International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance

Jonathan M. Vitale, Elizabeth McBride & Marcia C. Linn

To cite this article: Jonathan M. Vitale, Elizabeth McBride & Marcia C. Linn (2016) Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance, International Journal of Science Education, 38:9, 1548-1569, DOI: 10.1080/09500693.2016.1198969

To link to this article: http://dx.doi.org/10.1080/09500693.2016.1198969

Published online: 12 Jul 2016.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
VOL. 38, NO. 9, 1548–1569
http://dx.doi.org/10.1080/09500693.2016.1198969

Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance

Jonathan M. Vitale, Elizabeth McBride and Marcia C. Linn

Graduate School of Education, University of California, Berkeley, CA, USA

ABSTRACT

We compared two forms of automated guidance to support students' understanding of climate change in an online inquiry science unit. For *specific* guidance, we directly communicated ideas that were missing or misrepresented in student responses. For *knowledge integration* guidance, we provided hints or suggestions to motivate learners to analyze features of their response and seek more information. We guided both student-constructed energy flow diagrams and short essays at total of five times across an approximately week-long curriculum unit. Our results indicate that while *specific* guidance typically produced larger accuracy gains on responses within the curriculum unit, *knowledge integration* guidance produced stronger outcomes on a novel essay at posttest. Closer analysis revealed an association between the time spent revisiting a visualization and posttest scores on this summary essay, only for those students in the *knowledge integration* condition. We discuss how these gains in knowledge integration extend laboratory results related to 'desirable difficulties' and show how autonomous inquiry can be fostered through automated guidance.

ARTICLE HISTORY
Received 15 June 2015
Accepted 4 June 2016

KEYWORDS
Inquiry-based teaching; learning technologies; earth science education; automated assessment

Introduction

Automated scoring offers rich opportunities to explore adaptive forms of instruction. In this paper, we investigate two forms of automated guidance: *specific* and *knowledge integration*. *Specific* guidance directly communicates ideas that are missing or misrepresented in student responses. *Knowledge integration* guidance provides hints or suggestions to motivate learners to analyze features of their response and seek more information. In both cases, guidance may refer students back to previous material; however, only in *knowledge integration* guidance is revisiting necessary to acquire the information alluded to in the guidance prompt.

Much of the previous literature on automated, adaptive instruction has focused on tasks that can be evaluated with high accuracy and guided with accurate, highly specific suggestions (e.g. Anderson, Corbett, Koedinger, & Pelletier, 1995; Anderson & Schunn, 2000). Recently, the introduction of new approaches to automated assessment for open-ended problems and inquiry processes expands the range of tasks suitable for automated

CONTACT Jonathan M. Vitale jonvitale@berkeley.edu Graduate School of Education, University of California, 4407 Tolman Hall, Berkeley, CA 94720, USA

guidance (Gobert, Sao Pedro, Raziuddin, & Baker, 2013; Linn et al., 2014; Shute, 2008; Vitale, Lai, & Linn, 2015). In particular, online science curricula typically incorporate a range of complex, self-directed tasks where students could benefit from evaluation and guidance (Quintana et al., 2004).

According to recent standards, autonomous inquiry practices are not only a means for attaining knowledge, but also a goal of instruction (NGSS Lead States, 2013). By autonomous inquiry practices, we refer to the ability to monitor one's own learning and take actions to remedy gaps in understanding. In the context of challenging science topics, studies are needed to investigate how automated guidance impacts and is impacted by autonomous student inquiry practices. Because knowledge integration instruction was developed, in part, to foster autonomous inquiry (Linn & Hsi, 2000), we applied this framework to develop *knowledge integration* guidance. By comparing *specific* and *knowledge integration* guidance, we begin to investigate how guidance may foster inquiry, and what role (if any) inquiry-based guidance plays in supporting task performance and learning.

In this study, we compare *specific* and *knowledge integration* guidance methods, assigned at random, in a week-long curriculum delivered as part of students' primary science curriculum. We investigate the following research questions:

1. How do *specific* and *knowledge integration* guidance impact performance on the learning task?
2. How do *specific* and *knowledge integration* guidance impact retention of material and success on related posttest items?
3. How do students' autonomous inquiry behaviors differ by condition and impact performance?

Rationale

Instructional design decisions often require one to weigh competing values of autonomy and efficiency. Minimal guidance, including discovery-oriented approaches, places greater emphasis on autonomy with the aim of fostering motivation and more comprehensive learning (Bruner, 1961). Yet, a number of studies suggest that minimally guided instruction often leads to inefficient practices and the introduction of errors (see Mayer, 2004). In response, some researchers advocate for more direct instructional approaches (Klahr & Nigam, 2004). Other researchers suggest that minimal guidance and discovery can be valuable in initial stages of learning (Kapur, 2008) and warn that overly specific guidance can limit conceptual learning and transfer (Martin & Schwartz, 2005). Koedinger and Aleven (2007) label the challenge of navigating this continuum the 'assistance dilemma', and describe it as perhaps 'the fundamental open problem in learning and instructional science' (p. 261).

