

Designing Automated Guidance to Promote Productive Revision of Science Explanations

Charissa Tansomboon¹  · Libby F. Gerard¹ ·
Jonathan M. Vitale¹ · Marcia C. Linn¹

Published online: 24 April 2017

© International Artificial Intelligence in Education Society 2017

Abstract 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, we compare forms of automated guidance designed to motivate productive revision and help students integrate their understanding of science. Research shows the benefit of providing timely, transparent guidance to students and identifies some challenges. Specifically, (a) students often believe online guidance is generic rather than adapted to their response; and (b) students do not always engage effortfully with online guidance to improve their written responses. We conducted two studies to address these challenges. In Study 1, we created *transparent* guidance that clarified how the computer personalizes guidance based on the student response. We hypothesized that transparent guidance would be especially valuable for low prior knowledge students who might expect the computer guidance to be too difficult. We found that transparent guidance had a greater impact than typical guidance on low prior knowledge student revisions, suggesting that student beliefs about how guidance is designed influence their performance. In Study 2, implemented in six schools, we compared two specific guidance strategies: *revisiting* evidence and *planning* writing changes. We found that both revisiting and planning guidance resulted in significant improvement

✉ Charissa Tansomboon
charissa@berkeley.edu

Libby F. Gerard
libbygerard@berkeley.edu

Jonathan M. Vitale
jonvitale@berkeley.edu

Marcia C. Linn
mclinn@berkeley.edu

¹ University of California, Berkeley, CA, USA

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. These results illustrate ways to design guidance to strengthen student understanding of science. They raise important questions about when to encourage revisiting, how to design instruction focused on planning, and how to instill a lifelong practice of engaging in iterative refinement of scientific explanations.

Keywords Automated guidance · Writing revisions · Motivation · Science learning · Knowledge integration

Introduction

Writing is a critical aspect of scientific practice that allows scientists to communicate ideas, generate insights, and clarify ambiguities (Yore et al. 2004). Scholarship has demonstrated similar benefits of writing explanations for students of science (Hayes 1987; Rivard 1994). Writing explanations helps students reflect on what they know and integrate prior knowledge with new ideas (Fellows 1994). The revision process prompts students to revisit evidence, distinguish their ideas from the evidence, and refine their explanations (Rivard 1994). Contemporary education standards (i.e., Next Generation Science Standards 2013), place strong emphasis on iterative refinement of written materials.

Revising written explanations based on personalized guidance rarely occurs in science classrooms because teachers lack the time to respond to every student's work. (Most science teachers have 5 or 6 classes of 30 to 40 students.) New technologies that take advantage of natural language processing (NLP) tools can immediately score students' written explanations. When embedded in a web-based inquiry project, these scores can be used to assign adaptive guidance. Such guidance can prompt students to revise their explanations using evidence gathered in the inquiry project and strengthen their understanding.

In this paper we investigate how best to design guidance that motivates students to revise their explanations to build a more integrated understanding. We report on two comparison studies. In the first study we address the fact that some students dismiss automated guidance because they think it is not aligned with their current knowledge. This is especially likely for students who come to science class with low prior knowledge of the topic. In this study we compare *transparent* guidance that includes clarification of the relevance of the suggestions to the needs of the student, to *typical* guidance, that does not include the clarification. In the second study we compare two promising strategies for promoting knowledge integration: *revisiting* guidance that encourages students to get additional evidence by reviewing evidence in the unit and *planning* guidance that encourages students to plan their revisions based on identification of gaps in their explanations.

Promoting Learning with Writing in Science

Knowledge Integration Approach to Science Learning

In these studies we draw on the knowledge integration (KI) framework to inform the design of the short-answer questions and automated guidance to promote productive

revision. The knowledge integration framework is a constructivist approach to instruction that emphasizes eliciting student ideas, adding ideas using models, simulations, graphs, and other evidence, distinguishing ideas using strategies such as writing explanations, and reflection using strategies such as iterative refinement of written work (Linn and Eylon 2011). We test designs for automated guidance by adding them to proven online units designed using knowledge integration principles and the Web-based Inquiry Science Environment (WISE).

