UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Design of Automated Guidance to Support Effortful Revisions and Knowledge Integration in Science Learning

Permalink

https://escholarship.org/uc/item/4967n6n6

Author

Tansomboon, Charissa

Publication Date

2017

Peer reviewed|Thesis/dissertation

Design of Automated Guidance to Support Effortful Revisions and Knowledge
Integration in Science Learning

By

Charissa Tansomboon

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Marcia C. Linn, Chair
Professor Susan D. Holloway
Professor Sheri L. Johnson

Spring 2017

Design of Automated Guidance to Support Effortful Revisions and Knowledge

Integration in Science Learning

© 2016

by

Charissa Tansomboon

Abstract

Design of Automated Guidance to Support Effortful Revisions and Knowledge

Integration in Science Learning

by

Charissa Tansomboon

Doctor of Philosophy in Education

University of California, Berkeley

Professor Marcia C. Linn, Chair

Students studying complex science topics can benefit from receiving immediate, personalized guidance. Supporting students to revise their written explanations in science can help students to integrate disparate ideas and develop a coherent, generative account of complex scientific topics. Using natural language processing to analyze student written work, this dissertation compares forms of automated guidance designed to motivate productive revision and help students improve their science understanding. Online environments can support science learning by providing timely, personalized guidance to students, but challenges for effective implementation still exist. Specifically, (a) students often believe online guidance is generic rather than personalized for them; and (b) students do not always engage effortfully with online guidance and improve their written responses. This dissertation includes a series of three studies that address these challenges. A computerized learning environment is used to explore useful and motivating forms of automated guidance for middle school students learning challenging science topics such as thermodynamics.

Informed by the knowledge integration framework and established ideas about student motivation, these research studies examine effective designs for automated guidance. Study 1 demonstrates that automated knowledge integration guidance provided by a computer can promote integrated understanding of science as effectively as expert teacher guidance. In addition, students who began to distinguish scientific ideas after receiving guidance in the embedded assessment, as evidenced by adding either non-normative or normative ideas to their response, made greater gains over the course of the unit than those who did not add any new ideas in response to guidance. However, in this study some students discounted automated computer guidance, assuming it was generic rather than personalized. In Study 2, *transparent* guidance clarified to students how the computer generated personalized guidance based on their response. Results showed that

transparent personalized guidance had a greater impact than standard adaptive guidance on student revisions, suggesting that student beliefs about how guidance is designed influences their performance. Transparent guidance was particularly effective for students who started with a low initial score. This finding resonates with the idea that students who felt they were struggling may have particularly benefited from the reassurance that automated guidance was provided at a level they were expected to be able to achieve. Study 3 compares two specific guidance strategies: *revisiting* evidence or *planning* writing changes, prior to revision. Analysis of student actions after receiving guidance demonstrated that students in the *revisiting* evidence condition were more likely to revisit prior evidence, and students in the *planning* writing condition were more likely to make significant writing revisions. Both revisiting and planning guidance resulted in significant improvement in student knowledge integration, although neither guidance strategy showed a significant advantage over the other. In addition, we found that the form of guidance interacted with school, suggesting that teacher practices could reinforce a specific guidance strategy.

This sequence of studies shows that the design of online guidance is important in encouraging students to revisit dynamic models and make effortful revisions to their work. Carefully designed automated guidance can augment the effectiveness of teachers by motivating students to better use computer learning environments and make effortful revisions that ultimately improve science learning. The results also raise important questions about when to encourage revisiting, how to design instruction that best fits with individual classroom strategies, and how to instill a lifelong practice of engaging in iterative refinement of scientific explanations.

Table of Contents

Acknowledgements

I would like to thank my advisor, Marcia Linn, for her generous support and wisdom throughout my time at Berkeley. Her dedication to mentorship and fostering of a welcoming, collaborative research environment has made my graduate studies exceptionally rewarding. I feel particularly lucky to have worked as her mentee. I would also like to thank the other wonderful members of my dissertation committee. Susan Holloway, for always supporting what was best for me since day one and encouraging me to find and pursue my own path, and Sheri Johnson, for her meticulous insights and positivity, and for helping me to become not only a better researcher but also a better teacher. I would also like to thank Libby Gerard for being incredibly thoughtful, supportive, and encouraging at every step of my research journey. Working with all of you has changed not only the way that I think about education, but also the world.

This work would not have been possible without the great community of people working with the Linn research group and Web-based Inquiry Science Environment (WISE). I owe thanks to many wise WISE researchers, particularly Jonathan Vitale, Lauren Applebaum, Eliane Wiese, Jacquie Madhok, Beth McBride, and Jennifer King Chen, as well as the WISE technology team - Jonathan Breitbart, Hiroki Terashima, Geoffrey Kwan, and David Crowell, for providing the foundation of WISE and never saying no to any requests that were thrown at them. I am also extremely grateful for the many teachers who welcome us and WISE into their classrooms and constantly contribute their insights to improve WISE and inform our research.

Finally, I would like to thank my Berkeley friends and family for their constant support these past four years. To my mom, for listening to more stories of graduate school struggle and study designs than she may have ever wanted to hear. To Emily Campbell, Mike Chen, Steven Tang, Nics Theerakarn, Leon Lin, Claire Kunesh, Becky Hachmeyer, Qian Wang, Sira Park, Irenka Dominguez-Pareto, Varada Sarovar and countless others. You all made Berkeley feel like home.

Chapter 1: Introduction and Rationale

To promote a coherent understanding of science, effective teachers provide personalized guidance and encourage students to engage in activities such as revisiting difficult topics and revising scientific explanations. These activities align with the NGSS standards (2014) that emphasize iterative refinement as a key strategy for science learning. This research program employs natural language processing (NLP) in a web-based learning environment to design and refine automated guidance. Using NLP for automated analysis of student responses allows for immediate assignment of individualized guidance. Whether automated guidance effectively encourages meaningful revision and durable understanding depends on how it is designed (Andersen, Corbett, Koedinger, Petelletier, 1995; Bjork, Dunlosky, & Kornell, 2013; Shute, 2008) . Therefore, careful consideration and design of automated guidance may have large impacts on student learning.

This dissertation examines designs of automated guidance that promote knowledge integration and increase students' feelings of agency and motivation to learn science. The Web-based Inquiry Science Environment (WISE; http://wise4.berkeley.edu) was used to deliver inquiry instruction and guidance as well as conduct randomized comparison studies and log student activities. Design of the curriculum, assessment and guidance is aligned with the knowledge integration framework, a constructivist theory (Linn & Eylon, 2011). To design activities to increase motivation I draw on Dweck's social cognitive view of motivation (Dweck, 1986; Dweck & Master, 2007), and define agency as feeling empowered to take meaningful action and make consequential choices (Kramsch, A'Ness & Lam, 2000, Basharina, 2013). To assess the impact of instruction designed to promote motivation and agency, I examine students' learning strategies, the actions they take to make revisions, and their progress in integrating their ideas about complex science. The following research questions are addressed:

1. How effective is automated knowledge integration guidance from a computer in comparison to typical guidance given by a teacher?
2. How can we best design automated guidance that promotes deep understanding of science along with a propensity to iteratively refine one's understanding?
3. What types of revision strategies, or student-initiated activities such as revisiting scientific models, lead to improved understanding of science?

Writing in Science

Writing is a critical aspect of scientific practice that allows scientists to communicate ideas, generate insights, and clarify ambiguities (Yore, Hand, & Florence, 2004). Writing activities align with the knowledge integration perspective of science learning because it encourages students to reflect on their range of ideas and allows them to identify and reconcile inconsistencies and gaps in their thinking (Fellows, 1994; Hand, Lawrence, & Yore, 2010). Without such activities students may hold onto their multiple, often conflicting views. Prior research has revealed similar benefits of writing for students of science (Hayes, 1987; Rivard, 1994). In the revision process students clarify and refine complex ideas (Rivard, 1994). Writing helps students to reflect on what they know, integrate old ideas with new, and gather feedback on their ideas (Fellows, 1994).

In addition, through engaging with the authentic practice of iterative refinement, students have the opportunity to develop their identity as scientists. Contemporary education standards (i.e., Next Generation Science Standards, 2013), place strong emphasis on iterative refinement of written materials.

Several previous studies have found benefits from students writing in science contexts. Meta-analyses by Graham & Hebert (2011) found that writing about material enhances comprehension for middle schoolers, and that this applied across several subject areas including science. Inquiry lessons that included more opportunities for students to generate explanations using new ideas resulted in significantly greater science learning gains than lessons with the same content but fewer explanation writing opportunities (Fellows, 1994). Studies that compare generation of written explanations to more passive tasks (e.g., reading summaries) demonstrate stronger learning gains for students who write than those who do not (Richland, Linn, & Bjork, 2007). Ryoo and Linn (2014) found that students who generated written explanations about a visualization depicting the mechanisms of photosynthesis were better prepared to extract salient information about the visualization and make more connections between depicted energy transformations and other energy concepts than students who simply read summaries of the process. Students who wrote about science concepts after searching from experimental trials particularly show learning gains (Klein, 2000).

Writing about scientific concepts also aligns with the literature on self-explanations. Previous research finds that self-explanation is successful because it is a constructive activity that facilitates the integration of new information into existing knowledge (Chi, De Leeuw, Chiu, & LaVancher, 1989). Writing tasks can engage students in an active, constructive process of refining their ideas and clarifying their understanding (Fitzgerald, 1987). The writing activities that help students draw out connections between evidence to discover patterns has been shown to produce stronger conceptual understanding than more typical laboratory reporting tasks in which students produce only a final account for evaluation (Hohenshell & Hand, 2006). By manipulating written content, students are more likely to remember and understand the concepts (Langer & Applebee, 1987).

Technology to Support Writing in Science

Student writing can be supported by asking appropriate questions, providing opportunities for students to generate and organize their ideas, and promoting cross-referencing between diverse materials (Hand, Lawrence, & Yore 2010). Technological platforms, such as web-based learning environments, can provide students with ample opportunities to record and revise their ideas and navigational tools to help them revisit relevant materials. For example, activities designed in WISE (Linn, Clark, & Slotta, 2003; Matuk et al. 2015), typically use reflection prompts following visualization activities to help students organize and interpret evidence. Hints and hyperlinks are used to help students navigate to relevant pages that can support their reflection.

In addition to tools that scaffold writing tasks to help students organize ideas, new technologies can help guide revision of writing artifacts by providing formative guidance (Proske, Narciss, & McNamara, 2012). Guidance is valuable to encourage students to distinguish the new ideas from their existing ideas (Hagemans, van der Meij, & de Jong,

2013; Wichmann & Luetuner, 2009). Natural language processing (NLP) allows computers to rapidly analyze written sentences. While most teachers do not have the time to provide detailed, personalized guidance to each individual student in the classroom (Shepard, 2000), the use of NLP to analyze student essays can immediately score student work and allow for personalized assignment of appropriate guidance. Computer based tools also have the benefit of being able to evaluate many text responses consistently and objectively across students (Roscoe & McNamara, 2013). Many existing NLP tools however focus on writing mechanics rather than coherence or scientific accuracy (e.g., Warschauer & Grimes, 2008). New tools are emerging in science that allow for evaluation of student work on a more conceptual level.

Previous work has found that computer assigned guidance to support knowledge integration can improve students' writing revisions and improve science learning (Linn & Eylon, 2011). Guidance designed according to the knowledge integration perspective prompts students to distinguish between their own scientific ideas and new ideas introduced in instruction. This is consistent with how expert teachers guide student reasoning during inquiry (Herrenkohl, Tasker, and White, 2011; Minstrell & VanZee, 1997). In prior research, knowledge integration guidance led to significantly more productive essay revisions, and subsequently more coherent and accurate science essays than did generic guidance (e.g. "Add more evidence") or specific guidance (e.g. "Incorrect. Energy transforms from light energy into chemical energy") (Gerard et al., 2015). Studies also show potential benefits when automated scoring alerts teachers to students who would benefit from their help (Gerard & Linn, 2016).

Knowledge Integration Approach to Science Learning

In these studies we draw on the knowledge integration (KI) perspective to inform the design of writing activities and automated guidance to promote science learning. The knowledge integration framework is a constructivist approach to instruction that emphasizes reflection on one's repertoire of ideas, adding new scientific ideas, using evidence to distinguish accurate and relevant ideas, and forming links between ideas to explain a phenomenon (Linn & Eylon, 2011). Students often enter the classroom with preconceived notions, and in the process of learning science may develop multiple, often conflicting ideas (Smith, diSessa, & Roschelle, 1993). Students typically respond to instruction by adding the new ideas to their multiple and often conflicting views (diSessa, 2006; Osborne, 2000). The KI framework describes how students develop an integrated understanding of a domain by linking and connecting ideas. The framework calls for eliciting, adding, distinguishing, and sorting out ideas as they engage in challenging scientific activities, such as writing (Linn & Eylon, 2011). Student ideas are seen as "building blocks rather than impediments to understanding" and students are "encouraged to reconcile anomalies rather than just memorize isolated pieces of information" (Linn, 1995, p.4). According to the Knowledge Integration framework, "the goal of instruction is to motivate learners to consider new models and to integrate new models with their existing perspectives" (Linn, 1995, p.4).

Knowledge Integration Guidance

Guidance developed to implement KI acknowledges normative (scientifically valid and correct) student ideas, directs students to activities where they can add new ideas or distinguish between normative and non-normative ideas, and provides direction for integrating ideas into more coherent responses. A coherent response is a normative account where students use evidence to link scientific ideas together. The role of guidance, according to this framework, is to assist students in considering new information. The KI approach "stands in contrast to instruction designed to instill correct models by diagnosing weaknesses and correcting them" (Linn, 1995, p.4).

New standards for science education require that instruction goes beyond presentation of simple, isolated facts to encourage development of a coherent, integrated understanding of interrelated science concepts (NGSS Lead States, 2013). While fact comprehension and procedural mechanics (e.g. solving an equation) can be guided through directly mapping ideas or operations (Andersen, Corbett, Koedinger, Petelletier, 1995), an integrated understanding requires more elaborate response formats and a flexible guidance approach. Prior research suggests that within more complex contexts, guidance can be beneficial when it prompts learners to reconsider and refine their ideas (Butler & Winne, 1995).

