions that students made, scoring whether they added new ideas and whether their revisions were connected to their original written explanations. Students answered embedded essay questions throughout the unit, then participated in one of the two guidance activities (Figures 2.4 and 2.5 below), then were prompted to revise their initial essay response immediately after. The revision prompt automatically imported students' initial essays for students to revise.

Table 2.5. Outline of the sequence for *annotator* and *revisit* conditions.

Steps on Embedded Assessment	***Annotator***	***Revisit***
Essay Prompt	Prompt: Explain how you would use a Punnett square to figure out the probability of getting a certain genotype.	
Students split randomly by WISE ID	*Annotator*: Place pre-written labels on a fictional students' essay to show where ideas are missing.	*Revisit*: Place prewritten labels on a screenshot of a model which the students have already interacted with to show what parts of the model represent.
Revision Prompt	Revise: Now that you've learned a bit more about Punnett squares and probability, take some time to revise or improve your answers to these two questions from earlier.	

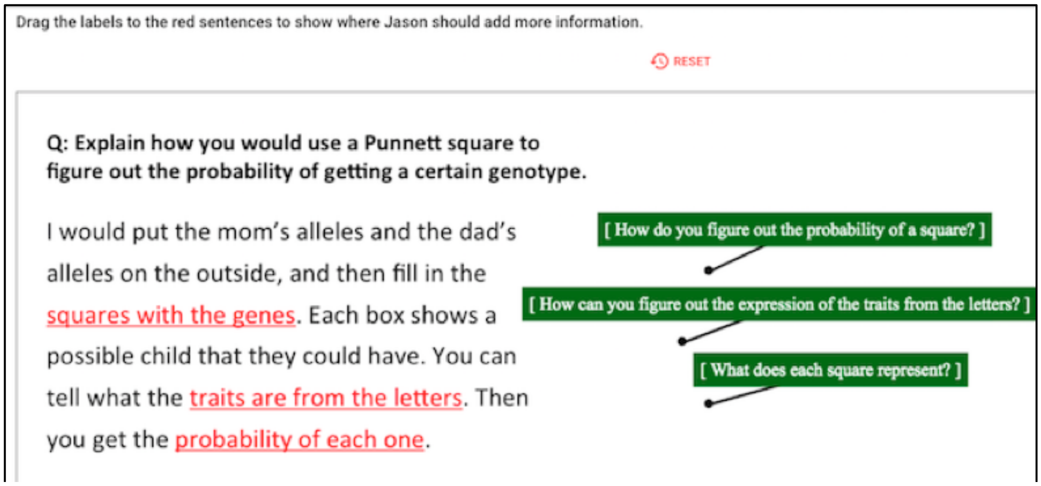


Figure 2.4: Example revision guidance activity for the *annotator* condition.

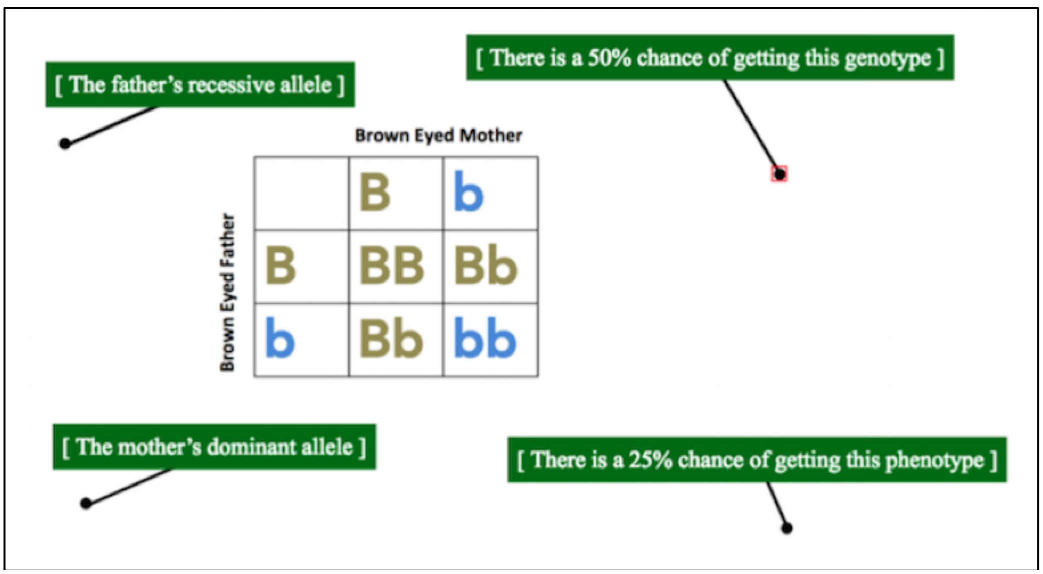


Figure 2.5: Example revision guidance activity for the *revisit* condition.

Participants

Six classes of 6th grade students (N=153) from one teacher participated in this study (94% non-white, 89% free/reduced lunch, 30% ELL). Students completed our 10-day WISE genetics unit during class periods in student teams created by the teacher (1-3 students per team, N=119 student teams).

Curriculum

This study employs the Web-based Inquiry Science Environment (WISE) curriculum on the topic of Genetics and Inheritance corresponding to NGSS (MS-LS3) (NGSS Lead States, 2013). Small modifications were made based on observations of student use from study 1. It is built according to the Knowledge Integration (KI) framework, which elicits and builds upon student ideas to integrate new content, making it an ideal platform to explore revision guidance strategies (Liu et al., 2008). Throughout the unit, students ask questions, use models, and construct arguments from evidence in accordance with NGSS science and engineering practices. The technological environment allows for interactive visualizations of concepts that are difficult to explore in real life, including phenotypic outcomes of various allele combinations. A significant feature of this unit is that it includes regular opportunities for students to revise their ideas after interacting with models, encouraging students to make significant connections.

Assessment

This study included a pre and posttest, as well as embedded writing and revising assessments. Students completed the pretest individually the day before beginning the unit, worked in teams (1-3 students) determined by the teacher on the unit itself, then individually completed a posttest the day after completing the unit.

Student essay responses on the pre/posttest and embedded items were scored using a 5-point Knowledge Integration (KI) scale (See study 1 for rubric). KI scoring rewards students for successfully linking scientific ideas (Liu et al., 2008).

A significant addition to this study is the inclusion of a new pre/posttest item which involved a revision prompt (Figure 2.6). Students were given an open-response item asking them to explain how a recessive trait is passed through generations based on a pedigree. In the following step, students are given a small new piece of information and asked if they would like to revise their explanation. This item was designed to assess ability to revise before and after completing the genetics unit. Students' final explanations on the pre and posttest were scored using a KI scale and used to calculate overall learning gains from the curriculum. Student revision gain scores for the pretest and posttest were also calculated (revised KI score minus initial KI score on both pretest and posttest) to compare how much students gained from their revisions before and after the unit. Additionally, revision codes (see rubric below) were used to see if the nature of students' revisions changed significantly before and after completing the genetics curriculum, specifically if they revised by adding new ideas and/or if they made connected revisions.

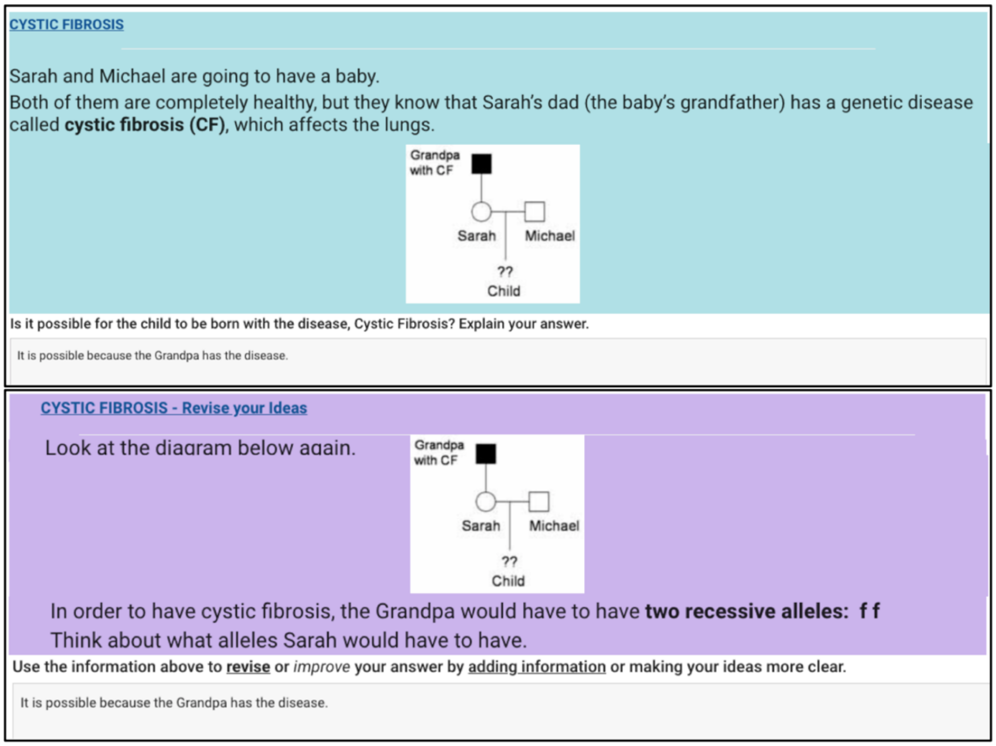


Figure 2.6: Pre/post revision assessment item: initial prompt and revision prompt

I analyzed how students revised to better understand the impact of guidance. I was interested in the *way* students revised on the two embedded essay questions: *Siblings* and *Punnett Square*. Specifically, I was interested in whether students made revisions that were connected to their prior knowledge, rather than simply tacking on to the end of their original explanation. Additionally, I was interested in whether students added new scientific ideas to their explanations, or if they simply elaborated on or clarified their original ideas. Revision codes were given to student responses based on the type of revision. Codes were given for connected (C) versus disconnected (D) revisions as well as for adding new ideas (N) versus expanding on existing ideas (E) (Table 2.6).

Table 2.6: Revision type rubric and examples for embedded essay revisions (student revisions underlined)

Initial Response	**Revision**	**Revision Score**	
		Connected (C) / Disconnected (D)	**New (N) / Expanded (E)**

You would use the method of counting by 25's. Each square is a 25% chance.	You would use the method of counting by quarters. The two letters from each parent would represent a quarter of the genotype and all the quarters combined would show what phenotype would be dominant over the other.	**C**	**N**
A method I would use is that I would choose the number of squares that have a certain genotype and find out what percent of the squares have that genotype.	A method I would use to calculate the probability of getting a certain genotype is that I would choose the number of squares that have a certain genotype and find out what percent of the squares have that genotype.	**C**	**E**
You would put the alleles of each parent on the outside of the square and the possible alleles for their children would be in the square.	You can find the dominant and recessive traits to calculate the probability out of four.	**D**	**N**
The figure shows how many possible genotypes that children can have.	The figure shows how many possible genotypes that children can have. D means dimples and d means no dimples.	**D**	**E**

Results

Overall, the curriculum was effective at teaching students genetics concepts. The Genetics curriculum helped students improve their ability to revise their scientific writing, and those that revised consistently throughout the unit gained better understanding by the posttest. The *annotator* condition was most helpful for students on the embedded assessments, suggesting students were encouraged to revise more often and add more new ideas to their explanations.

Pre/post Analysis (Individual Students)

After removing students with missing data, 153 students completed the pre/posttest. Across both conditions, students improved their average KI score from pretest (mean=1.89) to posttest (mean=2.61) [t(152)=9.66; p<0.001]. This suggests students effectively learned genetics concepts from our curriculum. Condition had no statistically significant effect on pre/post learning gains.

Looking at how students revised on the pre/post revision item (see Figure 2.6 above), it is possible to examine how students changed specifically in their ability to revise after completion of the genetics unit. On the pretest as well as the posttest, only approximately 27% of students chose to revise their explanations at all, with no statistically significant difference between conditions. On the posttest, a significantly greater proportion of students added new ideas to their revised essays (9% on the pretest compared to 21% on the posttest) [z(152)=-2.96; p<0.01]. This suggests that, while the curriculum did not encourage students to revise more frequently in general on the posttest, students that did revise added more new ideas. In terms of making connected revisions, 12% of students revised their ideas in a connected way on the pretest, and 10% revised this way on the posttest (not significantly different at the 5% level). Students continued struggling with connecting ideas to their prior knowledge, which is a difficult process. In terms of adding new ideas and making connected revisions, there was no statistically significant difference between guidance conditions.

I then looked at how revision on this item is related to students' learning gains. A paired t-test revealed that students gained significantly more from revision on the posttest than they did on the pretest [t(152)=3.77; p<0.001]. On the pretest, students gained, on average, 0.006 points (on the 5-point KI scale) after revising their essay. On the posttest, students gained an average of 0.162 points from their revisions. I then looked at only students that revised on both the pretest and posttest (N=20), since students that did not revise could not improve their score; within this group, students gained an average of 0.15 points from their revisions on the pretest, and an average of 0.65 points from their revisions on the posttest [t(19)=2.364; p<0.05]. Though there was no statistically significant difference across conditions, this suggests the curriculum helped students improve their ability to revise their scientific ideas.

Effect of Embedded Revision Behavior on PrePost Gains

Next, I looked at how students' revision behaviors during the unit influenced the revisions they made on the posttest. Students' posttest revision gain scores were calculated (the final KI score of their revised posttest explanation, minus their initial KI essay score on the posttest) in order to see how much students gained from their revisions on the posttest. Regression analysis revealed that students who revised on at least one of the embedded revision items during the unit gained, on average, 0.24 points more from their revisions on the posttest compared to students that did not revise on any of the embedded items during the unit [t(152)=2.02; p<0.05]. I then looked at how the types of revisions (new and/or connected) students made during the unit influenced their posttest scores. I calculated students' pre to post gain, subtracting students pretest KI score (after revision on this item) from their posttest KI score (after revision on this item), as a representation of their improvement after the unit. Then, regression analysis of this pre to post "gain" score revealed that students who made at least one connected revision during the unit gained an average of 0.57 points more from pre to post than students who did not make any connected revisions during the unit [t(152)=2.39; p<0.05]. Condition had no statistically significant effect on this gain score, as well as adding new ideas during the unit. This suggests that practice making connected revisions during the unit resulted in increased pre/post learning gains.

<u>*Embedded Revision Analysis (Student Teams)*</u>

I examined how student teams performed on two embedded essay revision items during the unit, the *Siblings* and *Punnett Square* items. First, I examined if the *annotator* or *revisit* condition influenced students to revise more often. For the *Siblings* item, logistic regression revealed that the odds of revising their explanation at all were 2.78 times as great for student teams in the *annotator* condition as compared to student teams in the *revisit* condition [$z(118)=2.38$; $p<0.05$]; 38% of student teams in the *annotator* condition revised their essay compared to only 18% in the *revisit* condition. For the *Punnett Square* item, logistic regression revealed that there was no statistically significant difference in odds of revising for student teams between the two guidance conditions; 34% of student teams in the *revisit* condition revised their essays, and 33% of student teams in the *annotator* condition revised their explanation on this assessment.

Next, I looked at how student teams revised in each condition for these 2 embedded assessments. First, I looked at whether students revised by adding new ideas. For the *Siblings* item, logistic regression revealed that the odds of revising by adding new ideas were 2.69 times as great for student teams in the *annotator* condition compared to student teams in the *revisit* condition [$z(118)=1.97$; $p<0.05$]. For the *Punnett Square* item, logistic regression revealed that there was no statistically significant difference in odds of adding new ideas for student teams between the two guidance conditions. Then I looked at which condition encouraged more connected revisions on each of these embedded items. Logistic regressions revealed that there was no statistically significant difference in odds of revising in a connected way for either embedded item for student teams between the two guidance conditions.

Significance

Overall, these results show that students can succeed at improving their understanding of genetics as well as their ability to use evidence to revise their explanations, which was captured by the new pre/post revision assessment item.

Genetics and inheritance are complex topics, especially for middle school students; passage of genetic material from one generation to another, and how alleles interact to produce a phenotype, are complex mechanisms that many adults still struggle to comprehend (Campanile et al., 2015; Jiménez et al., 2000; Jacobson & Wilensky, 2006). Similarly, revising ideas is difficult and uncommon, especially in science classrooms. Students often choose not to revise their ideas at all, or simply tack unrelated or contradictory ideas onto the end of their original explanations (Bridwell, 1980), often because they struggle to diagnose issues or inconsistencies with their understanding (Flower et al., 1986). This study shows similar findings, with only 27% of students revising their explanations of inheritance even after completion of the genetics unit. On assessment items presented in this study where students were given the opportunity to revise, the percentage of students that actually revised their answers ranged from 27-34% of total students across the 4 items.

However, as the results support, consistent revision of ideas is important for science learning. Students who revised during the unit ended up being able to construct better explanations about how inheritance works at the end of the unit, specifically how someone can inherit a recessive trait from two parents that do not exhibit that trait. This supports the KI perspective that consistently engaging with your own prior knowledge is a powerful learning strategy. Additionally, these students gained more from their revisions on the posttest than students who did not revise during the unit. Revision is important for learning, but is also a skill that requires practice.

The *annotator* guidance activity encouraged more students to revise their ideas, and encouraged students to add new ideas to their revisions. Considering how helpful it is for students to practicing revising, this is a promising form of guidance for students that might not otherwise revise their scientific ideas. The goal of comparing these two guidance conditions was to discover which part of the knowledge integration process students need the most support with, in terms of revision. Students were encouraged by the *annotator* activity to add new ideas even more so than with the *revisit* activity, which was designed to help with interpreting evidence that the interactive models provided. Similar to earlier findings (Harrison et al., 2018), we see that students instead need explicit help with *how* to revise, modeled by the *annotator* activity, in order to encourage revision. Students in this study did not necessarily need more support gaining ideas from evidence, but perhaps instead needed help distinguishing which of their ideas are relevant for their written explanations, resulting in more addition of new ideas in their revisions. Traditional instruction tends to suppose students need more information in order to progress, but our results suggest that modeling the revision process to support how their relevant ideas connect is more beneficial. The next study in this chapter builds on these findings, by exploring further how this annotator tool helps students revise more often and more successfully.

Study 3: Promoting more integrated revision of scientific ideas through guidance

The benefits seen from students engaging with either the critique or annotator activities in the previous two studies led to the design of this study, comparing these two successful activities in order to better deduce the mechanism of promoting revision. The *critique* activity was designed, very similar to the activity in study 1, to help students in distinguishing normative and more mechanistic ideas from vague understandings of genetics in order to promote knowledge integration. Students explained in open-response items what was wrong, missing, or too vague from common non-normative statements. The annotator tool was modified significantly from study 2, but maintained the goal of making visible the process of revision, specifically promoting the idea of integrating information throughout valid prior knowledge ideas, to demonstrate ways to revise ideas other than tacking unrelated ideas onto the end of a statement. In this study, I modified this activity to prompt more critique-style thinking by including pre-written labels with critical questions rather than statements of new ideas. The goal was to have students think critically in both guidance conditions to reveal if explicating critiques was more successful as compared to making the (critical) revision process more visible.

Methods

Participants

This study includes student data collected from the 10-day WISE Genetics and Inheritance online unit. Ten classes of 8th grade students from two teachers in one California school participated in this study (55% non-white, 40% free/reduced lunch, 13% ELL). 195 students completed a pre/post-intervention measure individually, one day before and after the unit, respectively. Students completed the unit itself in teams (1-3 students) determined by their teacher during 50-minute science classes (N=130 student teams). Analysis was done on one pre/post-test open-response revision item (same as study 2 above) and two embedded benchmark items (*Siblings* and *Punnett Square*, same as study 2 above).

Curriculum

The WISE Genetics and Inheritance unit was modified slightly from earlier iterations to meet the research questions in this study, as well as based on observation of student use. Two different guidance conditions were randomly assigned to student teams several times throughout the unit. The *critique* condition prompted students to explain what is incorrect about several non-normative commonly held student ideas (Figure 2.8). The *annotator* condition presented students with a fictional students' written response to the benchmark question and asked them to place labels to tell the student where their explanation could be improved (Figure 2.7). For these items, student pairs generated an initial essay response, then were randomly assigned one of two guidance activities, and were finally prompted to revise their initial essay response immediately after. The revision prompt automatically imported students' initial essays for students to revise.

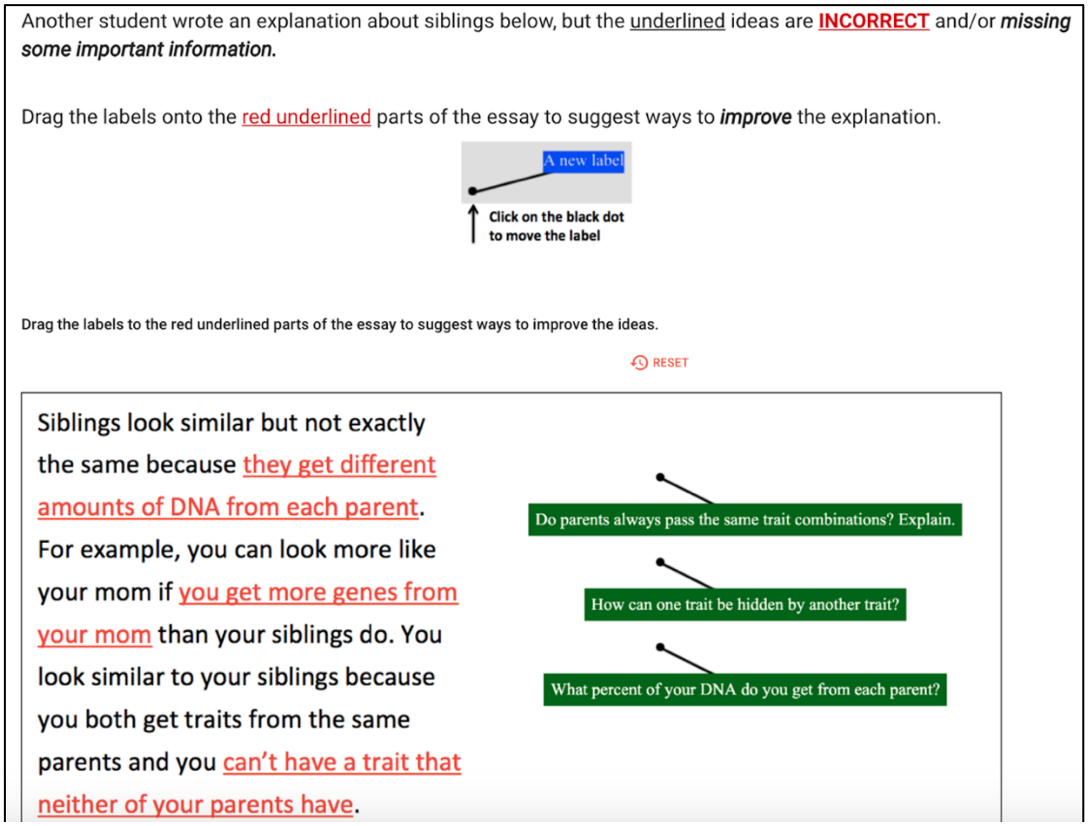


Figure 2.7: Example revision guidance activity for the *annotator* condition.

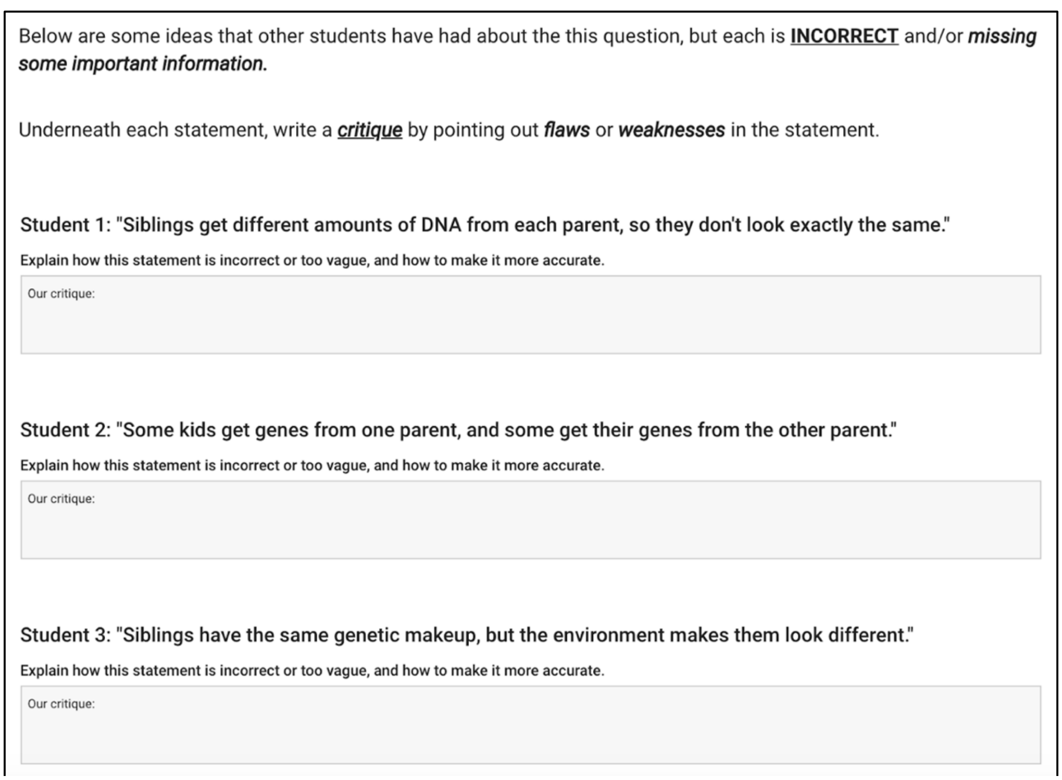


Figure 2.8: Example revision guidance activity for the critique condition.

Assessment

For the pre/post-test and embedded essay items, responses were scored using a 5-point Knowledge Integration (KI) scale (see study 1). KI scores were given for students' initial responses as well as their revised responses on the embedded assessments as well as both the pre and post-test, providing a measure of student learning from beginning to end of the intervention, as well as score increase due to revision on both the pre and post-test. As in earlier studies, student revisions were also categorized based on whether or not they were connected, and whether or not they added new ideas (see study 1 for revision rubric).

Results

Pre/post Learning Gains

Students across both the *critique* and *annotator* guidance conditions scored higher on the posttest compared to the pretest [$t(194)=10.85$; $p<0.001$], with a mean score of 2.19 on the pretest and 2.86 on the posttest. While there was no statistically significant difference in gain between conditions, students across conditions gained more from their revisions on the post-test assessment, gaining only 0.18 points from revision on the pretest and 0.52 points on the posttest, on average [$t(194)=5.92$; $p<0.001$]. This demonstrates that students improved not only in their content knowledge of genetics upon completion of this intervention, but also in their ability to integrate new information into their scientific explanations through revision. Students also revised more often on the posttest (71%) than on the pretest (58%) [$z(194)=-2.88$; $p<0.01$].

Pre/post Revision Type

Students across both conditions added more new ideas to their revisions on the posttest (62%) than they did on the pretest (36%) [$z(194)=5.28$; $p<0.001$]. They also had more connected revisions on the posttest (31%) than the pretest (13%) [$z(194)=4.70$; $p<0.001$], but it is important to note that, similar to findings in the previous two studies, connected revisions are much more infrequent overall. There was no statistically significant difference (at the 5% level) between effect of the two guidance conditions on students' pre/post gain, revising by adding new ideas, or revising in a connected way.

Effect of Embedded Revision Behavior on PrePost Gains

Across conditions and prior knowledge levels, students that revised on both embedded items in the unit scored, on average, 0.31 points higher on the post revision item than students that did not consistently revise throughout the unit [$t(198)=2.20$; $p<0.05$].

Embedded Revision Analysis (Student Teams)

I examined how student teams performed on two embedded essay revision items during the unit, the *Siblings* and *Punnett Square* items. First, I examined if the *annotator* or *revisit* condition influenced students to revise more often. For the *Siblings* item, logistic regression revealed that there was no statistically significant difference in odds of revising for student teams between the two guidance conditions [odds = 1.007; z(129)=0.02; p=0.98]; 56.3% of student teams in the *critique* condition revised their essay and 56.1% in the *annotator* condition revised their *Siblings* explanation. While the student teams in each condition revised in the same proportions, it is important to note that many more students revised here than in our previous studies in this chapter. For the *Punnett Square* item, logistic regression revealed that there was no statistically significant difference in odds of revising for student teams between the two guidance conditions [odds = 0.78; z(129)=-0.69; p=0.49]; 50% of student teams in the *critique* condition revised their essays, and 56% of student teams in the *annotator* condition revised their explanation on this assessment.

Next, I looked at how student teams revised in each condition for these 2 embedded assessments. For the *Siblings* embedded assessment item, 37.9% of students in the *annotator* condition added new ideas to their revisions, and 31.3% of students in the *critique* condition added new ideas to their revisions (Figure 2.9 (a); not significantly different at the 5% level). For this item, 40.9% of students in the *annotator* condition made connected revisions, and 42.2% of students in the *critique* condition made connected revisions (Figure 2.9 (a); not significantly different at the 5% level).

For the *Punnett Square* item, 33.3% of students in the *annotator* condition added new ideas to their revisions, and 25% of students in the *critique* condition added new ideas to their revisions (Figure 2.9 (b); not significantly different at the 5% level). For this item, 37.9% of students in the *annotator* condition made connected revisions, and only 14.4% of students in the *critique* condition made connected revisions (Figure 2.9 (b); z(129)=1.78; p<0.05).