Writing explanations promotes knowledge integration (Linn and Eylon 2011). Students enter the classroom with prior knowledge about phenomena to be studied and often have multiple, sometimes conflicting, ideas (Smith et al. 1993). Students typically respond to classroom instruction by adding the new ideas presented by instruction to their repertoire of ideas but not necessarily resolving inconsistencies among the new ideas and their prior views (diSessa 2006; Osborne and Collins 2000). Writing explanations encourages students to reflect on their repertoire of ideas, distinguish the relevant and accurate ideas to explain the event, and use evidence to create coherent explanations (Fellows 1994; Hand et al. 2010).

Previous studies illustrate how writing in science supports students to develop a more integrated understanding (e.g. Klein 2000; Richland et al. 2007; Ryoo and Linn 2014). Meta-analyses by Graham and Hebert (2011) for example found that writing about material enhances comprehension for middle school students, and that this applied across several subject areas including science. Writing activities that call for students to integrate multiple pieces of evidence to explain a phenomena 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 and Hand 2006). 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-type writing opportunities (Fellows 1994).

Students benefit from guidance on their explanations and the opportunity to revise (Ryoo and Linn 2014). Explanations often reveal the gaps and or inaccuracies in students' reasoning. Revision encourages students to draw on evidence to further distinguish among ideas presented in instruction and their existing views (Gerard and Linn 2016). A prior 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 et al. 2015). 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. By manipulating written content, students are more likely to remember and understand science concepts than if they only generate ideas (Langer and Applebee 1987). The centrality of revision in scientific practice is reflected by the Next Generation Science Standards. Students are expected to construct explanations and arguments, and, to “identify flaws in their own arguments and modify and improve them in response to criticism” (NGSS Lead States 2013).

Technology Advances and Challenges in Promoting Revision

New technologies can help teachers guide revision by providing students with immediate, adaptive guidance on their written artifacts (Proske et al. 2012; Shepard 2000). Computer

based tools also have the benefit of being able to evaluate many text responses consistently and objectively across students (Roscoe and McNamara 2013). Many existing NLP tools focus on writing mechanics rather than coherence or scientific accuracy (e.g., Warschauer and Grimes 2008). However, new tools are emerging in science that allow for evaluation of student explanations on a more conceptual level.

Previous work has found that computer-assigned guidance, which is designed to support knowledge integration, can improve students' explanation revision and improve science learning (Gerard and Linn 2016; Linn and Eylon 2011). Automated 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 et al. 2011; Van Zee and Minstrell, 1997). In prior research, computer-assigned knowledge integration guidance led students to make significantly more productive revisions, and subsequently, to produce more coherent and accurate science explanations than did generic guidance (e.g. "Add more evidence") or specific guidance (e.g. "Incorrect. Energy transforms from light energy into chemical energy") (Gerard and Linn 2016). Studies also show potential benefits when automated scoring alerts teachers to students who would benefit from their help on how to revise their explanation (Gerard and Linn 2016).

While NLP tools can analyze student writing, determine a score and assign guidance, choosing the optimal guidance for each score 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 to make surface-level changes instead of deeper conceptual changes (Cohen and Ball 2001). Likewise, many school tasks promote the idea that science is a "simple, algorithmic form of reasoning" (Chinn and Malhotra 2002). This belief may lead students to look for a correct answer when revising rather than seeking evidence to strengthen their argument (Berland and 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 et al. 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 et al. 2015). In the following studies, we explore designs of guidance that can better support students to engage in effortful revision of science explanations.

Study 1: Comparing Transparent and Typical Guidance

While students generally benefit from using automated guidance, not all students engage with and use the guidance they get. 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 and 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 and 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 et al. 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). Conversations with students using WISE have revealed that some may perceive guidance that comes from a computer as generic and unresponsive, especially in comparison to guidance from the teacher. 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. 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 et al. 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?

Study 1 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.