Existing literature on KI guidance builds upon Vygotsky's idea of the zone of proximal development, or ZPD. The idea of ZPD is relevant to consider, particularly in the design of guidance. The ZPD is defined as "the discrepancy between a child's actual mental age and the level he reaches in solving problems with assistance" (Vygotsky, 2012, p.187). This definition lends itself to applications in the realm of automated guidance. For the guidance provided within WISE to be effective, it must be providing suggestions that fall within the student's ZPD. In guidance designed based on the KI framework, students' levels of understanding are assessed through automated scoring done by a c-rater™ natural language processing model, and appropriate guidance comments are given based on students' score level. As noted by Linn, Davis & Eylon (2004), useful guidance provides students with "hints about what to think about; for example, 'what to include in the report' or 'pieces of evidence we do not understand.' These prompts may act as a 'more able other' (Vygotsky, 1978), encouraging the students to consider issues they may not have considered otherwise" (p.93). Guidance according to this framework works as a "more able other" by directing students to reconsider appropriate information based upon their current level of understanding; the guidance can scaffold students to reach a level of understanding that they could not achieve on their own, but that they can successfully complete with this provided assistance.

Guidance Design

Meta-analyses suggest that effective guidance includes verification of the correctness of a response, elaboration on why or why not a student's response is correct, and guidance on how to improve (Azvedo & Bernard, 1995; Hattie & Temperley, 2007). Providing an answer directly would not be an example of successful scaffolding, as the student could just copy the correct answer and would not necessarily be forced to

consider and attain the new ideas at hand. Guidance has been found to be significantly more effective when it gives details on how to improve the answer (Shute, 2008). KI guidance aims to do this by directing students towards a concept to explore further, but without telling students specifically what is wrong with their answer. KI guidance has been found to be more effective than typical guidance, that provides accurate scientific information directly ("the energy is transformed, not conducted") or motivational support ("try harder") (Linn et al, 2014). Guidance was found to be more successful when it did not give students the answer directly, but rather led them to reconsider ideas and offered suggestions for reviewing ideas.

Challenges in Student Use of Automated Guidance

While NLP tools can analyze student writing and assign scores, choosing optimal guidance for such scores is an active area of research. One challenge is designing guidance that motivates students to engage in substantial writing revisions. Students often follow classroom norms that support correctness instead of refinement. Thus, students who get guidance on their writing in the classroom tend towards surface-level changes instead of deeper conceptual changes (Cohen & Ball, 2011). In addition, less competent writers tend to undertake less meaningful revision than more competent ones (Fitzgerald, 1987). Many school tasks promote the idea that science is a "simple, algorithmic form of reasoning" (Chinn & Malhotra, 2002). This belief may lead students to look for a correct answer rather than seeking evidence to strengthen an argument (Berland & Reiser, 2011). Studies on computerized guidance to support writing have found that students most often make mechanical and surface level revisions instead of ones based on content (Roscoe, Snow, Allen & McNamara, 2015), supporting the idea that students view revision more as a form of proofreading than conceptual reconsideration. In one of our prior studies, over 50% of students who received automated guidance either did not revise their answers or only made surface-level changes without adding a new idea, meaning that less than 50% added a new (correct or incorrect) scientific idea in their revisions (Tansomboon, Gerard, Vitale, & Linn, 2015).

Student Motivation and Automated Guidance

Other studies illustrate the value of effortful engagement with revisions. One study found that students who make more effortful revisions (added either a correct or incorrect scientific idea) made larger gains from pre to posttest than students who did not add a new idea in their revisions (Tansomboon, Gerard, Vitale, & Linn, 2015). Effortful writing revisions may have the benefit of setting in motion a process of reconsidering ideas that is beneficial to science learning, even if students do not immediately process the distinction between correct and incorrect ideas at time of revision.

Whether automated guidance effectively encourages both meaningful revision and advancement in understanding may depend on subtle features of wording within guidance (Shute, 2008) that improve student motivation to make effortful revisions. The role of motivation can become especially important particularly in situations where students feel they are struggling (Dweck & Master, 2009), like the point when they are given guidance and asked to improve. Making strides in learning may not always be enjoyable for

students, but struggling for understanding can be an important part of the zone of proximal development (Vygotsky, 1967, Chaiklin, 2003).

The work in this dissertation addresses the challenge of motivating students to make revisions and improve learning in situations where they may feel like they are struggling and not feel particularly motivated. Students are given automated guidance on a question and then asked to revise their answer to that question. While this process is a valuable opportunity for learning, encouraging students to slow down and reconsider their ideas requires engagement and effort. In addition, motivating students within an online context is a unique challenge, without the affordances of interpersonal interaction that come from students' typical interactions with classroom teachers.

To address these motivational challenges in students' use of automated guidance in WISE, the studies in this dissertation build upon research on student agency. Agency is defined as the belief that one's actions will result in meaningful outcomes. Students who feel higher levels of agency are more likely to feel motivated and be engaged in the task (Bandura, 1989; Basharina, 2013; Kramsch, A'Ness & Lam, 2000). Self-determination theory (Deci, Vallerand, Pelletier, & Ryan, 1991) also posits that people "need to feel a sense of personal causation" as a basis for intrinsic motivation (Zuckerman, Porac, Lathn, Smith, & Deci, 1978). In online environments, students may often lack the feeling of agency because the tasks seem impersonal. Students feel that the activities they do are generic and unspecified, rather than adaptive or dependent on students' individual responses and actions. This misconception may lead students to not feel a sense of causation, which if remedied could increase students' feeling of intrinsic motivation. Existing research on online learning has addressed student motivation by increasing aspects such as contextualization and choice (Cordova & Lepper, 1996). Motivation has been shown to improve due to simply the illusion of choice, even when the choice is illusory or seemingly trivial (Langer, 1989).

Previous research has found that automated guidance technology can be helpful in supporting students' writing and revising in science. However, students do not always engage effortfully when they receive guidance. This dissertation explores designs of automated guidance that can better motivate and support student writing to develop a better understanding of science topics. The designs of automated guidance tested are derived from literature on the knowledge integration framework for science learning, as well as literature on supporting student motivation. The studies in this dissertation use student actions such as revision strategies of adding new written ideas or clicking to revisit prior evidence as markers of engagement and motivation. In addition to examining student learning gains, understanding the specific strategies and actions that are reflective of student motivation in learning can help researchers design more effective automated guidance.

In Study 1, we compare automated KI guidance to automated simulated teacher guidance, to validate that KI guidance content is valuable to students, in comparison to typical guidance content from a teacher. In Studies 2 and 3, we keep science content of the KI guidance the same across conditions but vary the presentation of guidance. In Study 2, we aim to increase students' feelings of agency, by ensuring they recognize that WISE automated guidance is actually adaptive to their responses. In Study 3, we increase student agency in using specific, previously proven strategies to make their revisions after receiving guidance.

6

Chapter 2: Curriculum Design

All materials used for curriculum and assessments were implemented in the Web-Based Inquiry Science Environment (WISE; http://wise4.berkeley.edu). WISE is an online platform for designing and implementing science activity modules. The software is free and open for public use, and has been used by teachers, researchers, and curriculum designers to reach over 100,000 students around the world. Each unit is designed around one scientific topic -- for example, thermodynamics, or global climate change -- and these units are implemented in middle school science classrooms. Within WISE, students view visualizations, conduct experiments, and respond to embedded assessments.

This chapter provides an overview of the WISE Thermodynamics unit, which has been tested and refined over several years of classroom implementation. The studies in Chapter 4 and 5 used the Thermodynamics unit, while the pilot study in Chapter 3 used the Global Climate Change unit. The Global Climate Change unit is described separately in Chapter 3. All 3 studies test varying designs for automated guidance by adding them to an existing WISE unit and determining how they affect student revision strategies and overall learning outcomes.

Curriculum Unit: Thermodynamics/Understanding Heat and Temperature

Thermodynamics curriculum

Varying designs of automated guidance were implemented within a WISE curriculum unit entitled, "Thermodynamics: Understanding Heat and Temperature". WISE Thermodynamics is a week-long unit completed in science class by 6th grade students. Important concepts taught in this unit include heating and cooling curves, conduction, equilibrium, and heat versus temperature. The thermodynamics unit provides instruction about conduction as students engage in visualizations to test heat flow through different materials. This concept is one that can be difficult for students to grasp due to their previous experiences (Clark & Jorde, 2004). For example, students who have felt a glass cup and a wooden cup in the refrigerator may assume that the glass is at a lower temperature than wood, while in reality they are at the same temperature but feel different because glass is a better conductor. Previous research conducted with this unit suggests that it helps students understand how the inherent conductivity of a material affects the speed in which energy transfers through it, and subsequently how hot an object will feel when touched briefly (Donnelly, Vitale, & Linn, 2015).

The thermodynamics unit includes two hands-on experiments in which students use USB temperature probes. The unit also employs NetLogo visualizations to demonstrate different forms of heat flow to students. WISE units are designed to help students compare and contrast their own ideas and others added by the unit. Some steps of the unit include automated guidance from the computer, which automatically scores short essay responses and gives guidance appropriate to their score level which directs them to consider new ideas and then revise their answers.

Activity 1: Temperature Graphs

Activity 1 introduces students to graphs of temperature, and leads them through an experiment to plot their own heating and cooling curves by placing a hot/cold temperature probe in water (Figure 2.1). Students initially predict on a graph whether the temperature change will form a straight line or a curved line, and then import their data onto the prediction graph so they can directly compare the two shapes.
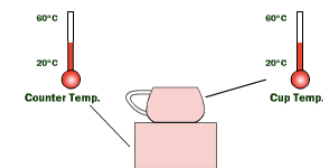

Figure 2.1 Temperature Experiment in Activity 1

Activity 2: Heat Transfer and Equilibrium

In Activity 2, students interact with a simulation of heat transfer between two objects. They are also introduced to the idea of equilibrium.
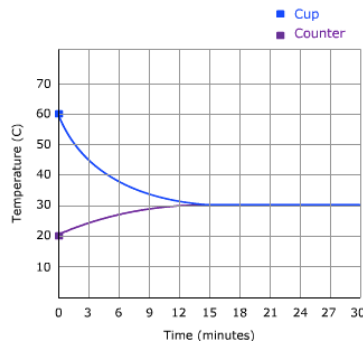

Figure 2.2 Equilibrium Simulation in Activity 2

Activity 3: Feels Warm, Feels Cool…?

In Activity 3, students complete the second experiment using temperature probes. Students are asked to take the temperatures of wood, metal, Styrofoam, glass, and plastic at room temperature. This experiment was intended to help students understand that different objects at the same temperature do not necessarily feel the same to touch.

Students also interact with visualizations intended to instruct students about conduction. In one visualization, students viewed an animation of fingers touching objects made of diverse materials (Figure 2.3). Depending on the type of material, the animation depicted the flow of energy into the finger at different rates. This visualization aimed to demonstrate that objects made of different materials may feel hotter, due to varied conductivity levels, even if they are at the same temperature. This visualization was targeted for revisiting later in the unit; if students did not demonstrate mastery of the concept of conduction, automated guidance suggested for them to revisit this step to gather more information.
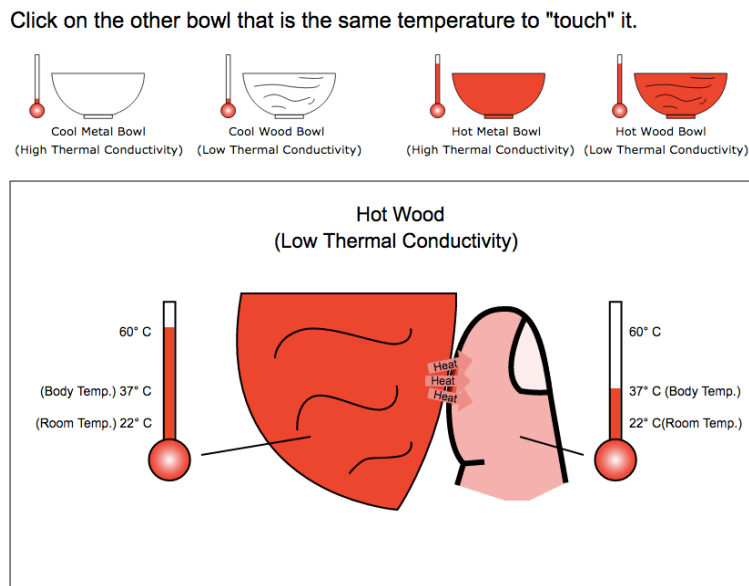
Figure 2.3 Interactive Simulations about Conduction in Activity 3

At the end of Activity 3, students answer a question about conduction called *Spoons*. The studies in this dissertation focus on the design of automated guidance for *Spoons*, which is discussed in further detail in the following section.

Automated Guidance in WISE

Spoons Item

The WISE thermodynamics curriculum includes an automatically scored short answer question called *Spoons* (Figure 2.4). *Spoons* prompts students to select and explain which of three spoons (metal, wood, plastic) would feel the hottest after being placed in a hot water after 15 seconds. This question helps students understand that

conduction varies in different materials and encourages students to distinguish between heat and temperature. After students' answers to *Spoons* are automatically scored, WISE assigns guidance based on the students' score level, which prompts them to consider missing scientific ideas and revise their answer.

A metal spoon, a wooden spoon, and a plastic spoon are placed in hot water.

After 15 seconds, which spoon will feel the hottest and WHY? Be sure to explain your ideas below.

Metal will be hottest because it is the best conductor

Figure 2.4. Spoons question with initial student response

Automated Guidance in WISE

Spoons is scored on a knowledge integration rubric (Table 2.1). In prior research, knowledge integration assessments of inquiry science learning have been found to be valid and adequately measure student understanding and explanation of scientific concepts (Liu, Lee, & Linn, 2011). Spoons was also chosen as the focus of our research on automated guidance design because its automated c-rater ML™ scoring has shown to be reliable and effective (Donnelly, Vitale, & Linn, 2015). The knowledge integration guidance developed for Spoons includes three components: (a) a question targeting a concept not ad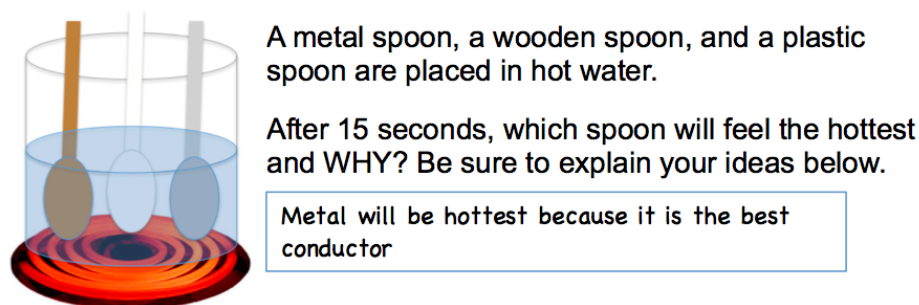dressed by the student response; (b) a prompt directing the students to revisit a visualization (the finger animation) in the unit to review evidence of key concepts, and (c) instruction telling the student to generate an improved explanation that distinguishes among new ideas and the ideas in the response.