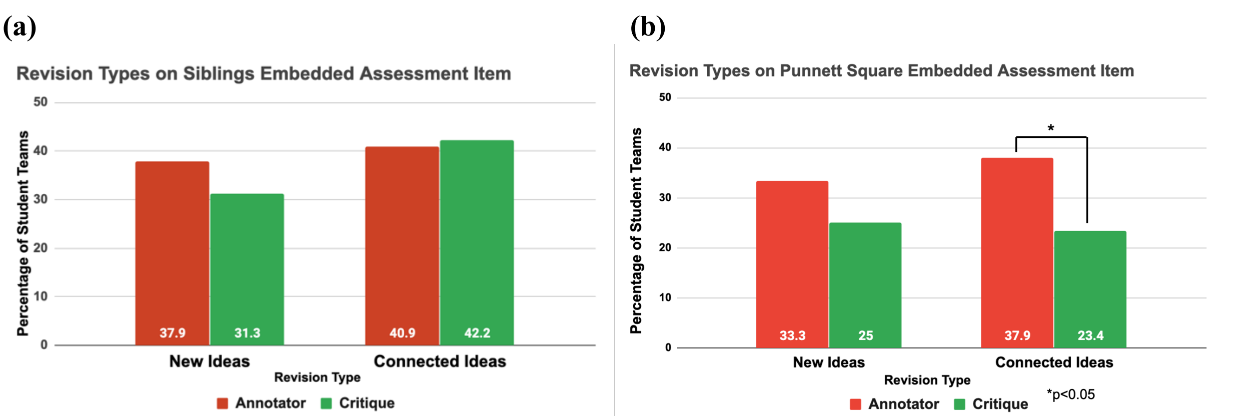


Figure 2.9: Student revision types by condition on the two embedded revision items, (a) *Siblings* and (b) *Punnett Square*

<u>*Embedded Item Revision (Student Teams): Influence of Prior Knowledge*</u>

Next, student teams were separated by prior knowledge to see if either guidance condition was more helpful for students coming into the unit with less content knowledge. Again, students that scored below a 3 on the initial essay were categorized as low prior knowledge (LPK), and those scoring a 3 or above were categorized as high prior knowledge (HPK). The following analyses use only students classified as having low prior knowledge (N=61 student teams).

For the *Siblings* item, 55.8% of LPK student teams in the *annotator* condition revised their explanation, and 54.2% of LPK student teams in the *critique* condition revised their essay (not significantly different at the 5% level). On this item, 36.5% of LPK student teams in the *annotator* condition added new ideas to their essay, and 27.1% of LPK student teams in the *critique* condition added new ideas to their essay (not significantly different at the 5% level). Looking at connected revisions, 40.4% of LPK student teams in the *annotator* condition revised their essays in a connected way, and 35.4% of LPK student teams in the *critique* condition revised their essay (not significantly different at the 5% level). Though these differences were not statistically significant, the *annotator* condition seemed to be slightly more helpful for LPK students in terms of revising more often, in a more connected way, and adding new ideas.

For the *Punnett Square* item, logistic regression revealed that the odds of a student team with LPK in the *annotator* condition revising their explanation at all were 3.1 times as great compared to LPK student teams in the *critique* condition [$z(60)=2.15$; $p<0.05$]. In terms of revising by adding new ideas, the odds of adding new ideas were not statistically different at the 5% level for the two guidance conditions. For this item, logistic regression also revealed that the odds of a student team with LPK in the *annotator* condition revising their explanation in a connected way were 3.6 times as great compared to LPK student teams in the *critique* condition [$z(60)=2.39$; $p<0.05$]. These findings are consistent with findings in study 1 that showed critique was difficult for students with low prior knowledge in genetics. LPK students in the *annotator* condition also gained more on the *Punnett Square* item, scoring an average of 0.2 points higher than LPK students in the *critique* condition [$z(60)=2.64$; $p<0.05$].

Discussion and Implications

Study 1 and 2 found that both the critique and annotator guidance activities helped students revise more often, and often by adding new ideas, than revisiting relevant material. This study was designed to compare these two activities to see which supported students more in revision of their scientific essays.

Both conditions helped students improve their revision abilities by the end of the unit, as demonstrated by the improvements (and KI score gain) from their posttest revisions. This is interesting because after writing their initial response to a question, they are immediately given a new small piece of information (that they have already learned in the unit), and asked to revise their answer. Many students took this opportunity to incorporate this new piece of information in their answer. This item was designed to demonstrate students' ability to revise when presented with new information before and after the unit. Maybe students did not include this information in their answer when first prompted on the posttest, but decided to revise to include it. Most did not revise to include the information, or did not include the new information successfully, in their revisions on the pretest. Instead of measuring students' revision abilities, this item may reveal the importance of prior knowledge in revision success and/or motivation. Students knew how to use the information by the posttest, and so were more likely to revise and include that new information in their explanation.

The annotator is designed to model the revision process, showing how to integrate ideas with prior knowledge. These findings show us that the annotator can help students connect their revisions to their original thoughts. The annotator particularly benefitted students with low prior knowledge; LPK students that used the annotator guidance had more connected revisions and revised their essay more frequently than LPK students that received critique guidance. This supports our earlier findings that students with low prior knowledge struggle with critique. Critique, in study 1, helped LPK students revise more often, presumably by challenging their prior knowledge, but this activity did not necessarily help them improve their KI scores on the posttest. This might be due to the fact that critique itself is challenging for students who come in with less content knowledge. Critique may have helped LPK students recognize gaps in their understanding, as compared to revisiting ideas they may have missed the first time through, it does not necessarily help them fill those gaps. Supporting students to make integrated connected revisions is very difficult, particularly those with lower prior knowledge about the content (Flower et al., 1986; Bridwell, 1980; Crawford et al., 2008). It is much more common that students tack new, even contradictory, information onto the end of their initial response. However, the annotator helped LPK students revise in a connected way, resulting in improved scientific explanations.

Additionally, across conditions and prior knowledge levels, students that revised on both embedded items in the unit scored higher on the post revision item than students that did not consistently revise throughout the unit. This demonstrates the benefit of consistently revising your scientific ideas; even if these revisions did not result in gain on embedded items, engaging with the material resulted in more learning by the end of the unit.

Combined Analysis of Studies 1, 2, and 3

Studies 1, 2, and 3 compare 3 different guidance conditions; in order to see if there were any patterns throughout the studies overall, a combined analysis was done. Students from all three studies were separated into 3 categories based on their guidance condition: Critique, Annotator, or Revisit.

Methods

Students were combined based on guidance condition across all three studies (*revisit, annotator,* or *critique*), (N=444 student teams). To see if conditions across all 3 studies showed differences in student performance, one embedded revision item that appeared throughout all studies was scored, students' revisions were scored, and revision gain was calculated. The item used is the *Siblings* item from all 3 studies, asking students to use their understanding of genetic inheritance to explain how siblings from the same parents look similar but not exactly the same. This item was chosen because it addresses essential genetics concepts, including inheritance of half of their alleles from each parent, and combination of alleles resulting in different phenotypes. For this embedded item, student teams write an initial explanation, learn the relevant concepts over the next several steps, then complete the condition-related guidance activity immediately before revising their original explanation, which is automatically imported from the initial step.

Results

Revision Gain on Embedded Item

Kruskal-Wallis Test showed significant differences between conditions ($p<0.001$) for revision gain, but also showed that the majority of students did not improve from initial to revised essay score on the item (Figure 2.10).

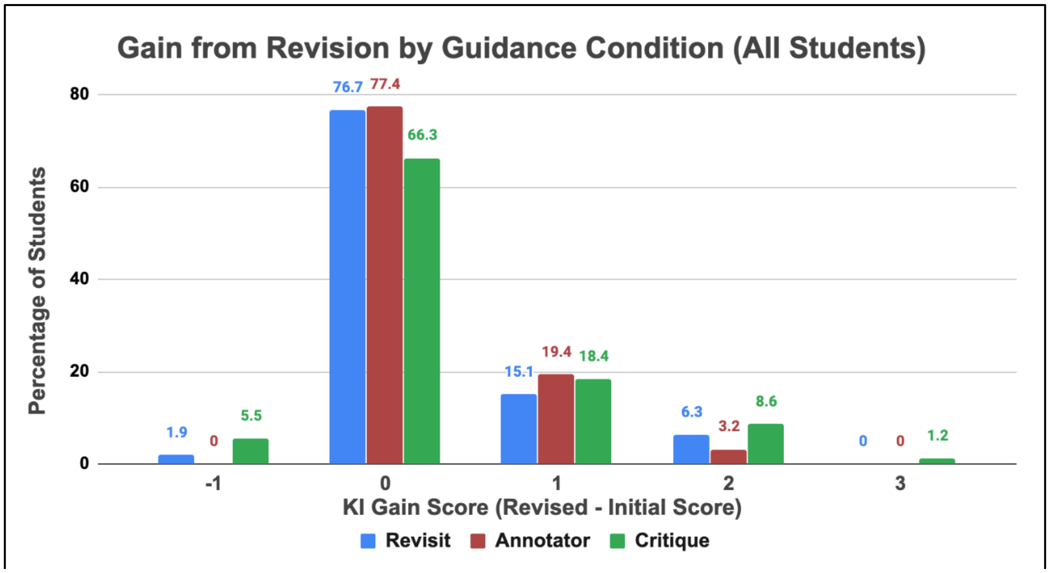


Figure 2.10: Gain in KI score from revision on one embedded item (revised KI score minus initial KI score) by revision guidance condition

In all conditions, the majority of students gained 0 points in KI score from their initial to revised essay (on a 5-point KI scale). As a reminder, KI rubrics reward students for making connections between valid scientific ideas. Therefore, an increase in score shows that students are integrating or connecting new science ideas, or are making stronger connections between existing ideas. Students that do not improve may be revising their answer to make their current ideas more clear, but are not adding or integrating new ideas in their revisions, and therefore receive the same score. Students whose scores decreased may have taken up incorrect ideas from the guidance activity, and changed their original response to include those ideas.

In the revisit condition, about 2% (N=3) of student lost one point from their revision, 77% (N=122) of students gained 0 points, while 15% (N=24) gained 1 point from their revision, and 6% (N=10) gained 2 points. No students in the revisit condition increased their score by 3 points.

In the annotator condition, no students decreased their score after revision. 77% (N=96) of students gained 0 points, 19% (N=24) gained 1 point, and only 3% (N=4) gained 2 points from their revision. No students in this condition gained 3 points.

In the critique condition, 5.5% (N=9) of students decreased by a point after revision. 66% (N=108) of students' scores were unchanged from revision, 18% (N=30) gained one point, and about 9% (N=14) gained 2 points from revision. This is the only condition where students gained 3 points, though only 1.2% (N=2) of students fall into this category.

However, this gain is measured from initial to revised essay; a major reason many students did not improve (gained 0 KI points) is that many students did not choose to revise their response, and therefore their score could not change. Table 2.7 shows what percentage of students revised their essays in each condition.

Table 2.7: Percentage of students that revised their scientific explanation in each guidance condition.

Guidance Condition	Percent of Students that Revised
Revisit	36.3% (N=59)
Annotator	47.6% (N=59)
Critique	56.4% (N=90)

The lowest frequency of revision was seen with students in the *revisit* condition, where only 36.3% revised their posttest answer when prompted. In the *annotator* condition, 47.6% of students revised, and the *critique* condition had the highest revision frequency at 56.4%. This suggests that students were encouraged to revise more often in the critique condition than the other conditions.

Figure 2.11 below shows students' gain in KI score on their revised response by condition, only for those students that did revise their essay.

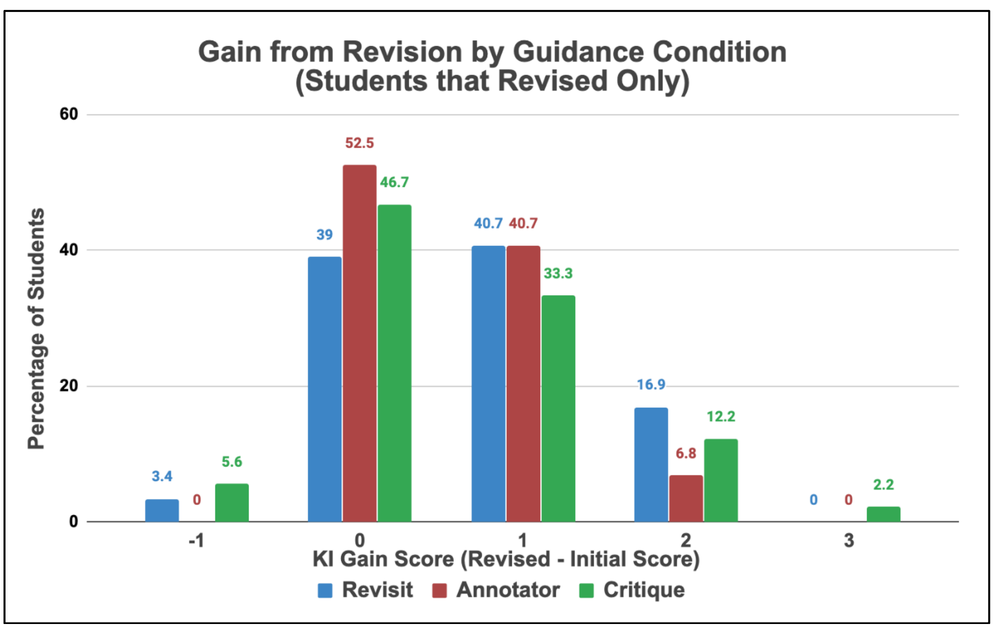


Figure 2.11: Gain in KI score from revision on embedded item (revised KI score minus initial KI score) by revision guidance condition for *only* those students that did revise their explanation on this item

When looking at only students that did revise, the revisit condition has 3% (N=2) of students that decreased their score, 39% (N=23) students whose score remain unchanged, 41% (N=24) gained 1 point, 17% (N=10) gained 2 points.

In the annotator condition, of students who revised, no students decreased in points, 53% (N=31) stayed at the same score, 41% (N=24) gained 1 point, and 7% (N=4) gained 2 points.

In the critique condition, 5.6% (N=5) decreased by 1 point, 47% (N=42) remained the same, 33% (N=30) gained one point, 12% (N=11) gained 2 points, and 2% (N=2) gained 3 points from their revisions.

Even when students that did not revise at all are removed from the analysis, across conditions, still a large percentage of students remained at the same score. Of students that did revise, students in the *annotator* condition had the least amount of gain overall from revisions. While they revised more often than students in the *revisit* condition in general, students in the *revisit* condition gained more when they did choose to revise. Students in the *critique* condition revised the most often, and generally benefitted from these revisions, though some students decreased their score. This is consistent with findings that critique is difficult for some students, particularly those with low prior knowledge; these students may have taken the ideas they were supposed to critique as fact and incorporated them into their responses.

Discussion and Implications

Overall, this analysis gives us a bigger picture of how students were revising during the unit. The critique guidance encouraged students to revise the most often, and many students increased their KI score from their revisions. This condition had the highest gain as well, though only a few students gained 3 points from their revisions. This condition also had the most students decrease their score after revision; this is consistent with findings earlier in this chapter that students with low prior knowledge struggled with critique. It helped them revise more often, but not necessarily in a productive way. However, study 3 shows that consistent revision throughout the unit can lead to improvement in learning by the end; students may not have directly improved on this embedded item through their revisions, but grappling with the material can lead to later improvement. Therefore, while critique is difficult and may lead to temporary confusion, it encourages students to revise more often, which is beneficial overall. Giving students more explicit support and practice with *how* to critique may help students benefit even more from this type of guidance.

The annotator guidance activity encouraged students to revise more often than the revisit guidance condition, which is beneficial, but the large majority of students in this condition maintained the same score from their revisions. As study 3 shows, this may be due to the fact that the annotator generally encourages more connected revisions, but does not necessarily promote addition of new ideas. However, encouraging students to grapple more with their present ideas is beneficial for learning in a way that integrates ideas with prior knowledge to create usable science understanding.

Interestingly, the revisit condition had the lowest percentage of students that chose to revise at all (~36%), but also had the lowest percentage of students whose scores remained the same after revision. The majority of students (by a small margin) increased by 1 point, which is significant on the KI scale of 5 points. This type of guidance does not necessarily encourage students to revise that would not have chosen to do so already, but those that *do* choose to revise, generally benefit from doing so in a way comparable to the other guidance conditions. This suggests that revision has a self-directed component to it; students that actually revisit and reassess material with minimal prompting also tend to revise more often and more productively. This idea is explored further in chapter 4.

Overall, this analysis shows the various benefits of each of these guidance conditions, with various teaching implications. Perhaps more self-motivated students will benefit more from simple prompts to revisit relevant material and to revise their ideas. Students (particularly those with low prior knowledge, as shown in study 3) benefit more from the annotator to help them make more connected revisions. Students that have a decent handle on the science content benefit a lot from critique, which encourages them to revise more often and more productively.

Additionally, a combination of these guidance conditions may prove the most useful. For example, students that come into a unit with low prior knowledge may benefit from first gathering new content knowledge, then grappling it with the annotator, and then practicing critique once they have a handle on the new material. Students with high prior knowledge may benefit instead from critique first, then annotator activities to help them better connect new content to their prior knowledge. Overall, guidance activities that encourage students to revise more often beneficial to help students engage with the material in an active way.

Conclusion

This research involves students' construction and revision of explanations and arguments about complex scientific ideas, and the exploration of various types of guidance activities that support students in the endeavor in different ways. Overall, this research supports that writing and revising ideas can help students more effectively integrate ideas with their prior knowledge (Linn et al., 2014), even if students do not immediately improve their explanation with their revision. Revising consistently while learning new material helped students improve their understanding by the end of the curriculum. This suggests that revision helps students grapple with the material in new ways, resulting in better scientific understanding.

Writing, as well as science, are often taught to students as linear processes (Flower & Hayes, 1981; Campanile et al., 2013), though in reality they are very much iterative. Encouraging students to engage in revision of their science explanations after learning new content can help students not only improve their comprehension, but also improve their understanding of the nature of science. The nature of scientific thinking involves continual revision of ideas as an individual encounters new evidence and experiences new phenomena and anomalies (Berland et al., 2016).

Writing to learn can enhance understanding of science ideas (Brownell et al., 2013; Rivard, 1994); revising their science explanations can help students further engage with their understanding and communication of complex ideas. However, students are rarely asked to revise, particularly in science class (Boote, 2012). Unsurprisingly, when they are asked to revise, students often make surface-level changes rather than evaluating their work and making substantial revisions to content (Bridwell, 1980). Our combined analysis supports this, showing that the large majority of students did not change their KI score with their revisions. Content-level revisions are even more important when writing scientific explanations and arguments since students are learning about and developing understanding of underlying mechanisms of real-world phenomena; surface-level revisions of their ideas do not show fundamental change in student thinking about how complex scientific phenomena occur.

Flower et al. (1986) tells us that revision is a complicated process, and for students to be successful we cannot simply prescribe expert actions to novices with less experience and content knowledge. This is consistent with the fact that having students revise more often did not always mean revising *better* by integrating new ideas on each item. However, more practice with revision can lead to more expert habits. Therefore, guidance that encourages students to revise can lead to long-term benefits. Indeed, like Tansomboon (2017) showed, this study supports that students who revised during the science unit, even if their revisions did not directly improve their responses, showed more improvement by the posttest.

Additionally, experts are capable of recognizing gaps in their understanding, thereby motivating a need for revision (Flower et al., 1986). Students that are novices at both genetics and revision are less capable of recognizing these gaps. Practicing critique can help students begin to recognize the gaps in their scientific understanding (Henderson et al., 2015; Jimenez-Aleixandre & Duschl, 1999; Zohar & Nemet, 2001). Our critique items were specifically designed to critique common student misconceptions about inheritance, which did promote revision. However, our findings are consistent with the fact that critique (and revision) depend heavily on prior knowledge (Donnelly et al., 2015; Flower et al., 1986); students with low prior knowledge were sometimes further confused by the critique guidance and benefitted more from the annotator tool.

Many studies over the past 30 years have shown that when students revise their writing, if they revise at all, they often maintain conflicting ideas, tack ideas onto their initial response without integrating information, or they only make surface-level grammatical changes without incorporating new content (Rivard, 1994; Bridwell, 1980; Crawford et al., 2008). While many students in this collection of studies had similar issues, providing students with guidance specific to their needs may prove useful in overcoming some of these problems. In general, critique encouraged students to revise more frequently, often by adding new ideas. Modeling how integrated revision looks with an annotator tool helped students make more connected revisions. By isolating which aspect of the revision process a student is struggling with, specific guidance sequencing can help students overcome their individual difficulties.

Our newly designed guidance, the critique and annotator activities, were compared to revisiting relevant material. Past studies in our group have explored this as a benefit of technology, accurately directing students back to relevant interactive models or multimedia, and has helped some students improve their ideas. Here, we see that the other two guidance activities prompt students to revise their ideas much more often. Curiously, however, those in the revisit condition that *did* choose to revise often were very successful. This suggests a self-directed aspect to the revision process. Self-directed learning (SDL) involves students taking greater ownership of their learning. SDL can include, but is not limited to, ability to set learning goals, independent pursuit of learning, student knowledge of what they do and do not understand, choose resources, and monitor and evaluate progress in learning (Hmelo, 2004; Loyens et al., 2008; Schmidt, 2000; Candy, 1991; Fisher et al., 2001). Successful revision easily incorporates many of these skills. Students need to evaluate their learning to recognize gaps in what they do and do not understand in order to integrate new ideas. Particularly in the revisit condition, they must choose to re-examine material from the unit. Less self-directed students may see a model they have already used and choose not to explore it further to see what they may have missed the first time, or they may struggle to determine what concepts they do not yet understand. Even rereading an initial explanation is a self-directed choice; Bridwell (1980) found that students who reread their drafts in their entirety had better, deeper-level revisions, but many chose not to evaluate their initial explanation at all. In some ways, the *revisit* activity is the most autonomous condition, and benefits students with more independent learning styles. These ideas, as well as ways to promote more independent learning, are explored in chapter 4.

Chapter 3: Guiding student construction, revision, and interpretation of Graphs

This chapter involves several studies that investigate student's graphing abilities. While graph interpretation is an important skill for scientific literacy, it is often not heavily featured in science classrooms. When students see graphs only in math classrooms, they may learn the skills involved but continue to struggle connecting graphs to the real-world aspects they represent. This may result in some graphing abilities, but not necessarily in competence related to graph construction or the ability to use data from a graph as evidence in an argument. Two studies are presented here with progressive redesigns of an online graphing curriculum to help students use and understand the purpose of graphs in science.

The first study involves a comprehensive redesign of the WISE Graphing Stories unit to include multiple types of data representations, but focused mainly on position-time graphs. A pre/post position-time graph revision assessment item was created to see how students revise their constructed graphs. Additionally, I looked at students' ability to critique others' graphs and their ability to use that skill to recognize gaps in their own graph construction. Lastly, in order to better understand why students choose to revise the way they do, students' open-response revision rationales were analyzed.

The second study includes the same unit with minor revisions. An analysis of the same pre/post revision graph item was given with new guidance to help students improve their revisions, as well as to provide more information about student thinking while constructing and revising these abstract representations of motion.

Introduction

Interpreting and constructing data visualizations is a skill necessary for scientific literacy; it allows for the elucidation of patterns and underlying processes and helps to reveal correlations between events (Friel et al., 2001; Wu & Krajcik, 2006). Additionally, graphing is heavily featured in the NGSS Science and Engineering Practices (SEP), including *analyzing and interpreting data*, *engaging in argument from evidence, and obtaining, evaluating, and communicating information* (NGSS Lead States, 2013). To be successful in these practices, students need to be able to do more than plot or read points on a graph; they need to be able to think with graphs and extract evidence from data. This involves not only reading and comprehending graphs, but also constructing and critiquing graphs (Lai et al., 2016). Construction and critical interpretation of graphs are important practices, however middle school students are not often asked to interpret, let alone construct, graphs in science classes (Boote, 2012). Even more rare is the opportunity for students to *revise* their own graphs. For this reason, the following set of studies explores guidance to help students improve their ability to critique, construct, and revise graphical representations.

Many descriptions have been given regarding what is meant by "graphing comprehension" for students (e.g. Shah & Hoeffner, 2002; Lai et al., 2016; Friel et al., 2001; Bertin, 1983). Friel et al (2001) describes a taxonomy of graph literacy, with the highest level of understanding being "moving beyond the data", for example predicting or inferring values and trends or aggregating the data on the graph to answer a question about the topic. However, missing from this taxonomy and from graphing literature in general is the ability to critique graphs or proposed interpretations of graphical data. Critique is an essential practice of scientists (Chang & Linn, 2013), and is an important element of metacognition. Graphical data is often used to make arguments, both in science as well as advertising; critique of graphs is therefore a necessary skill to evaluate those arguments and the evidence from which those arguments are drawn. Critique is rare in most literature on graphing, except to cite it as a gap and a recommended instructional implication (Donnelly-Hermosillo et al., 2020; Lai et al., 2016). This chapter aims to include practice in graph critique, both as its own activity as well as serving as guidance for graph construction.

Graphing literature also points to the fact that much graphing instruction is limited to rote plotting or interpretation of points, rather than "thinking with graphs" or interpreting data in light of real-world scientific phenomena (Shah & Hoeffner, 2002; Donnelly-Hermosillo et al., 2020). Therefore, having students construct graphs is a useful opportunity for students to reveal how they are thinking about scientific information (Lai et al., 2016). Furthermore, having students construct qualitative or semi-qualitative graphs, rather than simply plotting points from a table, allows students to demonstrate more complex combinations of their ideas, both normative and non-normative. This can reveal more about student thinking regarding scientific phenomena and how they represent it (Hattikudur et al., 2012; Stylianou et al., 2005; Vitale et al., 2015; DiSessa & Sherin, 2000).

This chapter takes these ideas one step further, and also includes opportunities for students to *revise* their constructed qualitative graphs after receiving feedback or guidance. This allows students to make predictions about scientific concepts and the general relationships between specific variables involved, and then reflect on their understanding and their representation of that understanding. The value of revision is widely accepted in writing (Flower et al., 1986), and I describe in chapter one the specific value of revision of students' scientific ideas in written form, especially due to the complex nature of scientific understanding (Berland & Reiser, 2009; Crawford et al., 2008). Students' graph construction can benefit from all the same processes involved in revision of written material, including improved ability to diagnose gaps in your own knowledge and other metacognitive skills, and improved communication of ideas. Additionally, graphs showing data of science or social-science phenomena are arguably even more complex; not only is the phenomenon likely abstract, students must also devise an abstracted way of representing the relationships involved. This added complexity makes opportunity for revision even more compelling, allowing students to distinguish their ideas, reflect on their thinking, and improve their graphical representations (Lai et al., 2016). Since student-constructed graphs often combine a complex combination of normative and non-normative ideas (Vitale et al., 2015), students can benefit from revisiting and reevaluating those ideas through revision.

Specific attention in literature is paid to position-time graphs because students have a great deal of difficulty with this type of graph (Boote, 2012; Brasell, 1987; McDermott et al., 1987). Students particularly struggle with accurately constructing position-time graphs (Vitale et al., 2015). A common non-normative idea around position-time graphs include slope-height confusion, where students interpret line segments higher on the graph as representing a faster moving object, rather than attending to the slope of the line (Bell & Janvier, 1981; Clement, 1985; Vitale et al., 2015). Another common misconception, known as graph-as-picture error, occurs when students interpret a graph as a drawing of an event rather than a relationship between variables representing an event abstractly (Clement, 1985). For example, when asked to interpret a graph showing the speed of a car around a racetrack, students interpreted the curve as the design of the racetrack itself (Vitale et al., 2015; Brasell, 1987). Here, I explore students' ability to construct and revise various position-time graphs and the effects of various forms of guidance.

Theoretical Framework

Knowledge Integration

This study employs a Web-based Inquiry Science Environment (WISE) unit, titled Graphing Stories, which was designed according to the Knowledge Integration (KI) framework, a constructivist framework that involves eliciting and building off of students' prior knowledge (Linn & Eylon, 2011). Students' issues with interpreting graphs often come from a difficulty in connecting the significance of graphical features, such as scale, slope, and direction of lines, to the thing they represent in reality. The KI framework encourages making these connections in our curriculum by *eliciting* students' initial ideas (for example, by having them draw prediction graphs), and then helping them *discover* new ideas and *distinguish* multiple new and prior ideas through various forms of guidance. Finally, *reflecting* on their graph construction or comprehension, especially through opportunities for revision, allows students to form *connections* between abstract aspects of graphs and what they represent.

This framework is based in part on the principle that students often hold multiple conflicting ideas, and can form more coherent scientific explanations by actively reasoning through a situation (Linn & Hsi, 2000). This is especially true when it comes to students' ideas about graphs; students often integrate multiple perspectives when interpreting the meaning of a graphical representation, including a confusion of speed and acceleration, or interpreting part or all of a graph as a picture of an event (Donnelly-Hermosillo et al., 2020). Knowledge Integration guidance can help students distinguish between these various interpretations while acknowledging that students are bringing valuable ideas to work with. KI rubrics are built to reward students for making connections between valid scientific ideas; this makes this rubric ideal for assessing students' graphing understanding, where it's necessary to make connections between parts of a graph and what they represent. KI has also been used previously to score students' constructed graphs (Vitale et al., 2015; Lai et al., 2016). Together, these aspects make knowledge integration an ideal framework for guiding and assessing students' graph knowledge.

Curriculum

In the studies presented below, all instructional materials were part of an online unit called *Graphing Stories* in the Web-Based Inquiry Science Environment (WISE) platform. WISE is an open-source platform ideal for designing interactive activities, giving immediate feedback on formative assessments, and capturing student thinking in a variety of ways with a multitude of item types. The *Graphing Stories* unit was first designed according to the KI framework as an introduction to interpreting and constructing data visualizations with a focus on position-time graphs. The goal was to introduce students to the mechanics and purpose of graphs. It addresses several NGSS science and engineering practices, specifically those related to using data as evidence (i.e. *Analyzing and interpreting data, Engaging in argument from evidence*; NGSS Lead States, 2013). The unit was redesigned by the author for study 1 to include many different types of data visualizations to convey various ways data can be represented, as well as to address research questions of the studies presented here.

Many factors have been identified that affect students graphing abilities. For example, visual characteristics of graphs can influence students' interpretation of them (Shah & Hoeffner, 2002). Additionally, students' general knowledge about graphs can influence the things they notice (Carpenter & Shah, 1998; Cleveland & McGill, 1985; Shah & Shellhammer, 1999); this includes the types of graphs they've been exposed to previously, what relationships they expect different types of graphs to show (for example, a line versus a bar graph), and visual aspects of data that draw attention to specific relationships (Pinker, 1990). To address this, the unit was redesigned to include several different types of graphs so students can see how all types show a relationship between the variables on the axes but may look different from each other.

Students' prior knowledge about the content of the graph or the context of the data can affect how they interpret a graph as well (Shah, 1995; Freedman & Smith, 1996); for example, students familiar with a subject may interpret the graph according to their expectations due to prior knowledge, rather than what the data is actually saying. Alternatively, more familiarity with a subject may reduce cognitive load and allow students to focus on graph interpretation itself (Shah & Shellhammer, 1999). To address this, the unit includes scenarios that middle school students will be familiar with (ex. A swimmer swimming back and forth across a pool), but where prior knowledge of the content plays little role. Students are also given ample revision opportunities, as well as practice with critique to encourage them to question their predictions and expectations about what a graph is communicating.