Materials and Procedures

All materials used for curriculum and assessment were implemented in the Web-Based Inquiry Science Environment (WISE; <http://wise4.berkeley.edu>). WISE is an online platform designed following the knowledge integration framework to promote integrated understanding of science. WISE curriculum units are typically designed around one scientific topic, such as thermodynamics. Within WISE, students view visualizations, conduct experiments, and respond to embedded assessments.

Some short-answer writing activities within WISE are scored by the NLP tool c-raterML™. 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 (Appendix Table 14). c-raterML™ works by building a model using a series of NLP steps based on human scoring of at least 1000 student responses on an item (Liu et al. 2016). c-raterML™ scoring shows satisfactory agreement with human scoring. The specific short-answer question scored by c-raterML™ in this study (*Spoons*) has a Pearson correlation of .72 between c-raterML™ score and human scores (Liu et al. 2014). 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 (Appendix Table 14).

Thermodynamics Curriculum

We implemented the study in a WISE curriculum unit entitled, “Thermodynamics: Understanding Heat and Temperature”. The thermodynamics unit provides instruction about conduction as students interact with visualizations to test heat flow through different materials. This concept can be difficult for students to grasp due to their previous experiences (Clark and 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

Table 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%

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 et al. 2015).

In one visualization (used in the automated guidance for student revisiting), students saw an animation of fingers touching objects to test the conductivity of diverse materials. The animation depicted the flow of energy into the finger at different rates, consistent with the nature of the material. This visualization illustrated that objects made of different materials may feel hotter, due to varied conductivity levels, even if they are at the same temperature.

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 et al. 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.

Spoons-Embedded Item and Guidance

To investigate the effect of transparent versus typical guidance on revision of written material we focused on an question from the thermodynamics curriculum called *Spoons*, which is automatically scored by c-raterML™ (Fig. 1). *Spoons* prompts students to select and explain which of three spoons (metal, wood, plastic) would feel the hottest after being placed in hot water. This question helps students understand that conduction varies in different materials, and encourages students to distinguish between heat and temperature. *Spoons* is scored on a knowledge integration rubric (Appendix Table 14). Prior research has found that knowledge integration assessments of inquiry science learning are valid and adequately measure student understanding and explanation of scientific concepts (Liu et al. 2011). *Spoons* was also chosen because its automated c-raterML™ scoring has shown to be reliable and previous research suggests that automated guidance is effective (Donnelly et al. 2015).

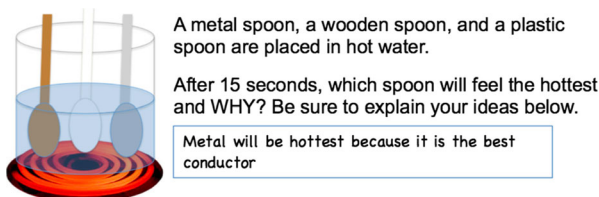


Fig. 1 *Spoons* question with initial student response

The knowledge integration guidance developed for *Spoons* includes three components: (a) a question targeting a concept not addressed by the student response; (b) a prompt directing the students to revisit a visualization (such as the finger animation) in the unit to review evidence of key concepts, and (c) instruction asking the student to generate an improved explanation that distinguishes among new ideas and the ideas in the response. In this study, all students were given two rounds of guidance, so they were prompted to revise their response twice.

In the transparent condition we revised both the guidance and the surrounding instruction for the *Spoons* item 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 (Fig. 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 3 shows an example of guidance tailored to a specific pair of students. Additionally, 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 Table 14). 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.