While extreme positions on this continuum are unlikely to be successful – for example, no guidance or directly providing students with solutions – the specific balance between autonomy and specificity requires exploration across various instructional contexts. We describe the development of our implementations of *specific* and *knowledge integration* guidance in the context of an instructional unit about climate change.

Design of specific guidance

A human tutor can significantly enhance student problem-solving by selecting appropriate exercises, providing example solutions, evaluating student strategies, and correcting errors (Anderson, Boyle, Corbett, & Lewis, 1987). Computerized tutoring systems can approximate human tutoring by evaluating the accuracy of student work and presence of specific ideas, particularly in well-defined domains where discrete symbol manipulation reveals clear strategies (e.g. algebra, Koedinger & Anderson, 1997). Within this context, an important role of guidance is rapidly identifying and correcting errors so that students will not reinforce them through continued application. Thus, in this case, 'telling' students how to solve problems reduces the risk that they will reinforce errant strategies (Anderson & Schunn, 2000).

As the instructional domain increases in complexity, the task of guiding the student becomes more challenging. As Sweller (1994) describes, students have limited cognitive resources with which to solve complex problems. Therefore, with tasks that have high intrinsic complexity it is the responsibility of the instructional designer to focus learners on germane concepts. In response, a number of researchers have advocated for instructional methods that reduce demands on students' cognitive resources. For example, 'worked examples' (Sweller & Cooper, 1985) introduce learners to the specific strategies prior to student engagement, thereby reducing reliance upon inefficient or irrelevant problem-solving strategies (van Gog, Paas, & van Merriënboer, 2006).

In the context of response-adaptive guidance, research provides a strong basis of support for specific guidance that provides information succinctly and directly (Anderson et al., 1995). For example, a number of studies show that automated, response-specific guidance supports performance better than praise or grades (Butler, 1987), general advice or vague guidance (Phye & Sanders, 1994), or simple correct/incorrect guidance (Pridemore & Klein, 1995). Conversely, specific guidance should avoid cumbersome text or unnecessary vocabulary (Kulhavy, White, Topp, Chan, & Adams, 1985). These guidelines indicate that specific guidance is successful when it provides immediate, clear advice for improving a response.

Design of knowledge integration guidance

While, in the context of complex science topics, specific guidance can be an efficient means of improving student responses, it is not typically designed to foster autonomous inquiry practices. In some cases, overly specific guidance may even impede autonomous engagement with a task. For example, students often take advantage of hints to 'game the system' in tutoring programs (Baker, Corbett, Koedinger, & Wagner, 2004). If supporting inquiry practices is a goal of instruction, then guidance that limits autonomy is not appropriate.

Rather, promoting student inquiry requires an explicit structuring of tasks and instructional guidance to motivate students to investigate phenomena and critique their thinking (Singer, Marx, Krajcik, & Chambers, 2000). For example, the knowledge integration framework (Linn & Eylon, 2011) recommends that instructional activities include making predictions about phenomena under investigation; following novel content with activities that help students distinguish between new ideas and their predictions; and reflecting upon their new ideas. The processes of distinguishing and reflecting may motivate students to revisit challenging materials to continue their investigation.

Beyond fostering inquiry, knowledge integration activities aim to help students construct a coherent understanding of the content (Linn & Eylon, 2011). Specifically, inquiry-related activities are typically more challenging than direct instructional approaches. While this may seem contrary to the goal of facilitating student learning, in a series of studies, Bjork and colleagues recognized that some impediments to performance during learning may actually contribute to greater retention (Bjork, 1994; Christina & Bjork, 1991). Specifically, by introducing 'desirable difficulties' that compel learners to engage in more active processing of information, learning tasks that may be perceived as challenging or inefficient may prove more beneficial than those completed with high fluency. In related studies researchers have found benefits for generating rather than reading text (DeWinstanley & Bjork, 2004), interleaving disparate concepts rather than blocking related concepts (Ziegler & Stern, 2014), and spacing study sessions rather than massing practice (Appleton-Knapp, Bjork, & Wickens, 2005).

From an instructional perspective, findings related to desirable difficulties suggest that effective guidance should promote elaborating on previous material, making connections through further inquiry, and reflecting on strategies (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Ryoo & Linn, 2014). For example, automated guidance may stimulate elaboration by addressing complex problem goals (McKendree, 1990), introducing metacognitive prompts (Roll, Aleven, McLaren, & Koedinger, 2011), or incorporating reflection prompts (Van Den Boom, Paas, Van Merriënboer, & Van Gog, 2004).