This unit employs the affordances of technology in learning to construct and interpret graphs, providing students with interactive animations and simple graphing tools of various types to aid in construction. The unit is organized into three major sections. A short introductory section has students participate in an online discussion with their classmates about the purpose of graphs in science and compare the purpose and structure of a graph they might see in a math class to one they might see in a science class. This section also allows students to practice using the online graphing tools; for example, students construct a graph of their average happiness throughout the week, and pairs compare similarities, differences, and general patterns on their graphs (Figure 3.1).

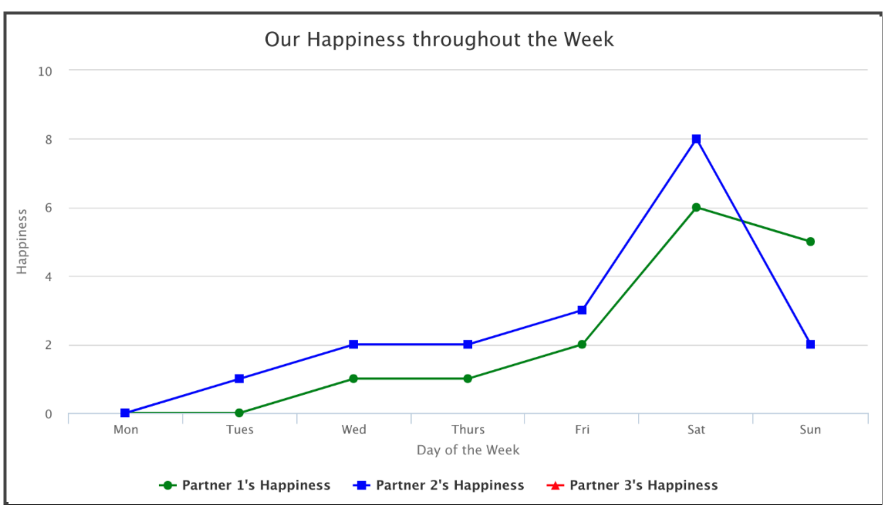


Figure 3.1: Example "Happiness Graph"; student pairs graph their own estimates of happiness on each day of the week and compare patterns in an open response item.

The next section of the curriculum deals with constructing and interpreting position-time graphs. Students are given short stories about a swimmer's motion and construct prediction graphs; then they receive feedback through animations that read the student-constructed graphs to see if they successfully graphed the swimmer's movement (Figure 3.2). Students then have several opportunities to revise their graphs after feedback.

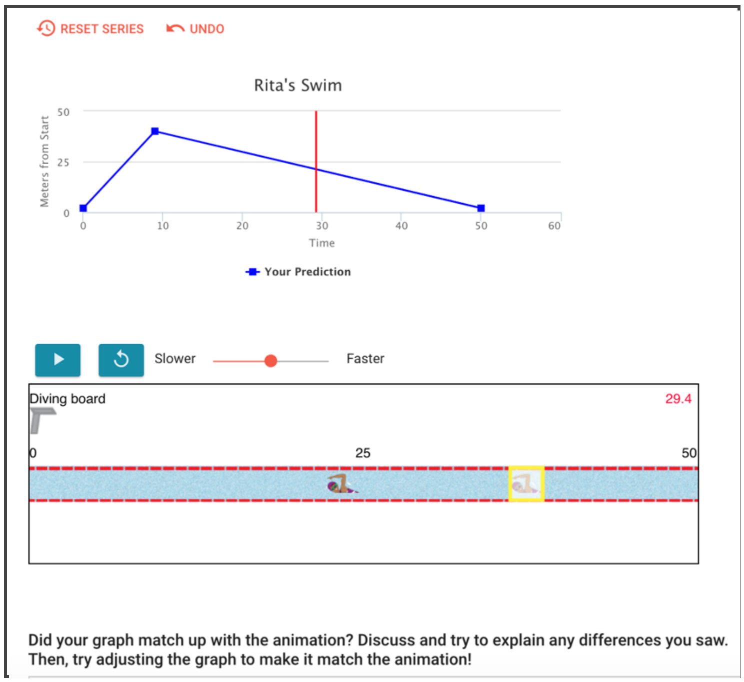


Figure 3.2: Animation feedback for student-constructed position-time graphs; yellow box shows the target and swimmer shows the student graph.

Students practice with several position-time graphs, revising after animation feedback. Students also answer several open response questions about aspects of position time graphs, including meaning of line steepness and direction, while providing evidence from their graphs. They also have opportunities to revise their explanations to these questions throughout this section of the unit.

The third section focuses on several other types of data visualizations, including scatter plots, to emphasize that different graph types show relationships between variables in different ways. Again, students draw prediction graphs regarding relationships familiar to middle school students, and receive feedback in the form of actual data to assess their predictions (Figure 3.3). Students then respond to an open-response item: "How closely did your predictions match the actual data? Explain any differences you see."

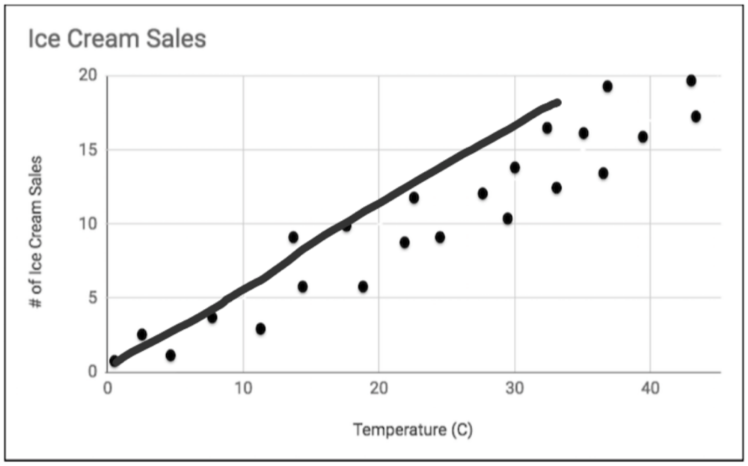


Figure 3.3: Scatter plot item showing example student prediction line and actual data (data appears after students draw and submit their prediction)

Lastly, students practice critique by pointing out flaws in multiple graphs taken from the internet, with issues ranging from using the incorrect graph type to containing impossible data. These items were designed to help students recognize poorly represented data and consider how to improve it (Figure 3.4, for example). They were also designed with the hope of helping students develop deeper metacognitive awareness about their own graphs and what issues to avoid.

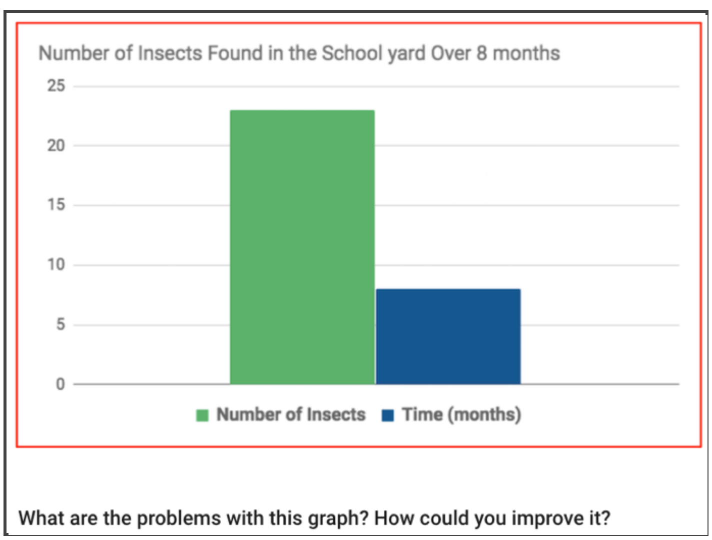


Figure 3.4: Example "bad graph" item where the wrong graph type was chosen for the data presented

Another purpose of these "bad graph" items was to illustrate to students the purpose of basic graph mechanics, such as the necessity of labeling the axes, to help with the interpretation of the data. The goal was to illustrate the purpose of these basic graph components rather than simply tell students they need to include appropriate labels, title, and scale, such as with Figure 3.5 below, which is uninterpretable without the y-axis label.

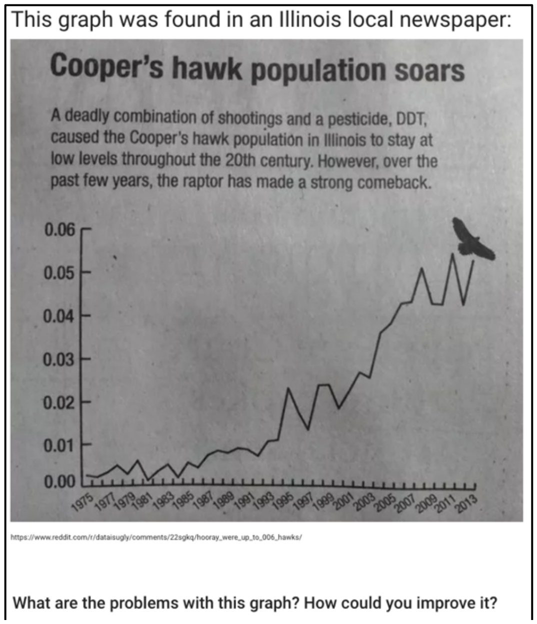


Figure 3.5: "Bad graph" item illustrating the importance and purpose of basic graph aspects, such as axis labels

Additionally, these critique items used graphs taken from real-life scenarios to illustrate to students that graphs they find on the internet or in newspapers may not be perfect, and need to be considered carefully rather than taken at face value. This skill is particularly important with data visualizations meant to advertise or misconstrue important information that students may encounter on the internet. The goal is to increase student agency with graphing by having them point out flaws and suggest improvements.

New pre/posttest items were developed for the studies below to assess students' improvements with constructing and revising position-time graphs. Students' ability to critique graphs is also assessed pre and post, as well as their ability to use that skill to improve their own graphs.

Study 1: Revision Analysis of Students' Position-Time Graphs

This study investigates how engaging students in critique can promote productive revision of student-constructed position-time graphs. I conducted the study in the context of the WISE Graphing Stories unit with 8th grade students. This study focuses on a pre/post item in which students construct a graph to represent a given story, then revise their graph after completing critique guidance. This set of items was designed to promote knowledge integration by encouraging students to identify gaps in their original understanding and representation of motion in graphical form and revise them to show new coherent connections (Linn & Eylon, 2011).

I categorized the types of revisions students made on their graphs as a result of the guidance, and look at explanations students gave for their revisions. I also looked at students' critiques in order to examine how they used the guidance to aid in their revision. Critique is an important scientific skill that can help students become more reflective of their work (Sato, 2015), and is meant to support the *distinguishing* step of the KI process. It is also a skill in itself that requires practice.

The goal of this study is to help students distinguish ideas regarding their representations of motion and to recognize and revise gaps or inconsistencies. Additionally, examples are presented to demonstrate student thinking throughout the assessment and guidance to help pinpoint difficulties and inform future studies and curricular design.

Methods

Participants and Procedures

Two teachers from one middle school (62% non-white, 22% free/reduced lunch, 12% ELL) participated in this study, with a total of 10 classes of 8th grade students (N=230). Students completed the 5-day Graphing Stories unit during 50 minute science class periods. Students worked in collaborative workgroups assigned by their teachers, mostly pairs with a few students working individually or in groups of 3. Students completed the pretest one day before beginning the unit, and the posttest one day after completing the unit. Both pre and posttest were completed individually. For this study, I focus on the types of revisions students made on a posttest item that asks them to construct a position time graph and then revise their graph after receiving guidance.

Assessments

Before and after completion of the Graphing Stories unit, students completed the pretest and posttest individually. This study focuses on a graph construction-revision item on the pre/posttest. The prompt states: "Karim wanted to bike up a big hill in his neighborhood. He went slowly up the steep hill, then really fast on the way down the other side. Use the graph below to sketch his ride. Think about the different speeds he went during his ride." Students used the WISE graphing tool to construct a position-time graph to correspond with this story. In the next step, all students examined and critiqued two graphs created by "fictional students" before revising their own graph. This item was designed to help students recognize a common error where students draw a hill rather than representing the speed and direction on the graph as indicated in the story. After completing this item, students are directed back to their own graph to make revisions. Students are then prompted to explain what they changed about their graph and why.

The types of revisions students made on their graph were categories, as well as whether they were correct or revised their graph to be correct. Students' graph revisions were compared pre and post as well to see if the nature of their revisions changed after completing the unit. In addition to this graph revision assessment, students' graph critiques on the guidance item were also scored on the pre and posttest to see how students were using the guidance, as well as to see if their ability to critique others' graphs improved. A rubric was also created for students' open response explanations about what they revised and why. This item was used to examine the range of rationales students have for revising (or not revising) their graphs. One embedded item was also given a binary score for correctness to help inform why students continued making certain errors; this item asked students: “What does it mean when the line on the graph is going down?”. Examples are presented to show how students interacted with the sequence of steps (graph construction, critique, graph revision) to help inform future design.

Results

PrePost Analysis: Graph Revision

Students’ graph revisions for the assessment item were categorized into six different groups based on aspects of the graph that they changed (Figure 3.6).

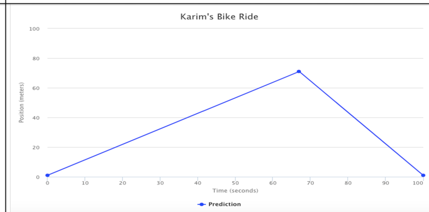
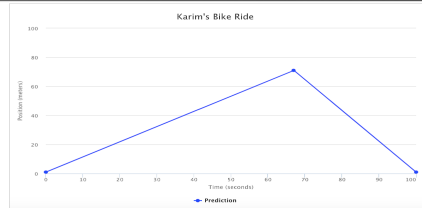
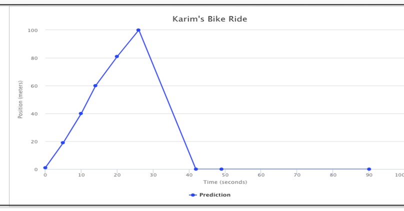
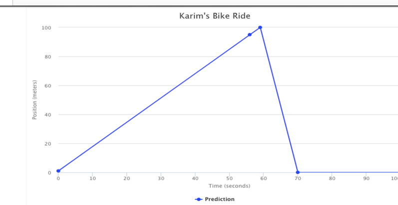
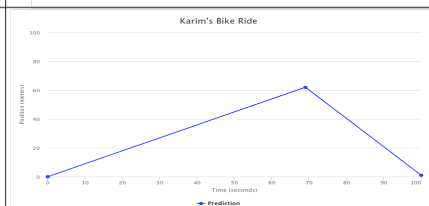
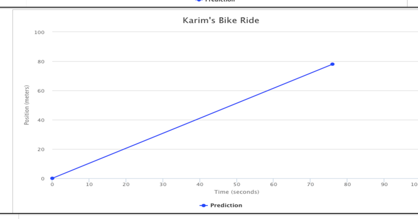
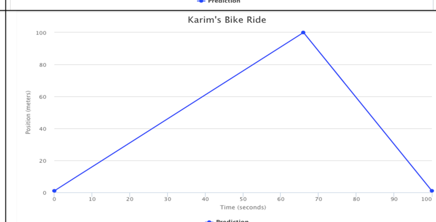
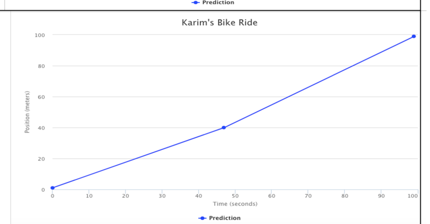
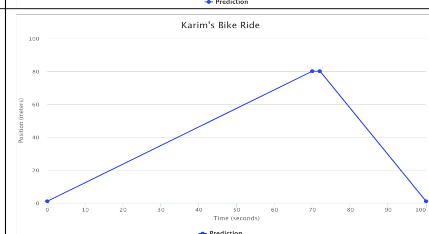
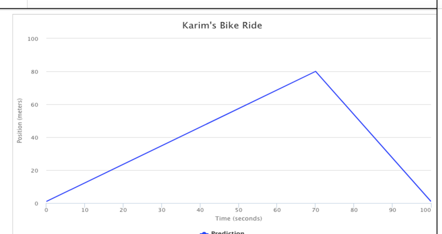
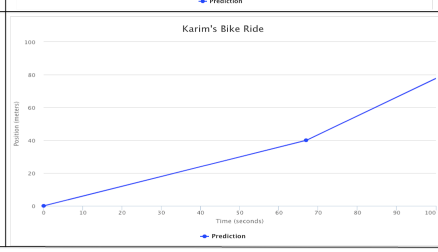
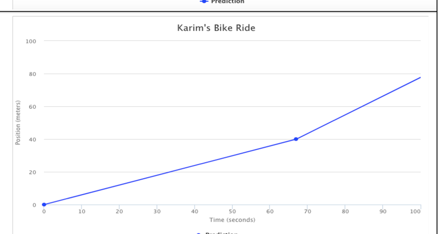
<table>
<tr><th>Category
(with example student explanation)</th><th>Initial Graph</th><th>Revised Graph</th></tr>
<tr><td><u>No Revision:</u>
Nothing really changed in my graph because, like I said he said he wanted to go up hill pretty slow, but super fast going down, so he's going to take more time to go up hill and take less time to go down hill.</td><td></td><td></td></tr>
<tr><td><u>Slope:</u>
I changed when Kareem went up because he must have went slower going uphill and faster going down.</td><td></td><td></td></tr>
<tr><td><u>Direction:</u>
I Change the graph so that it represents the distance is the y axis. I made these changes because my graph used to say he went back to the same place, I want it to show that he went somewhere.</td><td></td><td></td></tr>
<tr><td><u>Slope and Direction:</u>
i change the two points because the graph before shows him goes up hill very fast and down hill very fast</td><td></td><td></td></tr>
<tr><td><u>Arbitrary:</u>
I got rid of the small space of time where he isn't moving at the top of the hill.</td><td></td><td></td></tr>
<tr><td><u>Correct:</u>
I kept it the same because mine shows him going up the hill slowly, but then going back down faster, but still going in the same direction, not going back.</td><td></td><td></td></tr>
</table>

Figure 3.6: Graph revision categories rubric with student examples of their revision rationales (left column), and one example for each category showing initial graph (center) and revised graph (right)

A comparison of pretest graph revisions and posttest graph revisions are presented in Table 3.1.

Table 3.1: Number of students that made each graph revision type on the pretest and posttest assessment item.

Revision Category	**Pretest: % (N)**	**Posttest % (N)**
No Revision (incorrect)	41.3% (95)	39.1% (90)
Slope	13% (30)	15.7% (36)
Direction	9.1% (21)	10% (23)
Slope and Direction	0.9% (2)	0.4% (1)
Arbitrary	22.2% (51)	15.2% (35)
Correct	13.5% (31)	19.6% (45)

On the pretest, the majority of students (41.3%; N=95) made no revisions to their position-time graphs despite the need to improve their representations, even after critique guidance. Of the students that did make revisions, 13% (N=30) revised the slope on their graph, often making the speeds more distinct between the slow beginning and steep/fast ending of their graphs. 13.5% (N=31) of students constructed an accurate graph on the first try on the pretest. A large group of students attempted to revise, but made only arbitrary changes to their graph (22.2%; N=51). Fewer students (9.1%; N=21) adjusted the line direction of their graph, though this is the area where most students needed to revise. Only two students revised the direction and slope of their graph. On the pretest, 12% of students that did revise their graphs (N=13) revised to a correct graph.

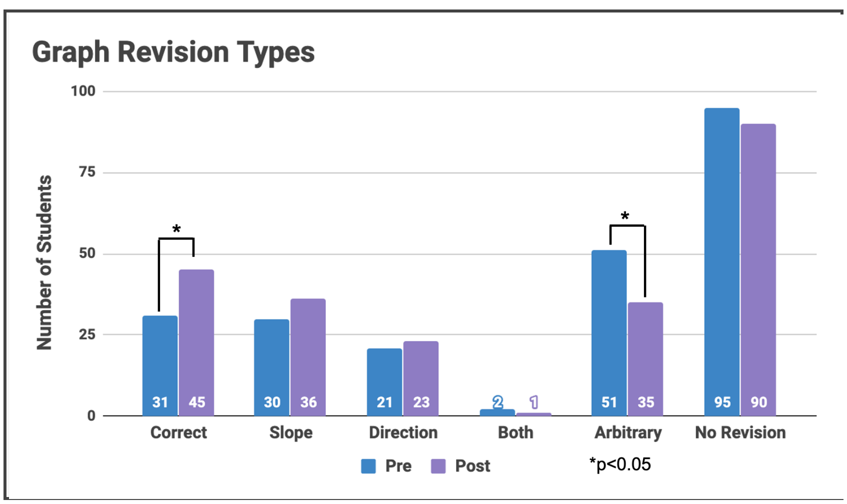


Figure 3.7: Graph of students' graph revision types, pre and post.

On the posttest, significantly more students (19.3%, N=45) [two-sample test of proportions: z(229)=-2.016; $p<0.05$] correctly drew the graph, and had no need to revise (Figure 3.7). Additionally, fewer students made arbitrary changes to their graphs on the posttest compared to the pretest [z(229)=1.73; $p<0.05$]. Of those that did not correctly draw the graph, many (N=90, 39.1%) chose not to revise their graph, generally citing that they believe they were already correct despite their graphs going back to the starting position. About 15.7% of students (N=36) revised the slope of their graph, either correcting the speed or making the difference between the two speeds more apparent. Many students on the posttest still needed to revise the direction of their line, but failed to recognize their error, keeping their "graph-as-picture" hill representation, even after guidance. Only 10% of students (N=23) revised their line direction. More students chose to revise their graphs on the posttest, but this did not translate to a significant improvement in graph construction overall. On the posttest, of the students that revised their graph, 18% revised to a correct graph (N=20).

Next, I examined how students' revisions shifted from pretest to posttest (Table 3.2). The largest category consisted of students that did not revise on the pretest, and again did not revise on the posttest (N=42). This is consistent with our previous findings (see chapter 2) that many students do not revise when prompted; the curriculum and the guidance was not enough to encourage these students to revise by the posttest.

Table 3.2: Graph revision shifts from pretest to posttest (only students that completed all items)

		Graph Revision Type on Posttest						
		Arbitrary	**Correct**	**Direction**	**No Revision**	**Slope**	**Both**	Total (Pretest)
Graph Revision Type on Pretest	**Arbitrary**	9	5	4	16	12	0	46
	Correct	1	19	1	7	2	0	30
	Direction	3	4	5	4	4	0	20
	No Revision	11	12	5	42	10	0	80
	Slope	7	1	3	5	5	1	22
	Both	0	0	1	1	0	0	2
	Total (Posttest)	31	41	19	75	33	1	

However, some students (N=26) that did not revise on the pretest attempted revision on the posttest. 5 of these revised the direction of their line to have a correct graph. 11 attempted revision, but only made arbitrary changes to their graph. 12 of these students drew the correct graph on their first try on the posttest, and 10 improved the slope/speed representation of their graph. There were also significantly fewer arbitrary revisions on the posttest than on the pretest (see Figure 3.7 above). These students moved to other types of revisions (though some decided not to revise). On the pretest, 13 students (6%) that drew an initially incorrect graph revised to a correct graph; on the posttest, only 20 students (9%) revised to a correct graph that did not already have a correct graph.

Graph critique analysis: Choose the best graph

To help further understand why so many students struggled with line direction errors on the posttest, student responses to the following critique item were analyzed (Figure 3.8). After constructing their own graph, students were asked to analyze graphs created by two other (fictional) students and to pick which one best represented the bike ride. This item was created to include the most common student graph, where they show some speed difference, but in general produce a picture of a hill with a "steep" edge on the far side. The correct graph is also presented to encourage students to think about what it means when the line keeps going up, but at a steeper angle. The goal of this guidance was to help students recognize the gaps in their own thinking and graph construction by encouraging them to distinguish between what each of these graphs represents in reality in terms of position over time. Students' critiques were scored to see if they improved in their ability to critique graphs from pretest to posttest.

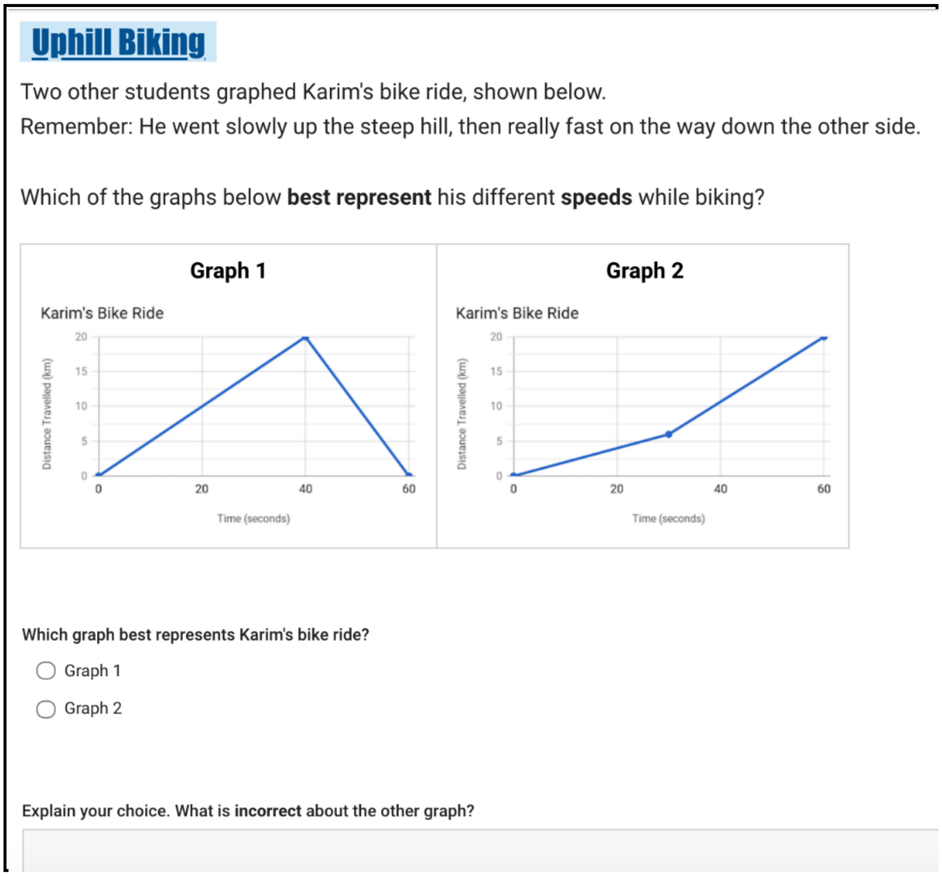


Figure 3.8: Critique guidance item completed by students after constructing their own graph and before being asked to revise their graph on the pre and posttest.

Students' graph critique responses were scored using a 1-5 scale KI rubric (Table 3.3), which rewards students for making connections between valid scientific ideas, such as a steeper slope representing a faster speed.

Table 3.3: Knowledge Integration rubric for critique of two different graphs

KI Score	Description	Examples
1	Irrelevant or off topic	"I don't know"
2	Non-normative ideas, only includes graph as picture without including anything about slope and speed	"I think the first one is correct because the seconds graph doesn't have the segment about Karim riding down the hill." "It shows him going up the hill, then keep going up the hill, not DOWN the hill."

3		"Karim is still coming back not keeping going. With that data he would be flying."
3	Includes at least one normative idea about speed/slope, even if graph-as-picture interpretation is included	"Graph 1 best represents Karim's bike ride because it shows how long he took when biking up and how fast he went when he biked down. Graph 2 however does not show Karim biking back down at all."
4	Expresses that the line going up on the graph means going further away from the starting point	"The other graph claims that he undid the distance that he went because it goes back down. The y-axis does not represent his position but the distance he has gone."
5	Includes both speed and direction in relation to the graph	"BEcause the first graph shows Karim going back towards where he started, even though he went down the other side of the hill. The second graph shows him going slowly up the hill, the quickly going back down the other side of the hill."

Students did not choose the correct graph significantly more often on the posttest than on the pretest; 36% of students chose the correct graph (graph 2) on the pretest, whereas 42% chose the correct graph on the posttest. However, students were significantly better at critiquing the graph on the posttest than on the pretest, with a mean KI score of 2.4 on the pretest, versus 3.0 on the posttest [$t(175)=7.10$; $p<0.001$]. This shows that students moved towards having a better understanding of the relationship between slope and speed on the posttest, but still likely struggled with the meaning of line direction versus a pictorial understanding of the graph.

Students' critique explanations show that students often chose the correct graph for the wrong reasons on the pretest, often interpreting a line moving upward as showing an increase of speed (for example, going up means increasing speed, going down means the speed is decreasing). Therefore, many students explained that they chose graph 2 (the correct graph) on the pretest because the line continuing upward represented speeding up. While overall students improved their critique abilities by the posttest, most students still struggled to recognize their graph-as-picture interpretation in order to correctly construct or revise their graphs.

To examine this further, one relevant embedded item was analyzed. Overall on the posttest, 72% of students still failed to correctly graph the line direction, or recognize and revise their line direction error. Only a total of 25.5% of students correctly graphed the line direction on their final graph. To investigate whether students understood the meaning of the line direction on a position-time graph, an embedded item was analyzed that asked students: "What does it mean when the line on the graph is going down." Student responses were coded as correct or incorrect based on whether they included the main idea, for example: "When the line on the graph is going down, it means that [they are] going back to the start". A total of 72% of students correctly answered this embedded essay question. Students were able to describe in words a correct understanding of this abstract graphical feature, but were largely unable to construct a graph with this feature correct on the posttest, even after revision.

Revision Rationale Analysis

After revising their pre/post graphs, students answered the open response question: "What did you change about your graph? Explain why you made these changes." A rubric was created to categorize students' rationales for revision of their graphs (Table 3.4).

Table 3.4: Revision rationale rubric with student examples.

Category	Description	Examples
No explanation	No Revision (no further explanation)	I didn't change my graph
No Revision - "I was correct"	Student explains that they chose not to change anything because they believed their original graph was accurate	i didnt change my graph because it's correct i didn't change anything because I thought it was accurate enough.
Superficial changes	Student attends to superficial visual aspects of the graph that they adjusted during revision. This includes graph-as-picture explanations (ex. Made it look more like a hill), as well as students that adjusted arbitrary aspects to make it more closely match the example graphs from the previous critique item.	i shaped it more like a hill and put time in to play more efficently I made him go up and down multiple mountains I changed the endpoint because the others' graphs ended at 60 seconds.
Distinguished Ideas - Line Direction	Student explains how they realized the line going down means the biker went back to his starting position, and adjusted their line to show the biker kept going further from the start	I realized that it does not come back to where he starts. I didn't put that he ended at 0 because its the distance he traveled not that he went from zero and returned to zero I changed the graph so it no longer showed Karim going back up the hill. I changed the direction of the second part of the line because at first I made it so that it said that he was biking back to the start, yet he was biking away from the start. I made these changes because I made the graph incorrectly and I wanted to show that

		he was biking away from the start, instead of biking towards it.
Distinguished Ideas - Slope	Student recognizes that their graph was not showing two different speeds accurately, and adjusts this	I made him take longer on the way up because I realized it would take him longer to go up the hill compared to going down it. I made the line where hes going up the hill less steep because the flatter the line the more time it takes to go up the hill.
Still confused	Student changed something (or didn't) because they were still unsure of whether or not they were doing it right, and did they best they could	i dont know Im still not sure if its right

Students' responses on the pre and posttest were sorted into one of these categories (Table 3.5).