Spoons automated scoring is done with the NLP tool c-raterML™. c-raterML™ works by applying a series of NLP steps that performs a series of linguistic analysis including correcting students' spelling and analyzing paraphrases in student responses, then examining student responses for the presence of specific concepts (Liu et al., 2016). The c-raterML™ system scores each response based on a 5-point knowledge integration rubric that rewards students for making coherent links between scientific ideas. After student answers are scored by c-raterML™, WISE instantaneously assigns automated guidance based on the score level, and prompts students to revise their answer. Human scored responses are used to inform the c-raterML™ model building process. c-raterML™ scoring shows satisfactory agreement with human scoring. The c-raterML™ scoring of *Spoons* has a Pearson correlation of .72 between c-rater ML™ score and human score (Liu et al., 2014).

WISE in the Classroom

Students completed the thermodynamics unit in pairs, with two students sharing one computer and working collaboratively throughout the unit. This follows the knowledge integration framework for science learning by taking advantage of the possibility that

students who work together can introduce each other to new concepts and critique each others' ideas (Linn, Clark, & Slotta, 2003). The method for forming pairs varied by classroom. In some classrooms, students were paired with those who were sharing a desk with them, while in others students were paired randomly by the teacher. Students were encouraged to discuss the materials with their classmates and ask the teacher for help. In some cases researchers were also available to answer student questions, but when students' questions involved automatically guided items both researchers and teachers prompted the students to follow the automated guidance.

In each school that ran the project, a researcher was present for classroom observations. Researchers spent 3-6 days in each classroom.

Assessments

Spoons Embedded Item

Table 2.1: Spoons knowledge integration (KI) rubric with corresponding guidance

	A metal spoon, a wooden spoon, and a plastic spoon are placed in hot water. After 15 minutes, which spoon will feel the hottest and why? Be sure to explain your ideas below.	Key Normative Ideas 1. Metal is/feels hotter 2. Metal is a better conductor, mentions rate of heat flow 3. Temperature of all 3 spoons is the same but metal feels hotter	
	Knowledge Integration Level & Description	**Student Examples**	**Knowledge Integration Guidance** **(assigned based on automated c-rater™ score)**
1	Off-task. Student writes, but it does not answer the question being asked.	"I don't know "	(Student Names), redo. Add evidence from step 4.6, the finger/bowl activity, to explain why one spoon feels hotter than the others.
2	Scientifically non-normative ideas or links; Vague Ideas	"A plastic spoon would feel the hottest, because plastic can melt easily."	
3	Partial link; Unelaborated connections using relevant features	"The metal spoon because it was in the cup for 15 seconds and the other ones are plastic and wood spoons."	Good start, (Student Names)! Add evidence from step 4.6, the finger/bowl activity, to explain why one spoon feels hotter than the others.
4	Full link; One scientifically complete and valid connection between two normative ideas	"The metal will be the hottest because it is the best conductor out of the three so heat moves the fastest."	Good start, (Student Names)! Now revisit 4.7, Metal and Wood Bowls and think about how the temperature of the metal spoon compares to the temperature of the wood and plastic spoon.
5	Complex links; Two or more scientifically complete and valid connections	"Metal conducts heat faster than plastic or wood so it FEELS the hottest, but it is actually the same temperature as the other spoons."	

Cups pre-post item

　　　　To investigate prior knowledge and learning gains we focused on a short-answer, pre-post item, *Cups* (Figure 2.4), that closely aligns with the thermodynamics concepts covered in the embedded *Spoons* item. Cups was scored on a KI rubric very similar to the one used for Spoons. Students completed the pre and posttests individually. In this item students were asked to identify which of three cups (metal, wood, plastic) would feel the hottest after hot liquid was placed inside.

Figure 2.4
Cups Pre-Post Item

Table 2.2
Assessment items, location in unit, and method of scoring for guidance and analysis

Item Location	Item Name	Pairs/Individual	Scoring
Embedded Within Unit Item	*Spoons-embedded*	Completed in pairs	Automated score by c-raterML™ for assignment of automated guidance Human scoring for analysis
Pre/Posttest Item	*Cups-prepost*	Completed individually	Human scoring for analysis

　　　　Table 2 includes a summary of the embedded Spoons and pre-post Cups measure. Throughout the current studies, analysis of the Spoons items and actions taken during the revision process applies to student pairs, because students completed the thermodynamics

unit in pairs. Analysis of pretest and posttest performance was done on individuals, since students worked separately on those portions.

Spoons was scored by c-raterML[TM] to allow for assignment of automated guidance based on score level. Both Spoons and Cups were scored by humans for purposes of analysis of students' learning gains on the item. Spoons and Cups were both scored by two researchers. Cohen's kappa calculated for a subset of 50 Spoons responses was .92.

Assessment of Revision Strategies

To examine students' specific actions during revisions, log files and specific writing changes during revisions were examined. KI guidance presents students with a link to a suggested step to visit to gather more information. Log files were used to determine whether students clicked to revisit the suggested step.

A rubric was also developed to categorize revision characteristics (Table 2.3), based on the patterns of revisions most commonly observed from students. Students initial and revised *Spoons* responses were compared. The lowest revision level includes students who did not make any changes to their initial answer. Level 2 included students who added words or rearranged words, but did not effectively change the meaning of their initial answer or add a new idea to the answer. Level 3 students had an additional or changed idea in their final answer, but the idea was non-normative, meaning that it was not scientifically correct and valid. Level 4, the highest level, included students who added at least one normative scientific idea from their initial to revised response. If students responses included edits that could be classified at several different levels, for example, having added both a non-normative and also a normative idea, they were given the score corresponding to the highest level revision that they made. A subset of student revisions were scored by two researchers, with the first researcher scoring responses and the second researcher checking the scores. The rubric was discussed and refined until they reached agreement. Table 2.3 includes detailed descriptions of each revision score level, along with example pairs of student initial/revised responses at each revision level.

Table 2.3
Rubric for revision characteristics and student examples

Revision Characteristic Level	Student Example
No Change	Initial Response: Metal will be hot. Revised Response: Metal will be hot.
Minimal Change: Words were changed, added, or deleted, but these changes did not add a new idea	Initial Response: Metal is the hottest. Revised Response: Metal is the most hot spoon of them all.
Substantial Non-Normative Change: At least one non-normative idea added	Initial Response: Metal is hotter. Revised Response: Metal is hotter in this case because it is best at keeping heat inside.
Substantial Normative Change: At least one normative idea added	Initial Response: I think metal is the hottest. Revised Response: I think metal is the hottest

	because it is the best conductor compared with wood and plastic.

Chapter 3: Comparing Knowledge Integration Guidance and Simulated Teacher Guidance

Introduction

In a computer-supported environment such as WISE, natural language processing tools allow for automated scoring and assignment of individualized guidance for student responses. Such forms of automated guidance can be beneficial in assisting the teacher in engaging students to revise written responses to complex science prompts, freeing teachers to work with students who most need individualized assistance, and providing all students with useful feedback regardless of teacher availability. The goal of this pilot study is to determine whether knowledge integration guidance, provided and assigned by the computer, can be as effective as typical teacher guidance provided to students at similar knowledge levels. This study contrasts two forms of automated guidance within a Web-Based Inquiry Science Environment (WISE) for student explanations of climate change based on scores generated using c-raterML™ natural language processing tools. Knowledge Integration (KI) guidance was designed following the knowledge integration framework (Linn & Eylon, 2011). Simulated teacher (ST) guidance was designed based on analysis of guidance given by experienced teachers.

Curriculum
WISE Global Climate Change Unit

Pilot Study 1 was conducted in a WISE unit titled "What Impacts Global Climate Change?". In this unit, students are asked to consider the effect of increased carbon dioxide on the planet's climate. One commonly held non-normative idea is that carbon dioxide itself is warm, and thus heats up the climate, or that increased carbon dioxide causes holes in the ozone that heat up the planet. In reality, carbon dioxide traps infrared radiation, which gets reabsorbed by the earth as heat.

The WISE Global Climate Change unit includes a short essay with automated guidance called *Coal* (Figure 3.1). *Coal* had been used in previous years' runs of this WISE unit, and accumulated a sufficient number of student responses and corresponding human scores to build a c-rater™ model. The c-rater™ model applies a sequence of NLP steps, including correcting students' spelling, determining grammatical structures, resolving pronoun reference, and analyzing paraphrases to identify concepts in students' responses (Sukkarieh & Blackmore, 2009). The c-rater™ scores for this item have been validated by comparing them to sets of human scores not used for modeling, resulting in an interrater agreement κ coefficient of 0.87 (10) (Linn et al, 2014).

Figure 3.1 *Coal* question with student initial response

Once a student submits an answer to *Coal*, the computer autoscores the response on a scale of 0-5, using a KI rubric specific to the question. After students submit their answer and it is autoscored, a pop-up appears with level-appropriate suggestions on how to improve their answer. The previous answer is retained in a greyed out box, and students are asked to write a new response in a white box below it (Figure 3.2).

Figure 3.2 Automated guidance *Coal* step in WISE

Research questions

This study investigates how best to design automated guidance to promote coherent understanding in inquiry science. We compare the impact of knowledge integration (KI) guidance, to guidance simulating that which was written by experienced teachers (ST) on students' understanding of climate change. KI guidance was developed based on the KI framework for science learning, while ST guidance was designed to mimic the feedback that previous classroom teachers had given to students in previous

years' iterations of this unit. We hypothesize that KI guidance may be more effective in directing students to add and distinguish new ideas, and may help students integrate these ideas into a coherent understanding and response. KI guidance prompts students to reconsider visualizations earlier in the unit and contrast these with their existing ideas, rather than providing them directly with the correct answer or reasoning as to why their own ideas were incorrect.

Outcome measures include embedded and pre-post assessments of climate change; logged navigation data; student perceptions of the value of the guidance; and analysis of the characteristics of student revisions based on guidance. Research questions include:

1. Which guidance condition (KI or ST) is most useful for helping students develop a deeper and more integrated understanding of global climate change, as measured by the WISE unit?
2. How does guidance condition impact the types of revisions students make?

Methods

Participants

177 6th grade students from one teacher's classes in a public school participated in this study. 26% of students in the school received a free or reduced priced lunch. Ethnicity breakdown for the school is 53% Caucasian; 28% Hispanic; 9% two or more races; 4% Asian. Students within each of this teacher's 6 class periods were randomly assigned to the KI or ST guidance conditions. Students using WISE in their classroom complete the unit in pairs and do pre/post tests individually. The main WISE projects is completed in pairs, as it builds upon a variety of collaborative tools that allow students to have discussions and learn from each other's ideas (Linn & Slotta, 2000).

Assigning guidance

KI guidance was designed by examining student responses at each level of the KI scoring rubric for the Coal prompt (Table 1). Table 1 also includes a conceptual description of each KI level, and a typical student response for that KI level. Guidance was constructed to target at each level the most basic concept not addressed accurately by the student response. The guidance followed the KI framework: make an observation to connect new information to initial ideas, ask a question about a missing or non-normative concept to direct attention for distinguishing ideas, suggest revisiting relevant parts of the unit, and ask students to generate an improved response.

In conjunction with the KI approach to guidance, students are directed in WISE to consider visualizations that demonstrate the interrelated movement of carbon dioxide, IR, and heat within the earth's atmosphere. Figure 3.3 is an example of a visualization that students who scored a 2, 3, or 4 were directed to revisit.
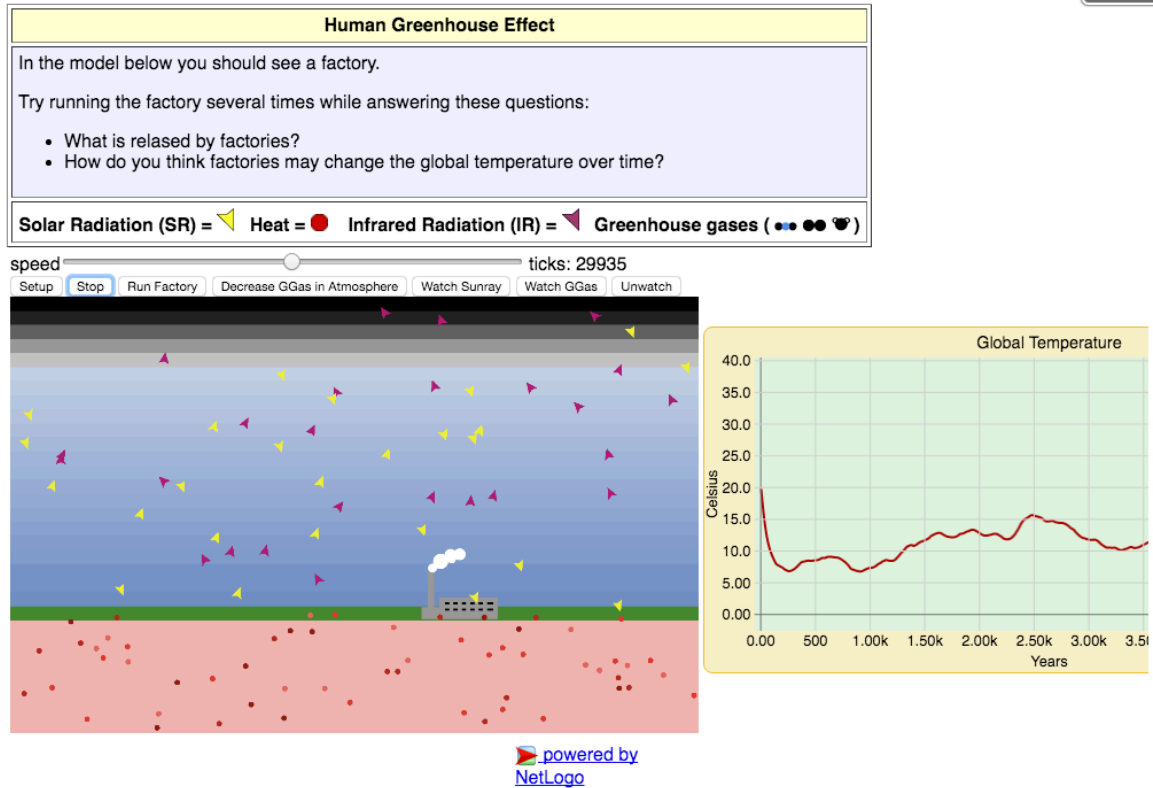
Figure 3.3. NetLogo simulation that students were prompted by KI guidance to revisit

To design ST guidance, we built upon guidance given by an experienced teacher who had previously taught this unit. WISE units are designed such that classroom teachers can sign into the system to view student work and also provide written comments to students on their answers to certain questions. In previous runs of the global climate change unit, when automated feedback for this Coal item were not implemented, teachers were able to view student work overnight and give feedback for students to improve their answer the next time they signed into the WISE unit. We identified patterns in guidance the teacher had given to students at each score level, and designed a representative typical comment to correspond to each KI score level of student responses (Table 1). For example, for students who gave a level 1 initial response, teachers typically gave short, non-descriptive feedback that did not mention specific aspects of the unit; for this score level, the typical teacher comment was encapsulated by the short statement of "Redo."