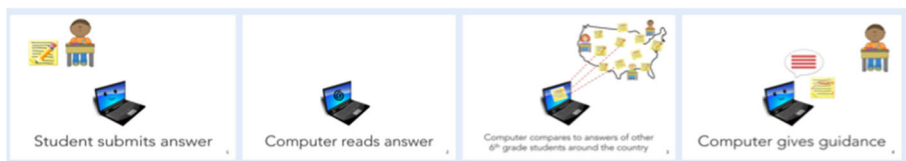


Fig. 2 Transparent condition informational page describing how automated scoring works. **a** Student submits answer. **b** Computer reads answer. **c** Computer compares to answers of other 6th grade students around the country. **d** Computer gives guidance

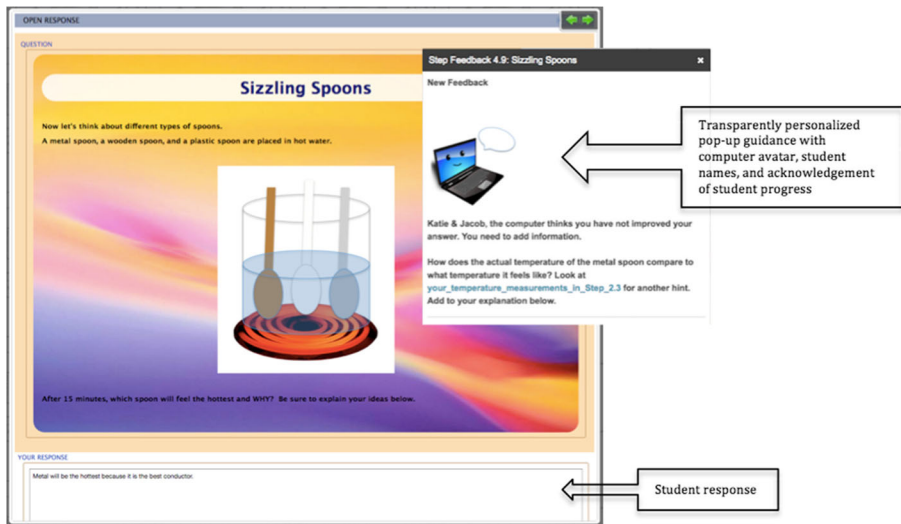


Fig. 3 Transparent condition automated guidance begins with student names and acknowledgement of progress: “Katie & Jacob, the computer thinks you have not improved your answer. You need to add information”

Cups-Prepost Item

To investigate prior knowledge and learning gains we focused on a short-answer, pre-post item, *Cups*, that addressed similar concepts as the embedded *Spoons* item. 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 when filled with hot liquid.

Data Sources and Analysis

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

Embedded and Pre/Posttest Items

Responses to *Spoons* were scored by c-raterML™ to allow for assignment of automated guidance based on score level. Both *Spoons* and *Cups* were scored by humans for purposes of data analysis (Table 2). *Spoons* and *Cups* were both scored using a

Table 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	<i>Spoons-embedded</i>	Completed in pairs	Automated score by c-raterML™ for assignment of automated guidance Human scoring for analysis
Pre/Posttest item	<i>Cups-prepost</i>	Completed individually	Human scoring for analysis

knowledge integration rubric by two researchers. Cohen's kappa calculated for a subset of 50 *Spoons* responses was .92. We used the *Cups* item to measure the impact of the conditions because it closely aligns with the thermodynamics concepts covered in the *Spoons* item. *Cups* was scored on a KI rubric very similar to the one used for *Spoons*. 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*.

Throughout both Study 1 and Study 2, 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 *Cups* performance was done on individuals, since students completed those portions individually. Students who did not complete the embedded *Spoons* step, and thus were not exposed to the experimental conditions, were dropped from analysis of both the *Spoons* embedded item and the *Cups* pretest-posttest. Students were also dropped from the pretest-posttest analysis when they did not finish the posttest, which occurred when classrooms ran out of computer time.

Specific Revisions

To understand more specific actions that students made during revision, log files and specific writing revisions were examined. Log files were used to determine whether students revisited the step within the unit suggested by KI guidance. Students' initial and revised *Spoons* responses were compared to examine their revision characteristics (Table 3). Students who added either a normative or non-normative new scientific idea in revision were classified as having made a substantial change. Changing words in their answer, but without adding a full idea, was classified as a minimal change. A final category of students copied over the exact same answer from their initial response without making any changes, and these were classified as no change. For the revision characteristics scoring, two researchers scored a subset of responses, with the first researcher scoring responses and the second researcher checking the scores. The rubric was discussed and refined until they reached agreement.