Table 3.5: Students' revision rationale on the pre and posttest

Rationale Category	**Pretest: % (N)**	**Posttest % (N)**
No Explanation	11.8% (33)	11% (30)
No revision - "I was correct"	12.5% (35)	20.1% (55)
Distinguished Ideas - Direction	4.6% (13)	7.7% (21)
Distinguished Ideas - Slope	12.1% (34)	16.1% (44)
Superficial	16.8% (47)	15.8% (43)
Still Confused	3.2% (9)	4.4% (12)

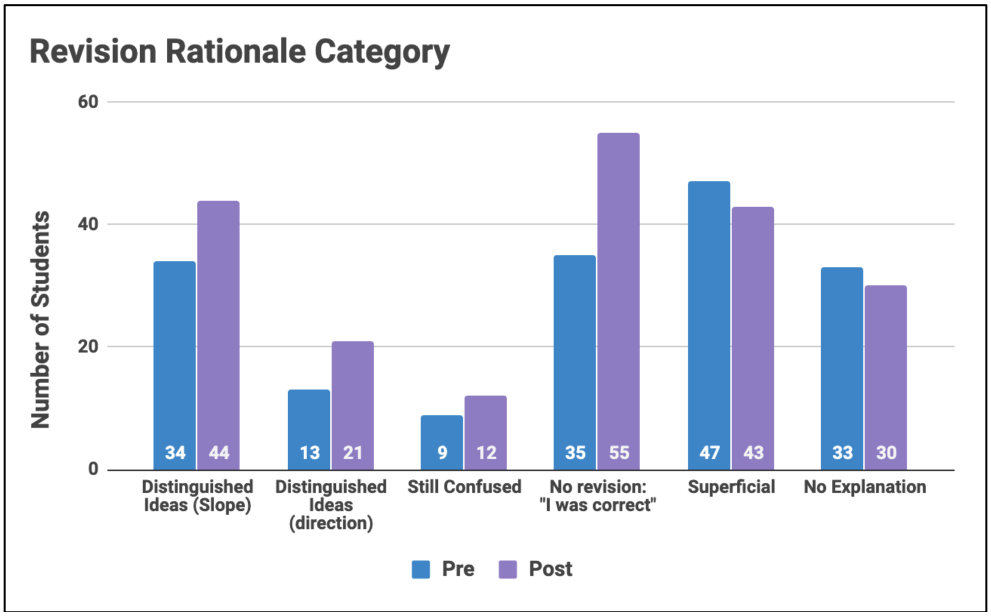


Figure 3.9: Number of students in each revision rationale category on the pretest and posttest

On the posttest, 20% of students (N=55) said they chose not to revise because they believed their original graph was already correct and could not be improved upon (Figure 3.9), but only 15 of those successfully constructed an accurate graph on the first try on the posttest. This is one of the biggest categories, and reveals the main reason that students choose not to revise is because they don't see a way of improving their answer. They have trouble distinguishing between new ideas and their prior knowledge, and the critique item was not helpful; several said they chose the (incorrect) graph on the critique item because it looked like the one they constructed.

Some students (15.8%, N=43) described superficial changes that they made to their graph on the posttest. Fewer students cited superficial reasons for revision on the posttest than the pretest, though this was not significant. Many students in this category explained that they wanted to make their graph look neater or more like the graph they saw on the page before. These are students that attempted revision, but did not change anything fundamental about their graph. However, they were attempting to make their graphs more readable and easier to interpret. These students are not necessarily recognizing gaps in their understanding, but are still revising for clarity.

On the posttest, 24% of students (N=65) distinguished ideas with the help of the critique guidance. 16% (N=44) distinguished ideas about slope, and explained how their revisions to their graph attempted to more accurately show how variations in line steepness represent different speeds. A smaller number of students (7.7%, N=21) distinguished ideas about line direction, describing their initial graph-as-picture hill error and realizing that the second half of their graph should continue upwards rather than showing that he returned to his original position, counter to the story.

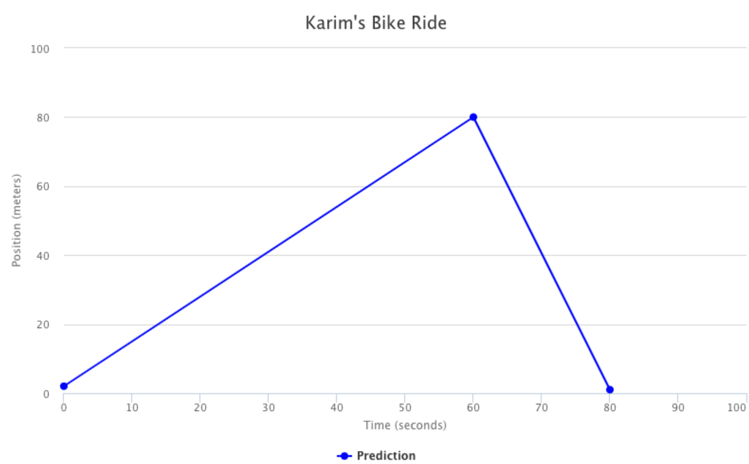
Students that distinguished ideas in their rationale for revising on the posttest also scored, on average, 0.4 points higher on the critique guidance item [$t(204)=2.21$; $p<0.05$]. This suggests that students that were successfully able to critique another student's graph were also better at distinguishing their own ideas and recognizing gaps in their understanding. Some students even cited the critique item as the reason they realized their errors (see examples below). This shows that while many students were still struggling with aspects of position-time graph construction, the curriculum helped students wrestle with their prior knowledge and integrate new ideas about abstract representation of motion.

Student Examples of Guidance Use

Examples were chosen to further explore the various ways that students used the guidance items to revise their graph and to explain their revision rationale.

Student 1: Misuse of guidance

Initial Graph:



Critique:
(Graph 1; incorrect) when someone goes downhill quickly on a graph, the line goes down quickly, graph one shows that, while graph 2 doesnt

Revised Graph:

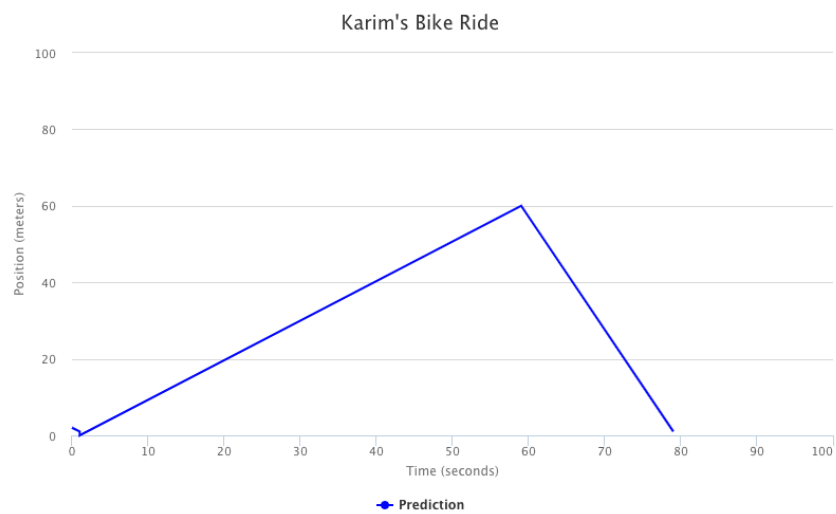
No Revision (maintained graph-as-picture error)
Rationale Explanation: i didnt change my graph because it's correct (No Revision - “I was Correct”)

This is a common “no revision” example, where the student expressed that they didn’t make any changes because they believed their graph was correct. This student has the common graph-as-picture error, as described in their critique: “when someone goes downhill quickly on a graph, the line goes down quickly”. This shows how the critique guidance did not help all students distinguish ideas or recognize the gaps in their understanding. This student requires further guidance to move past this difficulty.

Student 2: Arbitrary revision

Initial Graph:

Critique: (Graph 1; incorrect) The other graph does not show Karim going down.
Revised Graph:



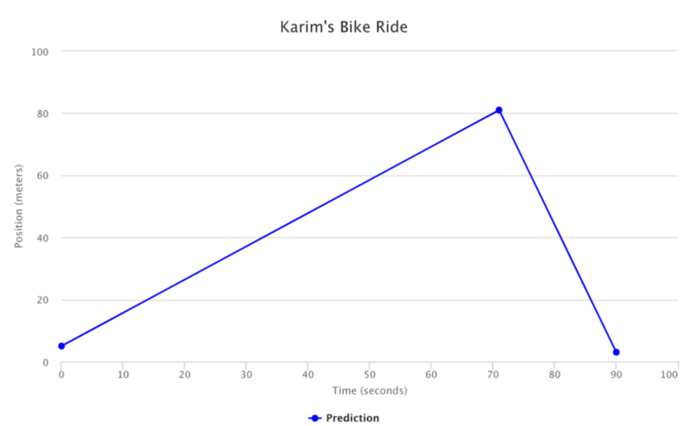
Arbitrary Revision (maintained graph-as-picture error)

Rationale Explanation:
I don't think I really changed anything; I just tried to make it look "neater". (superficial)

Similarly, this student maintained their graph-as-picture error on their revised graph. They revised, but only paid attention to superficial aspects of the graph, making it "neater". While they may have had difficulty with the technology, their revision makes the graph more readable based on what they intended to show. Improved communication of your ideas is an important aspect of revision, but students like this require more guidance to help recognize more fundamental issues in their graphical representations.

Student 3: Ideal use of critique guidance

Initial Graph:



Critique:

Graph 2 (correct): I think graph 2 is more accurate because the graph is following the distance he traveled total. Not the elevation. He didn't travel from zero to 60 to back to zero because he, in total traveled farther.

Revised Graph:

Revised Direction (correct)

Rationale Explanation:
I made it so that he traveled about 30 meters in roughly 60 seconds from zero then traveled 50 more meters in about 20 seconds showing when he went down the hill. I didn't put that he ended at 0 because its the distance he traveled not that he went from zero and returned to zero
(Distinguished Ideas - Slope)

This example shows the intended use of the critique guidance; the student examined the two graphs in the critique item and successfully explained their choice of the correct graph. This student also explained what was incorrect about the other graph: "..the distance he traveled total. Not the elevation". Their initial graph represented a graph-as-picture hill error, but with help from the critique item the student discovered their error and fixed it in the revised graph. The student further showed how they distinguished ideas in their rationale explanation: "I didn't put that he ended at 0", explaining that the revised graph corrected the line direction from the initial graph.

These student examples reveal various ways that students interpreted aspects of this position-time graph. Some are looking at the graph purely as a picture of a hill, showing a "steep" decline that Karim biked down on the other side.

Other students were not encouraged to distinguish ideas by the guidance item, and simply chose the graph that already looked like the graph they previously drew and citing that as correct. Instead of helping students distinguish between possible ideas, some students believed the critique item reinforced the ideas they already represented in their own graph construction.

Few students were able to use the guidance to distinguish ideas and realize that their graph was closer to a picture but actually showed Karim going backwards, which was not a part of the "story" represented in the graph. This was the goal of this guidance, but accomplished this for a relatively small portion of the students.

Interestingly, while many students exhibited a graph-as-picture representation of a hill, most of these graphs were not exclusively pictures. The most common graph shows differences in slope to represent speed, so the “slow” bike up the hill has a less steep slope compared to the fast speed down. If students were only drawing a hill, there would be no reason so many students draw the hill this specific way (Figure 3.10).

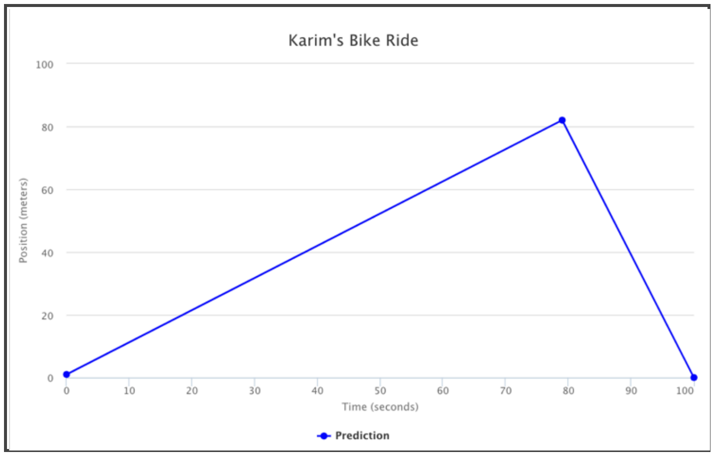


Figure 3.10: Most commonly drawn graph showing a partial graph-as-picture interpretation but including a speed/slope relationship

Discussion and Implications

Graphing, especially constructing position-time graphs, is difficult for middle school students (Bertin, 1983; Donnelly-Hermosillo et al., 2020), partially due to lack of practice in science classrooms (Boote, 2012) as well as the complex abstract nature of these types of graphs.

This study reveals that revision of their own graphs is difficult for students as well, even after guidance, as students have difficulty recognizing inconsistencies or gaps in their understanding. This is consistent with students’ difficulties recognizing inconsistencies in their ideas in their science explanations (see chapter 2). Most students (~85%) needed to revise their graphs on the posttest, but ~50% did not make any changes or changed arbitrary aspects of their graph rather than improving their representations.

The majority of graphs showed inaccurate line direction, and most students (~72%) were unable to recognize and revise this on their graph. Interestingly, while only ~25% correctly graphed the direction of the line on the posttest, 72% of students correctly answered the question “What does it mean when the line on the graph is going down?” on the embedded assessment item during the unit. This suggests students have an easier time explaining what a graph represents than they are at producing that graph; explanations are easier than constructing or revising graphs. Students’ posttest explanations often included their interpretation of the line direction, saying it represented “going down the hill” instead of back to the starting position, which is in direct contradiction to their explanations on the embedded item. The inconsistency here may have to do with the fact that the embedded item contained animation feedback for lines they drew on the graph, and they could watch the swimmer turn around and return to the start when the line was going down on the graph. Additionally, this embedded item was encountered several days before students completed the posttest, and many students may not have recalled this item. This inconsistency requires further study to learn more about student thinking in these contradictory scenarios.

In general, student construction and revision of qualitative position-time graphs did reveal more information about student thinking than simply plotting points (Lai et al., 2016; Shah & Hoeffner, 2002); many started out thinking that a line going down on a position-time graph represented a decrease in speed (slope-height confusion). By the posttest, this idea was all but gone in their representations and explanations. It was replaced with various new ideas, both normative and non-normative. For example, graph-as-picture interpretations and representations remained on the posttest, but some students were able to recognize their errors and revise their graphs correctly.

The revision rationale item gave us new insight into what encourages students to revise or not. Many students that chose not to revise their graphs gave the explanation that they believed their graph was already correct, despite the majority showing an incorrect graph-as-picture representation. This was a large category, and suggests that the main reason that students choose not to revise is because they don't see a way of improving their answer. This is consistent with literature (e.g. Crawford et al., 2008; Flower et al., 1986) that says revision is only necessitated when gaps in knowledge or communication are recognized, and if the student understands how to address those gaps. This requires a great deal of metacognitive monitoring, and is especially difficult when dealing with abstract graphical representations of motion.

However, students improved significantly in their ability to successfully critique others' graphs by the posttest. Learning to critique more effectively first might be more effective than using critique as guidance before students are capable and confident in this difficult scientific practice. A qualitative feature of students' critiques not reflected in their KI critique score came to light while scoring: on the pretest, students were much more likely to say "both graphs are correct, but I think ______ is better", whereas on the posttest, students were more confident in picking a single graph with which they agreed and saying that the other graph is outright incorrect. This suggests that practice graphing and critiquing graphs may help students develop agency, which is important in encouraging students to revise their ideas.

Looking more closely at some student examples also revealed how students were sometimes successful using the critique guidance, but this did not necessarily translate to their graph construction or revision. Other students interpreted the critique item as instantly confirming their initial ideas when they saw a graph in the item that matched their own. The span of student thinking revealed by this sequence of items can be helpful in designing future guidance to help students better distinguish ideas and critique their own initial understanding.

Graphing is an important skill, but students have few opportunities to construct, let alone revise, graphs. These results show that students need further guidance to recognize position-time graphs as a relationship between these two variables rather than a drawing of what is happening. Students also need practice evaluating their own work for correctness, in light of new information they have learned. Additionally, more information is needed regarding students' thinking that leads to their decision to revise or not.

Study 2 tests new guidance items designed to help students further distinguish ideas, as well as to provide us with more information regarding how students are thinking about these representations.

Study 2: Elaborating on student thinking about position-time graph construction and revision

In order to elucidate more information about why students struggled so much with constructing an accurate position-time graph representing a short scenario, I repeated the curriculum with a few changes to guidance, as well as more guidance on the pre/post assessment item to learn more about student thinking.

Students' science ideas are generally a mixture of normative and non-normative ideas, relating to prior knowledge and real-world experiences; students' ideas about graphs are similar. Study 1 showed us that having students construct qualitative position-time graphs resulted in a plethora of information about the many various ways students understood and represented aspects of motion on a graph. Students developed new ideas by the end of the unit and these ideas were visible in their posttest responses. However, several non-normative ideas were prevalent on the posttest, including continued convolution of a picture of a hill with the shape of the graph. In order to both understand more about how students are thinking, as well as to encourage students to distinguish this misunderstanding, new guidance items were added between their initial and revised graphs on the pre/posttest to help students revise their graphs and to provide us with new insights.

Methods

Nine classes of 8th grade students (N=250), taught by two teachers, completed a revised version of the Graphing Stories unit over the course of one week during their science class periods. Students participated in the unit in teams, and individually completed the pretest one day before the unit, and the posttest one day after.

Assessments

The same pre/post assessment was used, asking students to sketch a qualitative graph of a student biking up a hill and down the other side. For this iteration, new guidance was added to help students improve their revisions and to elucidate more about students' reasoning when drawing and revising the graph. This new guidance included an item specifically designed to help students distinguish between their ideas resulting in a picture of a hill on their graph rather than a representation of accurate speed and direction. A picture of a hill was presented, and after drawing their own graph students were asked to think about what information can be gathered from each, and to sort statements into the correct category (Figure 3.11).

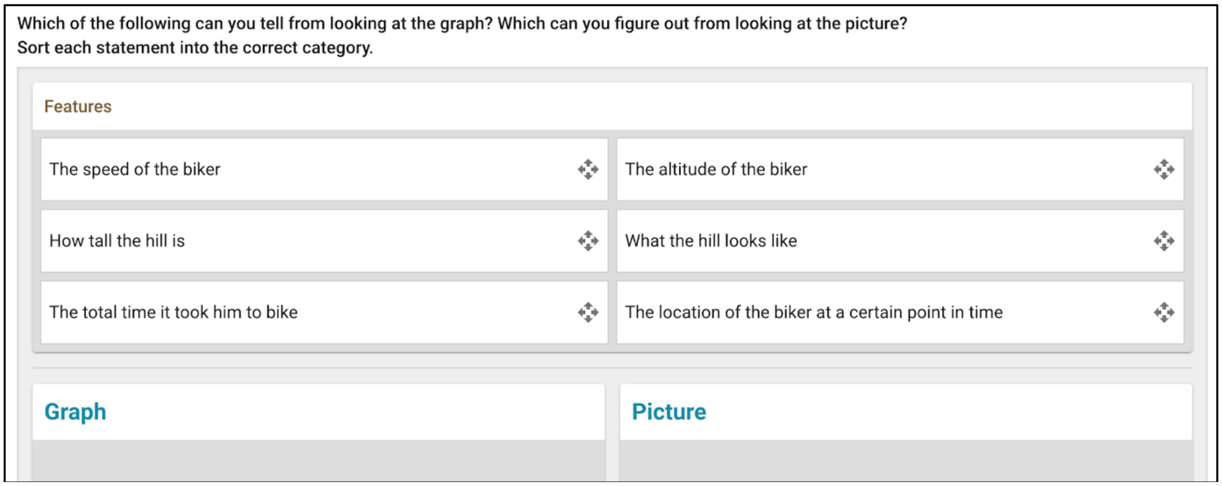


Figure 3.11: Sort and match item asking students to compare their graph to a drawing of a hill.

On the same page, students responded to an open-response item asking: In what ways do the graph and picture look different from each other?

On the next step, students were presented with a new critique guidance item. The critique item from the previous study showed students the most common incorrect graph as well as the correct graph for comparison to encourage students to rethink their own graph. This proved unhelpful for many students, who simply found the graph that looked most like their own and chose that as the correct option. A new critique item was created at this step (Figure 3.12) involving an incorrect graph to encourage students to point out errors, with the goal of helping them distinguish their own ideas and find gaps in knowledge. The instructions also directly addressed students' common graph-as-picture error and asked them to explain how to improve upon that type of error. The design of this item was motivated by success with previous similar critique items in other contexts (see chapter 2), which explicitly informs students that the following answer has incorrect aspects to be improved upon. In this case, the incorrect aspects of the graph include both slope and line direction. Additionally, due to findings from study 1, a question asking about the meaning of the line going down on the graph was included in this same step, in order to make it more salient for graph construction and revision.

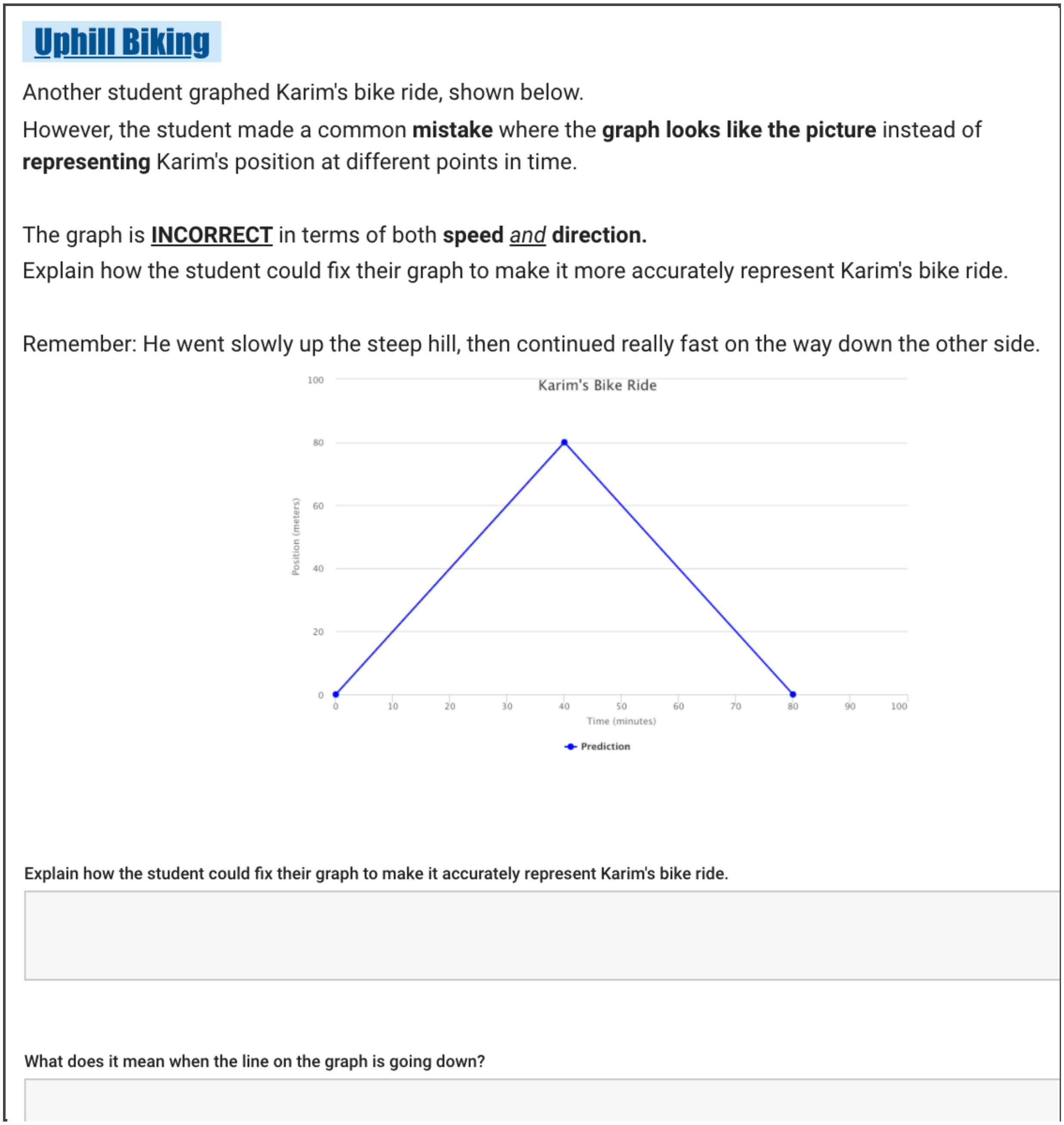


Figure 3.12: Critique guidance redesigned from study 1 to directly address student graph-as-picture error and connect the meaning of line direction to the graph itself

Results

PrePost Analysis: Graph Revision

Students' graph revisions for the hill assessment item were categorized into five different groups based on aspects of the graph that they changed (same as from study 1, but with the "slope and direction" category dropped because there were no students in this category). A comparison of pretest graph revisions and posttest graph revisions are presented in Table 3.6.

Table 3.6: Graph revision types on the pretest and posttest assessment item.

Revision Category	Pretest: % (N)	Posttest % (N)
No Revision (incorrect)	29.2% (71)	38.4% (94)
Slope	23.1% (56)	21.6% (53)
Direction	4.9% (12)	3.2% (8)
Arbitrary	32.5% (79)	26.9% (66)
Correct	10.3% (25)	9.8% (24)

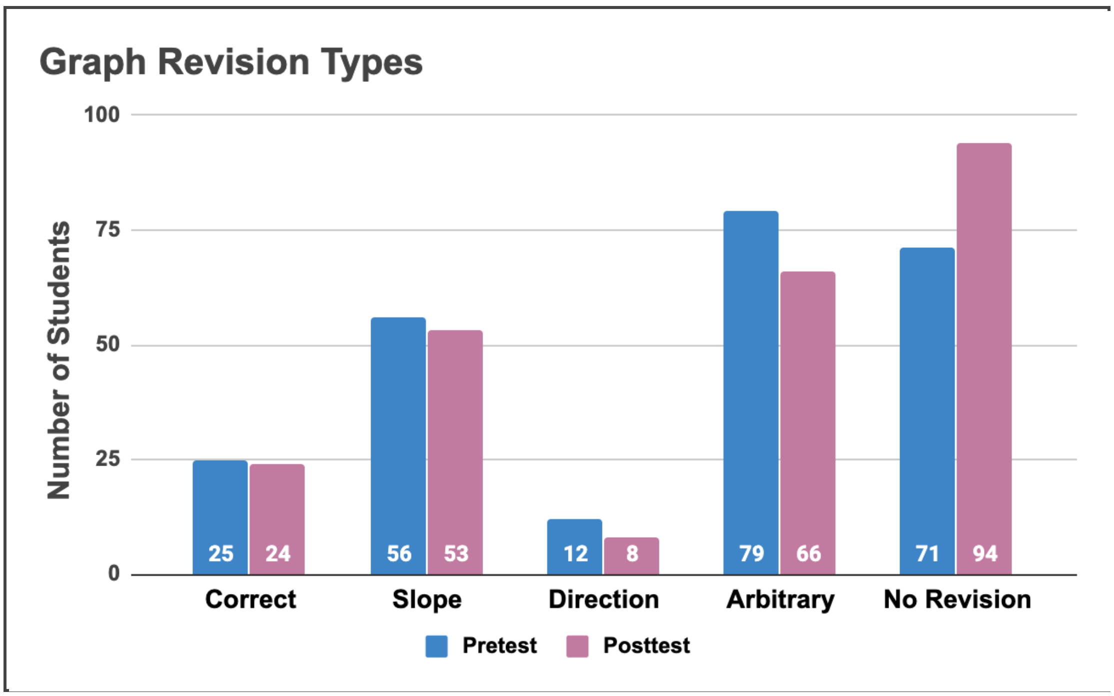


Figure 3.13: Graph of students' graph revision types, pre and post

Similar to study 1, a large percentage of students chose not to revise their graphs on the pretest (29.2%), and even more students chose not to revise on the posttest (38.4%). Many students that chose to revise made arbitrary revisions that did not fundamentally change their graphical representation (32.5% on the pretest, and 26.9% on the posttest; Figure 3.13). These students' revision generally did not improve their revised graph compared to their initial construction, but shows that students attempted to improve even on the pretest before learning about graphs. This may be due to the fact that they encountered more guidance between their original graph construction and revision in this iteration before revising. Only ~10% of students constructed a correct position-time graph on the pretest and the posttest. Students that chose to revise the slope of their graph often improved their representation, and generally made the differences in speed more apparent.

Table 3.7 below shows students' graph revision types from pre to posttest; similar to study 1, most students that did not revise on the pretest continued to choose not to revise on the posttest. Most of the students that correctly drew the graph on the posttest already succeeded at this on the pretest.

Table 3.7: Graph revision type shift from pre to post

		Graph Revision Type on Posttest					
		Arbitrary	**Correct**	**Direction**	**No Revision**	**Slope**	Total (Pretest)
Graph Revision Type on Pretest	**Arbitrary**	28	2	2	30	10	72
	Correct	0	18	0	5	0	23
	Direction	2	3	1	3	2	11
	No Revision	16	1	5	33	10	65
	Slope	14	0	0	17	22	53
	Total (Posttest)	60	24	8	88	44	

51.6% (N=110) of students on the pretest, and 56% (N=124) on the posttest expressed that the graph they constructed was showing the altitude of the biker (Figure 3.14), which may explain why many students drew something resembling a hill - rather than graph-as-picture error, students may have been interpreting the y-axis as distance from the original ground level rather than distance from the starting position. Having this as an option in the matching guidance item, however, may have encouraged students to see this in the graph rather than to distinguish between altitude and position.