18

Table 3.1: Coal KI Rubric with Corresponding KI Guidance and ST Guidance

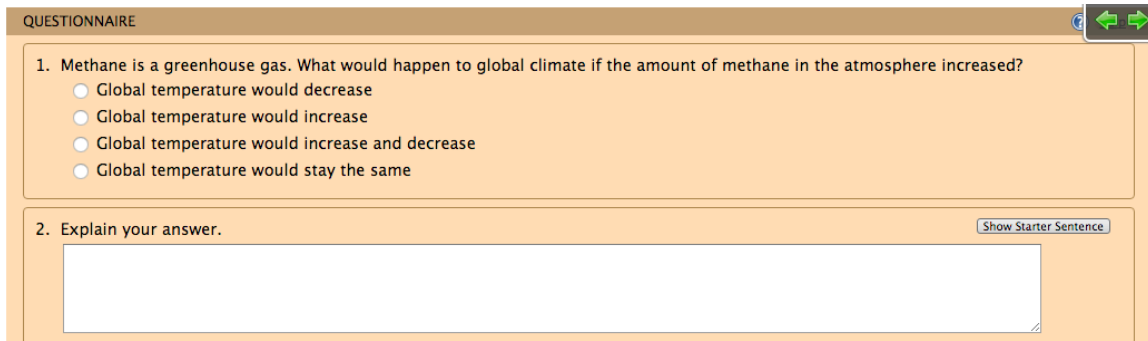|	**Burning coal for human use has dramatically increased the amount of carbon dioxide in Earth's atmosphere. What possible effect could the increased amount of carbon dioxide have on our planet?** Key Normative Ideas: 1. Carbon dioxide reflects/traps infrared radiation 2. Infrared radiation gets reabsorbed by the earth as heat 3. Carbon dioxide is a greenhouse gas/Carbon dioxide makes climate warmer			
	Knowledge Integration Level & Description	**Student Examples**	**KI Guidance**	**ST Guidance**
1	Off-task. Student writes, but it does not answer the question being asked.	Just guessed, idk	Think about a greenhouse. Add ideas from Step 4.3 about how increased carbon dioxide, a greenhouse gas, affects Earth's temperature. Write your explanation below.	Redo.
2	Scientifically non-normative ideas or links; Vague Ideas	It's like methane.	Good start. Now, look at the graphs in Step 4.4. Add evidence about how increasing carbon dioxide, a greenhouse gas, effects global temperature. Write your new explanation below.	Redo. What does increased carbon dioxide do to global temperature?
3	Partial link; Unelaborated connections using relevant features OR Scientifically valid connections that are not sufficient to solve the problem.	Burning coal increases greenhouse gases (CO2) and makes the climate warmer OR CO2 makes it/climate warmer. OR CO2 is like a greenhouse gas.	You are on the right track. Now revisit Step 4.4 and add details. **How** does carbon dioxide interact with infrared radiation to increase global temperature? Write your new explanation below.	Close, try a more complete answer. What does carbon dioxide do to infrared radiation?
4	Full link; One scientifically complete and valid connection between two normative ideas	CO2/Greenhouse Gases trap heat in the atmosphere.	Good progress. To improve your response return to Step 4.4 to find out what happens to energy from the Sun when it is absorbed by the Earth. Write a new explanation below.	Add detail. What happens to the infrared radiation?
5	Complex links Two or more scientifically complete and valid connections	As fossil fuels are released in the atmosphere the IR gets trapped/reflected in the atmosphere, heating up the earth.	Nice thinking. Now, go back to step 5.2 and add examples of how humans disrupt the natural greenhouse effect. Write your expanded explanation below.	Add detail. How else do humans contribute to climate change?.

The KI guidance was different from the ST guidance in several ways. KI guidance included a first sentence to recognize student progress with a comment such as "Good start", and it included a prompt/link for the student to revisit relevant evidence earlier in the unit (Figure 2). Students were prompted to visit a level-appropriate interactive visualization from which they could draw conclusions relevant to the assessment question. For example, students at KI level 4 were told to "return to Step 4.4 (link) to find out what happens to energy from the Sun when it is absorbed by the Earth." KI guidance was also more descriptive in comparison to ST guidance, particularly at the lowest KI level, in terms of which concepts the student should consider to improve their answer. At this KI level, where students did not make a response or wrote something off-task, the aggregate of teacher comments in ST condition was for students to simply "redo." In the KI condition, however, students were told to "Think about a greenhouse. Add ideas from Step 4.3 about how increased carbon dioxide, a greenhouse gas, affects Earth's temperature. Write your explanation below." These features of KI guidance were designed to promote student reflection and consideration of new ideas in their revisions, with the hopes that this would promote overall student learning and likelihood of integrating ideas into scientific understanding.

Assessments

To measure student learning in response to automated guidance, we examined students' initial and revised explanation for *Coal* after receiving KI or ST guidance. We also examined student performance on a pre/post test item called *Methane*, which addressed concepts similar to those in *Coal*. *Methane* was also scored on a KI rubric from 1-5, similar to the one used for *Coal*.

> QUESTIONNAIRE
>
> 1. Methane is a greenhouse gas. What would happen to global climate if the amount of methane in the atmosphere increased?
>   ○ Global temperature would decrease
>   ○ Global temperature would increase
>   ○ Global temperature would increase and decrease
>   ○ Global temperature would stay the same
>
> 2. Explain your answer.   [Show Starter Sentence]

Figure 3.4 Methane pre/post assessment item

All open response items were designed to measure students' integrated understanding of climate change mechanisms and were scored using KI rubrics. Research shows that questions designed to measure knowledge integration are also valid in assessing students' conceptual understanding (Liu, Lee, Linn, 2011). Studies show that these assessments have good psychometric properties including high reliability, lack of differential item functioning for subgroups, and satisfactory IRT scaling (Liu, Lee, Hofstetter, & Linn, 2008).

To assess the characteristics of students' revisions, we compared their initial to revised responses. Initial responses are the answers submitted by students when given the prompt: "You have 1 chance to receive feedback on your answer so this should be your best work! Are you ready to receive feedback on this answer?" and final responses were the responses students submitted after receiving pop-up guidance. Student revisions were scored based on the revision characteristics rubric presented in Chapter 2 (Table 2.3)

WISE also logs student navigation patterns, which allows researchers to determine whether students clicked on the link to revisit a prior step as suggested by KI guidance, as well as how long students spent on each page [Figures 1 and 2]. After revision, students were asked about their perceived utility of the guidance for strengthening their explanation.

On the step following the *Coal* revision step, students are asked to make comments about how helpful the automated guidance was. Students were asked, "Did the feedback help you write a better explanation? Explain how." A subset of students was also selected for interviews about the automated guidance step. Students were selected at random out of those who had turned in their parent permission slips to be interviewed and audio recorded.

Results

Score gains from KI vs. ST guidance

Automated assignment of KI guidance, in comparison to ST guidance, led to significantly greater improvement in students' revised *Coal* explanations (Table 3.2). Students in the KI condition had an average gain of .333 in KI score from initial to revised score, which was significantly different than those in ST condition, who had an average gain of .091 (t=3.58, $p < .001$, d=.54). Both approaches brought about equally significant pre- to posttest improvement (Table 3.2). Students' initial score on *Coal* was not a significant predictor of the effect of guidance condition on gain scores, suggesting that guidance was equally helpful for all levels of student responses. Students in all conditions showed a significant KI score gain from initial to final score on the embedded Coal item, as well as from pretest to posttest. On the embedded *Coal* item, students in KI gained an average of .333, while those in ST gained an average of .091. From pretest to posttest, students in KI gained an average of .237, while those in ST gained an average of .201.

Table 3.2: Mean Initial, Final, and Gain scores for Embedded Assessment and Pre/Posttest item, by knowledge integration (KI) and simulated teacher (ST) guidance.

	Guidance Type	n	Initial Score M(SD)	Final Score M(SD)	Gain Score M(SD)	KI vs. ST Gain Comparison	
						t	d
Embedded Assessment	KI	41 (groups)	2.28(.47)	2.61(.70)	.333(.521)***	3.58***	.54
	ST	42 (groups)	2.41(.69)	2.50(.73)	.091(.360)*		
Pre-	KI	77 (individuals)	2.29(.51)	2.60(.69)	.237(.814)*	.24	.05

Posttest	ST	79 (individuals)	2.37(.60)	2.51(.68)	.201(.695)*		

(* p<.05, ** p<.01, *** p<.001)
Gain scores for KI and ST conditions in both embedded assessment and pre-posttests are significantly different from 0.
Gain score on the embedded assessment significantly different between KI and ST condition.

In spite of being encouraged to revisit earlier steps, KI students who revisited did not make greater gains than those who did not revisit. Instead, revisiting was associated with smaller gains. This finding was in contrast to previous studies in WISE, where students who revisited key visualizations suggested by KI guidance were more likely to show an improvement in understanding (Ryoo & Linn, 2012). One explanation for the unexpected result in this current study may be that those who decided not to revisit were confident about their revisions, presumably based on their interpretation of the guidance prompt for this specific unit. Students who did not revisit may be benefiting from ideas suggested in the KI guidance, rather than from the links provided. Furthermore, a closer inspection of navigation patterns found that students who chose not to revisit spent nearly twice as long the first time they viewed the visualization, prior to the guidance prompt, than those who did revisit [$M_{no\ revisit}$ = 113 sec, SD = 110; $M_{revisit}$ = 211 sec, SD = 131; $t(36)$ = 2.3, $p < .05$]. This suggests that the students who chose not to revisit may have already felt confident about their interpretation of the visualization from their first pass through the unit, and they may have already gained ideas from their first visualization viewing.

Consistent with performance gains, significantly more students reported that KI guidance was helpful (70%) compared to those who reported ST guidance as helpful (29%) ($\kappa^2(1)$ = 25.66, $p < .001$). There was no significant relationship between reporting guidance as helpful and student improvement on the embedded assessment or pre to posttest. This suggests that students may not be able to accurately identify what helps them improve, which could be an interesting avenue to explore in future research.

Revision of explanations

Patterns of revision differed between the KI and ST conditions (Table 3.3). Over half of students in both conditions did not add new ideas, either normative or non-normative, to their explanation after guidance. In the KI condition, 51.8% of students had no or minimal change, corresponding to revision levels 1 and 2, 18.8% added a non-normative idea, corresponding to revision level 3, and 29.4% added a normative idea, corresponding to level 4. In the ST condition, 62.8% made no or minimal change, 30.2% added a non-normative idea, and 7.0% added a normative idea (Table 3.3).

When they did revise, significantly more students in the KI condition improved their explanation by adding at least one normative idea compared to those in the ST condition (X^2 (2, N = 83) = 30.84, $p < .001$). The majority of those in the ST condition who revised made minimal word substitutions, or added a non-normative idea, while this was not true for the KI condition.

Table 3.3: Distribution of Student Revision Characteristics by Guidance Condition

Revision Characteristics	KI	ST
no or minimal change	51.8%	62.8%

substantial change, added non-normative idea	18.8%	30.2%
substantial change, added normative idea(s)	29.4%	7.0%

Revision characteristics were distributed differently across guidance conditions (Chi sq = 30.84, p < .001).

These revision characteristic results support the KI framework by demonstrating that distinguishing ideas, whether by adding either a normative *or* a non-normative, is a valuable process. They also suggest that even guidance that does not necessarily direct students to adding normative ideas can still enable them to reconsider their ideas and may start in motion a process of distinguishing ideas that eventually leads to greater understanding (Figure 3.5). As supported by the KI framework for science learning, students can hold multiple, often conflicting ideas that become sorted out and distinguished over the process of learning. Students who are prompted by KI guidance to add even a non-normative idea may at least begin the process of considering and distinguishing ideas, and even if they are still wrong at the moment of revision, this process can eventually lead to a full, correct level of understanding.
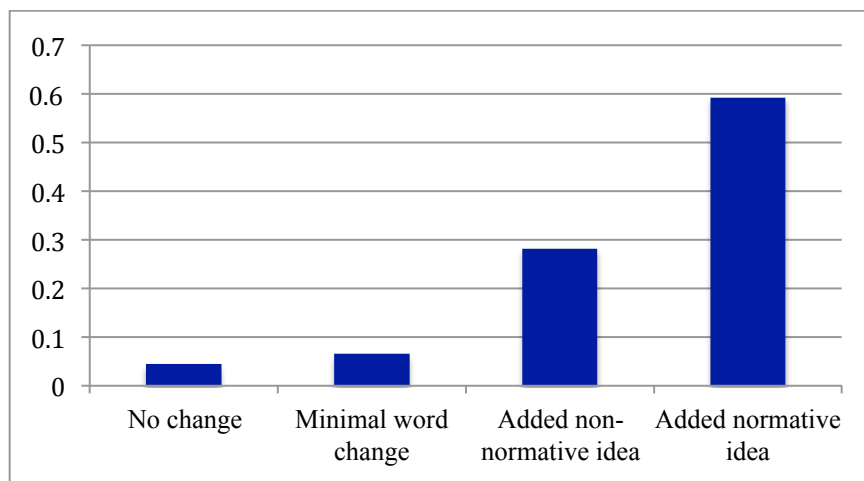
Figure 3.5: Average Gain from Pre to Post, by Revision Effort

Students at effort level 2 and 3 showed gains significantly different from 0 [* p < .05, *** p < .001], while those at effort level 1 and 2 did not show significant gains.

This argument is also supported by the finding that only those students who added a new idea to their response (normative or not) showed significant gains from the unit pretest to posttest, while those who changed only wording or made no revisions made limited pretest to posttest gains. Students who made no change or only wording changes to their responses after receiving guidance had pretest to posttest gains of .045 and .066, respectively, with neither of these gains being significantly different from zero. Students who added a non-normative idea within the Coal item revision had a significant average pretest to posttest gain of .282 (p < .05), and those who added a normative idea in their Coal revision had a significant gain of .592 (p < .001). See Table 3.4. The difference in gains between students who did not make significant revisions (levels 1 and 2, with an average pre to post gain of .060) was significantly different from those who made significant revisions (levels 3 and 4, with an average gain of .409), (p < .01).