In each of these instructional approaches, students are asked, in some way, to engage with the materials or their own knowledge critically. Likewise, rather than simply directing students to resources, knowledge integration guidance motivates critical engagement by asking students to evaluate their own ideas (Linn et al., 2014). For example, in cases where students incorporate a non-normative idea into their work (e.g. 'heat comes directly from the sun'), they may be asked to consider a relevant question ('i.e. what type of energy comes from the sun') prior to investigating a simulation. Effective use of this guidance requires students to actively consider these challenges and engage in the suggested inquiry tasks. Therefore, evaluating the effectiveness of *knowledge integration* guidance requires attention to both changes in responses and the use of inquiry resources. Online science curricula that track students as they navigate through challenges in autonomous groups is an ideal context to explore the relationships between inquiry, engagement, and learning.

Visualizing climate change

While students often regard science as a series of unrelated facts, in most cases, scientific content is characterized by complex systems of inter-related components and unseen dynamic processes (Hmelo-Silver & Azevedo, 2006). Despite their complexity, these systems often have clear effects on the lives of people, making them an important target for instruction. One such domain is atmospheric science, where predictive models are incomplete, but the effect of climate change is well established and beginning to impact the environment and lives of humans (Hansen, 2010). Yet, while many non-scientists are aware of climate change, they are typically unaware of or confused about specific mechanisms (Svihla & Linn, 2012). For example, many people confuse the 'greenhouse effect' (e.g. the interaction between greenhouse gases and infrared radiation) with ozone depletion (Hansen, 2010).

For these reasons, climate change is an important, but challenging topic of instruction. Addressing concepts in a complex system, like that of the atmosphere, can be facilitated with dynamic visualizations that clearly demonstrate the interaction of system components (Ryoo & Linn, 2014; Wilensky & Reisman, 2006). Yet, designing visualizations and supporting guidance that makes these interactions salient and tractable is challenging (McElhaney, Chang, Chiu, & Linn, 2015). While experts are adept at attending to appropriate features of a complex visualization, novice attention is often misdirected toward irrelevant features (Jarodzka, Scheiter, Gerjets, & van Gog, 2010). Additionally, novices may be unaware of their own limitations, incorrectly believing that they understood a complex visualization (McElhaney et al., 2015).

To address these challenges, in many cases, dynamic visualizations can be developed or modified to cue or signal important features or events (e.g. highlighting a feature in bold color) (Lin & Atkinson, 2011). For example, in the climate change simulation described below, students have the opportunity, by pressing a button, to follow a single ray of light through its various transformations through the atmosphere and earth. In cases where a visualization cannot be adapted to signal important features, supplementary guidance or advanced organizers can serve to direct student attention on visual features (van Gog, Paas, & van Merriënboer, 2004). For example, prior to engaging with a climate change simulation, students view an annotated image of the simulation to identify important components.

Finally, following a visualization, embedded assessment with automated scoring and guidance can be used to evaluate student knowledge, make suggestions, clarify non-normative ideas, and provide direction for review of the visualization. Written open response or diagraming tasks, which reveal complex student ideas, may provide particular insight about missing or non-normative student ideas (Linn et al., 2014). Both *specific* and *knowledge integration* guidance were designed to address student ideas in these challenging, constructed-response activities.

Method

Participants

We invited science teachers from three middle schools in a suburban region of the western United States to participate in this study. Two schools were chosen because their lead teachers had run a WISE climate change unit in previous years and were willing to participate in further research. However, because these schools served a predominantly middle-to-upper-middle SES population, we also invited teachers from a third school, serving a lower SES community, to ensure greater representation of regional diversity and test the generalizability of results.

From these three schools, five sixth-grade teachers, with a total of 283 students, chose to participate. School A (school demographics: 38% White, 30% Asian, 17% Hispanic, 5% Black, 13% English learners, 25% socioeconomically disadvantaged[1]) included three teachers (Teacher A_1: White female, 2 classes, over 20 years teaching experience; Teacher A_2: Black female, 1 class, 2 years teaching experience; Teacher A_3: White male, 1 class, 5 years teaching experience) and their 107 students (52% female) participated. School B (school demographics: 64% White, 23% Asian, 9% Hispanic, 3% Black; 6% English learners; 9% socioeconomically disadvantaged) included one teacher (Teacher B: White female, 3 classes, over 20 years teaching experience) and her 80 students (45% female). School C (school demographics: 5% White, 8% Asian, 73% Hispanic, 12% Black; 57% English learners; 92% socioeconomically disadvantaged) included one teacher (Teacher C: Hispanic male, 6 classes, 3 years teaching experience) and his 96 (47% female) students.

Materials

All tasks were performed using the *Web-based Inquiry Science Environment* (WISE), which is an online platform that supports common testing formats (e.g. multiple choice and open response) and instructional tools (e.g. system diagraming, dynamic visualizations) (Linn, Clark, & Slotta, 2003). Each web page interface within a WISE unit is referred to as a step. Each step may contain instructional content, assessment prompts, or a combination of the two. Figure 1 displays a screenshot of the learning environment.