Table 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.

Study 1 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. The learning gains for both conditions are reported below.

Table 4 shows scores by condition for Revision 1, Revision 2, and Revised *Spoons-embedded*. Across both conditions, students who completed one round of revision on *Spoons-embedded* did not have a significantly different revised score from those who completed two rounds of revisions [$M(\text{One revision}) = 3.52$, $SD = .89$; $M(\text{Two revisions}) = 3.63$, $SD = .84$; $t(239) = .97$, $p > .05$]. Revised *Spoons-embedded* scores in Table 4 and in future analysis refers to the final submitted revision for each student pair, regardless of whether it was their first revision or their second 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

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

	Initial <i>Spoons-embedded</i>	Revision 1 <i>Spoons-embedded</i>	Revision 2 <i>Spoons-embedded</i>	Revised <i>Spoons-embedded</i>
Transparent personalized	3.50(.79)	3.31(.85)	3.51(.84)	3.79(.82)
Typical adaptive	3.24(.86)	3.70(.77)	3.74(.82)	3.41(.85)

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 [$M(\text{Transparent}) = 3.50$, $SD = .79$; $M(\text{Typical}) = 3.24$, $SD = .86$; $t(239) = 2.52$, $p < .05$, $d = .33$]. 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 5 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.

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 [$F(1238) = 454.59$, $p < .001$, $\eta^2 = .66$], and a main effect for condition emerged [$F(1238) = 5.56$, $p < .05$, $\eta^2 = .02$]. 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 [$F(1237) = 122.48$, $p < .001$, $\eta^2 = .34$], a significant main effect of condition [$F(1237) = 9.68$, $p < .01$, $\eta^2 = .04$], and no significant interaction of condition and prior knowledge [$F(1237) = 2.29$, $p > .05$, $\eta^2 = .01$]. 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 [$M(\text{Pretest}) = 2.76$, $SD = .70$, $M(\text{Posttest}) = 3.38$, $SD = .65$, $t(351) = 13.56$, $p < .001$, $d = .91$].

Table 5 Mean initial and revised *Spoons-embedded* scores, by condition

	Initial <i>Spoons-embedded</i>	Revised <i>Spoons-embedded</i>
Transparent personalized	3.50(.79)	3.79(.82)
Typical adaptive	3.24(.86)	3.41(.85)
Comparison between conditions (t-test)	$t(239) = 2.52^*$	$t(239) = 3.44^*$
	$d = .33$	$d = .45$
	$p < .05$	$p < .001$

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 [$F(1352) = 14.40, p < .001, \eta^2 = .04$], no effect of experimental condition emerged [$F(1352) = 1.29, p > .05, \eta^2 = .004$]. Table 6 shows mean pretest and posttest scores for students in each condition.

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 [$F(1352) = 3.58, p > .05, \eta^2 = .01$], no significant main effect for condition [$F(1352) = 2.99, p > .05, \eta^2 = .008$], and a significant interaction of condition and prior knowledge [$F(1352) = 5.00, p < .05, \eta^2 = .014$]. 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 [$M(\text{Transparent}) = 3.43, SD = .63, M(\text{Typical}) = 3.15, SD = .70, t(131) = 2.40, p < .05, d = .42$]. High prior knowledge students across both conditions had similar posttest scores (Table 7).

Study 1 Discussion

Study 1 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 revised *Spoons-embedded* scores,

Table 6 Mean *Cups-pretest* and *Cups-posttest* scores, by condition

	<i>Cups-pretest</i>	<i>Cups-posttest</i>
Transparent personalized	2.78(.72)	3.33(.67)
Typical adaptive	2.73(.69)	3.42(.61)

Table 7 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*	t(217) = .43
	d = .42	d = .06
	p < .05	p > .05

controlling for initial *Spoons-embedded* scores, while students in the two conditions did now show significantly different *Cups-posttest* scores, controlling for *Cups-pretest* scores. No significant difference in revised *Spoons-embedded* score 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 1 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 posttest. 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 2 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.