Table 3.4: Explanation Revision Characteristics and Student Performance

Level of Revision	% of Students	Average Gain Pre- to Posttest M(SD)	Average Gain Pre- to Posttest, Comparing No/Minimal vs. Substantial Change
No change	17.0%	.045(.58)	.060[+]
Minimal change	40.4%	.066(.77)	
Substantial change: Added non-normative idea	24.6%	.282(.72)*	.409[+]
Substantial change: Added normative idea(s)	18.1%	.592(.80)***	

(* p<.05, ** p<.01, *** p<.001)
Gains from pre to posttest score were significant for students who made substantial changes, but not for those who made no/minimal changes.
[+]Mean pretest to posttest gain was significantly higher (p<.01) for students who made substantial changes (level 3/4) compared to those who made no/minimal changes (level 1 and 2).

Table 3.5: Explanation Revision Characteristics and Student Performance, by Guidance Condition

Level of Revision	% of Students		Average Gain Pre- to Posttest M(SD)		Average Gain Pre- to Posttest, Comparing No/Minimal vs. Substantial Change	
	KI	ST	KI	ST	KI	ST
No change	25.88%	8.14%	0(.52)	.167(.75)	.026	.090
Minimal change	25.88%	54.65%	.045(1.0)	.077(.62)		
Substantial change: Added non-normative idea	18.82%	30.23%	.33(.72)	.25(.74)	.444[+]	.370
Substantial change: Added normative idea(s)	29.41%	6.98%	.52(.81)***	.833(.75)*		

Discussion

This study reinforces the advantage of designing instruction, automated scoring, and guidance to promote the process of KI learning in science. Guidance based on automated scoring of student explanations, motivated approximately half the students to revise their responses by considering and adding a new scientific idea. KI guidance was more effective than ST guidance at helping students integrate normative ideas into their revised responses. In addition, students who began the process of distinguishing ideas after receiving guidance in the embedded assessment, as evidenced by adding either non-normative or normative ideas, made greater gains over the course of the unit than those who did not add any new ideas in response to guidance. This suggests that automated KI guidance can initiate a process of considering scientific ideas and strengthening student explanations, which can lead students to improve their scientific understanding over the course of the unit.

Furthermore, the results of this study demonstrate that the role of scientific inquiry is not simply to convey specific facts but to provide students an opportunity to generate and reflect upon their own ideas. From this perspective, engagement in inquiry may be beneficial even when it leads to short term difficulties. This idea ties into our counterintuitive result that students who revised by adding either non-normative or normative ideas to their embedded Coal item revision still demonstrated significant progress at posttest. Further studies may help distinguish whether this difference in overall learning based on revision type is a result of student characteristics such as overall attentiveness and effort to learn, or whether certain types of guidance that better prompt students to make a revision can help students to begin thinking about and processing ideas that benefit their learning. If the latter were true, this would be particularly good support for the importance of prompting students to grapple with new ideas themselves while learning new concepts, in hopes of long-term development and understanding.

While the current study uses automated scoring as a way of giving guidance, it is likely that if the same types of guidance, KI and ST, were delivered in person by a teacher, the outcomes of this study would still be similar. This study's success in improving student performance is due not necessarily to the automated aspect of guidance, but rather to the careful and complete design of guidance. At the same time, the results of this study do support the value of automated guidance as a tool for student learning. At KI score level 1, for example, where a typical student answer is "I don't know" or an irrelevant response, the classroom teachers generally responded with a comment such as "redo", without further ideas for the student to build upon. This type of guidance is representative of the problem of time constraints for teachers in the classroom; teachers may simply not have time to give detailed comments to every student response. This may hold particularly true in cases where there is not a clear error in student thinking for the teacher to point out, yet there still exists room for improvement and additional ideas that the student has not mentioned. With automated guidance, the time constraint is no longer a factor. The value of automated guidance is further demonstrated by the fact that student responses were successfully read and classified by the computer into KI score levels, and that guidance designed specifically for each KI

score level was beneficial to students. If automated scoring and guidance can continue to be refined such that it accurately supports student learning, it could be more helpful in certain situations than the guidance that a teacher would realistically be able to give students, given their time constraints in the classroom.

Limitations

While this study shows promising results within the context of the global climate change unit in WISE, further work needs to be done to explore whether these results are generalizable to other online learning contexts. One factor to consider may be the nature of the visualizations or other pages that the guidance prompts students to revisit; more work needs to be done to determine what types of revisits may be useful, depending on student level of understanding. For example, sending students back to a previous step in the unit that did not make sense to them the first time around may still be ineffective in the second visit. In addition, students completing the global climate change unit rarely receive a maximum score of 5 on their initial Coal response; the most common score on this question is initially a 3. It would be interesting to determine whether KI guidance remained equally effective for questions where students' initial scores average either higher or lower. Important future steps include testing the effectiveness of KI guidance in different questions from different topic areas as well. If automated KI guidance can be designed and generalized to strengthen student learning as effectively as simulated teacher guidance, the use of automated KI guidance could free classroom teachers from this task and allow them valuable time to work with individual students who need help.

Another aspect to consider in this study is the role of student engagement and motivation. Guidance itself can only be useful if students take the time to process and follow it. Thus, the motivational aspect is also important to consider when designing guidance. Previous research on feedback has found that comments that are not specific enough may lead students to find feedback frustrating or useless, and they may be more uncertain about how to respond (Fedor, 1991; Shute, 2008). According to Shute (2008), this "uncertainty and cognitive load can lead to lower levels of learning" (p. 158) and result in lower motivation and utilization of the guidance. Following these ideas, one possible explanation for why students in the KI condition showed more addition of ideas, both non-normative and normative, in comparison to the ST condition, may be because the KI guidance is sufficiently specific and directed to reduce the cognitive load of uncertainty and increase student motivation to respond. In the next chapter, I attempt to increase student motivation in revision by making it more transparent to students that automated guidance is personalized to their responses.

Chapter 4: Transparent Personalization of Automated Guidance

Introduction

While students generally benefit from using automated guidance, not all students engage with and use the guidance they get. Some students, accustomed to completing single drafts, may resist revision. Others may find the perceived challenge of revising a written artifact to require substantive effort in revisiting materials and performing novel inquiry tasks, and avoid the task altogether (Chaiklin, 2003). Still other students may simplify the task and provide superficial arguments rather than engaging critically with new ideas (Dweck & Master, 2009). Students who fear that they will be unable to reach their goals may avoid trying in an attempt to protect their beliefs about their own ability (Nussbaum & Dweck, 2008).

Teachers can increase student motivation for difficult tasks by providing help for struggling students and expressing a belief in students' ability to reach academic standards (Cohen, Steele, & Ross, 1999). Encouragement from teachers can motivate students of all prior knowledge levels to revise (Beason, 1993). Research suggests that low performing students in particular are more likely to disregard guidance on explanations because they feel like they cannot succeed (Shute, 2008). Students who feel that guidance is not appropriate for their skills or ideas are unlikely to try their hardest at a task (Shute, 2008). Low performing students who receive guidance from a teacher may be inherently assured that the guidance is at a level they can achieve, because the teacher knows them. However, automated guidance that comes from a computer does not have this same assurance. Students may perceive guidance that comes from a computer as generic and unresponsive, especially in comparison to guidance from a teacher. For this reason, low performing students in particular may benefit from the knowledge that guidance coming from a computer has been personalized for their current score level.

In this study, we compared transparent guidance that pointed out the personalized nature of automated guidance to typical automated guidance that did not emphasize this alignment. Both conditions utilized the same conceptual guidance, with the only difference being that the transparent condition included features that assured students of the personalized nature of the guidance. We hypothesize that the transparent condition may reassure students that guidance is at an appropriate and attainable level, so it may be particularly beneficial for students who start off at a lower performance level. Transparency can also communicate to students that effortful revisions will result in higher scores from the automated system. This information may promote agency, the belief that student actions will result in meaningful outcomes and lead to improved effort (Bandura, 1989; Basharina, 2013; Kramsch, A'Ness & Lam, 2000). By supporting student agency in our guidance we hypothesized that students would respond with more effortful revision. Study 1 examines the impact of transparent personalization on student revisions and science learning. Research questions include:

1. Does transparent personalization of automated guidance compared to typical guidance improve students' overall performance and learning gains in an inquiry science unit?
2. Do low prior knowledge students particularly benefit from transparent personalization of automated guidance?

Methods

Participants

Participants included 482 sixth grade science class students taught by four teachers in three different public schools. Due to absences, only 323 students completed the full set of items analyzed in this study, which includes the automated guidance item in the unit as well as the corresponding item in pretest and posttest. Students within each class period were randomly assigned to either the transparent or typical adaptive guidance conditions. Students did not self-report demographic information, but we include overall demographic information for each school (Table 1). Gender information was not available for a large percent of students in this study.

Table 4.1
Student Demographics by School

	School 1	School 2	School 3
Free/Reduced Price Lunch	93%	24%	26%
Caucasian	6%	35%	53%
Hispanic	69%	17%	27%
Asian	8%	31%	6%
African American	12%	4%	3%
Multiracial	0%	10%	8%
English not primary language at home	74%	36%	21%

Students completed the thermodynamics unit in pairs, with two students sharing one computer and working collaboratively throughout the unit. This aligns with the knowledge integration framework for science learning, as students who work together can introduce each other to new concepts and critique each other's ideas (Linn, Clark, & Slotta, 2003). Two versions of the thermodynamics unit were made for this experiment, and student pairs in each class period were assigned randomly to one of the two conditions.

Guidance design
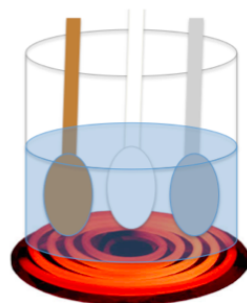
A metal spoon, a wooden spoon, and a plastic spoon are placed in hot water.

After 15 seconds, which spoon will feel the hottest and WHY? Be sure to explain your ideas below.

Metal will be hottest because it is the best conductor

Figure 4.1. Spoons question with initial student response

In the transparent condition we revised both the guidance and the surrounding instruction for the *Spoons* item (Figure 4.1) to make personalization within WISE more transparent. We increased transparency by: (a) explaining to students how automated scoring of their responses works, (b) integrating student names into automated guidance, and (c) explicitly indicating individual progress on revision. Specifically, prior to instruction students were presented with an informational page that described how WISE automatically scored their responses (Figure 4.2). Students clicked through an animation that explained the c-raterML™ process in age-appropriate terms. The animation showed that when the student submits an answer, the computer reads the answer, and the computer compares their answer to that of thousands of other 6th grade students around the country before assigning them guidance. The automated process was explained with a personified computer avatar.

Automated guidance adapted for the transparent condition also incorporated students' names into the text. Figure 4.3 shows an example of guidance tailored to a specific pair of students. In addition, the personified computer avatar, introduced in the prior step, was included to help students recall how the automated scoring process is conducted. For the second round of guidance, the transparent condition included student names and also a comment about student progress. Progress was measured by automatically comparing the score of the initial response to the first revision. If WISE detected that the students had not improved their KI score after the first round of guidance, the second round of guidance began with "(student name), the computer thinks you have not improved your answer. You need to add information." If the student's KI score had improved after the first round of revision, they were told "Good work (student name)! The computer thinks you added a correct scientific idea and explained your reasoning. Now consider this." After this header, students were presented with conceptual KI guidance appropriate to their score level.

Students in the typical condition received two rounds of adaptive KI guidance based on c-raterML™'s scoring of the student's response (Appendix 1). Typical guidance did not include a description of the scoring process with computer avatar, personalized introduction with student names, or acknowledgement of students' individual progress in revision.
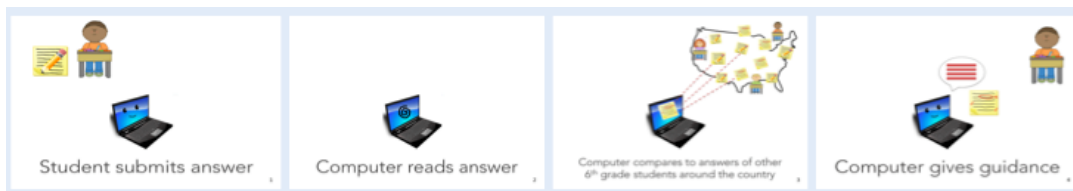

Figure 4.2. Transparent condition informational pages describing how automated scoring works
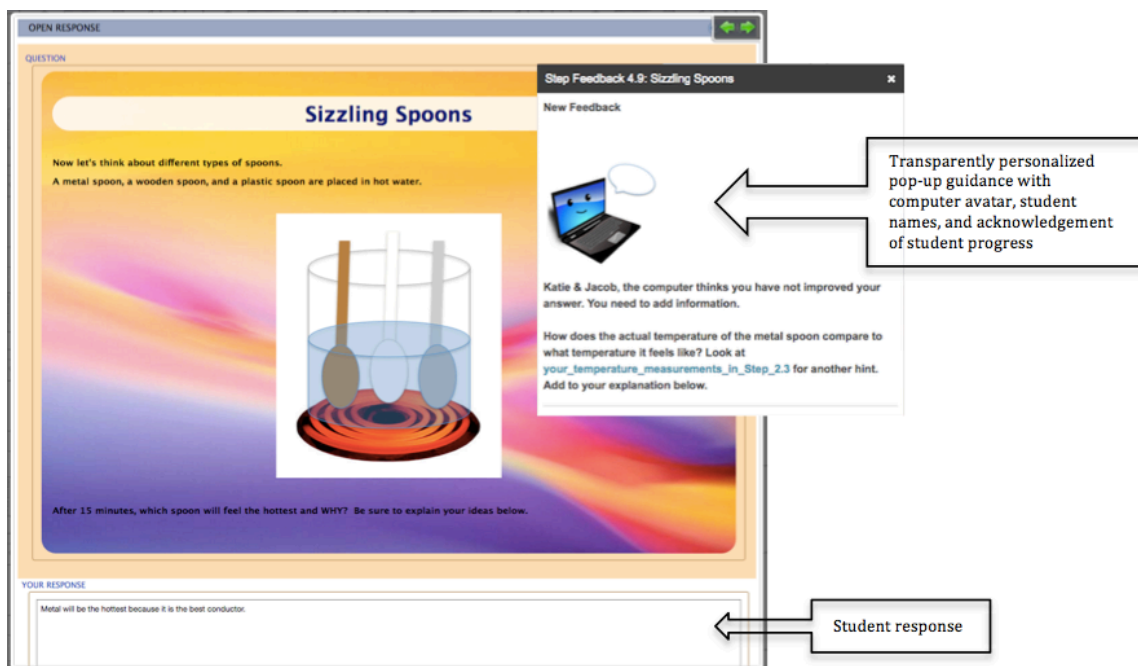
Figure 4.3. Transparent condition automated guidance pop-up

Assessment Items

Student performance was analyzed on the embedded *Spoons* item and the pre/posttest item *Cups* (Table 2.2). Throughout this study, analysis of the embedded *Spoons* items and revision process applies to student pairs, because students completed the thermodynamics unit in pairs. Analysis of pretest and posttest performance was by individuals. Students who did not complete the *Spoons* step, and thus were not exposed to the experimental conditions, were dropped from analysis. For clarity, throughout the results, the within-unit *Spoons* item will be referred to as *Spoons-embedded* and the pretest-posttest *Cups* item will be referred to as *Cups-prepost*.