Pretest and posttest

We focused on three items in the pretest and posttest that were aligned with the learning activities within the scope of this study. These items were validated in prior research on the *climate change* unit (Svihla & Linn, 2012). These items address several standards

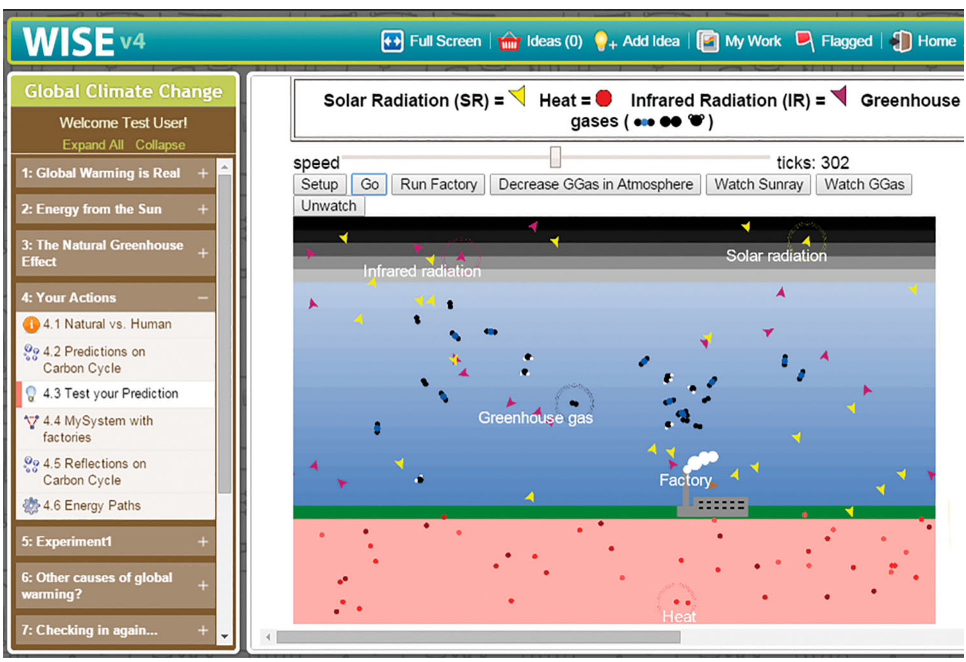


Figure 1. Screenshot of global climate change unit in WISE. Main panel displays a NetLogo simulation. The left panel displays the sequence of activities in an inquiry map.

established in the NGSS (NGSS Lead States, 2013), including, but not limited to: MS-ESS3–4 (‘Construct an argument supported by evidence for how increases in human population and per-capita consumption of natural resources impact Earth’s systems’) and MS-ESS3–5 (‘Ask questions to clarify evidence of the factors that have caused the rise in global temperatures over the past century’). These items, described below, include two items that were presented within the curriculum to assess retention (*Coal*, *MySystem*), and one item that was not addressed directly with the curriculum to assess general comprehension and transfer (*Energy Story*).

Coal

This mixed multiple-choice, open-response item prompted participants to consider how the increased carbon dioxide resulting from burning coal to produce electricity may affect the planet. Participants first chose from a list of four options (‘a warmer climate’, ‘a cooler climate’, ‘lower relative humidity’, ‘more ozone in the atmosphere’), and then explained their choice. To avoid confusion between *coal* presented during the curriculum unit and *coal* presented at pre- or posttest, we label the latter as *coal-prepost* (or *coal-post* in cases where we analyzed posttest only). In all cases, the text of the item was identical.

MySystem

In addition to written text, energy concepts may assessed with energy flow diagrams (Ryoo & Linn, 2014). Like concepts maps, more generally, an energy system allows students to depict the flow and transformation of energy as links between nodes (icons). For example, the greenhouse effect can be represented by connecting icons representing the Earth’s surface and greenhouse gases with arrows that represent infrared radiation (Table 2).

In the *MySystem* workspace (Figure 2), participants had access to a vertical menu of icons representing components of the natural system (e.g. Sun, space, surface of Earth, beneath surface, greenhouse gases, and ozone layer). Participants’ placed these icons on an empty canvas and connected them with arrows. Immediately after connecting icons, participants chose a label for the arrow to represent the type of energy flowing from one component to the other (e.g. solar radiation, heat, infrared radiation, or ultraviolet radiation). Participants were prompted to show how the Earth is warmed by energy, while considering where energy comes from, how it moves, and how it transforms. Like *coal*, we label *MySystem* given at pre- or posttest as *MySystem*-prepost.

Energy story

This open-response transfer item prompted participants to explain how the Earth is warmed by energy, while considering where energy comes from, how it moves, and how it transforms.

Global climate change curriculum

The climate change curriculum is a WISE unit that was developed and refined by multiple researchers and teachers (Svihla & Linn, 2012). Like other WISE units, student work-groups proceeded through *global climate change* at a self-directed pace.