Study 2: Revisiting versus Planning Writing Revisions

Study 1 shows that transparent KI guidance appears to promote more successful revisions and science learning, particularly among low prior knowledge students. In Study 2 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 et al. 2006; Chiu and Linn 2012; Gerard and Linn 2016) 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 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.

Study 2 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 8). 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.

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

Table 8 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%

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 9). 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 9). 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 Fig. 4.

Table 9 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 <i>the finger/bowl activity</i> suggested above? Pick all that apply.	How do you plan to change your answer? Pick all that apply.
	Yes, because I wanted more information	I plan to add another scientific idea
	Yes, because the computer suggested it	I plan to add evidence to support my idea
	No, I looked back at a different step than the one suggested	I plan to remove an incorrect idea
	No, I remember the information in that step already	I don't plan to change my answer

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).

Study 2 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%; $X^2(3, N = 464) = 14.86, p < .01, d = .36$] (Fig. 5). 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.

Students across both conditions did not show significant gains from initial to revised *Spoons-embedded* score. Additionally, neither condition demonstrated an advantage in performance on the *Spoons-embedded* item (Table 10).

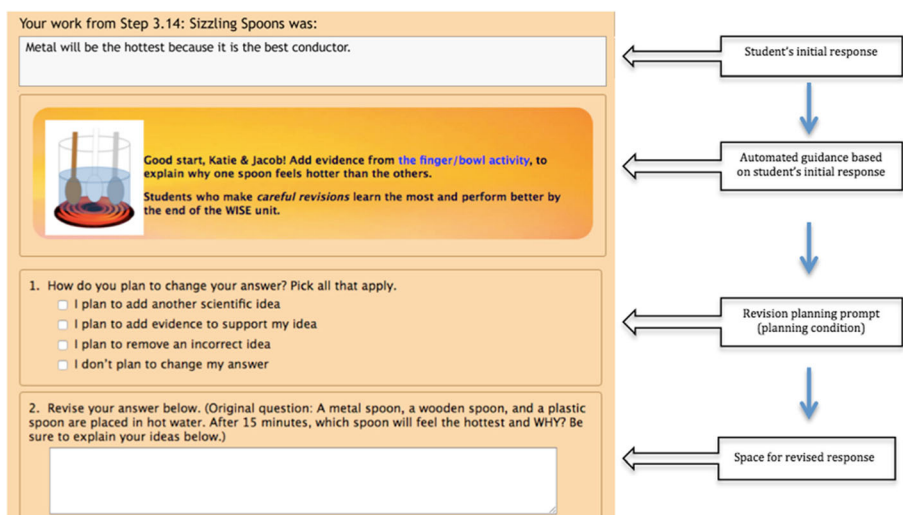


Fig. 4 Planning condition automated guidance and revision page

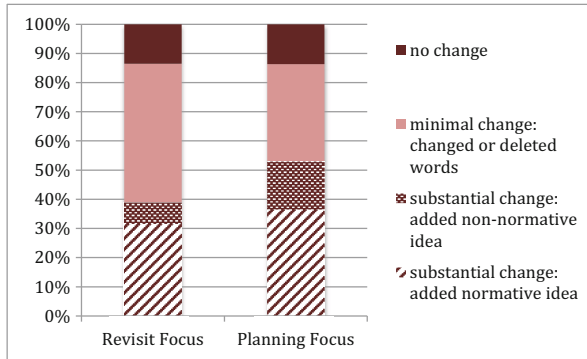


Fig. 5 Revision characteristics by condition

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 10).

To investigate our hypothesis that students with lower prior knowledge would benefit more from revisiting to add ideas, we performed an ANOVA to determine if there was an interaction effect between prior knowledge and condition following the procedure from Study 1. We categorized students as having high or low prior knowledge based on whether their pretest response did or did not include a scientifically valid idea. No significant interaction between prior knowledge and condition was found, suggesting that intervention conditions did not impact gains differently depending on prior knowledge level.