Results

Participation in revision

106 student pairs in the transparent condition and 142 student pairs in the typical condition wrote an initial response to *Spoons-embedded*. While over 90% of students in both conditions made one revision, only 59% (63 students) in the transparent and 47% (68 students) in the typical condition submitted a second revision. A chi square test of independence, performed to examine the relationship between guidance condition and submission of a second revision, shows a trend towards significance [$\chi^2(1)=3.27$, p=.07, V=.02]. This suggests that students in the transparent condition may be more likely to submit a second revision than those in the typical adaptive condition, although further study is needed. In addition, students across both conditions who completed both rounds of revision showed comparable learning gains to those who did not complete both rounds of revision.

Students' revision strategies were also examined to determine whether those in the transparent condition were more likely to add new ideas or revisit the suggested page mentioned in guidance. There were no differences in either revision strategy used in each

condition. Also, across both conditions, students who added new ideas or revisited the page suggested by guidance did not perform better than those who did not adopt these strategies. This suggests the potential need for more specific guidance on revision strategies. We investigate this question further in Study 2.

Transparent personalization and student revisions

Students in the transparent condition received one additional feature prior to writing their initial response to the *Spoons-embedded* item, which was the page explaining how automated guidance works. To examine whether the transparent instruction had any immediate impact on students' initial *Spoons-embedded* response, we ran a t-test on initial score, by condition. Our *Spoons-embedded* analysis includes responses from all student pairs who wrote at least an initial response and one revision to *Spoons-embedded*. All students who completed the *Spoons-embedded* step were included in the *Spoons-embedded* analysis, even if they did not complete both the pretest and posttest. This resulted in a sample size of 102 pairs in the transparent personalization condition and 139 pairs in the typical adaptive condition. A t-test of initial score by condition finds that students in the transparent condition had a significantly higher initial KI score on *Spoons* compared to students in the typical adaptive condition [M(Transparent)=3.50, SD=.79; M(Typical)=3.24, SD=.86; t(239)=2.52, p<.05, d=.33]. While students in the two conditions showed significantly different initial scores on the *Spoons* item embedded within the unit, they did not show a significant difference in KI score at pretest, suggesting the difference in initial scores was due to the additional features that the transparent condition received prior to the *Spoons-embedded* step. Table 4.4 shows the average score for each condition on the initial and revised *Spoons-embedded* response, both of which demonstrate an advantage for the transparent condition.

Table 4,4
Mean initial and revised Spoons-embedded scores, by condition

	Initial *Spoons-embedded*	Revised *Spoons-embedded*
Transparent Personalized	3.50(.79)	3.79(.82)
Typical Adaptive	3.24(.86)	3.41(.85)
Comparison between conditions (t-test)	t(239) = 2.52* d = .33 p<.05	t(239) = 3.44 d = .45 p<.001

To investigate the effect of the transparently personalized condition on students' writing revisions we ran an ANCOVA on revised *Spoons-embedded* scores, controlling for initial *Spoons-embedded* scores. Revised *Spoons-embedded* scores were defined as students' last *Spoons-embedded* submission, whether it was their initial response, first revision, or second revision. Initial scores were significantly related to revised scores [F(1,238) = 454.59, p<.001, η^2 = .66], and a main effect for condition emerged [F(1,238)

= 5.56, p<.05, η^2 = .02]. This suggests that the transparent condition was more effective in helping students revise their responses.

To examine whether prior knowledge moderated the effect of condition, we categorized students as having low or high prior knowledge on their initial response. Those who did not include a scientifically valid idea in their initial response were categorized as "low", and those who included at least one correct idea were "high". 38 pairs had a low initial score, and 203 pairs had a high initial score. An ANOVA was run on revised *Spoons-embedded* scores with prior knowledge and condition as predictors. The analysis revealed a significant effect of prior knowledge classification [F(1,237) = 122.48, p<.001, η^2 = .34], a significant main effect of condition [F(1,237) = 9.68, p<.01, η^2 = .04], and no significant interaction of condition and prior knowledge[F(1,237) = 2.29, p>.05, η^2 = .01]. This suggests that the transparent personalization condition was not differentially effective depending on whether or not students' initial responses included a valid scientific idea.

Transparent personalization and pretest to posttest learning gains

Students in both conditions showed similar performance on the pretest. Across both conditions, students had a significantly higher KI score on the posttest than the pretest [M(Pretest) = 2.76, SD = .70, M(Posttest) = 3.38, SD=.65, t(351) = 13.56, p<.001, d = .91].

To determine if experimental condition had a significant impact on learning we performed an ANCOVA on posttest scores, controlling for pretest scores, with condition as a predictor. While, pretest scores were significantly related to posttest scores [F(1,352) = 14.40, p<.001, η^2 = .04], no effect of experimental condition emerged [F(1,352) = 1.29, p>.05, η^2 =.004].

To determine whether prior knowledge had a moderating effect upon condition we again categorized students based upon having high or low relevant prior knowledge at pretest. Students who did not express a relevant, valid scientific idea on the *Cups-prepost* item at pretest were coded as "low", and those who expressed at least one valid idea were coded as "high". 133 students were classified as low prior knowledge, and 219 as high prior knowledge. We then performed an ANOVA on posttest scores with both prior knowledge and experimental condition as predictors. This analysis revealed no significant main effect of prior knowledge classification [F(1,352) = 3.58, p>.05, $\eta2$ = .01], no significant main effect for condition [F(1,352) = 2.99, p>.05, $\eta2$ = .008], and a significant interaction of condition and prior knowledge [F(1,352) = 5.00, p<.05, $\eta2$ = .014]. These results suggest that the effectiveness of the transparent condition was different depending on whether students started out with relevant knowledge at pretest.

To further investigate the moderating effect of prior knowledge on condition, we compared conditions at low and high prior knowledge levels with Bonferroni-adjusted alpha levels for multiple comparisons. Low prior knowledge students in the transparent condition had a significantly higher posttest score than their counterparts in the typical condition [M(Transparent)=3.43, SD=.63, M(Typical)=3.15, SD=.70, t(131) = 2.40, p < .05, d = .42]. High prior knowledge students across both conditions had similar posttest scores (Table 4.5).

Table 4.5
Posttest score by condition for low versus high prior knowledge students

	Low prior knowledge	High Prior Knowledge
Transparent Personalized M(SD)	3.43 (.63)	3.41 (.61)
Typical Adaptive M(SD)	3.15 (.70)	3.45 (.63)
Comparison between conditions (t-test)	t(131) = 2.40* d = .42 p<.05	t(217) = .43 d = .06 p>.05

Discussion

Study 2 aimed to determine whether transparent personalization of automated guidance improves students' learning gains. Students in both conditions made significant improvements on *Spoons-embedded* and *Cups-prepost* after receiving guidance. Results show that students in the transparent condition had a higher score on their initial *Spoons-embedded* explanation than students in the typical adaptive condition. This difference can be attributed to the informational page before *Spoons-embedded* that explained to students how WISE automatically scores students' responses and assigns guidance personalized to their responses. Students may have taken this page as a signal that the computer would be evaluating and providing guidance for their particular response. Results also showed that students in the transparent condition had higher *Cups-posttest* scores, controlling for *Cups-pretest* scores, and higher revised *Spoons-embedded* scores, controlling for initial *Spoons-embedded* scores. No significant difference in scores was found between students who completed two rounds of revisions and those who only completed one round, suggesting that the informational page and the use of student names in round 1 of guidance likely influence learning more than the indicator of progress in round 2 of guidance.

Study 2 also investigated whether transparency features were particularly beneficial for students who begin with a low initial score. Transparent guidance compared to typical guidance led to a higher level of understanding at posttest for students who began with low prior knowledge. The transparent condition emphasized personalization of guidance to increase students' feelings of agency during the revision task, by reinforcing the link between the student's writing and revision actions and the computer's response. Seeing their names and acknowledgement of how well they had changed their response from revision 1 to revision 2 also gave concrete evidence that the computer was adapting to their responses. Students in the transparent condition who started out with a low initial score had higher *Cups-posttest* scores, but not higher revised *Spoons-embedded* scores. A possible reason for this may be that the transparent condition may have set into motion the process of reconsidering ideas for these students, but they did not fully integrate them until the pretest. Another reason for this might be that the

transparent features increased student engagement and attention to the unit overall, not only on the automated guidance step, and that this engagement throughout the unit led to higher scores by posttest.

The effect of transparent guidance on low prior knowledge students demonstrates that automated guidance can motivate revision, consistent with the impact of effective teachers who consider relevant information about individual students and personalize instruction to increase motivation and learning (Shepard, 2000). Teachers can recognize each student as an individual, and evaluate student's progress on a task, rather than only the final state of their work. Likewise, automated guidance is most effective when it uses students' names and acknowledges the progress they have made in each refinement to their writing, rather than only assessing the final state of their work. These results suggest that online guidance can capture some of the elements of effective guidance used by teachers.

In Study 3 we expand upon these findings by focusing student effort not only on revising in general, but specifically on revision strategies that have been found effective to improve science writing.

Chapter 5: Supporting Specific Revision Strategies to Improve Conceptual Understanding

Study 2 shows that transparent KI guidance appears to promote more successful revisions and science learning, particularly among low prior knowledge students. In this study we examine whether presenting guidance along with a suggestion of either revisiting evidence or planning writing revisions is more effective for knowledge integration. Revisiting previous evidence (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Chiu & Linn, 2012; Gerard et al., 2015) and planning writing revisions (Rivard, 1994) are both well-documented strategies for improving scientific explanations through revision. Both of these strategies can be beneficial for science learning because they allow students to consider and sort through new and old ideas. Whether it is more beneficial to focus on revisiting or planning may hinge on whether students need more ideas (remedied by a focus on revisiting) or need to strengthen their writing to make their ideas more coherent (remedied by a focus on planning).

The two conditions in this study allow us to examine the impact of revisit focused versus planning focused guidance on student revision strategies and understanding of thermodynamics concepts. We hypothesize that students who begin with low prior knowledge will benefit more from revisit guidance, while students who begin with a high prior knowledge will benefit from writing guidance. We predict that students who start off with low prior knowledge may need to interact with the dynamic model to gather new ideas and improve their understanding, while those with high prior knowledge may already have a grasp of the scientific content and instead benefit more from planning how to link and connect their ideas in writing.

Method

Participants

A total of 551 students from 11 teachers' sixth grade science classes in five different public schools participated in this study. A subset of students self-reported gender, including 128 females and 144 males. Students did not self-report demographic information, but school demographics are reported (Table 5.1). School 1 was distinctly different in demographics from Schools 2-5, with the student population being comprised of more low-income students and more English language learners. Some teachers participated in both Study 1 and Study 2, however, the studies were conducted in different school years so no students participated in both studies.

Table 5.1
Student Demographics by School

	School 1	School 2	School 3	School 4	School 5
Free/Reduced Price Lunch	93%	34%	24%	26%	3%
Caucasian	6%	43%	35%	53%	74%
Hispanic	69%	32%	17%	27%	7%
Asian	8%	20%	31%	6%	10%

African American	12%	2%	4%	3%	0%
Multiracial	0%	0%	10%	8%	6%
English not primary language at home	74%	41%	36%	21%	15%

Materials and procedure

We implemented this study in the same Thermodynamics unit used in Study 1. Students completed the pre and posttest individually, and the Thermodynamics unit in pairs. Two versions of the Thermodynamics unit were created, and students within each class period were assigned randomly to one of the two conditions.

Study 2 retained aspects of the transparent personalization guidance tested in Study 1. All students were shown an instructional page informing them about how automated guidance works, and guidance addressed students by name. Due to the more extensive nature of guidance in Study 2, we chose not to include the pop-up guidance format used in Study 1, in favor of directing (branching) students to new web pages based upon both their randomly assigned experimental condition and initial response score. This procedure allowed us to apply more elaborate guidance, which included prompts for student responses within the guidance pages. However, because the current software limited this branching procedure to a single iteration, students only received one round of guidance. Therefore we did not investigate students' progress over multiple revisions, as we did in Study 1.

On the guidance page students in both conditions were presented KI guidance appropriate to their score level, referencing students by name and including a suggested link to revisit for additional information. In the revisit condition students were prompted again to revisit the target visualization and then asked to respond to a multiple choice item to report and justify their chosen behavior (Table 5.2). In the planning condition students were prompted to make careful revisions of their response and then respond to a multiple choice item to report their plan for revision (Table 5.2). By drawing student attention to a revision strategy, but still giving them a choice as to which specific action they were going to take regarding it, we attempt to increase student motivation to perform the specific revision strategy. An example of the guidance page in WISE that students are directed to with automated guidance and either revisit or planning instructions is shown in Figure 5.1.

Table 5.2
Randomly assigned focus for using guidance

	Revisit Focus Prompt	**Planning Focus Prompt**
Statement reinforcing focus of revisit or planning	"Students who revisit previous steps and carefully gather evidence learn the most and perform better by the end of the WISE unit."	"Students who make careful revisions learn the most and perform better by the end of the WISE unit."
Multiple choice question to promote student agency in revisit or planning strategy	Did you revisit *the finger/bowl activity* suggested above? Pick all that apply. Yes, because I wanted more information	How do you plan to change your answer? Pick all that apply.