Three simulations were constructed using NetLogo, a tool for depicting complex systems with numerous independent agents (Wilensky & Reisman, 2006). NetLogo users can interact with the simulation by moving sliders and or clicking buttons. For

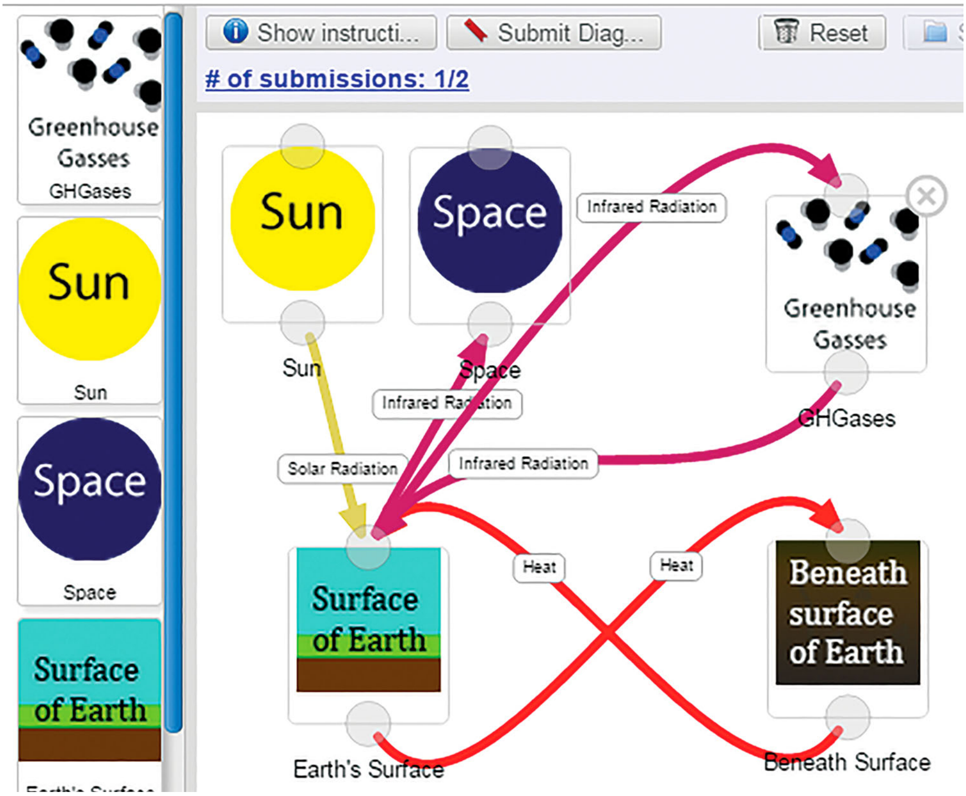


Figure 2. Screenshot of *MySystem*. Left, vertical panel displays available icons to use as nodes in diagram. On the right is an example diagram depicting the greenhouse effect.

global climate change, participants observed the motion of energy and its interaction with other materials depicted as arrows or simple shapes (Figure 1).

Following interactions with each simulation, workgroups engaged in guided *MySystem* diagram activities. The first two activities (*MySystem 1*, *MySystem 2*) used only a subset of the icons available for *MySystem 3* and *MySystem-prepost*. After *MySystem 1*, subsequent diagrams imported the work from the previous diagram, thus allowing for a progressive accumulation of concepts. Workgroups received automated guidance according to their experimental condition and the automated score representing the diagram's completeness. For each *MySystem,* workgroups had the opportunity to receive guidance up to three times.

Guidance prompts, in both guidance conditions, presented students with a screen capture image of the previous NetLogo simulation with text overlaid (Figure 3). In the *specific* condition, participants were asked to recall the visualization, observe the image, and revise the *MySystem*. In the *Knowledge Integration (KI)* condition, participants were prompted to observe the image and revisit the simulation to answer questions given in the text. Additionally, in the *KI* condition, at least some part of the image was occluded, thus requiring participants to revisit the visualization to acquire the same information as the *specific* condition.

Finally, students responded to the *Coal* item and received automated guidance at two points in the unit. The first *Coal* opportunity within the curriculum unit (*Coal 1)* was

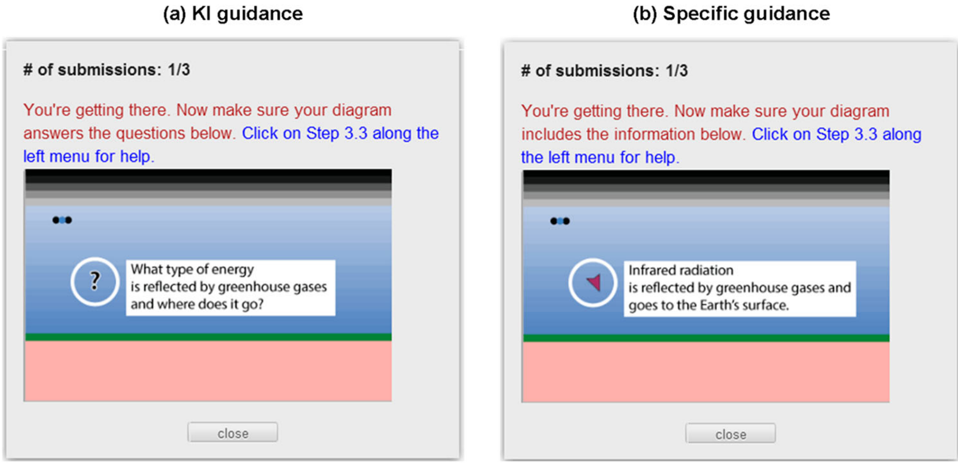


Figure 3. Examples of (a) KI and (b) specific *MySystem* guidance for students missing the idea that gases trap infrared.