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 et al. 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 11), we performed ANCOVA with pretest score as the covariate. An initial test of

Table 10 Mean embedded and pretest and posttest scores, by condition

	Initial <i>Spoons-embedded</i>	Revised <i>Spoons-embedded</i>	<i>Cups-pretest</i>	<i>Cups-posttest</i>
Revisiting	3.37(.76)	3.32(.93)	2.84(.66)	3.41(.82)
Planning	3.37(.78)	3.35(.91)	2.93(.68)	3.49(.90)

Table 11 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)

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 $>.1$].

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 et al. 2011).

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.

Student A and B were paired together to work on the Thermodynamics unit. Table 12 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 13 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.

Study 2 Discussion

We found that students who received guidance directing them to either revisit scientific models or plan substantive writing changes improved overall learning outcomes from pretest to posttest. Analysis of student actions demonstrates that

Table 12 Examples of initial and revised student responses after receiving planning condition KI guidance

Planning condition		KI Score	KI Score	KI Score
Pair: Student A and Student B	<i>Spoons-embedded</i> initial response After 15 min metal spoon will feel the hottest because it is a good conductor	3	<i>Spoons-embedded</i> revised response After 15 min metal spoon will feel the hottest because it is a good conductor and it transfers heat faster than all of the materials do	4
Student A	<i>Cups-prepost</i> pretest response Because the molecules will have space to move	KI Score 2	<i>Cups-prepost</i> posttest response The metal cup will feel hottest because it is a good conductor and will transfer heat quicker	KI Score 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

Table 13 Examples of initial and revised student responses after receiving revisit condition KI guidance

Revisit condition	<i>Spoons-embedded</i> initial response	KI Score	<i>Spoons-embedded</i> 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 min the metal spoon will be the hottest. It will feel the hottest because metal conducts heat the fastest.	4
Student A	<i>Cups-prepost</i> pretest response I think the plastic cup would feel the hottest because plastic materials are more easily melted	2	<i>Cups-prepost</i> posttest response 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 feel 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

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.

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. Students with low or high prior knowledge made similar pre to posttest learning gains in the two conditions.

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.

Revisiting Patterns In contrast to prior studies, students' actual revisiting patterns did not show correlations with their score gains (e.g., Ryoo and Linn 2014). Since

students were directed to revisit a step that occurred only 2 or 3 steps before the Spoons step, it is possible that students did not make significant improvements from revisiting because they remembered the information or had already acquired sufficient knowledge from this visualization. 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 photosynthesis in the Ryoo and Linn 2014 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.

General Discussion

While automated guidance can help students make revisions on science explanations, not all students make effortful revisions. Our findings suggest ways to design automated guidance to improve student agency in making revisions, and subsequently to improve student learning in science writing tasks.

In Study 1, we found evidence that increasing the transparency and personalization of automated guidance particularly improved performance for low prior knowledge students. One reason for this may be that transparent personalization motivated 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 2, we found that across schools, both revisiting and planning strategies were equally effective in motivating revision of student written explanations. 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 and 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 2 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 pretest and posttest item that we examine aligns very closely with the concept addressed in the automated guidance step, which suggests that the revisions they 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 1). We also find differences in student actions after being presented with different revision strategies (Study 2). 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. Next steps include finding ways to refine understanding of 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 such that they effectively resonate with the supports provided by teachers.

Acknowledgements Funding provided by National Science Foundation grant DRL-1119670, Continuous Learning and Automated Scoring in Science (CLASS).