	Yes, because the computer suggested it No, I looked back at a different step than the one suggested No, I remember the information in that step already	I plan to add another scientific idea I plan to add evidence to support my idea I plan to remove an incorrect idea I don't plan to change my answer
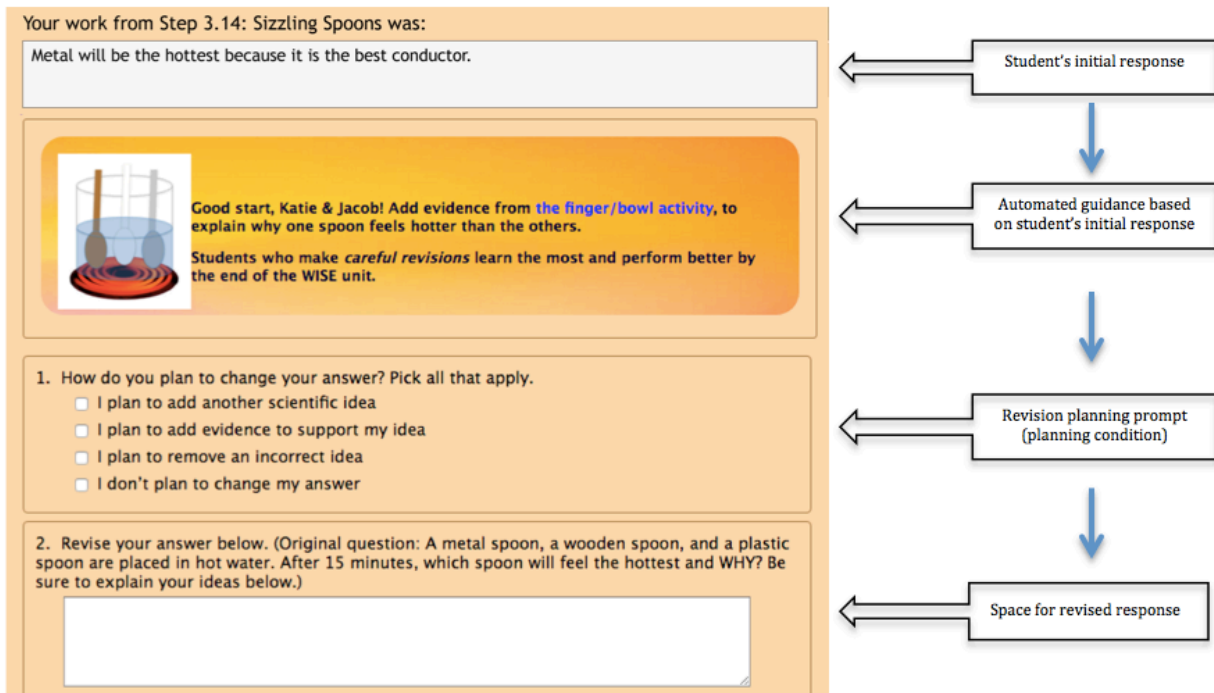
Figure 5.1. Planning condition automated guidance and revision page

Data sources and analysis

Similar to Study 1, researchers performed classroom observations in every school. We scored students' initial and revised *Spoons-embedded* explanations, and their pretest and posttest explanations on the *Cups-prepost* item, which tested similar thermodynamics concepts. To examine students' effort in the targeted revision strategies we analyzed the log files that show whether students revisited or not, and analyzed the changes students made from initial to revised *Spoons-embedded* explanation using the revision characteristics rubric (Table 3).

Results

Revision Strategies

Students' revision strategies aligned with the guidance condition. Students in the revisit condition were 27% more likely to revisit the step suggested by the guidance than those in the planning condition [revisit: 61%; planning: 34%; $X^2(1, N = 465) = 31.82$, $p <$

.001, d=.54]. Conversely, students in the planning condition were 14% more likely to make substantial writing revisions to their responses (i.e., add a new normative or non-normative idea) than students in the revisit condition [revisit: 39%; planning: 53%; X^2(3, N=464) = 14.86, p<.01, d=.36] (Figure 5.2). These results suggest that students were attentive to task demands and engaged with the revision process. Furthermore, the finding that the planning group added more non-normative ideas suggests that a combined intervention promoting both revisiting evidence and planning writing changes could potentially be an effective future intervention to test.
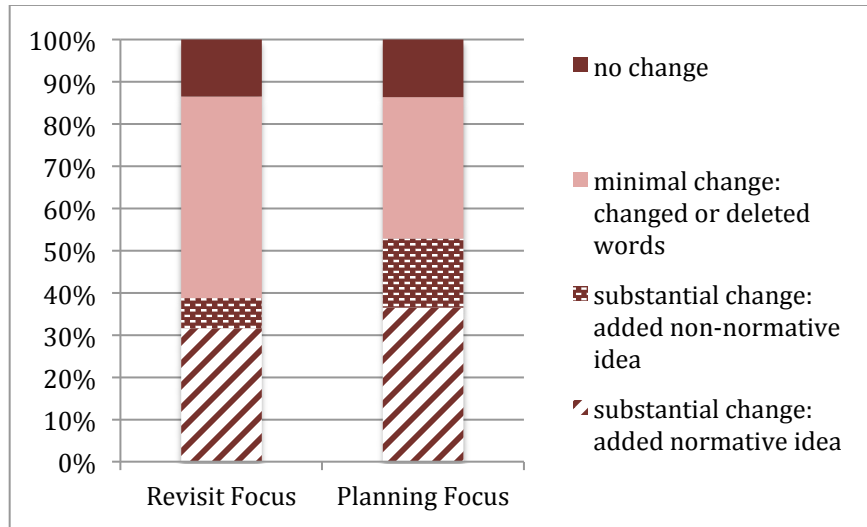
Figure 5.2. Revision Characteristics by Condition

Students across both conditions did not show significant gains from initial to revised *Spoons-embedded* score. In addition, neither condition demonstrated an advantage on the *Spoons-embedded* item.

Learning Outcomes

Overall, students in both conditions showed significant gains on from pretest to posttest [M(Pretest) = 2.88 KI points, SD = .67; M(Posttest) = 3.45 KI points, SD = .86; t(550) = 14.10, p<.001, d=.74] (Table 5.3).

Table 5.3
Mean Pretest and Posttest Scores, by Condition

	Pretest	Posttest
Revisiting	2.84(.66)	3.41(.82)
Planning	2.93(.68)	3.49(.90)

To investigate our hypothesis that students with lower prior knowledge would benefit more from revisiting to add ideas we performed an ANCOVA to determine if there was an interaction effect between prior knowledge and condition following the

procedure from Study 1. No significant interaction was found, suggesting that intervention conditions did not impact gains by prior knowledge.

Given that Study 2 involved five schools serving very different student populations, we chose to investigate the interaction between school context and condition. Informed by prior research, we hypothesize that student willingness to revisit prior material or plan writing changes may be dependent on teacher approach and classroom culture. Studies in WISE have found that teaching contexts, including teacher beliefs about inquiry teaching practices, impact student knowledge integration outcomes (Lee, Linn, Varma, & Liu, 2010). Students may be reluctant to take the time to revisit or plan writing changes for their revisions because they want to keep progressing forward, so classroom culture may make a difference in how students engage with an autonomous inquiry-learning unit such as WISE.

To investigate the consistency of the conditions across schools (Table 5.4), we performed ANCOVA with pretest score as the covariate. An initial test of assumptions demonstrated no significant interaction between the covariate, school, and condition [$F(9, 531) = 1.1$, $p > .1$)], indicating that the homogeneity of regression slopes assumption was not violated. The ANCOVA revealed a significant association between the covariate pretest score and posttest score [$F(1, 540) = 36.57$, $p < .001$, $\eta^2 = .06$], no significant main effect for condition [$F(1, 540) = 0.54$, $p > .1$, $\eta^2 = .00$], a significant main effect for school [$F(4, 540) = 12.69$, $p < .001$, $\eta^2 = .09$], indicating that schools differ on posttest performance. We also found a significant interaction of school and condition [$F(4, 540) = 3.35$, $p < .05$, $\eta^2 = .02$], controlling for pretest, indicating that the relative effectiveness of each condition differed by school.

To investigate this interaction of treatment condition and school further, we tested simple effects of condition, for each school, on adjusted posttest scores (measured at mean pretest level), by performing t-tests with Bonferroni-adjusted alpha levels for multiple comparisons. We found a significant advantage for the planning condition at School 1 [$t(75) = 3.24$, $p < .01$]. For all other schools, no significant differences emerged between conditions [all p values > .1]. A similar ANCOVA analysis performed on *Spoons-embedded* revision did not reveal any significant main effects or interactions [all p values > .1].

Table 5.4
Mean pretest and posttest scores, by school and condition

		School 1	School 2	School 3	School 4	School 5
Revisit M(SD)	Pretest	2.70(.72)	2.61(.69)	2.92(.51)	2.83(.71)	3.10(.57)
	Posttest	3.59(.95)	3.14(.74)	3.51(.77)	3.38(.78)	3.55(.84)
Planning M(SD)	Pretest	2.77(.14)	2.75(.59)	3.04(.66)	2.93(.68)	3.10(.63)
	Posttest	4.19(.15)	3.09(.79)	3.48(.80)	3.11(.97)	3.65(.78)

To examine whether School 1 may have performed differently due to the high population of English language learners, we examined whether English language learners across all schools benefited more from the planning condition than the revisit condition.

We defined English language learners as students who selected the survey response "At home, my parents mostly or only speak a language other than English." We did not find that the planning condition was more effective for English language learners across all schools.

Gender Analysis

Since some previous research has shown gender differences in student performance on writing tasks, we examined student performance on written assessment items, as well as their likelihood to make written revisions to *Spoons-embedded* by gender. No significant differences were found between females and males on *Spoons-embedded* or *Cups-prepost* KI scores. There was also no difference between genders for written revision characteristics on *Spoons-embedded.* These findings align with previous research that has not found gender differences in performance on KI items (Liu, Lee, & Linn, 2011).

Word Count Analysis

Lengths of student responses were analyzed to determine whether word count on *Spoons-embedded* responses correlated with KI score. Word Count on revised *Spoons-embedded* responses did not significantly predict students' gain scores from pretest to posttest. Word count on *Spoons-embedded* accounted for only 10.22% of the variance in pretest to posttest gain scores. [$F(1,550) = .69$, $p>.05$, $\eta^2=.1022$].

Student Examples

Analysis of student KI scores finds that students in both conditions were equally likely to make pretest to posttest gains. In this section we present case studies of a student pair in each guidance condition that illustrate the effect of our guidance. To find student responses that illustrate the revision patterns, we searched for responses that were of sufficient length, more than 10 words, began with a typical initial KI score of 3, and finished with a typical gain score of 1. Of the groups that met our criteria, we selected two examples that best illustrate the intended use of the guidance in each condition.

Table 5.5
Examples of initial and revised student responses after receiving planning condition KI guidance

Planning Condition				
	Spoons-embedded Initial Response	KI Score	*Spoons-embedded* Revised Response	KI Score
Pair: Student A and Student B	After 15 minutes metal spoon will feel the hottest because it is a good conductor	3	After 15 minutes metal spoon will feel the hottest because it is a good conductor and it transfers heat faster than all of the materials do	4
	Cups-prepost	KI	*Cups-prepost*	KI

	Pretest Response	Score	Posttest Response	Score
Student A	Because the molecules will have space to move	2	The metal cup will feel hottest because it is a good conductor and will transfer heat quicker	3
Student B	the metal cup because it absorbs the heat	3	The metal cup will feel the hottest because it is the best conductor and it will transfer heat to roger's hand faster than the other cups	5

Student A and B were paired together to work on the Thermodynamics unit. Table 5.5 shows their initial and revised responses after receiving planning guidance, as well as their individual pretest and posttest scores. After writing their initial response and receiving guidance, students selected the option "I plan to add evidence supporting my idea" and added an idea about how metal "transfers heat faster than all of the materials do." This is an example of a substantial revision, as the students both elaborate on the rate of heat transfer and add a comparison between the metal and other spoon materials. On the posttest, both students carried over the idea of heat transfer rate, with Student A stating that metal "will transfer heat faster" and Student B stating that "it will transfer heat to roger's hand faster." Several groups in the planning condition followed this similar pattern, where students added an idea to the revised response and carried this idea over to their posttest response as well. This pattern of student responses supports the idea that the planning condition may be successfully encouraging students to integrate new ideas in their revised responses, and students are deeply integrating this idea into their knowledge that is then reflected on the posttest.

Table 5.6 shows a pair of students in the revisit condition. Their initial answer to *Spoons-embedded* correctly identifies that the metal spoon will be the hottest, but does not give an explanation why. After receiving guidance, the students were asked "Did you revisit the finger/bowl activity suggested above?". The students selected the answer "Yes, because I wanted more information" and then revisited the simulation. This simulation allows students to experiment with different materials and visualize heat flowing through them at different rates. In their revised response to *Spoons*, the students correctly include the reasoning that "metal conducts heat the fastest." By the posttest, one of the students carried over the idea that metal heats up the fastest. This pattern of student responses supports the idea that the revisit condition may be successfully encouraging students to revisit simulations and incorporate ideas from them into their *Spoons-embedded* revisions, and that this knowledge can also be carried over to their long-term understanding as reflected in the posttest.

Table 5.6
Examples of initial and revised student responses after receiving revisit condition KI guidance

Revisit Condition				
	Spoons-embedded Initial Response	KI Score	*Spoons-embedded* Revised Response	KI Score

Pair: Student A and Student B	The metal spoon will be the hottest and the wood spoon will be the second hottest and the plastic will be the least hot.	3	After 15 minutes the metal spoon will be the hottest. It will feel the hottest because metal conducts heat the fastest.	4
	Cups-prepost Pretest Response	KI Score	*Cups-prepost* Posttest Response	KI Score
Student A	I think the plastic cup would feel the hottest because plastic materials are more easily melted	2	I think that the plastic or the metal cup will be the hottest. It depends on the temperature of the cup before it was put in the sink	2
Student B	I think the metal cup will fell the hottest because it can produce heat	2	The metal cup will be the hottest because in some recent projects I learned if you put metal bar and chocolate it melted the fastest	3

Discussion

We found that guidance directing students to either revisit scientific models or plan substantive writing changes improved overall learning outcomes from pretest to posttest. Analysis of student actions demonstrates that students in the revisit focus condition were more likely to revisit the step suggested by KI guidance, and students in the planning condition were more likely to make significant revisions to their initial answer, suggesting that both conditions were successful in motivating students to take relevant actions to improve their initial responses. This effort was translated into more coherent, integrated understanding of thermodynamics for students in both conditions.

Embedded item
On the embedded item, we did not find a significant gain from initial to revised score. This may have occurred because some students did not rewrite all their initial ideas on the revised response page. In contrast to Study 1, where students' initial response remained in the answer box and students simply added/removed ideas from their initial response, the technological design of Study 2 required students to navigate to a different textbox and rewrite their answer. Several students wrote only new ideas in the revised response textbox (rather than building on their initial response), which may account for the lack of significant improvement in score from initial to revised *Spoons-embedded* response.

Prior knowledge
Our hypothesis that revisiting would be more effective for low prior knowledge students who need additional ideas was not confirmed. The two conditions improved student learning gains from pretest to posttest equally.

42

School effect

We found an interaction between school and condition. In School 1, the planning condition made significantly larger gains on the embedded item than students in the revisiting condition. This advantage was not limited to just the embedded item, but also carried over to the posttest. Although School 1 has a large population of English language learners, further analysis of the effects on English language learners across schools did not support a benefit of the planning condition for this population. Our measure of English language learners, which does not completely distinguish between bilingual students and those who primarily speak English at home, is a limitation to fully understanding the effect of being an English language learner.