placed directly following activities that addressed the normative mechanisms of global warming. *Coal 2* was placed following an activity that addressed non-normative ideas about global warming (e.g. the effect of the ozone layer). Automated scores were determined using the c-Rater-ml system developed by the Educational Testing Service (Liu et al., 2014). The c-Rater-ml system utilizes an associative model developed from previously scored responses based on a knowledge integration rubric (Table 1).

Automated guidance was developed for each score level and differentiated by experimental condition. In the *specific* condition, participants were told to integrate an explicit description of an energy process into their revision. In the *KI* condition, participants were asked to revisit a visualization to explore this energy process, and then revise their response (Table 1). For *Coal 2,* additional differentiation of non-normative ideas was conducted, and in some cases, students were provided with guidance addressing misunderstandings (e.g. ozone layer depletion increases global warming). Students had one opportunity to receive guidance and revise on *Coal 1* and on *Coal 2*.

Rubrics and coding

Open response

All open response items, including *Coal* and *Energy Story*, were coded with a knowledge integration rubric (Table 1 for *Coal* example) to determine a score that reflected the coherency of the response. *Coal* was scored on a 5-point scale, while *Energy Story* was scored on a 6-point scale to account for its additional conceptual breadth. All work was scored by at least one of two coders, with at least 30% overlap on each item to determine inter-rater agreement. Inter-rater agreement between the two coders reached acceptable levels for all three items [*Energy Story*: $n = 208$, $\kappa = .91$; Coal: $n = 206$, $\kappa = .89$].

MySystem

Pretest and posttest diagrams were scored using an automated, rubric-based system that evaluated the presence of specific icons and links. Like the 6-point KI score for *Energy*

Table 1. Rubric for *Coal* item with guidance.

Score	Description	Example	*KI* guidance	*Specific* guidance
5	Two or more scientifically valid links between ideas	*Greenhouse gases reflect IR back to the surface, where it transforms into heat. This causes global warming.*	Great Job, no revision is necessary.	Great Job, no revision is necessary.
4	One scientifically valid link between ideas	*CO2 traps heat from the surface in the atmosphere*	*You are almost there!* To improve your response return to Step 2.8 to find out what happens to energy from the Sun when it is absorbed by the Earth.	*You are almost there!* To improve your response recall from Step 2.8 that solar radiation is transformed into heat when it is absorbed by the Earth.
3	Unelaborated links between ideas, or partial idea	*Greenhouse gases make the climate warmer*	*Good progress, but your answer can be improved.* To improve your response return to Step 3.3 to find out how carbon dioxide in the atmosphere affects the global temperature by interacting with energy released by the surface of the Earth.	*Good progress, but your answer can be improved.* To improve your response recall from Step 3.3 that carbon dioxide in the atmosphere increases the global temperature by trapping infrared radiation released by the surface of the Earth.
2	Scientifically non-normative ideas or links between ideas	*CO2 is warm*	*Your answer needs more ideas* To improve your response return to Step 4.3 to find out how increased carbon dioxide in the atmosphere affects the global temperature.	*Your answer needs more ideas* To improve your response recall from Step 4.3 that increased carbon dioxide in the atmosphere increases the global temperature.
1	Irrelevant or off task	*I don't know*	Same as 2 (above)	Same as 2 (above)

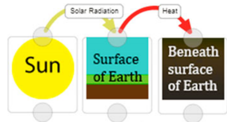
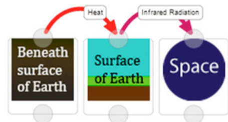
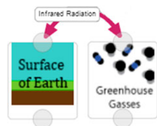
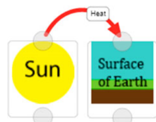
Story, *MySystem* scores were determined by the presence of scientifically normative links. Scoring was determined by accounting for similar ideas and links between ideas as expressed in *Energy Story*. Table 2 provides examples of normative links and their expression in both formats.

Revisiting

The WISE system provides researchers with a log of student navigation, which can be used to analyze revisiting of previous steps. Specifically, for each *MySystem* and *Coal* step, we identified all earlier simulation steps that were visited between the initial and final response submissions. By identifying these simulation revisits, we could categorize work-groups as revisiting/not-revisiting and calculate total time spent revisiting, per item. We chose to categorize revisiting only for step visits greater than 5 seconds, which is approximately the amount of time needed to begin meaningful observation of the simulation.