Appendix

Table 14 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 min, 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	Round 1 automated KI guidance	Round 2 automated KI guidance
		Transparent Personalization group received italicized and non-italicized portion	Transparent Personalization group received italicized and non-italicized portion
		Typical adaptive group received only non-italicized portion	Typical adaptive group received only non-italicized portion
1 Off-task. Student writes, but it does not answer the question being asked.	“I don’t know”	(<i>Student Names</i>), redo. Add evidence from step 4.6, the activity, to explain why one spoon feels hotter than the others.	<i>{{studentFirstNames}}</i> , the computer thinks you did not include a valid, scientific idea in your explanation. You need to change your answer.
2 Scientifically non-normative ideas or links; Vague Ideas	“A plastic spoon would feel the hottest, because plastic can melt easily.”		Redo. Observe how heat moves inside the metal, wood and plastic bars in Step 4.3. Based on your observation, which spoon will feel the hottest and why? Write a new explanation below.
3 Partial link; Unelaborated connections using relevant features	“The metal spoon because it was in the cup for 15 s and the other ones are plastic and wood spoons.”	Good start, (<i>Student Names</i>)! Add evidence from step 4.6, the activity, to explain why one spoon feels hotter than the others.	Good work, <i>{{studentFirstNames}}!</i> You added a valid, scientific idea. Now, take your response one step further. OR <i>{{studentFirstNames}}</i> , the computer thinks you did not add a new, valid, scientific idea to your explanation. You need to change your answer

Table 14 (continued)

A metal spoon, a wooden spoon, and a plastic spoon are placed in hot water. After 15 min, 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	Round 1 automated KI guidance	Round 2 automated KI guidance
		Transparent Personalization group received italicized and non-italicized portion	Transparent Personalization group received italicized and non-italicized portion
		Typical adaptive group received only non-italicized portion	Typical adaptive group received only non-italicized portion
			<i>and make the links between your ideas clearer:</i>
			How does heat move differently through the metal spoon compared to the wood and plastic spoons? Take a look at Step 4.3 for another hint. Then revise your explanation below.
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, <i>(Student Names)!</i> 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.	<i>Good work {{studentFirstNames}}! You added a correct scientific idea and explained your reasoning. Now consider this.</i>
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.”		OR <i>{{studentFirstNames}}, the computer thinks you did not add a new, valid, scientific idea to your explanation You need to add information.</i>
			How does the temperature of the metal spoon compare to the temperature of the wood spoon and the plastic spoon? Check out Step 3.2 for a hint. Then, add to your explanation below

References

- 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.
- Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191–216.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354.
- 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.
- 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.
- Chiu, J. L., & Linn, M. C. (2012). The role of self-monitoring in learning chemistry with dynamic visualizations. In *Metacognition in science education* (pp. 133–163). Netherlands: Springer.
- 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.
- diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 265–281). Cambridge, MA: Cambridge University Press.
- 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.
- Dweck, C. S., & Master, A. (2009). Self-theories and motivation. *Handbook of Motivation at School*, 123–140.
- 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.
- 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.
- 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.
- Hayes, D. A. (1987). The potential for directing study in combined reading and writing activity. *Journal of Reading Behavior*, *19*, 333–352.
- Herrenkohl, L. R., Tasker, T., & White, B. (2011). Pedagogical practices to support classroom cultures of scientific inquiry. *Cognition and instruction*, *29*(1), 1–44.
- 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.
- Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction*, *18*(3), 317–348.
- Kramsch, C., A'Ness, F., & Lam, E. (2000). Authenticity and authorship in the computer-mediated acquisition of L2 literacy. *Language, Learning and Technology*, *4*(2), 78–104.
- Langer, J. A., & Applebee, A. (1987). *How writing shapes thinking: a study of teaching and learning* (NCTE Research Rep. No. 22). Urbana: National Council of Teachers of English.
- Lee, H. S., Linn, M. C., Varma, K., & Liu, O. L. (2010). How do technology-enhanced inquiry science units impact classroom learning? *Journal of Research in Science Teaching*, *47*(1), 71–90.
- Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: taking advantage of technology to promote knowledge integration*. Routledge.
- Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, *87*(4), 517–538.
- 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, 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, 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.
- 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.
- Osborne, J., & Collins, S. (2000). *Pupils' & parents' views of the school science curriculum*. London: Kings College.
- 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* (pp. 553–583). Second Edition: Handbook of Applied Cognition.
- 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*, 29, 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.
- 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.
- Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences*, 6(2), 227–269.
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36.
- 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.