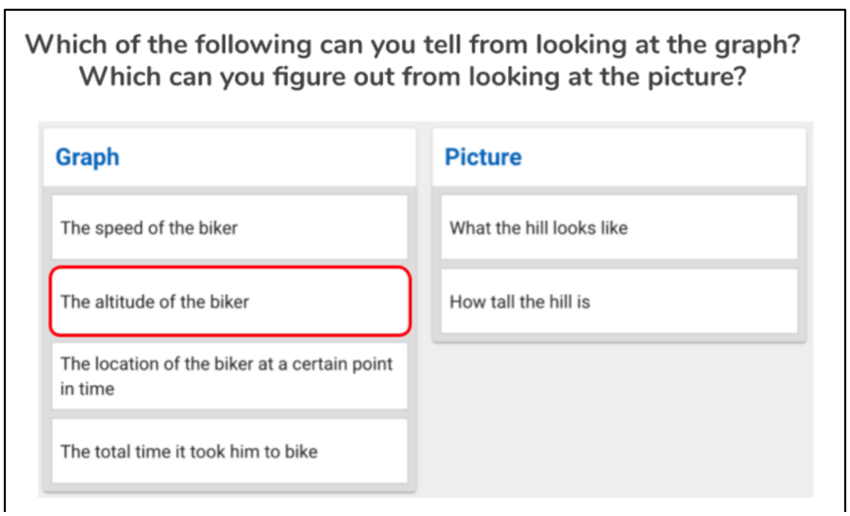


Figure 3.14: Common student answer to matching item asking students to identify what the graph shows versus what the picture shows

This might instead support the idea that students exhibit feature correspondence error, and map visual features of a situation (ex. height/altitude of a hill) onto a similar feature of the graph (Clement, 1985), while showing slope represented by steepness of the line. Similar to study 1, the graph many students produce is not entirely a picture of a hill; most students successfully show steeper slopes representing faster speeds.

Students' explanations on the item following this matching item also show that they understand differences between the picture of the event and a graph representing that event

(Table 3.8). On the pretest, students often cited visual differences between the graph and the picture, whereas on the posttest more students drew on informational differences between the two.

Table 3.8: Example answers: "In what ways do the graph and picture look <u>different</u> from each other?" Students compared a picture of a hill to their own initial graph.

Pretest	Posttest
My hill is much bigger and it much more taller, but the one in the picture the hill is not that big and also the person on the bike is really big.	the picture is just the hill with nothing to tell you how fast he is going and the graph has a flatter line at the start and a very vertical line at the end
The graph is straighter, with fewer curves. also, the graph shows where the biker is at certain points in time. The hill does not.	The graph shows lines and the manipulated and responding variable to tell us how fast he biked up the hill while the picture just shows how it would look like but is not accurate.
The picture is more round at the top while the graph is pointy.	the picture has no way to tell us how long it took for him to get there. while the graph on the other hand shows us where he is on the hill at any point in time.

Graph critique analysis

Students improved significantly on the critique item asking all students to critique a graph-as-picture hill-shaped graph, though students only scored, on average, 2.85 on the posttest compared to 2.76 on the pretest [$t(204)=1.88$; $p<0.05$]. Most students continued to perform close to a level 3, only citing that the slope should be improved on the graph to show variations in speed. While this is an important normative idea, this is consistent with other evidence that students did not improve in terms of line direction understanding.

Revision Rationale Analysis

After revising their pre/post graphs, students answered the open response question: "What did you change about your graph? Explain why you made these changes." The rubric from study 1 was used to categorize students' rationales for revision of their graphs (see Figure 3.6 from study 1 above). Table 3.9 below shows the number of students in each category.

Table 3.9: Students' revision rationale on the pre and posttest

Rationale Category	Pretest: % (N)	Posttest % (N)
No Explanation	9.1% (18)	14.5% (31)

No revision - "I was correct"	13.2% (26)	24.8% (53)
Distinguished Ideas - Direction	2.5% (5)	0.9% (2)
Distinguished Ideas - Slope	35% (69)	30.4% (65)
Superficial	36.6% (72)	28% (60)
Still Confused	3.6% (7)	1.4% (3)

Figure 3.15 represents these findings graphically.

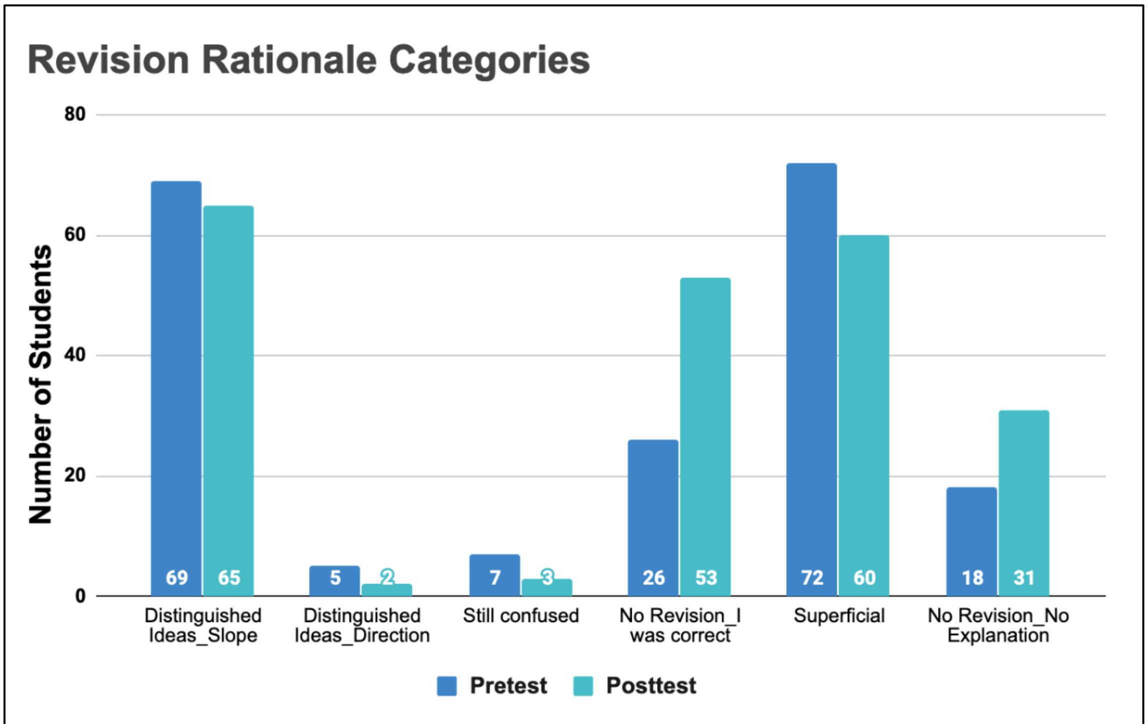


Figure 3.15: Number of students in each revision rationale category on the pretest and posttest.

On the posttest, about 25% of students said they chose not to revise because their graph was already correct, despite only 7 of these students drawing a correct graph. Students cited already drawing a correct graph as a reason for not revising significantly more often on the posttest [$z(188)=-2.97$; $p<0.01$]. Despite not improving significantly at correctly constructing this position-time graph, students became more sure of their answers. Students continued to give superficial reasons for revising on the posttest, such as making their graph neater. A large number of students were able to distinguish ideas about slope, stating that they made lines steeper or less steep to better represent a fast or slow speed respectively. This is consistent with students' critique scores, where they were able to recognize errors with slope and speed on another's graph. The fact that so few students distinguished ideas about line direction is consistent with students struggling to correct this on their graphs.

Overall, whether it's from a graph-as-picture standpoint or mapping altitude onto the graph, students continued to struggle most with line direction on the graph. To investigate this further, a pre/post item was included asking “What does it mean when the line on the graph is going down?”, right after students critiqued an incorrect version of the graph. On the pretest, 85% of students (N=188) did not answer this question correctly, instead saying that the line going down represented decreasing speed or “the person going down the hill”, and these ideas were heavily represented in their pretest graphs. This is expected, since students did not yet learn about position-time graphs.

This question was asked again on an embedded item during the unit in another position-time graph situation, and 90% of student workgroups answered correctly, saying “it means they turn around and go back to the start”.

While students did significantly better on this question on the posttest [$t(199)=4.54$; $p<0.001$], still only 28.5% of students answered this item correctly, despite many more answering this correctly during the unit. This could be due to the fact that the embedded item had accompanying animation feedback, where students saw a swimmer turn around in a pool lane and swim in the opposite direction corresponding to the line on the graph turning downward. Students also encountered this item several days before completing the posttest, and perhaps did not retain this information. However, students may also view the meaning of the line direction as situation-dependent; several student explanations present this as an alternative explanation, where the meaning of line direction changes depending on the situation (Table 3.10). Having seen many different types of graphs throughout the unit, that is generally an accurate assessment - line position and direction mean various things depending on the variables on the x and y axis, such as with a negative-relationship scatter plot encountered in the unit.

Table 3.10: Example students posttest responses about how line direction meaning “depends” on the situation.

Well it all depends on what the graphs about. For example the swimming graph that we did earlier in this test it showed when going down the swimmer was going backwards. But in this case the biker is going fast or slow when the graph is going down. I say "fast or slow " because it all depends when going down how steep the line is, if it is very steep than it means that is going fast, if it is not steep that means that is going slow.
The line going down means in this situation that the biker is going down the mountain, but in other situations in could mean that someone or something is going back to the start.

These items together, while not sufficient to help students move to a more sophisticated understanding of position-time graph construction, help us learn more about student thinking about this complex scenario. Students are not simply drawing a picture of a hill, or misrepresenting the variables on the axes. Some students use their graph to represent altitude (another interpretation of position), which would include the biker going back down on the y axis when he goes down the hill. Students chose not to revise their graphs a majority of the time because they believed they were correct, that they drew the best graph they possibly could on the initial item.

Discussion and Implications

Similar to study 1, a large percentage of students chose not to revise their graphs on the pretest, and even more students chose not to revise on the posttest. Again, many students that chose to revise made arbitrary revisions that did not fundamentally change their graphical representation. While these revisions did not generally improve their revised graph, it shows that students were attempting to grapple with their graph construction, and could potentially lead to better understanding later down the road. The new matching item, designed to elicit more information about student thinking, showed us that approximately half the students on the pre and posttest expressed that the graph they constructed was showing the altitude of the biker, which may explain why many students drew something resembling a hill. Students may have been interpreting the y-axis as distance from the original ground level rather than distance from the starting position. This guidance did not help students recognize that altitude was not actually represented on either of the axes, but did show us more about student thinking in constructing their graph. Having this as an option in the matching guidance item, however, may have encouraged students to see this in the graph rather than to distinguish between altitude and position.

Consistently, students that do attend to a component of the graph in their revisions, rationales, or critique, they tended to only cite the slope as something that should be adjusted. This is likely because this is an easier abstract representation than thinking of how direction on a graph corresponds to direction of movement in reality. Their resulting hill-shaped graphs may exhibit feature correspondence error, mapping visual features of the story (ex. height of a hill) onto a similar feature of the graph (Clement 1985), while still accurately showing slope represented by steepness of the line.

Similar to study 1, students' revision rationales help us understand why students chose to revise the way they did, or chose not to revise their graph. On the posttest, about 25% of students said they chose not to revise because their graph was already correct, despite most needing to revise some aspect of their graph. This reveals further that students struggle to recognize gaps in their understanding and representation, thereby not motivating revision. Flower et al. (1986) says we cannot simply prescribe expert revision habits to novices; it's clear that having students revise more did not help students revise better. Even new guidance, specifically designed to help students overcome their preconceived notions about how a position-time graph should look, was not enough to help students recognize these gaps and revise.

Interestingly, this study shows how context-dependent students' graph interpretations are. When asked what it means when the line is going down on the graph, during the unit 90% students answered correctly, saying "it means they turn around and go back to the start". This is similar to study 1, so for this study I added this question again on the posttest right after revising their own position-time graphs. On that same item on the posttest, only ~30% of students answered this item correctly, some even citing that they remembered answering differently during the unit, but the current situation is different. This suggests students might benefit from exploring different contexts through graphs side by side to see similarities and differences.

Conclusion

Overall, these studies give us more insight regarding why and how students revise, specifically when constructing their own position-time graphs. Study one shows that students largely do not revise their answers, and critique does not help if they are already struggling with the material. Critique is an advanced skill that requires more practice in itself. To investigate this further, study 2 added more guidance specifically designed to support students with recognizing contradictions in their understanding of position-time graphs. However, even with more guidance, students struggled to revise their graphs successfully. These additional guidance items, however, did allow us to see more of what students were thinking while constructing their graphs, such as including altitude as a feature in their representation.

The importance of graphing skills in science is well documented in literature (e.g. Clement, 1985; DiSessa & Sherin, 2000; Shah & Hoeffner, 2002; Hattikudur et al., 2012; Boote, 2012; Vitale et al., 2015) as well as current middle school science standards (NGSS Lead States, 2013). Graph skills are necessary for elucidation of patterns and underlying processes of complex scientific phenomena (Friel et al., 2001; Wu & Krajcik, 2006). This set of studies supports the idea that having students construct graphs is a useful way to reveal student thinking about scientific information (Lai et al., 2016). Having students construct qualitative graphs, rather than simply plotting points, allowed students to demonstrate complex combinations of normative and non-normative ideas (Hattikudur et al., 2012; Stylianou et al., 2005; Vitale et al., 2015; DiSessa & Sherin, 2000). Specifically, students revealed ample attention to the meaning of slope in reference to speed on a position-time graph, while struggling with the meaning of line direction. Students sometimes conflated the y-axis of their graph with altitude rather than position from an origin point, and often produced graphs with a combination of graph-as-picture aspects and normative representations.

In order to be successful in graphing practices students must be able to think with graphs; this includes constructing and critiquing graphs as well as interpreting them (Lai et al., 2016). Critique is rare in most literature on graphing (Donnelly-Hermosillo et al., 2020; Lai et al., 2016) and likely rare in classrooms. This set of studies reveals that students succeeded in critiquing incorrect graphs and improved at this skill by the end of the curriculum. However, students often only critiqued the aspects of the incorrect graphs with which they were already familiar, emphasizing that critique as a skill depends heavily on prior knowledge (supported by chapter 2 findings). Students often critiqued the slope of the graph because this is a graph aspect they could draw and recognize successfully, but failed to point out errors they themselves also struggled with.

Students have very little opportunity to construct, let alone to *revise,* their own graphs. This chapter explores various types of guidance to help students revise their position-time graphs. The value of revision of writing is well known (Flower et al., 1986) and the revision of students' written science explanations is explored in chapter 2. Here, we see that students' graph constructions can benefit from the same processes. This includes opportunities for students to recognize and diagnose gaps in their understanding or representation of ideas. The abstract nature of graphs, especially position-time graphs, makes them even more difficult for students than writing explanations; this is supported by study 1, where students successfully answer the question about line direction, but fail to incorporate that idea onto their constructed graph. The added complexity of graphing makes revision even more necessary for students to have time to distinguish their ideas, reflect on their thinking, and improve their graphical representations (Lai et al., 2016). Students in this study benefitted from revisiting and reevaluating their graphs, even if they did not end up succeeding at understanding all aspects of their graphs.

However, similar to our chapter 2 findings, students often do not revise their graphs when prompted, even after guidance. Students' revision rationales help us understand student thinking during the revision process. Many students chose not to revise because they believed their original graph construction was correct. Many of these students did not have correct initial graphs, meaning they were unable to diagnose the gaps in their understanding. It is likely this is the general rationale for not revising in other contexts as well; a student that believes they have produced their best possible response has no need to revise. Students that chose to revise substantial portions of their graph often did so after using the guidance to distinguish ideas, though a small percentage of students in both studies were able to do this. This suggests that students would benefit from more activities that help them distinguish ideas before revising their graph, such as explicit instruction on how to critique these types of graphs.

Graph construction and critique are difficult for students; our curriculum helped them improve these skills, but not become proficient at position-time graph construction, critique, or revision. Students continued to struggle with construction and revision of qualitative position-time graphs by the end of the curriculum in both studies presented here. However, this task is difficult, and even older students and adults make some of the common errors observed in these studies (Clement, 1985). Students were able to improve on some aspects of this task, including (partially) successfully critiquing an incorrect representation of the graph in question. Students' revision rationales are extremely valuable in revealing a major reason students failed to revise their graphs successfully. Like students' essay revisions in chapter two, revision of graphs may involve a self-directedness. Recognizing your own gaps in knowledge involves metacognitive techniques (Zohar & Barzilai, 2013; Brandsford et al., 2000; Hattie, 2009; Kuhn, 1999), but these techniques are often themselves complex and also require practice before students become proficient. Self-directed learning (SDL) involves students taking greater ownership of their learning, including self-evaluation of what you do and do not know, and the independent pursuit of learning those things which you do not already know (Hmelo, 2004; Loyens et al., 2008; Schmidt, 2000; Candy, 1991). The students that revised the most successfully were those that used the given guidance to distinguish new ideas, and incorporated those new ideas into their graph revisions. However, the critique, matching, and graph comparison guidance provided was not enough to help the majority of students distinguish their ideas. Future curricula can be designed to support specifically helping students plan and investigate those things they do not already understand. Consistent critique of their own work throughout the unit may be beneficial as well.

Chapter 4: Promoting questioning and revision to support Self-Directed learning skills for middle school students

A major finding from chapters 2 and 3 is that, while the ability to critique ideas and recognize gaps in your own understanding helps with revision of ideas in science, a majority of students still do not attempt to revise their science ideas. Instead, students often explained that they decided not to revise because they were happy with their existing response, even when their responses had clear errors. Even when encouraged to critique, students failed to diagnose errors in their understanding and revise their explanation. To respond, in this chapter I investigate ways to promote self-directed learning (SDL) and examine its relationship with students' ability to revise, critique, and investigate their own questions.

In this chapter, study 1 examines the importance of student-generated questions, a prominent aspect of SDL. Students were asked to generate questions they still had about genetics at the end of the unit, and these questions were categorized based on whether they went beyond the scope of the unit and addressed bigger ideas. The relationship between question generation and revision throughout the unit is examined as well. In addition, self-directed learning involves more than simply asking questions; students also need to be able to investigate their questions. Thus, in study 2, I investigate students' ability to use a dataset to develop and investigate their own question about the data. Students generate, critique, and revise their question, then generate a graph to test their prediction. I look at the quality of student questions based on whether they can be investigated, and capture students' revision habits throughout this activity. I interpret these findings in terms of self-directed learning and knowledge integration.

Introduction

Substantial research illuminates the importance of self-directed learning (SDL), particularly in technological learning environments (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Bannert et al., 2015). Historically referred to as "intentional" or "autonomous" learning as well as "self-directed", this capability is popular in adult education (Gatewood, 2019). It has also become recognized as an important goal in science classrooms more broadly (Hmelo, 2004), as well as within the current wave of reform for national science standards (NGSS Lead States, 2013). There are many definitions of SDL. Generally, it involves having students take a broader role in the selection and evaluation of learning materials (Loyens et al., 2008), giving the student, rather than the teacher, the power to define their own learning activities (Schmidt, 2000). This can include the freedom to evaluate learning needs, decide on the content of "learning issues", and implement learning strategies (Fisher et al., 2001). Much research employs Candy's (1991) framework for SDL, which includes: personal autonomy, self-management, independent pursuit of learning, and learner control of instruction. Hmelo (2004) breaks SDL down into four subskills: metacognitive awareness, student knowledge of what they do and do not understand; ability to set learning goals, student ability to identify what they need to learn more about; ability to plan and select appropriate learning strategies, student ability to decide on course of action to reach their goals; ability to monitor and evaluate progress, student ability to determine whether or not their goals have been attained. These definitions and subskills encompass aspects of self-directed learning and self-regulated learning (SRL), which focuses more heavily on metacognitive monitoring of learning, intentional learning (Scardamalia & Bereiter, 2006), and autonomous learning (Linn & Eylon, 2011). For example, Loyens et al. (2008) describes how SDL can include SRL, but not the other way around. SDL additionally deals with *what* learning issues students investigate as well as how they regulate their learning. In other words, students formulate their own questions and "learning issues" to investigate.

Our research group devised a definition of SDL appropriate for middle school students grounded in literature, NGSS science and engineering practices, as well as consultation with a group of middle school teachers with whom we work (Bradford & Gerard, 2020). We asked: what would self-directed learning look like with your middle school students? What practices would show you that they are self-directed learners? After some discussion, the following perspective emerged, and has been revised several times. Self-directed learning that leads to knowledge integration occurs when a learner is motivated by their personal experiences to:

1. Generate their own questions
2. Gather evidence to reason about evidence-based conclusions to their questions
3. Select strategies that work for them
4. Determine when and from where to seek guidance and feedback
5. Engage in self-assessment to reflect on, evaluate, and modify their strategies and products to meet their learning goals, and
6. Communicate their ideas and understanding

In the following studies, I use this breakdown of SDL-related practices, and specifically focus on the first and fifth components: students generating their own questions, and engaging in self-assessment to reflect on, evaluate, and modify their strategies and products.

Generating Questions

A prominent indication of learners who are self-directed is the ability to generate meaningful questions to both check their understanding and pursue more personally meaningful learning goals (Biddulph et al., 1986). Generating, as opposed to only answering, questions has, additionally, been recognized as a skill that promotes agency and engagement (King, 1992), particularly when science and engineering practices are involved (Barton & Tan, 2018). Indeed, student questions that address scientific concepts have been shown to promote enhanced understandings of the content and better integration of ideas (Chin & Brown, 2002; Linn & Eylon, 2011). The following studies in this chapter investigate students' question quality as a measure of self-directed learning, and to see what other self-directed behaviors co-occur. As the debate about Bloom's Taxonomy suggests, the definition of a "good question" in science is complicated. Knowledge integration research emphasizes questions that can be resolved with evidence to help students distinguish ideas (Linn & Eylon, 2011). Chin and Brown (2002) developed a rubric for students asking "wonderment" questions, which go beyond factual or procedural concepts, to include comprehension, prediction, anomaly detection, and application. This definition of a "wonderment" question is very useful for showing whether students are thinking beyond the material presented in the curricula, in a way where they can apply new information. However, "wonderment" questions, specifically where middle school level genetics is concerned, are very difficult for students to investigate without intense research and scaffolding. Another way to define a "good question", specifically in science as well as in self-directed defining of "learning issues" (Hmelo, 2004), is the ability to investigate that question.

Can you form a hypothesis, research, and answer that question? If not, can you narrow your question to something that is possible to investigate with available resources? In the following studies, I examine students' abilities to generate both wonderment questions and narrow research questions that they can investigate through data manipulation.

Revision and SDL

While the various definitions of SDL explicate different components or sub skills, all in some form support the idea that reflection and self-assessment are crucial components (Candy, 1991; Blumberg, 2000; Hmelo, 2004). Our definition includes this component as well: "engage in self-assessment to reflect on, evaluate, and modify their strategies and products to meet their learning goals." This involves cycles of critique and reflection, processes necessary for revision of students' ideas. Revision involves the need to assess the gaps in your own knowledge (Flower et al., 1986) and investigate these further in order to revise by integrating newly learned ideas with prior knowledge (Hmelo, 2004; Liu et al., 2008). I investigate students' abilities to reflect on and improve their understanding of scientific concepts by revising their arguments and explanations at various points throughout our curricula, and examine how revision frequency and success are related to other measures of SDL.

SDL for Middle School Students

Self-directed learning has historically been designed for and studied within adult education, specifically medical school, and students still reported doubting their self-efficacy in determining their own questions to investigate, also termed "learning issues" (Dahlgren & Dahlgren, 2002; Loyens et al., 2008). Even upper-level medical school students struggled with what necessary knowledge they had and still needed to investigate further, what incorrect ideas they still held, and what resources were reliable/useful to learn that information (Dahlgren & Dahlgren, 2002). One can imagine that middle school students would have an even more difficult time with these aspects of SDL, and would require more scaffolding (Hmelo, 2004). Technology has been shown to have a positive effect on self-directed learning and student engagement (Rashid & Asghar, 2016) and can provide useful scaffolds for students struggling with self-efficacy. Bannert et al. (2015) found that metacognitive prompts in a tech-based learning environment helped students form their own metacognitive scaffolds and continue using them afterwards. Muller and Seufert (2018) found that SDL-related prompts increased self-efficacy in a hypermedia learning environment. Therefore, our WISE online learning environment is ideal to scaffold questioning and critique to encourage middle school students to conduct their own investigations. In the same vein, Candy (1991) considers SDL a learning goal as well as a process, and others found that SDL abilities have a learning curve, with upper-classmen exhibiting greater abilities to define their learning goals and higher feelings of self-efficacy than lower-classmen (Loyens et al., 2008; Hmelo, 2004). Therefore, our middle school curricula include heavier scaffolding for the aspects of SDL we wish our students to practice, including explicit critique and revision steps to help with the generation of narrow scientific questions and investigation of data.

Theoretical Framework: Knowledge Integration and SDL

These studies, as well as the Web-based Inquiry Science Environment (WISE) units employed here, were designed according to the Knowledge Integration (KI) framework. This is a constructivist framework built on the idea that students learn by successfully integrating new information with their existing understanding of how the world works (Liu et al., 2008). KI design involves cycles of *eliciting* student ideas, *adding* new information, helping students *distinguish* between normative and non-normative ideas, and finally *reflect* and *connect* their new knowledge to their prior understanding (Linn et al., 2014). These KI steps require active engagement, self-assessment, and reflection on the part of the student, making this theoretical framework appropriate for investigating measures of SDL. Specifically, questioning is an important aspect of KI; eliciting student questions (in addition to their prior knowledge) allows students to actively and more knowingly address gaps in their knowledge, and makes reflection on learning more meaningful. Asking questions at the end of the unit, after students have had opportunities to make connections, promotes continual building on their developing understanding of scientific phenomena, which is a hallmark of self-directed learning.

Study 1: Factors affecting student-generated questions in an online genetics curriculum

Introduction

Generating questions is an important scientific practice and component of self-directed learning, yet students have limited opportunities to practice this skill. In this study, I examine the relationship between various aspects of our online inquiry Genetics unit, including students' revisions, and the quality of student-generated questions. Historically, most SDL measures are based on student self-report, where students' respond to Likert-style items about their initiative, openness to challenging opportunities, and feelings of self-efficacy (Teo et al., 2010; Guglielmino, 1989; Kim et al., 2019; Fahnoe & Mishra, 2013). In this study, I look at the nature of questions that student teams generate at the end of an open-source unit on genetics and inheritance as evidence of engagement and agency, as well as examining the quality of explanations, the practice of critique, and revision of ideas as alternative measures of SDL. Students' ability to generate high-quality questions can serve as both a reflection of their engagement and comprehension, as well as a springboard for future interest and investigation in science. Critique is also a self-directed activity to identify gaps; revision is the SDL process of analyzing the gaps and improving your explanation. In addition, the goal is to use these capabilities to create coherent understanding. In order to help support teachers regarding curricular decision, this work aims to identify factors that support students in these SDL-related skills.

Methods

This study includes student data collected from a 10-day WISE Genetics and Inheritance online unit (https://wise.berkeley.edu/project/25176). Ten classes of 8th grade students from two teachers in one California school participated in this study (55% non-white, 40% free/reduced lunch, 13% ELL). 195 Students completed a pre/post-intervention measure individually, one day before and after the unit, respectively. Students completed the unit itself in teams (1-3 students) determined by their teachers during 50-minute science classes. Two different guidance conditions were randomly assigned to student teams several times throughout the unit (named *critique* and *annotator* conditions, see description below). Analysis was done on one pre/post-test revision item and two embedded benchmark items, as well as two embedded items asking students to generate questions about genetics, one at the beginning and one at the end of the curriculum. For this paper, I present pre/post learning gains, and a logistic regression analysis of a subset of this embedded item data (N=130 student teams).

Curriculum

The WISE Genetics and Simple Inheritance unit used in this study was designed to correspond to NGSS (MS-LS3) (NGSS, 2013). Students ask questions, use models, and construct arguments from evidence in accordance with NGSS scientific practices. The technological environment provided interactive visualizations of concepts that are difficult to explore in real life, such as phenotypic outcomes of various allele combinations.

The pre/post-test item analyzed here included a revision prompt (Figure 4.1), where students wrote an initial answer to the question, then immediately received a new piece of information and were prompted to revise their response.

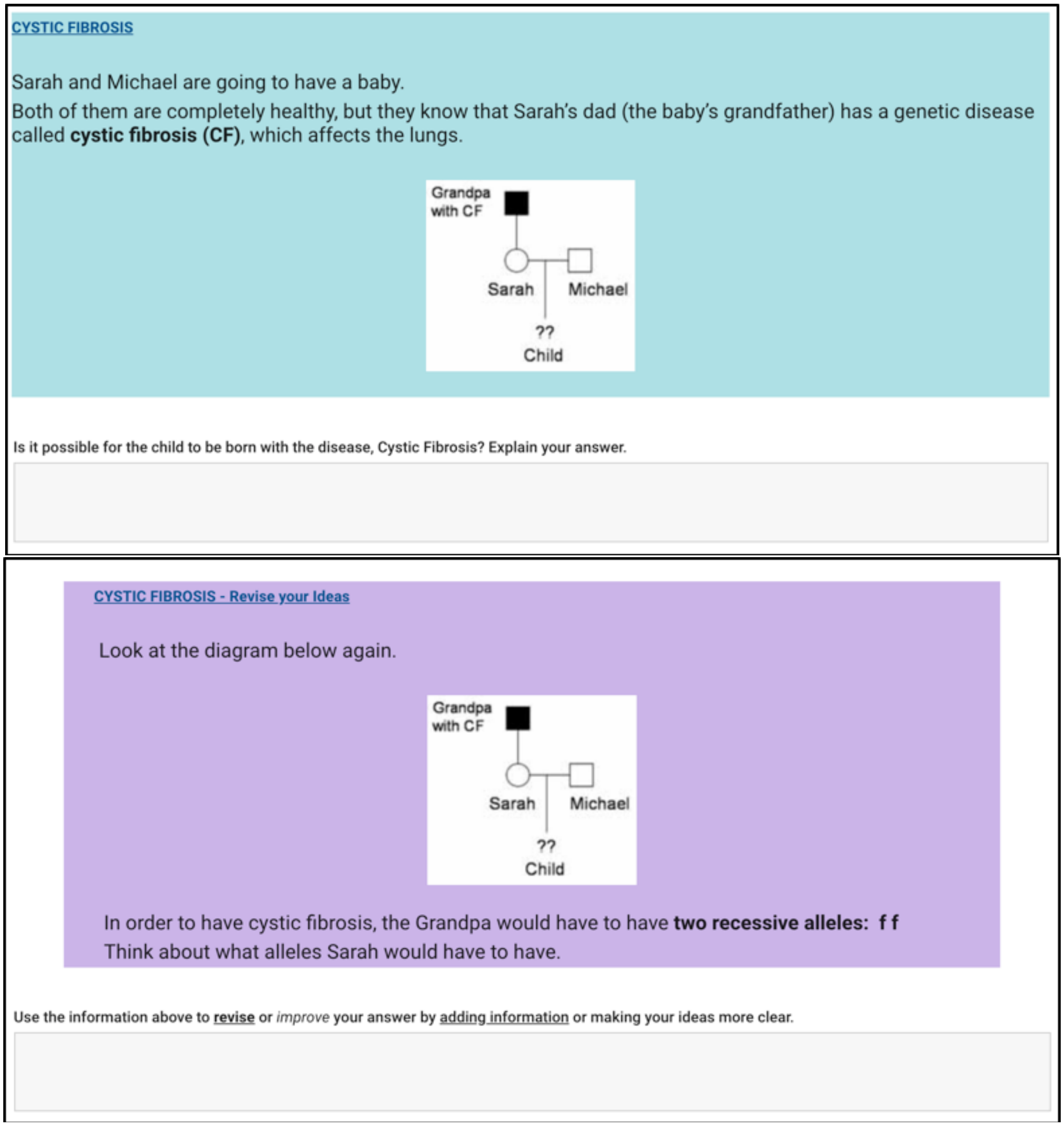


Figure 4.1: Pre/post revision assessment: initial prompt and revision prompt

Two guidance conditions were developed, based on our previous revision-guidance work, to aid students in distinguishing normative and more mechanistic ideas from vague understandings of genetics in order to revise their essays on two benchmark questions embedded in the unit. The *critique* condition prompted students to explain what is incorrect about several non-normative commonly held student ideas (Figure 4.3). The *annotator* condition presented students with a fictional students' written response to the benchmark question and asked them to place labels to tell the student where their explanation could be improved (Figure 4.2).