This approach of identifying all simulation visits between first and final revisions of an automated item was intended to support analyses of complex patterns of revisiting, but, in some cases, may have led to identification of simulation revisits that were not intentional or played little role in *MySystem* or *Coal* item revision. Another approach would be to identify only simulation revisits that immediately followed a *MySystem* or *Coal* item; however, in this case, idiosyncratic navigational patterns would have obscured some legitimate revisiting behaviors. While either approach is valid, we

Table 2. Similar ideas in *Energy Story* and *MySystem*.

Idea	*Energy story* example	*MySystem* example
Solar radiation transforms to heat	*When the energy reaches the Earth it either gets absorbed or reflected. If the energy gets absorbed, it turns into heat energy*	
Heat transforms to infrared radiation.	*Then earth releases the heat in the form of IR and back into space*	
Greenhouse gases prevent energy from escaping	*The infrared is trapped by greenhouse gasses and turned into heat, keeping the Earth warm*	
Heat travels from the Sun to the Earth. (non-normative)	*The sun gives off heat onto the surface of the earth*	

chose the former, a more sensitive analysis method to ensure that all potentially valuable revisits were counted.

Procedure

Participants performed the individual online pretest; however, the teacher in School B had students complete the pretest in pairs due to time restrictions and computer availability. The pretest was completed in a single class period. Students then completed the online *global climate change* unit in teacher-determined workgroups (163 groups in total). These workgroups were typically dyads (138), but in some cases, students worked in triads (7) or alone (18). Each workgroup was assigned to either the *specific* or *KI* guidance condition randomly. All students in the selected classroom participated in the study according to institutional review board procedures.

While we recommended that teachers provide at least four full periods for testing and instruction, we encouraged the teachers to implement the curriculum unit according to their time constraints and pacing requirements. As such, the duration of the curriculum unit varied by teacher. Several teachers allowed students to complete the unit at their own pace and complete the posttest immediately upon reaching the conclusion of this unit. Teachers A_1, A_3, and B allotted 5, 6, and 6 full periods, respectively, for the unit and posttest. Those who completed the posttest early were given supplementary work to complete. On the other hand, teachers A_2 and C prompted all of their students to begin the posttest simultaneously on the same day of instruction (5th day, 4th day, respectively). Some students in these classes did not complete the unit, although most were able to complete at least one submission of the embedded *coal* item.

During the *global climate change*, unit students were asked to work closely with the other members of their group, and could request assistance from the teacher or any available researcher in the classroom. In cases where questions addressed automatically guided items, students were first told to read the computer guidance. If students continued to request assistance, the teacher or researcher helped the students to understand and follow the guidance suggested by the computer, thereby reinforcing the guidance. Otherwise, teachers were not asked to alter their instructional approach. In response, several different instructional practices emerged in the classroom, including class-wide discussion activities.

Results

Research questions

RQ1: How do *specific* and *knowledge integration* guidance impact performance on the learning task?

We addressed these questions separately for the *MySystem* and *Coal* items embedded within the curriculum. In the case of *MySystem*, students engaged in three versions of the task and had multiple opportunities to revise (and receive guidance) within each version. We label each saved *MySystem* diagram as a 'submission' (i.e. submitted for guidance). To narrow our focus on submissions common to most workgroups, we included scores for the following submissions: (1) initial *MySystem 1* (produced prior to any guidance); (2) final *MySystem 1* (after guidance); (3) final *MySystem 2*; (4) final *MySystem 3*. To compare conditions across these four submissions, we performed repeated-measures ANOVA with score as dependent variable, submission number as within-subjects variable, condition as between-subjects variable, and the interaction of condition and submission

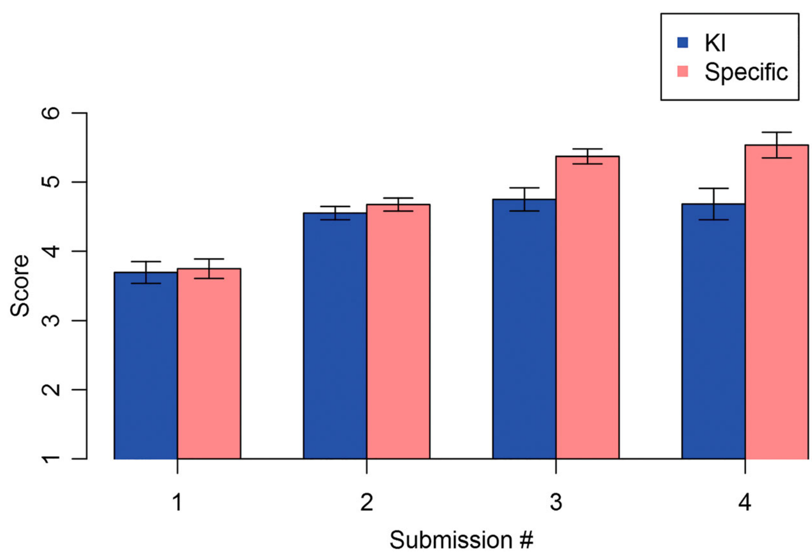


Figure 4. Knowledge integration scores for embedded *MySystem* items, by condition. First submission indicates the score on the initial diagram produced for *MySystem* 1. Submissions 2, 3, and 4 indicate the score for the final diagram produced for *MySystem* items 1, 2, and 3, respectively.