The advantage of planning for School 1 may reflect the effect of the classroom teacher and the overall school climate. From our classroom observations, we noted that the teacher in School 1, in response to her students' needs, specifically uses strategies such as providing sentence starters and guiding questions to prompt her students on what sentences to add or change. Through her teaching strategy, she emphasizes breaking down the language demands of the task. One reason that planning guidance may have led to a long-term effect in this school is that it resonated with the teacher's approach.

To further explore the role of classroom context in School 1, one of the teachers was asked for comments on which strategy she typically employs with her students. When asked about which kinds of revision strategies she encourages in her classroom, she notes:

> I did request my students to answer in complete sentences and even start a sentence with a capital letter. Some of the students did have scientific ideas, but the majority of them needed constant encouragement and guidance […] I created posters and displayed them in my room. Instead of going back and re-visiting the previous steps, they could look around the room and find the information needed for their essays. (Email communication, May 16, 2016)

The additional focus that this teacher puts on proper writing, even in a science classroom, may have contributed to greater success for her students in the planning writing condition. This teacher also frequently used sentence starters in multiple WISE units throughout the year (either by printing out example response paragraphs with content missing, or by entering in sentence starters within WISE response boxes) to assist her students in constructing full written responses to open-response questions.

Revisiting patterns

In contrast to prior studies, students' actual revisiting patterns did not show correlations with their score gains (e.g., Ryoo & Linn, 2012). Since students were directed to revisit a step that occurred only 2 or 3 steps before the Spoons step, it is possible that students believed they remembered the information or had already acquired sufficient knowledge from this visualization, consistent with some other revisiting studies (Svihla, et al., in press). In studies where the visualization is more complex, revisiting with specific goals in mind may be more beneficial. For example, in simulations that depict complex systems, with many variables that generate emergent phenomena, students may not notice behaviors on their first viewing. For topics in which students are guided to revisit complex visualizations with specific questions in mind (such as for

photosythesis in the Ryoo & Linn, 2012 study) students may benefit more than when the simulation is simple. Future studies may investigate how the role of revisiting is impacted by the complexity of the revisited materials.

Chapter 6: Conclusion

This dissertation explores designs of automated guidance to support knowledge integration and increase students' agency and motivation in making revisions in response to automated guidance. Prior research shows that automated guidance can be useful in providing students with immediate, adaptive feedback (Linn & Eylon, 2011). However, students often do not fully utilize automated guidance that is provided from a computer.

While automated guidance can help students make revisions on science explanations, not all students make effortful revisions. Existing research demonstrates that students are more likely to be motivated when they have a strong feeling of agency, meaning that they feel their actions will impact a meaningful response (Basharina, 2013, Zuckerman, Porac, Lathin, & Deci, 1978). The findings from this dissertation suggest ways to design automated guidance to improve student agency in making revisions, and subsequently to improve student learning in science writing tasks. The following research questions were answered through design studies conducted in the WISE Global Climate Change and Thermodynamics units.

The research studies in this study investigated the following research questions:
1. How effective is automated knowledge integration guidance from a computer in comparison to typical guidance given by a teacher?
2. How can we best design automated guidance that promotes deep understanding of science along with a propensity to iteratively refine one's understanding?
3. What types of revision strategies, or student-initiated activities such as revisiting scientific models, lead to improved understanding of science?

Study 1 investigates the first research question. Results show that automated KI guidance presented from a computer was more effective in encouraging students to add a new normative idea to their response than automated simulated teacher guidance presented from the computer. Also, students who added a new idea at the time of revision, whether scientifically valid or not, made greater gains by posttest than students who did not add any idea in response to guidance. This suggests that automated KI guidance can be effective in initiating a process of students considering new ideas, setting in motion a revision process that when students engage effortfully with, improves their performance by the end of the unit.

In Study 2, we increased the transparency and personalization of the automated guidance. We found evidence that these transparent personalization features motivated low prior knowledge students to expend more effort on their revisions, and, as a result to develop a more robust and lasting understanding of thermodynamics concepts than low prior knowledge students who received the typical adaptive guidance. This supports the view that students, particularly low prior knowledge students, benefit from insights into how computers can generate guidance and suggests that students may not have a full view of the computer's sophistication. Strengthening the computer guidance for low prior knowledge students can allow the teacher to work with and spend more time with fewer students who require additional help.

In Study 3, we found that across schools, both revisiting and planning strategies were equally effective in motivating revision and improving students' understanding of thermodynamics concepts from pretest to posttest. Both approaches aimed to strengthen student agency by offering students specific actions and supporting them to act on the suggestions. Revising can be a daunting process that requires students to incorporate many complex actions such as reconsidering previous ideas, collecting new ideas, distinguishing between ideas, and integrating the new ideas with the old ideas. By guiding students to plan their revision and supporting them to successfully locate relevant evidence, we may encourage students to engage in revision in the future. As previous research has found, students are more likely to persist in challenges if they feel that the next steps are manageable and their actions have the potential to result in meaningful outcomes (Nussbaum & Dweck, 2008).

We also found an interaction showing that students in School 1 benefited more from support on planning writing revisions than from revisiting. Students in this school may have benefited particularly from the planning condition because the condition reinforced the teachers' focus on support for planning writing actions. The guidance functioned to break down the process of writing into smaller and more manageable suggestions, consistent with the teachers' emphasis in instruction. This finding suggests a potential benefit of meeting with teachers to determine which forms of guidance align with their student needs and their teaching strategies.

One limitation that exists in our findings for Study 3 is that since all students improved from pretest to posttest, and there was no control condition in which students did not receive a revision strategy, we cannot determine for certain whether the students improved by the posttest because of the revision process or from increased understanding from the overall unit. However, the pre/posttest item examined aligns very closely with the concept addressed in the automated guidance step, which suggests that the revisions students made on the automated guidance step would contribute to their understanding of this specific thermodynamics concept.

As technological advances allow designers to create automated guidance that accurately adapts to student ideas in science writing, it is important to consider not only the content accuracy of guidance or the materials that students revisit for better understanding, but also the cognitive motivational processes that allow students to fully benefit from the guidance. In our studies, we find differential effects for student learning outcomes in response to guidance that has the same science content, but differs in levels of transparency concerning how the computer works (Study 2) or is presented with different revision strategies (Study 3). These findings illustrate the potential of research that clarifies how variations in guidance influence how students benefit from revising activities.

These results suggest benefits for studies that combine the revision strategies of revisiting and planning to help students integrate their understanding of complex science topics. They illustrate potential benefits of exploring ways to align automated guidance with the strategies used by individual teachers in the classroom. Advances in computer automated guidance technologies support investigations not only of the content of the guidance but also the methods for ensuring student engagement in using the guidance. Choosing to engage effortfully with revision strategies may lead students to acquire more skills, which leads students to feel efficacious. This could lead to long-term benefits such

as students being more likely to choose these complex strategies in the future (Katz & Assor, 2007, Pintrich & Schunk, 2002)

Next steps include finding ways to better understand specific learners' main challenges in revising explanations and developing guidance that supports them in making meaningful revisions. In addition, future studies can explore ways to customize guidance strategies to best resonate with supports provided by teachers and to support students in developing a long-term practice of iteratively refining scientific understanding.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences, 4(2)*, 167–207

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111–127

Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist, 44*(9), 1175.

Basharina, O. (2013). Student agency and language-learning processes and outcomes in international online environments. *CALICO Journal*, 26(2), 390-412.

Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the Teaching of English*, 395-422.

Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. Hillsdale, NJ: Erlbaum.

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

Butler, D. L., & Winne, P. H. (1995). Feedback and self regulated learning: A theoretical synthesis. *Review of Educational Research, 65,* 245–281

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. Vygotsky's educational theory in cultural context, 1, 39-64.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439-477.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175-218.

Clark, D., & Jorde, D. (2004). Helping students revise disruptive experientially supported ideas about thermodynamics: Computer visualizations and tactile models. *Journal of Research in Science Teaching*, *41*(1), 1-23.

Cohen, D. K., & Ball, D. L. (2001). Making change: Instruction and its improvement. *The Phi Delta Kappan*, *83*(1), 73-77.

Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, 25(10), 1302-1318.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology*, 88(4), 715.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational psychologist*, 26(3-4), 325-346.

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.

Durso, F. T., Rawson, K. A., & Girotto, S. (2007). Comprehension and situation awareness. *Handbook of applied cognition*, *2*, 163-193.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048.

Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95, 256-273.

Dweck, C. S. & Master, A. (2009). Self-theories and motivation: Students' beliefs about intelligence. K. Wentzel & A. Wigfield (Eds.), Handbook of Motivation at School (pp. 123-140). New York: Routledge.

Ebenezer, J. V., & Zoller, U. (1993). Grade 10 students' perception of and attitudes toward science teaching and school science. *Journal of Research in Science Teaching*, 30, 175–186.

Fedor, D. B. (1991). Recipient responses to performance feedback: A proposed model and its implications. Research in Personnel and Human Resources Management, 9, 73–120.

Fellows, N. (1994). A window into thinking: Using student writing to understand conceptual change in science learning, *Journal of Research in Science Teaching*, 31(9) 985-1001

Fitzgerald, J., (1987). Research on revision in writing. *Review of Educational Research*, 57(4), 481-506.

Gerard, L. F., & Linn, M. C. (2016). Using automated scores of student essays to support teacher guidance in classroom inquiry. *Journal of Science Teacher Education*, *27*(1), 111-129.

Glynn, S. M., & Muth, K. D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of research in science teaching*, 31(9), 1057-1073.

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710-744.

Hand, B., Lawrence, C., & Yore, L. D. (2010). A writing in science framework designed to enhance science literacy. *International Journal of Science Education*, 21(10), 1021–1035.

Hattie, J. & Timperley, H. (2007) The power of feedback. Review of Educational Research. 77(1), 81-112

Hayes, D.A (1987). The potential for directing study in combined reading and writing activity. *Journal of Reading Behavior,* 19, 333-352.

Hohenshell, L. M., & Hand, B. (2006). Writing-to-learn Strategies in Secondary School Cell Biology: A mixed method study. *International Journal of Science Education*, *28*(2-3), 261-289.

Katz, I., & Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review*, 19(4), 429-442.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching,* 36(10), 1065–1084.

Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction*, 18(3), 317-348.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119(2), 254–284.

Kramsch, C., A'Ness, F., & Lam, E. (2000). Authenticity and authorship in the computer-mediated acquisition of L2 literacy. *Language Learning & Technology*, 4(2), 78-104.

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, 32(2), 311.

Langer, J.A., & Applebee, A. (1987). How writing shapes thinking: A study of teaching and learning (NCTE Research Rep. No. 22). Urbana, IL: National Council of Teachers of English.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology*, *4*(2), 103-126.

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

Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). Internet environments for science education. Routledge.

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

Linn, M.C., Gerard, L., Ryoo, K., McElhaney, K., Liu, L., & Rafferty, A., (2014). Computer-guided inquiry to improve science learning. *Science*.

Linn, M. C., Lee, H. S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and Assessing Knowledge Integration in Science. *Science, 313*, 1049–1050

Linn, M. C., & Slotta, J. D. (2000). WISE science. *Educational Leadership,58*(2), 29-32.

Liu, O. L., Brew, C., Blackmore, J., Gerard, L., Madhok, J., & Linn, M. C. (2014). Automated scoring of constructed-response science items: Prospects and obstacles. *Educational Measurement: Issues and Practice*, 33(2), 19-28.

Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, 13(1), 33-55.

Liu, O. L., Lee, H. S., & Linn, M. C. (2011). Measuring knowledge integration: Validation of four-year assessments. *Journal of Research in Science Teaching*, 48(9), 1079-1107.

Liu, L., Rios, J., Heilman, M., Gerard, L., & Linn, M. (2016). Validation of automated scoring of science assessments. *Journal of Research in Science Teaching*, 53(2), 215-233

Matuk, C., Linn, M. C., & Gerard, L. Supporting the WISE Design Process: Authoring Tools that Enable Insights into Technology-Enhanced Learning.*Design Recommendations for Intelligent Tutoring Systems*, 95.

NGSS Lead States (2013). Next generation science standards: For states, by states. Washington, DC: National Academy Press

Nussbaum D. & Dweck C. S. (2008). Defensiveness Versus Remediation: Self-Theories and Modes of Self-Esteem Maintenance, *Personality and Social Psychology Bulletin, 34*, 599-612. doi:10.1177/0146167207312960

Pintrich, P. R., & Schunk, D. H. (2002). Motivation in education: Theory. Research, and Applications, Second Edition, Merrill Prentice Hall, Columbus, Ohio.

Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading*, *35*(2), 136-152.

Richland, L. E., Linn, M. C., & Bjork, R. A. (2007). Instruction. *Handbook of Applied Cognition, Second Edition*, 553-583.

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), 969–983

Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, *105*(4), 1010.

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: applications for intelligent feedback in a writing tutor. *Technology Instruction, Cognition, and Learning*, 10, 59-79.

Ryoo, K., & Linn, M. C. (2014). Designing guidance for interpreting dynamic visualizations: Generating versus reading explanations. *Journal of Research in Science Teaching*, 51(2), 147–174.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, 4-14.

Shute, V., (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 152-189.

Smith, J. P., DiSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115–163.

Sottilare, R., Graesser, A., Hu, X., & Brawner, K. (Eds.). (2015). *Design Recommendations for Intelligent Tutoring Systems: Authoring Tools and Expert Modeling Techniques*. Robert Sottilare.

Sullivan, N., & Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System*, *24*(4), 491-501.

Svihla, V., Wester, M. J., & Linn, M. C. (in press) Distributed Revisiting: An Analytic for Retention of Coherent Science Learning. *Journal of Learning Analytics*.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251–296.

Tansomboon, C., Gerard, L. F., Vitale, J. M., Linn, M. (April 2015) Impact of Knowledge Integration and Simulated Teacher Guidance on Inquiry Science Learning. Poster presentation at annual meeting of American Educational Research Association, Chicago, IL.

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. Harvard university press.

Vygotsky, L. S. (2012). Thought and language. MIT press.

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22-36.

Wichmann, A., & Rummel, N. (2013). Improving revision in wiki-based writing: Coordination pays off. *Computers & Education*, *62*, 262-270.

Yore, L. D., Hand, B. M., & Florence, M. K. (2004). Scientists' Views of Science, Models of Writing, and Science Writing Practices. *Journal of Research in Science Teaching*, 41(4), 338–369.

Zuckerman, M., Porac, J., Lathin, D., & Deci, E. L. (1978). On the importance of self-determination for intrinsically-motivated behavior. *Personality and Social Psychology Bulletin*, 4(3), 443-446.