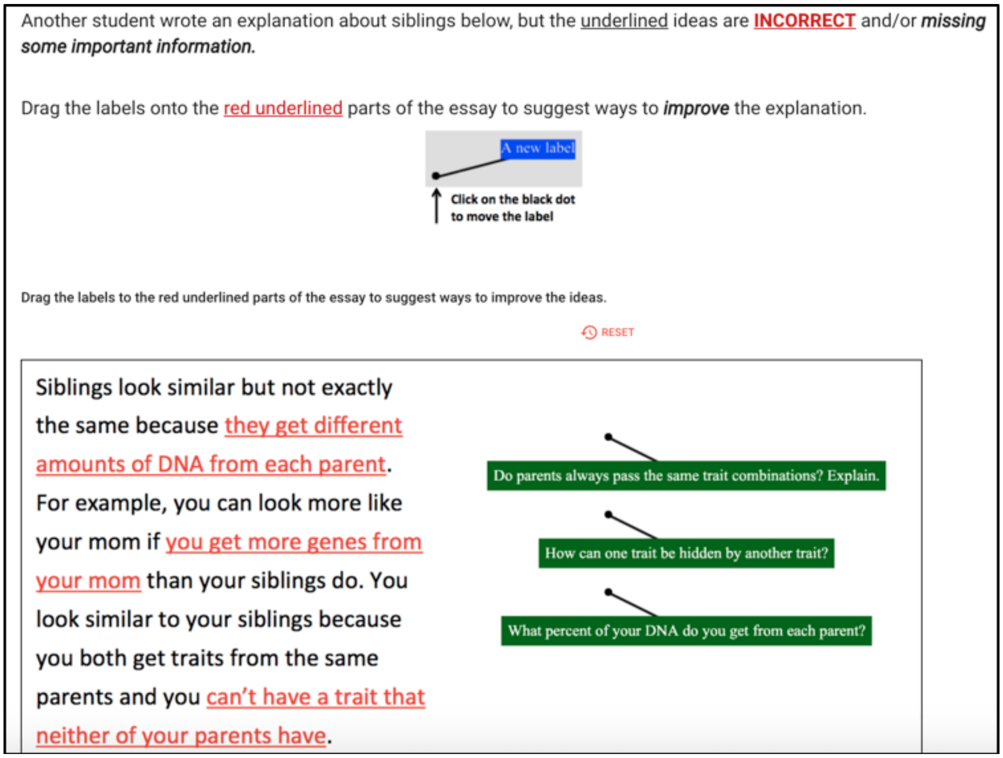


Figure 4.2: Example revision guidance activity for the *annotator* condition.

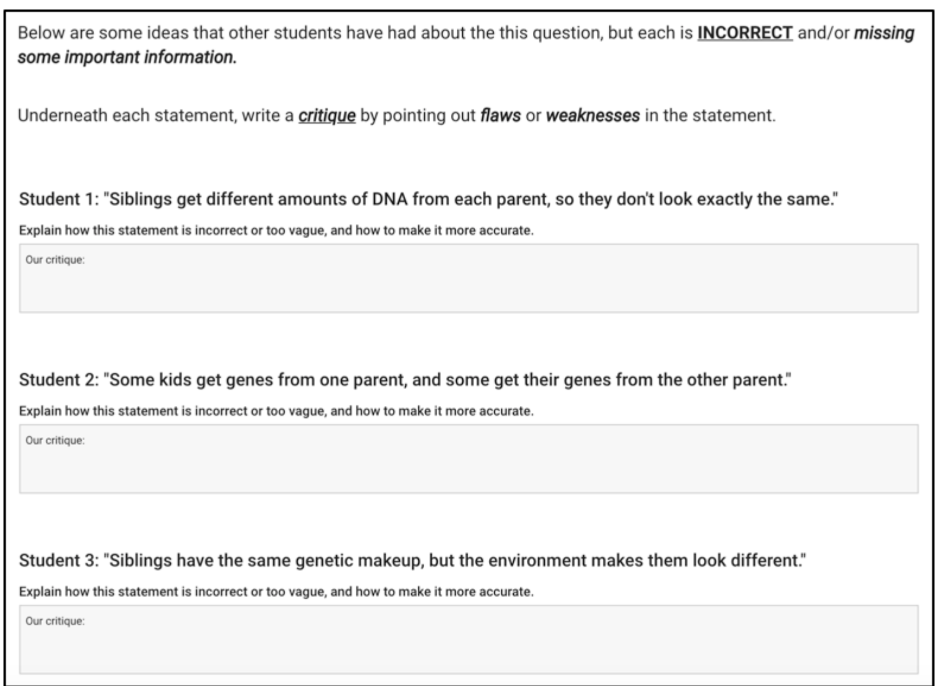


Figure 4.3: Example revision guidance activity for the *critique* condition.

For these items, student teams generated an initial essay response, then were randomly assigned one of two guidance activities (see table 4.1 for sequence), and were finally prompted to revise their initial essay response immediately after. The revision prompt automatically imported students' initial essays for students to revise.

Table 4.1: Outline of the sequence for *annotator* and *critique* conditions.

Steps on Embedded Assessment	***Annotator***	***Critique***
Essay Prompt	**Prompt**: Explain how you would use a Punnett square to figure out the probability of getting a certain genotype.	
Student teams assigned randomly by WISE ID	***Annotator***: Place pre-written labels on a fictional students' essay to show where ideas are missing.	***Critique***: Explain what is incorrect about several non-normative statements regarding the topic.
Revision Prompt	**Revise**: Now that you've learned a bit more about Punnett squares and probability, take some time to revise or improve your answers to this question.	

Student teams also responded to embedded prompts asking them to describe questions they had about genetics before beginning the unit, and again at the end of the unit.

Data Analysis

For the pre/post-test item, responses were scored using a 5-point Knowledge Integration (KI) scale (Table 4.2). KI scoring rewards students for successfully linking scientific ideas (Liu et al., 2008). This rubric was co-developed within our group, and revised until an inter-rater reliability of 0.9 or above was reached. KI scores were given for students' initial responses as well as their revised responses on both the pre and post-test, providing a measure of student learning from beginning to end of the intervention, as well as score increase due to revision on both the pre and post-test.

Table 4.2: KI Rubric Example: "Why do siblings from the same parents look similar but not exactly the same?"

KI Score	**Description**	**Examples**
1	No Answer	"I don't know"
2	Non-normative/irrelevant: Token mechanism only ("skips a generation") with no elaboration. Incorrect ideas: "you get different amounts of DNA from each parent"	Because you and your sibling have close genes but they are not the same genes. You inherit similar amounts of the same traits from the same parents at slightly different amounts. Because it's not exactly the same, you look a little different.

3	Partial link (one correct statement, but not connected to other scientific ideas, or student does not elaborate)	They get different parts of dna from their parents. You get a different set of genes then your sibling.
4	One full link between normative scientific ideas	Because you get half of your parents DNA but it does not specify which half you will inherit from them. This means that the half that you might get will not be the same that your sibling will get.
5	At least two full links	Siblings do not look exactly the same because they have slightly different alleles. Each child has a chance of receiving a different allele from its parents than its sibling because of probability.

For the purpose of this study, the embedded benchmark items were coded for whether student teams revised their explanations on both items, as opposed to revising only one or neither of these benchmark essays.

Student-generated questions from the beginning of the unit were assigned a binary code ("Answered") based on whether or not their question was answered by the material covered in the genetics unit. For example, the answer to the student-generated question "Is it possible to have a gene that neither of your parents have?" is covered in the curriculum. This was done to investigate the effect of students exploring their own questions during the intervention.

Questions generated by students at the end of the unit were also assigned to two categories. The first category, labeled "basic questions", included shallow questions that were factual or procedural in nature; this category also included responses stating they did not have questions, irrelevant questions, or questions that were previously answered by the curriculum (for example, repeating "Is it possible to have a gene that neither of your parents have?"). The second category, labeled "further questions", included those that addressed comprehension, prediction, and application of the knowledge gained from the curriculum. For example: "How did the first people alive get there traits from?"; "Are there specific genes that apply more change to a species?". (Chin & Brown, 2002; see Table 4.3 for rubric and examples).

Table 4.3: Student-generated questions rubric and examples for end-of-unit questions.

Question Classification	Criteria	Examples
Basic Question	No question Questions that address only factual or procedural aspects of genetics Questions that were answered fully in the recently completed unit	"We have no questions." "How can you tell if a trait is dominant or recessive?" "We still don't know exactly what a genetics is." "How do you use a punnett square to find genes?"
Further Question (based on Chin & Brown, 2002)	Questions that employ: ● Comprehension ● Prediction ● Anomaly detection ● Application ● Planning Questions that go beyond the scope of the material learned in the recently completed unit	"Can we take human DNA and combine it with animal DNA, and make a new species?" "How much DNA do we share with other species?" "Are there specific genes that apply more change to the species?" "How did the first people alive get there traits from?"

Results

Students across both the *critique* and *annotator* guidance conditions scored higher on the posttest compared to the pretest [t(194)=10.85; p<0.001], with a mean score of 2.19 on the pretest and 2.86 on the posttest. In addition, students across conditions also gained more from their revisions on the post-test assessment, gaining only 0.18 points from revision on the pretest and 0.52 points on the posttest, on average [t(194)=5.92; p<0.001]. This demonstrates that students improved not only in their content knowledge of genetics upon completion of this intervention, but also in their ability to incorporate new information into their scientific explanations through revision.

For our embedded assessments, logistic regression was run with a binary outcome variable for whether or not student teams produced higher-quality "further questions" at the end of the genetics unit. As shown in Table 4.4, the odds of student teams assigned to the *critique* guidance condition generating further questions at the end of the genetics unit were 1.9 times as great compared with student teams in the essay *annotator* guidance condition (approaching significance; p=0.08). This suggests that critiquing incorrect ideas about genetics was more effective in encouraging further questions than was labeling another students' essay with suggested revisions.

Table 4.4: Logistic regression analysis on students-generated questions at the end of the unit.

Further Questions	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**[95% Conf. Interval]**	
Condition	1.97	0.77	1.73	0.083	0.92	4.22
DidReviseOnBoth	2.91	1.16	2.69	0.007	1.34	6.35
Answered	0.79	0.33	-0.57	0.57	0.35	1.79
_cons	0.34	0.13	-2.78	0.006	0.16	0.73

Note: FurtherQuestions=dummy variable for whether or not students' end-of-unit questions built on their knowledge, 1=high-quality further question, 0=basic question (see rubric Table 3); Condition=dummy variable for randomly assigned guidance condition, 1=critique; 0=annotator; DidReviseOnBoth=dummy variable for whether student team revised on both embedded revision items, 1=revised on both, 0=revised on only one or neither item; Answered=dummy for whether students' initial questions at the beginning of the unit were addressed in the content of the unit, 1=question answered, 0=question not answered

Across both conditions, the odds of having further questions were 2.9 times as great for students that revised both of their benchmark essays embedded throughout the curriculum when compared with students that revised just one or neither of their embedded benchmark explanations ($p<0.01$; Table 4.4).

Interestingly, the odds of generating further questions for students that generated questions at the beginning of the unit that were subsequently answered in the curriculum ("answered") were only 0.79 times as great compared to teams whose initial questions were not directly addressed by the material, though this result was not significant at the 5% level ($p=0.33$).

Discussion and Implications

This study reveals several components of our web-based genetics inquiry curriculum that leads to generation of quality science questions and increases in knowledge integration. Our results show that the practice of critiquing common student misconceptions about genetics (the *critique* guidance condition) was helpful in promoting higher-quality student-generated "further" questions, as opposed to basic lower-level questions. This is consistent with literature describing the importance and benefits of critique in other related scientific practices, such as argumentation (Osborne, 2010; Berland & Reiser, 2009). Lower-level questions may lead students to tacking ideas onto their responses, whereas more abstract questions can lead to reformulation of a response and integration of ideas in a more connected way (Linn & Eylon, 2011).

Additionally, student teams that consistently revised their scientific explanations throughout the unit were more likely to generate genetics-related further questions at the end of the project. One possible explanation for this is that consistent revision helped these student teams get a better grasp on the material, and better enabled their questions to build off of this newly-gained knowledge. Wrestling more deliberately with the material through revision may have empowered students to generate questions related to prediction, anomaly detection, and application with regards to genetics (Rivard, 1994). Alternatively, students that are more prone to revise their ideas on a consistent basis may also already be the type of students to engage deeper with the material, and therefore generate higher quality further questions when prompted. Whether practicing revision influenced students' questions, or these were simply concurrent, this result is consistent with the notion that student-generated questions are related to student engagement in scientific inquiry (Biddulph et al., 1986; King, 1992; Chin & Brown, 2002).

The value of generating, as opposed to simply answering, good questions is essential to scientific discourse. Asking questions of a more sophisticated nature that build on knowledge can stimulate more productive discussion, encourage students to predict and hypothesize, and generate more in-depth explanations regarding complex scientific phenomena (Chin & Brown, 2002).

For teachers that wish to promote SDL and agency in their students, this study endorses the value of revision and critique as promising practices to better engage students in their own science learning. In addition to the self-report measures widely used to assess SDL, revision frequency and question quality can serve as quick and useful measures of self-directed learning, particularly as formative assessment for teachers conducting inquiry investigations in their classrooms.

Study 2: Scaffolding Self-Directed Learning Skills for Middle School Students through use of an online Data Tool

Students' ability to generate questions that go beyond the scope of what they learned is important, but "wonderment" questions can be difficult for students to investigate in a reasonable amount of time. Having many students develop and investigate their own questions at the same time has proven difficult in a typical middle school classroom. However, technology can help scaffold this process.

This study involves the Graphing Stories WISE curriculum, adapted to include a data tool that allows students to use data to ask questions about the variables and find an answer by producing a graph. This was designed to guide students through question development, narrowing down their question so they can investigate, and support students who may initially respond that they have no questions. Scaffolds were designed to help students successfully navigate the questioning process and the data tool. Activities were also included to help students critique both their own and others' questions. I look at students' question quality, and whether they were able to investigate their question successfully, as well as their critiques and revision behavior throughout the unit.

Introduction

Self-directed learning involves students participating more heavily in their learning by determining what to investigate. However, even upper level medical school students had issues with their self-efficacy in determining the learning issues to investigate and the resources to answer their questions (Dahlgren & Dahlgren, 2002; Loyens et al., 2008). It is also difficult to have 30 middle school students engaging in different investigations during the same class period without some scaffolding. An easy and useful way to constrain this is to provide students with a rich set of data on which to base their investigations. Students can develop questions about the data, choose the variables to investigate, and draw data-based conclusions. This provides students with a reliable resource for investigation, providing some constraints and scaffolding for asking questions.

Additionally, using data as a resource encourages students to reach evidence-based conclusions. Data literacy is a life skill that is increasingly important as interactions with data become more common and people make judgements from data more frequently (Haddadi et al., 2015). With ubiquitous access to information (and misinformation) through the internet, the ability to make data-informed decisions is more important than ever. Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process (Wolff et al., 2016; Schutt, 2013). It generally includes the ability to develop and answer questions using evidence from data, produce and interpret data representations to support conclusions, and use data to solve real problems and communicate their solutions (Vahey et al., 2006).

Datasets also serve as a useful constraint for middle school students investigating their own questions. In the following study, students generate questions based on a given data set. With scaffolding, students critique and revise their questions to make sure they are narrow enough to investigate using the provided data. Students make a prediction, then use the tool to generate a scatter plot, and interpret the graph to determine whether their hypothesis was correct. Students' questions, along with their revision behaviors, were analyzed.

Methods

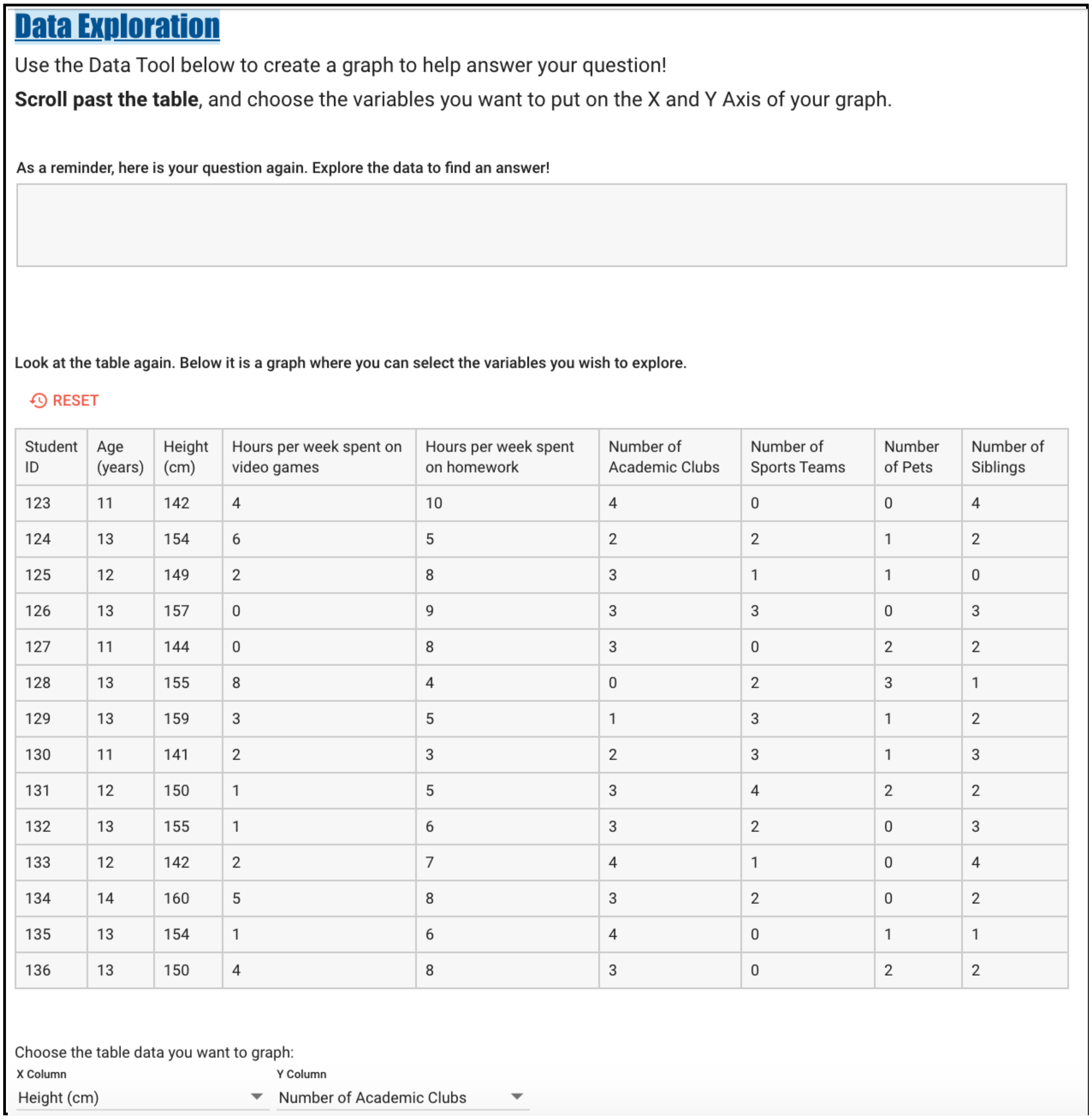
This study includes student data collected from a 5-day WISE online introductory graphing unit, Graphing Stories (https://wise.berkeley.edu/preview/unit/30375/node2). Six classes of 6th grade students from three teachers in one California school participated in this study (55% non-white, 40% free/reduced lunch, 13% ELL). 152 students completed a pre/posttest measure individually, one day before and after the unit, respectively. Students completed the unit itself in teams (1-3 students, determined by the teacher) during 50-minute math classes.

Analysis was done on pre/post question-generation items, and a series of items embedded in the unit. Specifically, I looked at the way students used a data tool that allowed them to ask, investigate, and answer their own questions using a data set in order to determine if the amount of scaffolding is enough, and if middle school students can practice and improve these self-directed learning skills.

Curriculum

This study was conducted in the context of the Web-based Inquiry Science Environment (WISE) *Graphing Stories* unit. This unit was designed for middle school students (grades 6-8) as an introduction to graphing. The original design of this unit included a variety of graphing tools to help students practice constructing and interpreting several types of graphs, including scatter plots, line graphs, and position-time graphs. This unit addresses several NGSS science and engineering practices (SEPs), specifically those related to using data as evidence (i.e. Analyzing and interpreting data, Engaging in argument from evidence; NGSS Lead States, 2013).

This iteration includes the major addition of a new data tool that allows students to use larger data sets to ask and investigate their own questions. This was created to address further NGSS SEPs, including *Asking question* and *planning and carrying out investigations* (NGSS Lead States, 2013). These practices are compatible with subskills associated with self-directed learning, which this tool was designed to foster (Loyens et al., 2008; Hmelo, 2004). The tool includes a table with any number of variables; students select which variables they wish to graph from a drop-down menu for each axis, and a scatter plot with trend line is automatically generated (Figure 4.4).



Data Exploration

Use the Data Tool below to create a graph to help answer your question!

Scroll past the table, and choose the variables you want to put on the X and Y Axis of your graph.

As a reminder, here is your question again. Explore the data to find an answer!

Look at the table again. Below it is a graph where you can select the variables you wish to explore.

RESET

Student ID	Age (years)	Height (cm)	Hours per week spent on video games	Hours per week spent on homework	Number of Academic Clubs	Number of Sports Teams	Number of Pets	Number of Siblings
123	11	142	4	10	4	0	0	4
124	13	154	6	5	2	2	1	2
125	12	149	2	8	3	1	1	0
126	13	157	0	9	3	3	0	3
127	11	144	0	8	3	0	2	2
128	13	155	8	4	0	2	3	1
129	13	159	3	5	1	3	1	2
130	11	141	2	3	2	3	1	3
131	12	150	1	5	3	4	2	2
132	13	155	1	6	3	2	0	3
133	12	142	2	7	4	1	0	4
134	14	160	5	8	3	2	0	2
135	13	154	1	6	4	0	1	1
136	13	150	4	8	3	0	2	2

Choose the table data you want to graph:

X Column: Height (cm)

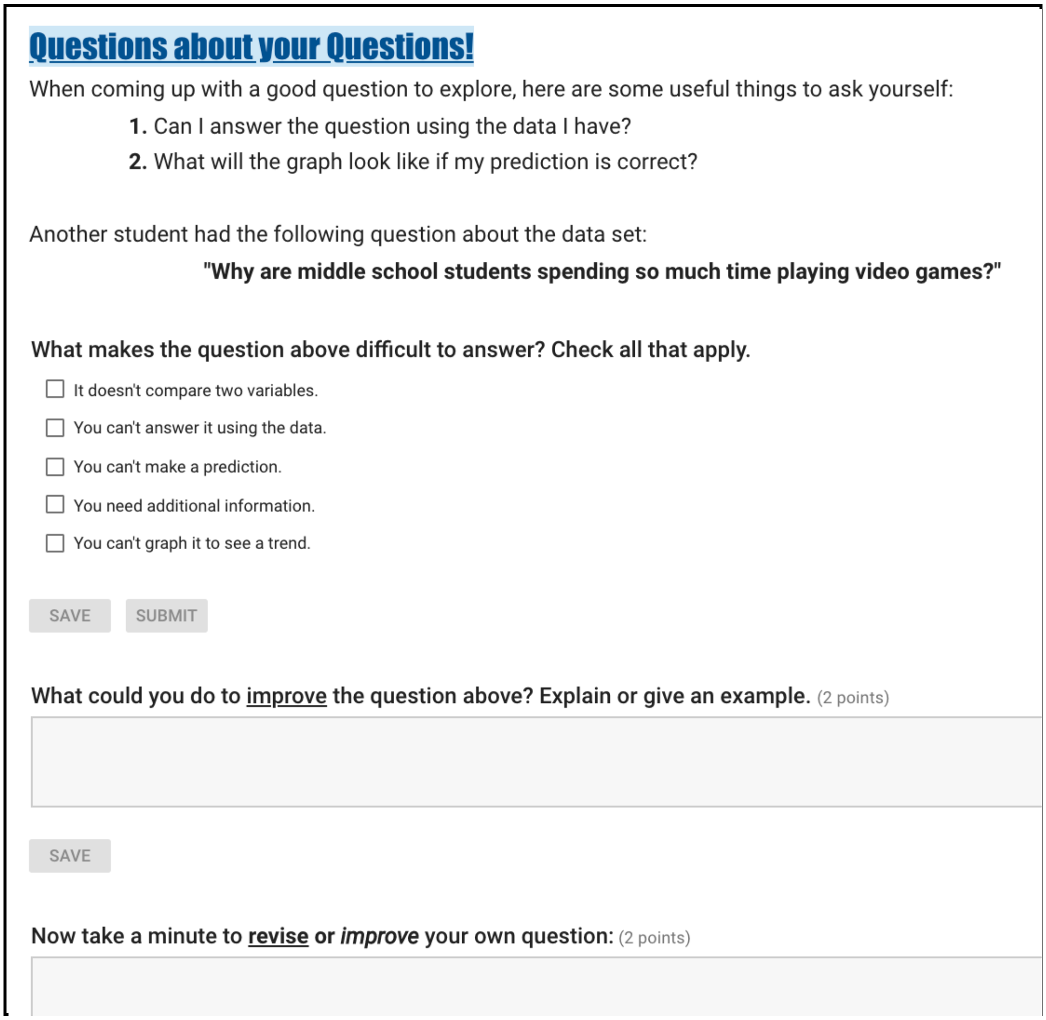
Y Column: Number of Academic Clubs

Figure 4.4: Data tool with a table of variables from which students can choose to graph, and the auto-generated scatter plot with these variables.

This tool was designed to encourage students to think of graphs as a tool for understanding phenomena in the world, rather than only as a final product. It makes clear the affordances of graphs, consolidating large amounts of data into readable patterns off of which students draw conclusions. It also allows students to look at various relationships relatively quickly and make inferences from data without the difficult mechanics of graphing the data by hand, which is similar to how professionals use graphs. Generating the graphs automatically allows students to focus on data interpretation. All of these attributes allow students to ask their own questions about the data and what it says, and then investigate those questions. This is a huge aspect of self-directed inquiry that is often difficult for teachers to implement or achieve due to the chaos of having every student or student team investigate something different. Here this idea is simplified by providing the data as the main resource, but with many options to explore various relationships.

The tool contains options to allow students to choose the type of graph they produce, but in this iteration, the tool was limited to produce scatter plots due to the nature of the data provided as well as the students' previous limited exposure to different types of graphs. For similar reasons, the data used deals with middle school student activities to reduce the cognitive load of students understanding new content and/or the influence of prior knowledge (Shah & Shellhammer, 1999; Freedman & Smith, 1996). The specific variables chosen were things middle school students would relate to in order to increase engagement, and included positively and negatively correlated components, as well as some variables with no clear relationship (for example, height and number of hours playing video games per week). This allows for students to encounter a variety of patterns in data. Students are also expected to have ready predictions about relationships between these variables due to experience.

Students develop their own question with the help of several rounds of scaffolding, and several opportunities to revise their question. Students explore the dataset (Figure 4.4 above) with their partner, and formulate a question based on the variables given. They are then prompted to critique their question, by asking "questions about their question" (Figure 4.5).



Questions about your Questions!

When coming up with a good question to explore, here are some useful things to ask yourself:

1. Can I answer the question using the data I have?
2. What will the graph look like if my prediction is correct?

Another student had the following question about the data set:

"Why are middle school students spending so much time playing video games?"

What makes the question above difficult to answer? Check all that apply.

- ☐ It doesn't compare two variables.
- ☐ You can't answer it using the data.
- ☐ You can't make a prediction.
- ☐ You need additional information.
- ☐ You can't graph it to see a trend.

SAVE SUBMIT

What could you do to improve the question above? Explain or give an example. (2 points)

SAVE

Now take a minute to revise or *improve* your own question: (2 points)

Figure 4.5: Students ask questions about their own question in order to determine if they can investigate and answer it using the data given.

After revising their question, and before investigating it with the data, students are asked to make a prediction. Students are also asked to draw a prediction graph: what would the graph look like if your prediction were correct? Then, students select variables and generate a graph to address their question (see Figure 4.6 for an example). Finally, students answer their question based on the data, and discuss whether their prediction was correct.

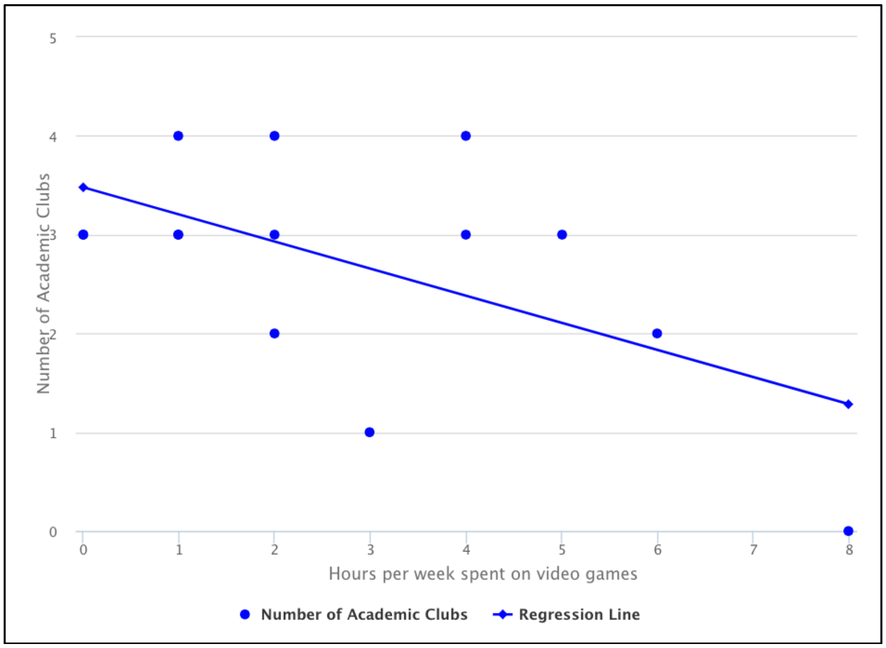


Figure 4.6: Example graph produced by the data tool when variables are selected, in this case: "Hours per week spent on video games" on the x-axis, and "Number of Academic clubs" on the y-axis.

After generating and interpreting a graph to answer their question, students are asked to critique a fictional students' interpretation of a graph produced to address their question (Figure 4.7).

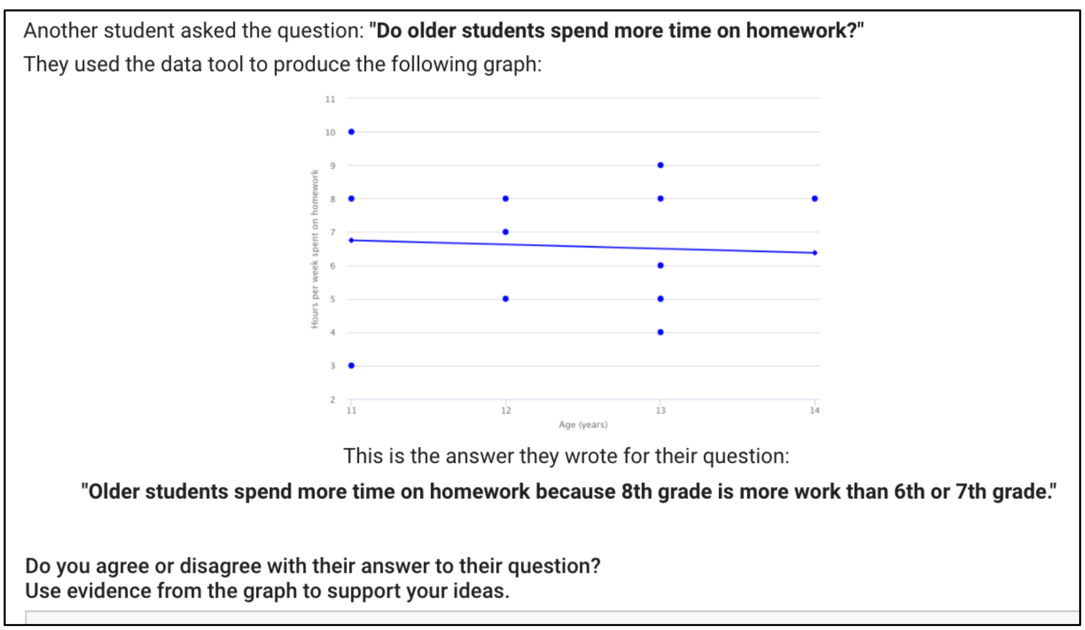


Figure 4.7: Open-response critique item of a fictional students' interpretation of the data tool output.

The purpose of this item is to see if students can critique an interpretation that is based on personal experience rather than based on the data in the graph.

Results

Pre/post

Students tended to have better data-related questions on the posttest; the average KI score for questions on the posttest was 3.3, versus 2.9 on the pretest. Students also tended to have more further questions on the posttest; 48% of students had further questions on the posttest, compared to 34% on the pretest [$z(151)=2.37$; $p<0.05$]. This shows that the curriculum helped students develop their question-generation abilities, both in terms of "wonderment" further questions, as well as questions that can be investigated using data visualizations.

Embedded Analysis

Data tool steps were analyzed to see how student teams moved through the tool, and which sections were more helpful with promoting revision and question generation.

Data Tool Questions and Revision

Students' questions about the data were scored using a KI rubric similar to the rubric used to score the pre/post data questions, based on whether they compared two variables that they could investigate using the data given. Student teams did very well at generating questions based on the dataset; average score on their initial question was 3.56 out of 4. Figure 4.8 has a breakdown of students' initial question scores. A score of 1 on the questions KI rubric indicates students that wrote "I don't know" or wrote something irrelevant; no student teams scored a 1 on their question, which indicates that all students understood the task and attempted to write questions about the data. The majority (70%) of student teams attained the maximum score on this item on their initial attempt.

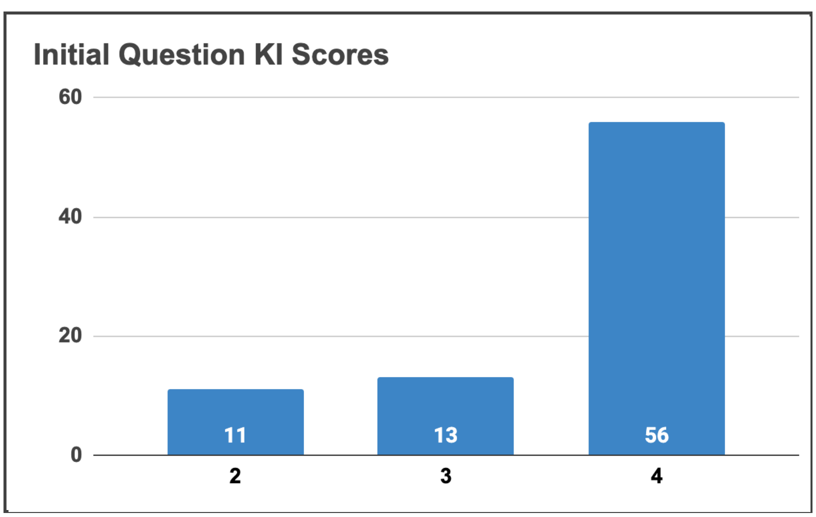


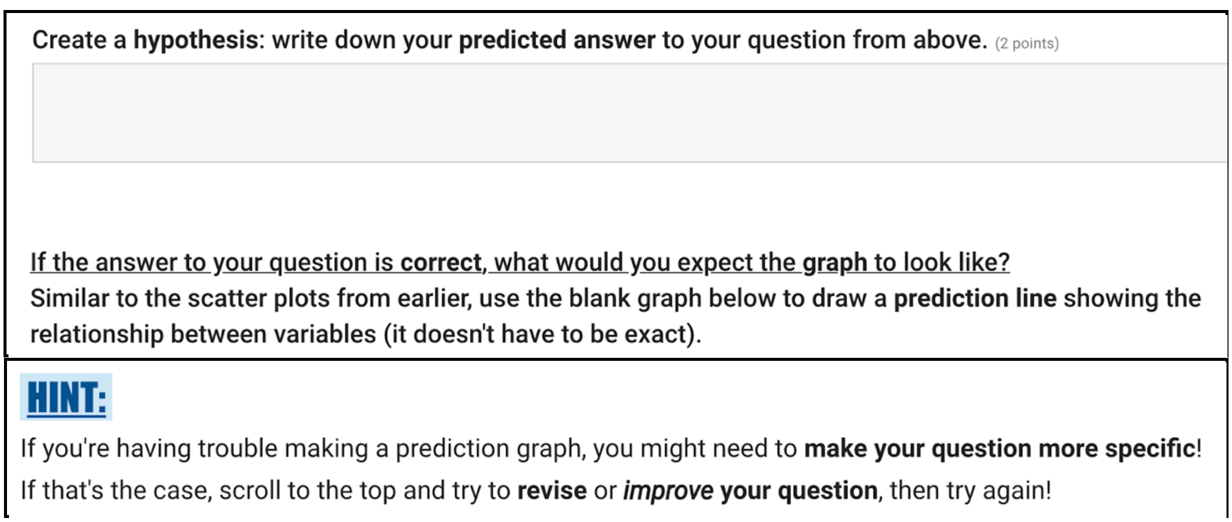
Figure 4.8: Number of students with each KI score on their initial questions about the dataset.

This shows us that students can readily develop questions about data to steer their investigation using our data tool. 52% of student teams (N=41) revised their question during the first revision opportunity, after the questions critique item; 75% (N= 18) of the students at a 2 and 3 level revised their questions during this first revision opportunity (Table 4.5), with all but 4 students moving up past a level 2.

Table 4.5: First revision of data tool question, by KI question score.

Question 1 KI Score	Revised	Did Not Revise
2	8	3
3	10	3
4	23	33
Total	**41**	**39**

At the second revision time point, 30% of student teams revised their question again after attempting to make a prediction (Figure 4.9) and before investigating their question with the data tool. Table 4.6 shows the breakdown of student teams' revisions on the first and then the second question revision timepoints.


Create a **hypothesis**: write down your **predicted answer** to your question from above. (2 points)

If the answer to your question is **correct**, what would you expect the **graph** to look like?

Similar to the scatter plots from earlier, use the blank graph below to draw a **prediction line** showing the relationship between variables (it doesn't have to be exact).

HINT:

If you're having trouble making a prediction graph, you might need to **make your question more specific**!

If that's the case, scroll to the top and try to **revise** or ***improve* your question**, then try again!

Figure 4.9: Prompt for students to make a prediction based on their question, then consider revising again if necessary.

Table 4.6: Question revisions after question critique, then after prediction.

	Question 2		
Question 1	**Did Not Revise**	**Did Revise**	**Total**
Did Not Revise	32	7	39
Did Revise	25	17	42
Total	57	24	81

The largest population here is students that did not revise during the first opportunity then continued to choose not to revise during the second opportunity. This is consistent with previous findings that most revision behavior is consistent, and we only help a small percentage of students revise that wouldn't have already done so. However, on this item, the majority of students' question scores were high. Table 4.7 has the breakdown of student question scores at each time point.

Table 4.7: KI question scores at each revision time point.

	Question KI Score			
	1	**2**	**3**	**4**
Initial Question	0	11	13	56
Revision 1	0	5	16	60
Revision 2	0	5	14	62

By the last revision opportunity, 77% of student teams (N=62) had a level 4 question, comparing 2 variables that could be investigated using the data table provided in the data tool. 82% of student teams (N=61) chose the correct variables in the data tool to produce a graph and investigate their question. 71% of student teams (N=52) interpreted the graph correctly, answering their question.

Students' critique item responses were scored using a KI rubric; again, no students scored a 1 meaning all groups attempted a relevant critique of the fictional students' interpretation of the graph. Table 4.10 shows that 29% (N=21) of students scored a 2, which generally included agreeing with the fictional students' interpretation based on personal experience rather than on the data. 23% of students (N=17) scored a 3, meaning they focused on a single point rather than on a trend as whole. The majority of students scored a 4 (40%, N=29) by disagreeing with the statement, and citing that the line on the graph does not support that conclusion. Only 6 student teams scored a 5 by asserting that the graph shows no relationship between the variables presented, represented by a large scatter of points and a trend line that is mostly horizontal.

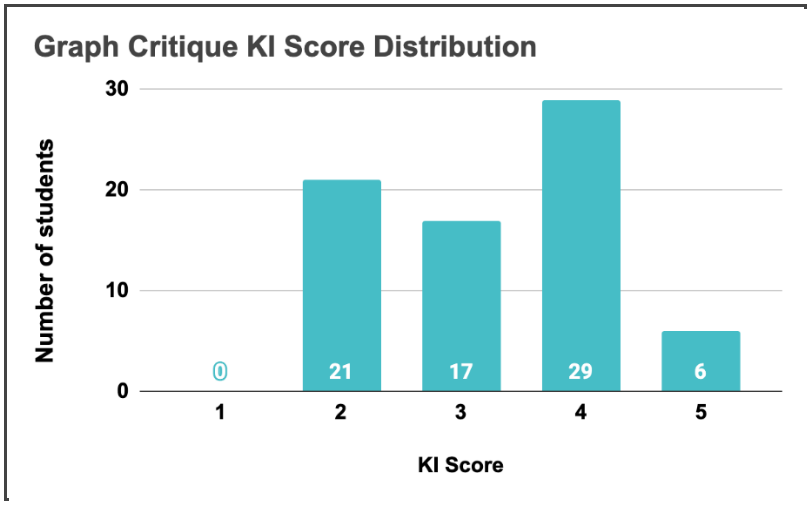


Figure 4.10: Distribution of students' KI scores on the graph critique item.

These critique scores show that approximately half the students scored either a 4 or a 5, meaning they successfully interpreted the graph and critiqued the statement. The other half of the students need more support, either in graph interpretation or specific support for critique.

While most students were successful at developing and investigating a question about the data, examples were chosen to see what less successful students struggled with during the investigation process.

Student Examples

Example 1: A student team's Data Tool Use
Initial question:

> "How tall is some one who is 12 years old?"

This question received a KI score of 3, because while it includes the variables of height and age, it focuses on a specific point or set of points rather than asking about a trend between the two variables.
On the critique item, they suggest an improvement to the given question:

> "A better question would be: how much time does an average middle school student spend on video games a day? this answer is better because you can graph it. Whereas the other question you can't graph.

Here, they cite being able to graph a prediction as criteria for a good question, but their suggested question would be impossible to graph with the given data (i.e. it only provides hours of video games per week). This suggests they are developing their criteria for a good question, but did not look closely enough at the data table while formulating their own question.

On the next step, this team revises their question:

“How tall is the average 12 years old?”

This question is an improvement in that it involves more than one point, but is still not asking about a relationship between two variables. On the KI scale, this is still at a level 3.
Next, the prediction step asks them to create a hypothesis, as well as draw a prediction graph. Students with questions such as this one are encouraged to revise their question again if they are having trouble drawing a prediction.

This student team revises their question again:

“as students get older do they get taller?”

This revision moves them up to a KI score of 4, looking for a trend in the relationship between 2 variables, height and age.

This team then draws their prediction graph (Figure 4.11):

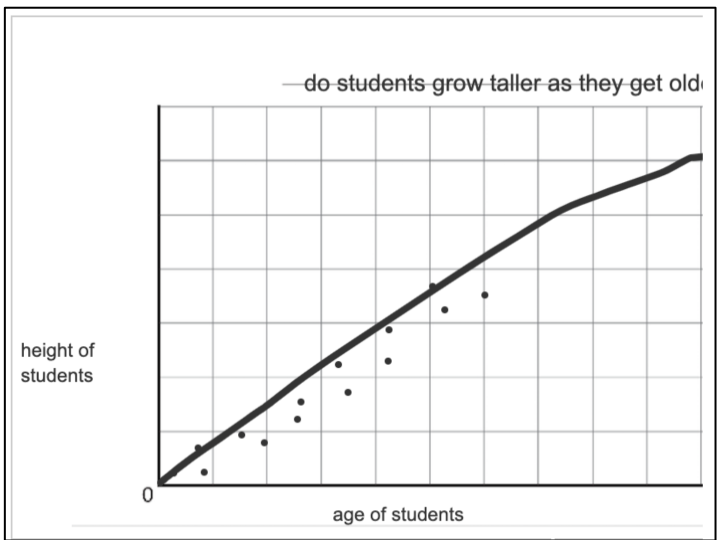


Figure 4.11: Prediction graph for the question: “Do students grow taller as they get older?”

They successfully select their variables in the data tool and produce a graph comparing these 2 variables (Figure 4.12).

They interpret their graph correctly, saying:

“it closely matches our prediction. there are some areas where our prediction wasn't correct but majority was close.”

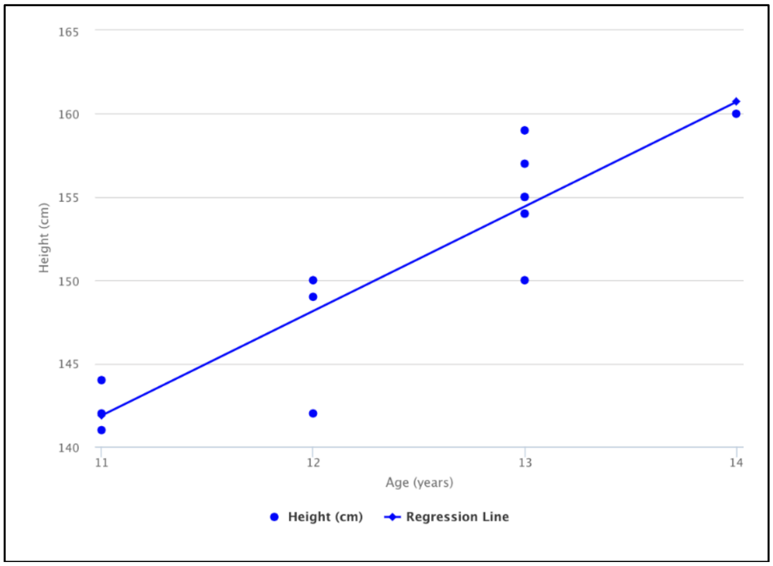


Figure 4.12: Graph generated by this pair of students using the data tool to investigate their question.

After completing their investigation using the data tool, students complete a critique question about another students' interpretation of a graph (Figure 4.13):

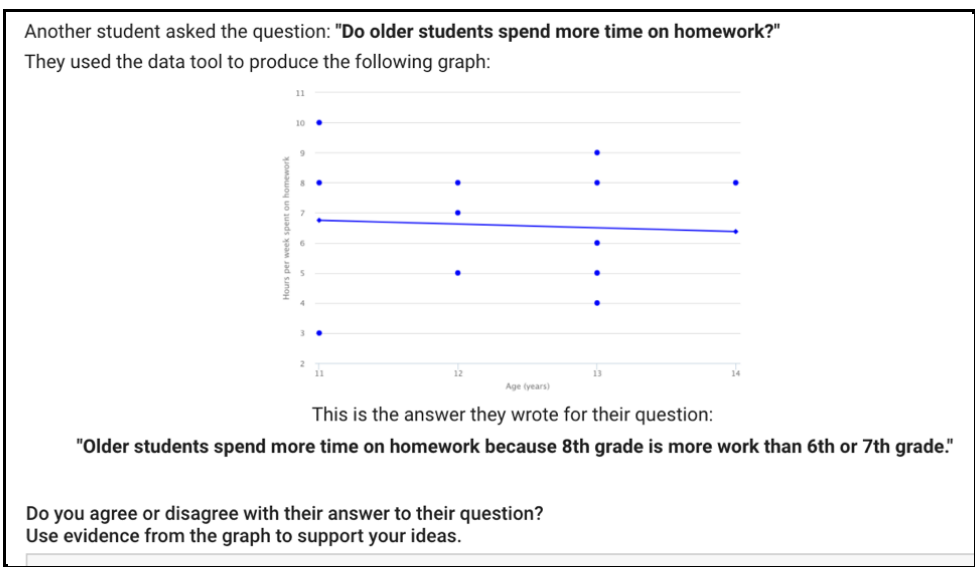


Figure 4.13: Critique item about graph interpretation after students complete their own investigation with the data tool.

This student team responds to this question:

> "I disagree because it states that eight graders have less homework than sixth graders and that is not true."

They interpret the graph correctly, but continue to rely on their personal experience. This suggests that these students need more support in interpreting graphs, especially when the data conflicts with their prior knowledge or predictions.

This example shows how the scaffolding worked to help students improve their questions before investigating them in the data tool. This student team began developing their question criteria on the critique step, but did not fully flesh out their question until they attempted to make a prediction and realized they needed to expand their question to include all of the second variable (age).

Example Critique Struggles

Student critique responses (Table 4.8) were examined to determine what students had difficulty with to determine how to support them in future curricular developments.

Table 4.8: Students' open responses critiquing a students' misinterpretation of a data tool graph

Student Pair Critique Response	**Description/Observation**
"yes we agree because as you get older you get into harder and advance work than when you were in elementary school"	Students respond to the item based on their personal experience rather than referencing the graph
"I agree because the graph shows that the 8th graders have more points showing"	Students find points on the graph that support the fictional student's interpretation rather than looking at the overall trend
"I disagree because older kids are also smarter and they can spend less time on homework. On the graph, 14 year olds spend a little more than 11,12,and 13 year olds. As the age increases, the hours spent on homework decreased."	Students disagree with the incorrect interpretation, but instead of citing the data they rationalize based on personal experience and look at specific points on the graph rather than the trend.
"We disagree, because the person that spends the most time on homework is 11, which is the youngest age on the graph. It also seems like it depends how responsible the child is, or how much homework they get."	Students focus on extreme points (highest or lowest) to interpret a graph with a lot of noise; speculated about the reasons the data might not match the interpretation.

These responses reveal various ways students are interpreting a noisy scatter plot with no clear trend; this is not surprising, since noise in scatter plots was not explicitly covered in the curriculum. Future guidance can be designed to help students focus less on individual points and more on overall trends being represented in the graph, as well as a lack of relationship between variables. Additionally, explicit critique can be encouraged for drawing conclusions based on personal experience that are not supported by the data. Many students reasoned about why the data did or did not support their conclusions or personal experience, including factors not present in the graph such as how quickly students of different ages can complete homework; future items can be developed to help students use these speculations to develop further questions for investigation about mechanism.

Discussion and Implications

Overall, the majority of students in this study were able to develop adequate questions about the given data set, even before encountering scaffolds or guidance. These 6th grade students had relatively limited experience with graphs previous to completing this curriculum. Their success with the data tool is likely due to the fact that the context was simple and readily familiar, leaving them able to develop questions for which they had predictions ready. Additionally, the scaffolding allowed students to think about the variables first and make predictions before encountering a data visualization of those variables, which may have been a major factor in their success as well. The majority of students also improved in their ability to both ask questions of data and ask questions that go beyond the scope of information provided by the posttest, even on other topics less directly relevant to them.

Similar to study 1, as well as previous studies in chapters 2 and 3, only about half the students revised their question as a result of the first prompt. However, in this case, there were many students that had constructed a useful question on their first attempt, so many students had no need to revise. These students may have successfully completed the question critique item, and used those criteria to establish that their question already met the requirements. This is supported by the fact that, of the students with a score of 2 or 3 on their initial question, 75% of them revised their question after completing the critique item. This is much higher than previously observed, possibly due to the easily relatable content of the dataset (Shah & Shellhammer, 1999). Students had predictions ready, for instance, about how playing many hours of video games may affect time for other things, such as homework or academic clubs, and vice versa. This may have also partially impeded their interpretations of their final graphs, however, since students familiar with a subject may interpret the graph according to their expectations due to prior knowledge, rather than what the data is actually showing (Shah & Carpenter, 1995; Freedman & Smith, 1996).

While students were generally successful at using the data tool to develop and answer their questions, there were a few areas where some students still had difficulty. For example, students had difficulty when they encountered scatterplots where the variables had no relationship, sometimes interpreting the noise as a relationship that supported their prediction. This is consistent with other studies conducted in our group, on topics such as global climate change (McBride, 2018; Vitale et al., 2016). More support is needed for students to interpret a lack of relationship, or a graph with too much noise to isolate a relationship between variables.

Since these 6th grade students with relatively limited graphing background were successfully able to navigate through this data tool, future studies can feature data involving more conceptually difficult scientific or real-world content. Additionally, the data tool options can be expanded to include different types of variables (e.g. categorical variables) and allow students to choose the type of graphical representation they want to display the data (e.g. bar graph, line graph, or scatter plot), with varying levels of scaffolding for each of these aspects based on students' prior graph experience.

While middle school students, compared to undergraduates, need more scaffolding with generating questions about data, as well as other self-directed skills (Hmelo, 2004), this study suggests that we can likely engage students in more challenging contexts with this data tool. The majority of students generated data-related questions on their first try, and those that did not were supported by the questions critique item, and 2 rounds of revision. These scaffolds might be more necessary if students are presented with more conceptually challenging relationships between variables. Future curricular revisions can include the version of the data set and scaffolds presented here to help students become familiar with the questioning process and the tool, and follow with another more challenging set of data.

Conclusion

Developing self-directed learning skills is important in a world where information, and misinformation, is ubiquitous on the internet. Developing and investigating your own questions, and reasoning towards a conclusion with evidence, is important for becoming a lifelong learner both in and out of school environments, and is a major goal of national science standards (NGSS Lead States, 2013). Self-evaluation and reflection of your scientific understanding are crucial processes to integrate new information with prior understanding. Students that exhibit these skills tend to also have higher academic achievement (Loyens et al., 2018), but these skills are an end goal in themselves (Candy, 1991; Hmelo, 2004). However, development of these skills is difficult and involves a learning curve, even with upper-level medical school students. Therefore, developing and implementing scaffolding that can foster these skills is necessary, especially for younger students.

In this set of studies, I focus on several SDL components determined by our group through interviews with middle school teachers, including having students ask their own questions, investigate these questions, and reflect on their learning. In study 1, students develop further questions about the complex topic of genetics after learning about simple inheritance. Students that revised their ideas consistently throughout the unit had more sophisticated questions by the end of the unit. This may be because students that engage more with the material are more capable of applying that knowledge to ask further questions. Future research can be done to determine the mechanism of the relationship between revision and questioning; specifically, if you increase students' revision frequency, do you also help them develop more sophisticated questions?

In the second study, students used data to investigate their own questions with our newly developed online data tool. While graph construction and interpretation of data visualizations are difficult for middle school students, these students showed unexpected aptitude for developing questions about a given data set. This may be due to the highly relevant context of the variables, or the scaffolding possible in this technological environment (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Bannert et al., 2015; Shah & Carpenter, 1995). While this study had students participate in questioning, critique, and revision throughout the use of the data tool, the relationship between these practices still requires some clarification. Next steps include breaking down if and how these practices influence each other, or if they are simply concurrent in students who are already self-directed. This study was limited in that it was the first test of our online data tool; a more longitudinal study would reveal how students fare with larger datasets related to more complex scientific topics, as well as the trajectory of middle school students' SDL learning curve.

Overall, this set of studies supports the notion that we can support students' ability to generate meaningful questions to both check their understanding and pursue more personally meaningful learning goals (Biddulph et al., 1986). Self-assessment is a crucial component of self-directed learning (Candy, 1991; Blumberg, 2000; Hmelo, 2004); practices such as critique can help students begin to recognize gaps in their understanding and promote more frequent revision of their scientific ideas. Questioning is an essential science practice (Barton & Tan, 2018) and questioning at a higher level has been shown to promote enhanced understanding of scientific content (Chin & Brown, 2002). The studies presented here show that engaging in revision of scientific ideas, revision of questions, and exploration of data can help students generate more sophisticated questions by the posttest. Future studies can combine these measures with other SDL measures of interest, including self-efficacy, metacognitive monitoring, and academic achievement.

Chapter 5: Conclusion

This dissertation research investigates how to encourage students to revise their scientific explanations and graphical representations, particularly through design of critique activities to help students to critically re-examine their prior knowledge. Critique is an important part of scientific thinking, and revision emphasizes the iterative process of science and of learning. While the critique guidance encouraged some students to revise, many students continued to choose not to revise at all. This revealed very clearly the self-directed nature of revision. Students with self-directed tendencies revised more often regardless of the guidance presented to them, and these students often generated more sophisticated questions, likely due to the development of greater understanding of the content material through their engagement with the curriculum. This is not entirely surprising, but helps reveal the role of revision in self-directed learning more explicitly.

Summary

In this work, I explore various ways to help students critique and distinguish science ideas, revise their ideas at a deeper level, generate scientific questions, and do these things iteratively in a way that promotes self-directed learning habits. Writing, as well as science, are often taught as linear processes (Flower & Hayes, 1981; Campanile et al., 2015), though in reality they are both very much iterative. Therefore, this dissertation set out to design curricula that emphasizes this iterative process by providing ample opportunity for revision, critique, and reflection, and designing guidance to scaffold these processes for students.

Research Questions

This dissertation investigated the following research questions:

1) How can critique guidance encourage students to distinguish science ideas as reflected in revisions of their written explanations?

2) With insights regarding critique guidance gained from chapter 2:
   a) Can critique guidance help students improve their qualitative position-time graphs?
   b) What rationales do students have for their graph revisions?
   c) How do these compare to, or differ from, critique and revision of written science explanations?

3) What is the relationship between consistent revision of science ideas and student-generated questions?
   a) How can critique guidance support students to investigate their own questions in self-directed explorations of data?

Research Question 1:

Research question one is addressed by the set of studies in chapter 2, which investigated the effects of three different forms of guidance on students' revisions of their science explanations. Overall, this research shows that writing and revising scientific ideas can help students more effectively integrate new information with their prior knowledge (Linn et al., 2014). Revising is important for learning even if students do not immediately improve their explanation with their revision; revising consistently while learning new material helped students improve their understanding by the end of the curriculum. Findings from chapter two show that students who revised consistently throughout the genetics unit performed better on various types of outcome measures by the end of the unit. Students who revise grapple with the material in new ways, resulting in better scientific understanding. Students also improved specifically in their revising abilities, gaining more from their revisions about genetics on the posttest.

However, like previous research on revision, we found that many students chose not to revise when prompted, or made surface-level changes rather than evaluating their work and making substantial revisions to content (Rivard, 1994; Crawford et al., 2008; Bridwell, 1980; Flower et al., 1986). Chapter two findings show that the large majority of students did not change their KI score with their revisions due to not fundamentally changing the content of their explanations. Content-level revisions are even more important when writing scientific explanations and arguments since students are learning about and developing understanding of underlying mechanisms of real-world phenomena. Surface-level revisions of scientific explanations and ideas do not show real change in student thinking about how complex scientific phenomena occur, resulting in coexistence of conflicting ideas. Therefore, guidance is needed that helps students revise more often, add more new content to their explanations, and revise in a way that connects new information to their prior knowledge.

Specifically regarding the effect of critique, the combined analysis of all three studies in chapter two found that critique guidance encouraged students to revise the most often, and students with high prior knowledge were able to use this guidance to improve their explanations. Critique encouraged students with low prior knowledge (LPK) to revise more frequently (compared to LPK students sent back to revisit material), but they often struggled with the critique and sometimes took incorrect ideas from the activity. Students with low prior knowledge often benefited more from the annotator tool, which helped them make more connected revisions. This guidance may have helped them sort through the ideas they already had more effectively. This suggests that students may benefit from explicit instruction on critique to use this guidance effectively, or that critique may be more helpful later in the unit after students gain more content knowledge. Experts are capable of recognizing gaps in their understanding, which motivates revision (Flower et al., 1986); students with low prior knowledge are novices at both genetics and revision, and these students continued to struggle to recognize these gaps or inconsistencies in their science ideas. Practicing critique can help students begin to recognize the gaps in their scientific understanding (Henderson et al., 2015; Jimenez-Aleixandre & Duschl, 1999; Zohar & Nemet, 2001), however, our findings are consistent with the fact that critique (and revision) depend heavily on prior knowledge (Donnelly et al., 2015; Flower et al., 1986). Therefore, while critique is an essential scientific practice, it may be more practical to build up students' content knowledge first.

Flower et al. (1986) says for students to be successful with revision, we cannot simply prescribe expert actions to novices with limited content knowledge. This is consistent with the fact that encouraging students to revise more often did not always result in revising *better* by integrating new ideas on each item, and sometimes resulted in disconnected revisions. However, students with low prior knowledge benefitted from the essay annotator activity in this sense, which encouraged them to revise in a more connected way as compared to the critique guidance. Bringing in new science concepts in a way that is connected with prior knowledge is essential for knowledge integration (Linn & Eylon, 2011).

Redirecting students back to relevant material to help them revise their explanations encouraged only a third of students to actually revise. However, the students who did revise generally improved their science explanations. This type of guidance, therefore, proved useful for students that already had self-directed tendencies, but did not encourage students to revisit and revise if they were not predisposed to do this.

In general, critique encouraged students to revise more frequently, often by adding new ideas, and students that used this guidance effectively were able to use this to improve their revisions. Modeling how integrated revision looks with the annotator tool helped students make more connected revisions, especially students with low prior knowledge. Potentially, by isolating which aspect of the revision process a student is struggling with, specific guidance sequencing can be designed to help students overcome their individual difficulties. Future research can examine if combining critique and essay annotation might be the most beneficial for distinguishing ideas, then integrating them into explanations in a connected way.

From this collection of studies, we found that critique guidance was somewhat successful in helping students distinguish complex scientific ideas, and encouraged students to revise. Students that revised their ideas more frequently learned more from the curriculum. However, many students still chose not to revise. For these reasons, the next set of studies investigated the value of critique guidance in the context of graphing, and probed students for their revision rationales to get more insight about when and how students choose to revise.

Research Question 2:

The set of studies in chapter 3 addressed the second set of research questions by investigating several sequences of guidance designed to help students recognize inconsistencies in the construction of their position-time graphs, and successfully revise these graphs. Graphing skills are necessary in science for elucidation of patterns and underlying processes of complex phenomena (Friel et al., 2001; Wu & Krajcik, 2006). This involves learning to think with graphs, which includes constructing and critiquing graphs as well as interpreting them (Lai et al., 2016). However, students have very little opportunity to construct, let alone to *revise,* their own graphs in science class. Chapter two explored the revision of students' written explanations, and revealed that engaging consistently with their ideas through revision was helpful for student learning. The work presented in chapter three reveals that students' graph constructions can potentially benefit from the same processes, and the added complexity and abstract nature of graphs makes revision even more essential for students to have opportunities to distinguish their ideas, reflect on their thinking, and improve their graphical representations (Lai et al., 2016; Flower et al., 1986). This includes opportunities for students to recognize and diagnose gaps in their understanding through critique.

This work had students construct their own position-time graphs representing a given scenario, and then revise their graphs after various types of critique guidance. Having students construct qualitative graphs, rather than simply plotting points, allowed students to demonstrate complex combinations of normative and non-normative ideas (Hattikudur et al., 2012; Stylianou et al., 2005; Vitale et al., 2015; DiSessa et al., 1991). After constructing their initial qualitative position-time graph, students critiqued graphs with common graph-as-picture errors, compared their graph to a picture to discuss differences, and answered several open-response items about the meaning of abstract graph features. Students that revised their graphs often improved aspects of them, but even all of the guidance provided did not help most students revise to a correct position-time graph. This is likely due to the fact that this particular item is difficult, even for adult learners (Clement, 1985). We found that students largely did not revise their answers. Critique guidance did not significantly help students that were already struggling with the material. Critique, for science ideas as well as graphs, is an advanced skill that requires more practice in itself. This is consistent with findings from chapter two about how critique is difficult for students that lack content knowledge and still struggle with the core ideas involved. This is evident in students' graph critiques; they paid ample attention to the meaning of slope in reference to speed on a position-time graph because this concept was overall better understood, while they continued struggling with the meaning of line direction after critique. Additionally, students successfully answered the open-response essay question about line direction, but failed to incorporate that idea onto their constructed graph, suggesting that graph critique and revision is even more difficult than revision of essays.

Students' revision rationales gave us more insight regarding why students choose to revise the way they do, or choose not to revise. Consistent with chapter 2 findings, students often do not revise their graphs when prompted, even after guidance. Students' open-response rationales revealed to us that the majority of students that choose not to revise did so because they believed their original graph was already correct, and therefore there was no need to change it. Many of these students did not have correct initial graphs, revealing that indeed students are struggling to recognize gaps in their understanding. This information can help inform the design of future guidance. Students that did successfully use the guidance to distinguish ideas about abstract aspects of the graph and what those aspects represent in terms of motion in real life were the most successful at revising and improving their graphs. While critiquing graphs was difficult for students, they did improve at this skill by the end of the curriculum. This suggests that practice with critique can help students improve their ability to recognize inconsistent ideas. Like students' essay revisions in chapter two, revision of graphs seems to involve a self-directedness. Recognizing your own gaps in knowledge involves metacognitive techniques (Zohar & Barzilai, 2013; Brandsford et al., 2000; Hattie & Temperley, 2007; Kuhn et al., 2011); this may also apply to students' use of critique guidance to reexamine their ideas. Future curricula can be designed to specifically support students in investigating those things they do not already understand and can include continual critique of their own work.

Research Question 3:

The third set of research questions is addressed by the studies in chapter 4. The findings from chapters 2 and 3 reveal a self-directed aspect to revision, and inspired the final empirical chapter, connecting revision and critique to other aspects of self-directed learning.

Chapter two revealed that, while the critique and annotator guidance conditions promoted students to revise more frequently, students who revised spontaneously with the least amount of prompting (the *revisit* guidance condition) had a lot of success improving their science explanations. Additionally, chapter three revealed that most students chose not to revise their graphs because they believed their original graph was already correct, emphasizing that the ability to reflect and reevaluate prior knowledge and critique your own ideas is a significant part of revision. Students that struggled with seeing inconsistencies in their ideas saw no need to revise their graphs. These findings motivated the last set of studies, designed to further examine the connection between critique, revision, and other aspects of SDL, particularly generating and investigating questions.

Our research group collaborated with our group of middle school teachers to determine several self-directed learning practices at which they would like to see their students succeed. These included students generating and investigating their own questions, and engaging in self-assessment to reflect on, evaluate, and modify their ideas and their communication of those ideas. Students' critiques and revisions in each of these studies provided evidence of their level of self-assessment and reflection. The last set of studies, therefore, investigated the relationship between these practices and students' ability to ask their own questions and investigate those questions. Developing and investigating your own questions, and using evidence to reach a conclusion, is important for becoming a lifelong learner both in and out of school, and is a major goal of national science standards (NGSS Lead States, 2013). I also chose question-generation as an outcome measure because it is a skill that requires practice to develop, and can involve a learning curve; this skill, and other SDL practices, are end goals in themselves (Candy, 1991; Hmelo, 2004). It was also a priority for our middle school teachers.

The first study in chapter 4 reveals that students who revised their ideas continually throughout the genetics unit were able to generate better questions at the end, which went beyond the scope of what they had learned. These questions were impressive, asking complex questions about human evolution and mechanisms of genetic engineering. Students that engaged continually with the content material through consistent revision of their explanations during the curriculum were able to generate questions at the end of the unit that went beyond procedural concepts to include prediction, anomaly detection, and application of their knowledge. Critique may have some benefit for generating questions as well, though more research is needed to explore this relationship further. Chapter two findings revealed that critique guidance can help students revise their ideas more often, and chapter four findings show that more frequent revision can lead to better questions, but the relationship between critique guidance and questioning was only verging on statistical significance in the analysis from chapter 4, and requires further study.

While students, especially those that revised consistently, were able to generate very interesting questions about genetic inheritance, it is not easy for teachers to have several classrooms full of middle school students investigating completely different questions. This is particularly true when students' questions involve complex mechanistic investigations of genetic anomalies and laboratory techniques. While it is inspiring to see students come up with these kinds of questions, it is also valuable to have students investigate their own questions in more constrained circumstances. This more constrained context was also designed to communicate to students the importance of generating more narrow scientific questions for which they can make precise predictions. For this purpose, in collaboration with our WISE tech team, for the second study in chapter four I designed a new Data Tool that students could use to develop questions about a collection of variables, make predictions, and then investigate those predictions by using the tool to produce a graph with the relevant variables.

While data literacy is an important scientific skill, our earlier studies show just how difficult graph construction and interpretation are for middle school students. Indeed, these skills are difficult even for older students (Hmelo, 2004). Therefore, our data tool graphing curriculum incorporated considerable scaffolding for middle school students to help them develop questions based on a given dataset, then investigate those questions with the data tool, which automatically produced a scatterplot with the variables selected by the students. The findings revealed that the 6th grade students that participated in this study were quite capable of generating narrow researchable questions about a given dataset. They were also successful at critiquing and revising their questions with the help of guidance. Students were generally also able to successfully investigate their predictions by generating and interpreting a scatterplot with the relevant variables. These students' great success with this may be due to the highly relevant context of the variables in the dataset, as well as the scaffolding possible in this technological environment (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Bannert et al., 2015; Shah & Carpenter, 1995).

Literature on SDL suggests that when students investigate their own questions, they are more likely to take ownership of their learning (Hmelo, 2004; Loyens et al., 2008). Supporting students to generate meaningful questions can help them both check their understanding and pursue more personally meaningful learning goals (Biddulph et al., 1986). These findings suggest that datasets can serve as useful constraints for middle school students to generate and investigate questions and draw evidence-based conclusions. Design of future curricula can include datasets with more complex scientific phenomena for students to investigate.

Overall, chapter 4 studies reveal a meaningful relationship between revision and self-directed learning, particularly generation of more sophisticated scientific questions. Revision and question generation go hand in hand; both are important components of self-directed learning, and students that engage in one are often more proficient at the other. But further research is needed to determine whether improving a student's skill in one area increases their success at the other, or if there are other self-directed learning behaviors that are influencing both of these practices. These studies also reveal that middle school students can investigate their own questions successfully with enough scaffolding. Critique of your own ideas and the ideas of others are crucial components of self-directed learning (Candy, 1991; Blumberg, 2000; Hmelo, 2004) and essential for knowledge integration. These studies set up research questions for future investigations of the potential for critique guidance to support students in further developing self-directed learning practices.

Design Implications

This work has several implications for design of future guidance. For example, when critiquing either explanations or graphs, students often attend to the parts of the critique item they already know. Future critique guidance can be designed to more specifically focus students on ideas or graph features they are most struggling with. Additionally, this research shows that successfully using critique to distinguish ideas requires a fair amount of prior knowledge of the content. Therefore, probing students' prior knowledge on a topic can help assign the most effective guidance. A student with low prior knowledge can be directed to explore more ideas first, or to use the annotator tool to help them organize and connect their ideas before engaging in critique. Explicit instruction on how to critique may also be helpful for all students.

Limitations and Future Research

This work is limited by the fact that all studies involved relatively short classroom interventions (5-10 day, 50-minute periods), and skills such as critique, revision, and other aspects of self-directed learning take much longer to develop. Additionally, the assessments used in these studies were designed to capture acute improvements, and do not necessarily account for the benefits of engaging in these curricular activities for future learning. Practicing critique more regularly might help students better develop internal metacognitive prompts for diagnosing inconsistencies in their thinking. Practicing revision may help students improve the connectedness of their ideas. Future studies can be designed to better capture students' development at these skills by having them engage in critique, revision, graphing, and question-generation more times throughout the school year.

For the graphing assessments, the item was notoriously difficult and students' "altitude" interpretation can be argued as a possible interpretation of "position" in that particular position-time graph. Therefore, future studies can use other position-time scenarios to explore if students can construct and revise graphs effectively in a different context, and if critique is more helpful in less ambiguous situations. Students' revision rationales were greatly insightful regarding when and how they choose to revise. Future studies can be designed to utilize these rationales more effectively, perhaps by giving guidance to students after they write their rationale. Particularly, students that write "I was correct" but do not have an accurate representation can be directed to further guidance and prompted to revise again.

The data tool used in the *Graphing Stories* unit was limited in that it was piloted for the first time in this work. Future studies can include larger datasets with more complex scientific data to see how students generate and investigate questions in a self-directed manner on more conceptually demanding content, and scaffolding can be adjusted accordingly.

Additionally, the pre/post assessments used in this research provided information about students' individual improvement after the curriculum, but revisions during the unit were often done in teams. Therefore, it is feasible not all students in a team participated equally. Future studies can take this into account, either by implementing team-level revision assessments, or recording more information about team behavior during the unit.

Concluding Remarks

Literature on science learning says that students need help critiquing and distinguishing their ideas to integrate new information with their prior knowledge (Campanile et al., 2013; Osborne, 2010; Berland & Reiser, 2009; Linn & Eylon, 2011; Berland et al., 2016; Henderson et al., 2015). Literature on revision in other contexts also acknowledges that novices require assistance with recognizing and diagnosing inconsistencies in their ideas and communicating those ideas in order to revise in a meaningful way (Flower et al., 1986; Crawford et al., 2008; Sato, 2015; Zheng et al., 2015; Bridwell, 1980). The middle school students that engaged with the curricula in this research are novices, in genetics and graphing as well as writing, and as educators we cannot simply prescribe expert habits, such as revising more frequently, and expect students to make substantial revision (Flower et al., 1986). Therefore, it's no surprise that when prompted to revise, students often added unrelated ideas or made surface level changes; students have more experience and confidence with spelling and grammar than they do with Mendelian genetics. Literature on writing and revision enumerates several underlying expert practices related to expert habits, including knowing when to reread their work and asking themselves metacognitive questions to re-examine whether their ideas make sense or if there are better ways to communicate those ideas (Flower & Hayes, 1981; Flower et al., 1986; Bridwell, 1980; Brownell et al., 2013; Bannert et al., 2015). Therefore, I aimed to design guidance activities to promote these underlying expert habits and thought processes, particularly through critique.

Building knowledge requires a cycle of construction and critique of explanations (Ford, 2008; Henderson et al., 2015), particularly in science. Critique is an essential mechanism for identifying flawed reasoning in science, which is a crucial step in revision. Critique involves a process of interrogation, seeking to explore why an explanation or argument is inconsistent or invalid, which motivates exploring alternative explanations (Henderson et al., 2015; Sato, 2015; Berland & Reiser, 2009; Mercier & Sperber, 2011). The critique guidance used in this work was designed to help students in this process, to scaffold re-examining their prior knowledge in the light of new ideas and help students recognize inconsistencies and gaps in their understanding.

Together, this work shows the value of guidance, particularly critique, for encouraging revision of scientific ideas and graphical representations, as well as generating relevant scientific questions. Engaging students in explicit critique of their own ideas and the ideas of others is difficult, but can help promote revision leading to coherent knowledge integration. By designing activities that encourage students to reflect on their own learning, distinguish ideas, and reevaluate their understanding, we can help them become more self-directed and take greater ownership of their science knowledge.

References

Bannert M., Christoph Sonnenberg, Christoph Mengelkamp Elisabeth Pieger. 2015. Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. Computers in Human Behavior. 52: 293-306. https://doi.org/10.1016/j.chb.2015.05.038

Barton, A.C., & Tan, E. 2018. A longitudinal study of equity-oriented STEM-rich making among youth from historically marginalized communities. American Education Research Journal 55:4 (761-800).10.3102/0002831218758668

Bell, A., & Janvier, C. (1981). The interpretation of graphs representing situations. For the learning of mathematics, 2(1), 34-42.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. Science Education, 93(1), 26–55. http://doi.org/10.1002/sce.20286

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. Journal of Research in Science Teaching, 53(7), 1082-1112.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. Science Education, 95(2), 191-216.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. Science Education, 94(5), 765–793. http://doi.org/10.1002/sce.20402

Bertin, J. (1983). Semiology of graphics: Diagrams, networks, maps. (W.J. Berg, Trans.). Madison, WI: University of Wisconsin Press. (Original work published 1967)

Biddulph, F., Symington, D., & Osborne, R. (1986). The place of children's questions in primary science education. Research in Science & Technological Education, 4(1), 77-88. http://dx.doi.org/10.1080/0263514860040108

Blumberg, P. (2000). Evaluating the evidence that problem-based learners are self-directed learners: A review of the literature. In D. H. Evensen & C. E. Hmelo (Eds.), Problem-based learning: A research perspective on learning interactions (p. 199–226). Lawrence Erlbaum Associates Publishers.

Blumberg P. & Michael J.A. (1992) Development of self-directed learning behaviors in a partially teacher-directed problem-based learning curriculum, Teaching and Learning in Medicine, 4:1, 3-8, DOI: 10.1080/10401339209539526

Boote, S. (2012). Assessing and understanding line graph interpretations Using a scoring rubric of organized cited factors. Journal of Science Teacher Education, 1–22. doi:10.1007/s10972-012-9318-8

Bradford, A., & Gerard, L. (2020). Supporting teachers to customize science curriculum for self-directed learning: customization approach and impacts on teacher and student learning. Paper presented at the annual international conference of the National Association of Research in Science Teaching (NARST). Portland, OR.

Brasell, H. (1987). The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. Journal of Research in Science Teaching, 24(4), 385-395.

Breakey, K. M., Levin, D., Miller, I., & Hentges, K. E. (2008). The use of scenario-based-learning interactive software to create custom virtual laboratory scenarios for teaching genetics. Genetics, 179(3), 1151-1155.

Bridwell, L. (1980). Revising strategies in twelfth grade students' transactional writing. Research in the Teaching of English, 14, 197–222.

Brownell, S. E., Price, J. V., & Steinman, L. (2013). Science Communication to the General Public: Why We Need to Teach Undergraduate and Graduate Students this Skill as Part of Their Formal Scientific Training. Journal of Undergraduate Neuroscience Education, 12(1), E6–E10.

Campanile, M. F., Lederman, N. G., & Kampourakis, K. (2015). Mendelian genetics as a platform for teaching about Nature of Science and Scientific Inquiry: The value of textbooks. Science & Education, 24(1-2), 205-225.

Candy, P. C. (1991). Self-direction for lifelong learning: A comprehensive guide to theory and practice. San Francisco: Jossey-Bass.

Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K-12 Science Contexts. Review of Educational Research, 80(3), 336–371. http://doi.org/10.3102/0034654310376953

Chang, H. and Linn, M. C. (2013), Scaffolding learning from molecular visualizations. J Res Sci Teach, 50: 858-886. doi:10.1002/tea.21089

Chin C., & Brown D.E. (2002) Student-generated questions: A meaningful aspect of learning in science, International Journal of Science Education, 24:5, 521-549, DOI: 10.1080/09500690110095249

Chinn, C. A. and Malhotra, B. A. (2002), Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Sci. Ed., 86: 175-218. doi:10.1002/sce.10001

Clement, J. (1985, July). Misconceptions in graphing. In Proceedings of the ninth international conference for the psychology of mathematics education (Vol. 1, pp. 369-375). Utrecht, The Netherlands: Utrecht University.

Crawford, L., Lloyd, S., Knoth, K.: (2008) Analysis of student revisions on a state writing test. Assessment for Effective Intervention, 33, 108–119.

Dahlgren, M. A., & Dahlgren, L. O. (2002). Portraits of PBL: Students' experiences of the characteristics of problem-based learning in physiotherapy, computer engineering and psychology. Instructional Science, 30(2), 111-127.

DeBoer, G. E. 1991. A history of ideas in science education, New York: Teachers College Press.

DiSessa A., D. Hammer, B. L. Sherin, and T. Kolpakowski, Inventing graphing: meta-representational expertise in children, J. Math. Behav. 10, 117 (1991); https://eric.ed.gov/?id=EJ443633.

Donnelly, D. F., Vitale, J. M., & Linn, M. C. (2015). Automated Guidance for Thermodynamics Essays: Critiquing Versus Revisiting. Journal of Science Education and Technology, 24(6), 861–874. doi:10.1007/s10956-015-9569-1.

Donnelly-Hermosillo, D. F., Gerard, L. F., & Linn, M. C. (2020). Impact of graph technologies in K-12 science and mathematics education. Computers & Education, 146, 103748.

Fahnoe, C. & Mishra, P. (2013). Do 21st Century Learning Environments Support Self-Directed Learning? Middle School Students' Response to an Intentionally Designed Learning Environment. In R. McBride & M. Searson (Eds.), Proceedings of SITE 2013--Society for Information Technology & Teacher Education International Conference (pp. 3131-

3139). New Orleans, Louisiana, United States: Association for the Advancement of Computing in Education (AACE).
Fisher, M., King, J., Tague, G., 2001. The development of a self-directed learning readiness scale for nursing education. Nurse Education Today 21, 516–525.
Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. College composition and Communication, 37(1), 16-55.
Flower, L. & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. College Composition and Communication, Vol. 32 , No.4, pp 365-387.
Ford, M. J. (2008). Disciplinary authority and accountability in scientific practice and learning. Science Education, 92(3), 404–423.
Freedman, E. G., and Smith, L. D. (1996). The role of data and theory in covariation assessment: Implications for the theory-ladenness of observation. J. Mind Behav. 17: 321–343.
Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, 32(2), 124–158. doi:10.2307/749671
Gatewood, E. 2019. Use of Simulation to Increase Self-Directed Learning for Nurse Practitioner Students. J Nurs Educ. 58(2):102–106.
Guglielmino, L. M. (1989). Manual self-directed learning readiness scale. Norman, OK: University of Oklahoma.
Haddadi, H., Howard, H., Chaudhry, A., Crowcroft, J., Madhavapeddy, A., & Mortier, R. (2015). Personal data: Thinking inside the box. arXiv preprint arXiv:1501.04737.
Hattie, J., & Temperley, H. (2007). The power of feedback. Review of Educational Research, 77, 81–112.
Harrison, E.J. Gerard, L. F., Linn, M. C. (2018). Encouraging Revision of Scientific Ideas with Critique in an Online Genetics Unit. Paper in proceedings and presented at the annual meeting of International Conference of the Learning Sciences (ICLS).
Harrison, E. J. (2020, Apr 17 - 21) Factors Affecting Student-Generated Questions in an Online Genetics Curriculum [Poster Session]. AERA Annual Meeting San Francisco, CA; http://tinyurl.com/v5s84jn (Conference Canceled)
Harrison, E.J. Gerard, L. F., Linn, M. C. (2018). Using Guidance that Models Revision to Encourage Integrating New Ideas in an Online Science Unit. Paper presented at the annual meeting of the American Education Research Association (AERA), NY, NY.
Harrison, E.J. McBride, E., Linn, M. C. (2019). Revision Analysis of Students' Position-Time Graphs. Paper in proceedings and presented as a poster at the annual meeting of Computer-Supported Collaborative Learning (CSCL).
Henderson, J. B., MacPherson, A., Osborne, J., & Wild, A. (2015). Beyond Construction: Five arguments for the role and value of critique in learning science. International Journal of Science Education, 37(10), 1668–1697. http://doi.org/10.1080/09500693.2015.1043598
Hmelo-Silver, C.E. Educational Psychology Review (2004) 16: 235. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3
Jacobson, M.J & Uri Wilensky (2006) Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences, Journal of the Learning Sciences, 15:1, 11-34, DOI: 10.1207/s15327809jls1501 4
Jiménez Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). "Doing the lesson" or 'doing science': Argument in high school genetics. Science Education, 84(6), 757–792. http://doi.org/10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Johnson, S. K., & Stewart, J. (2002). Revising and assessing explanatory models in a high school genetics class: A comparison of unsuccessful and successful performance. Science Education, 86(4), 463-480.

Kim, D., In-Heok Lee, & Joo-Ho Park. 2019. Latent class analysis of non-formal learners' self-directed learning patterns in open educational resource repositories. BJET. https://doi.org/10.1111/bjet.12746

King A. 1992. Facilitating Elaborative Learning Through Guided Student-Generated Questioning, Educational Psychologist, 27:1, 111-126, DOI: 10.1207/s15326985ep2701_8

Kuhn, D., Wang, Y., & Li, H. (2011). Why argue? Developing understanding of the purposes and values of argumentive discourse. Discourse Processes, 48, 26–49.

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. Journal of Science Education and Technology, 25(4), 665-681.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. Science Education, 87(4), 517–538.

Linn, M. C., Gerard, L., Ryoo, K., McElhaney, K., Liu, O. L., & Rafferty, A. N. (2014). Computer-guided inquiry to improve science learning. Science, 344(6180), 155-156.

Linn, M. C. & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: Lawrence Erlbaum.

Linn, M. C., & Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander, & P. H. Winne (Eds.), Handbook of educational psychology (2nd Ed., pp. 511-544). Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., & Eylon, B.-S. (2011). Science learning and instruction: Taking advantage of technology to promote knowledge integration. New York: Routledge.

Liu, O. L., Lee, H.-S., Hofstetter, C., & Linn, M. (2008). Assessing Knowledge Integration in Science: Construct, Measures, and Evidence. Educational Assessment, 13(1), 33–55. http://doi.org/10.1080/10627190801968224

Loyens, S. M., Magda, J., & Rikers, R. M. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. Educational Psychology Review, 20(4), 411-427.

McBride, E. (2018). Connecting Science Concepts and Engineering Practices: Supporting Student Understanding of Energy Transformations (Doctoral dissertation, UC Berkeley).

McDermott, L. C., Rosenquist, M. L., & Van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. American Journal of Physics, 55(6), 503-513.

McElhaney W., Kevin & Chang, Hsin-Yi & Chiu, Jennifer & Linn, Marcia. (2014). Evidence for effective uses of dynamic visualisations in science curriculum materials. Studies in Science Education. 51. 49-85. 10.1080/03057267.2014.984506.

Mercier H, & Sperber D. 2011. Why do humans reason? Arguments for an argumentative theory. Behav Brain Sci 34(2):57-74.

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. Learning and Instruction, 58, 1-11.

National Research Council. 2006. America's Lab Report: Investigations in High School Science. Washington, DC: The National Academies Press. https://doi.org/10.17226/11311.

National Research Council. 2007. Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, DC: The National Academies Press. https://doi.org/10.17226/11625.

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.

O'Neill, T. B. (2010). Fostering Spaces of Student Ownership in Middle School Science. Equity & Excellence in Education, 43(1), 22-45.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. Science, 328(5977), 463-466.

Rashid, T. & Asghar, HM. 2016. Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. Computers in Human Behavior. 63:604-612. https://doi.org/10.1016/j.chb.2016.05.084

Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. Journal of Research in Science Teaching, 31(9), 969–983. http://doi.org/10.1002/tea.3660310910

Sato, M. E. (2015). Designing critique for knowledge integration (Doctoral dissertation, UC Berkeley).

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (pp. 97-118). New York: Cambridge University Press.

Scheuer, O., McLaren, B., Weinberger, A., & Niebuhr, S. (2014). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. Instructional Science, 42(2), 127-157. Retrieved from http://www.jstor.org/stable/43575221

Schmidt, H. G. (2000). Assumptions underlying self-directed learning may be false. Medical education, 34(4).

Schutt, R. (2013). Taking a Chance in the Classroom: Embracing the Ambiguity and Potential of Data Science, CHANCE, 26:4, 46-51

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. Educational psychology review, 14(1), 47-69.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. Journal of Experimental Psychology: General, 124(1), 43.

Shah, P., & Shellhammer, D. (1999). The role of domain knowledge and graph reading skills in graph comprehension. In meeting of the Society for Applied Research in Memory and Cognition, Boulder, CO.

Shepard, L. (2000). The Role of Assessment in a Learning Culture. Educational Researcher, 29(7), 4-14. http://www.jstor.org/stable/1176145

Slotta, J. D., & Linn, M. C. (2000). The knowledge integration environment: Helping students use the Internet effectively. Innovations in science and mathematics education: Advanced designs for technologies of learning, 193-226.

Starbek, P., Starčič Erjavec, M., & Peklaj, C. (2010). Teaching genetics with multimedia results in better acquisition of knowledge and improvement in comprehension. Journal of Computer Assisted Learning, 26(3), 214–224. http://doi.org/10.1111/j.1365-2729.2009.00344.x

Stylianou, D. A., Smith, B., & Kaput, J. J. (2005). Math in motion: Using CBRs to enact functions. Journal of Computers in Mathematics and Science Teaching, 24(3), 299–324.

Tala, S., & Vesterinen, V.-M. (2015). Nature of Science Contextualized: Studying Nature of Science with Scientists. Science & Education, 24(4), 435–457. http://doi.org/10.1007/s11191-014-9738-2

Tansomboon, C. (2017). Design of Automated Guidance to Support Effortful Revisions and Knowledge Integration in Science Learning (Doctoral dissertation, University of California, Berkeley).

Teo, Timothy & Tan, Seng Chee & Lee, Chwee Beng & Chai, Ching & Koh, Joyce & Wen Li, Chen & Cheah, Horn. (2010). The self-directed learning with technology scale (SDLTS) for young students: An initial development and validation. Computers & Education. 55. 1764-1771. 10.1016/j.compedu.2010.08.001.

Vahey, P., Yarnall, L., Patton, C., Zalles, D., & Swan, K. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In Annual Meeting of the American Educational Research Association, San Francisco, CA.

Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. Journal of Research in Science Teaching, 52(10), 1426-1450.

Vitale, J. M., McBride, E., & Linn, M. C. (2016). Distinguishing complex ideas about climate change: Knowledge integration vs. specific guidance. International Journal of Science Education, 38(9), 1548-1569.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. Educational technology research and development, 53(4), 5-23.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics, 12(3).

Wu, H.-K., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh graders' use of data tables and graphs. Journal of Research in Science Teaching, 43(1), 63–95.

Zheng, B., Lawrence, J., Warschauer, M., & Lin, C.-H. (2015). Middle School Students' Writing and Feedback in a Cloud-Based Classroom Environment. Technology, Knowledge and Learning, 20(2), 201–229. http://doi.org/10.1007/s10758-014-9239-z

Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. Studies in Science education, 49(2), 121-169.

Zohar, A., & Nemet, F. (2001). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching, 39(1), 35–62. http://doi.org/10.1002/tea.10008