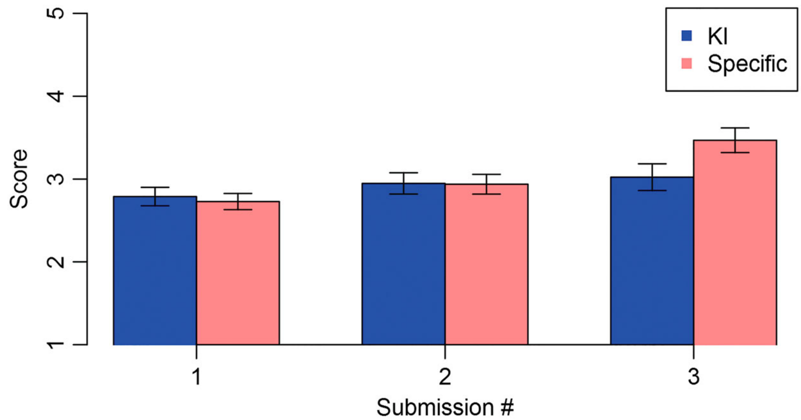


Figure 5. Knowledge integration scores for embedded Coal items, by condition. First submission indicates the score on the initial response produced for Coal 1. Submissions 2 and 3 indicate the score for the final revisions of Coal items 1 and 2, respectively.

number to determine whether the effect of condition was impacted by progress in the curriculum.

ANOVA results confirmed a significant effect of condition [$F(1, 137) = 5.4$, $p = .02$, $\eta_p^2 = .04$], a significant effect of submission number [$F(3, 411) = 50.6$, $p < .001$, $\eta_p^2 = .27$], and a significant interaction between condition and submission number [$F(3, 411) = 3.4$, $p = .02$, $\eta_p^2 = .02$]. The effect of submission number indicates that scores increased as students progressed through the unit – in part due to guidance and in part due to other sources of learning. Yet, as Figure 4 highlights, the interaction with condition indicates that score growth was impacted by experimental condition. Specifically, students in the *specific* condition made larger gains from their initial to final submission than students in the *knowledge integration* condition.

To compare conditions for embedded *Coal* responses (Figure 5), we performed a similar repeated-measures ANOVA on scores for the following submissions: (1) initial *Coal 1* response (prior to guidance); (2) final *Coal 1* revision (after guidance); (3) final *Coal 2* revision. ANOVA results revealed a nonsignificant effect of condition [$F(1, 91) = 5.4$, $p > .1$, $\eta_p^2 = .02$], a significant effect of *Coal* submission number [$F(2, 182) = 6.5$, $p = .002$, $\eta_p^2 = .07$], and a trend toward a significant interaction between condition and submission number [$F(2, 182) = 2.6$, $p = .08$, $\eta_p^2 = .03$]. In this case, the effect size for the interaction was similar to that found for *MySystem*, although the effect did not achieve full significance due to participant attrition.

Overall, these findings indicate some consistent advantages for the *specific* guidance condition on performance during the learning task. For both *MySystem* and *Coal* items, participants who received *specific* guidance were more likely to produce more accurate final responses than participants who received *KI* guidance.

> RQ2: How do specific and knowledge integration guidance impact retention of material and transfer to related posttest items?

Consistent with literature on desirable difficulties, we would expect that specific guidance would lead to better performance during learning but less retention. To test this question, we look at total score and then distinguish between novel items (not seen in the unit) and items that receive guidance during instruction. We calculated a *Total* score, out of 18, and conducted t-tests on both posttest *Total* and individual items. We used this approach, rather than calculating gain scores, because, in some cases, the pretest was conducted in pairs rather than individually. To ensure that no significant prior differences existed between the groups, we compared pretest means for *Total* and each individual KI-scored item by condition. No significant differences emerged [*Total-pre*: $t(323) = -1.6$; *MySystem-pre*: $t(312) = 0.1$; *Coal-pre*: $t(271) = -0.2$; *Energy Story-pre*: $t(280) = -1.2$, $ps > .1$]

Analyzing posttest results (Figure 6), there was no significant effect of condition on *Total*, *Coal*, or *MySystem* posttest scores [*Total-post*: $t(280) = 0.8$, $p > .1$, $d = 0.1$; *Coal-post*: $t(237) = 0.2$, $p > .1$, $d = 0.03$; *MySystem-post*: $t(276) = -0.8$, $p > .1$, $d = 0.1$]. However, for the novel *Energy Story-post,* participants in the *KI* condition scored significantly higher than participants in the *specific* condition [$t(255) = 2.5$, $p = .01$, $d = 0.31$]. Therefore, for the embedded items, the benefits accrued by the *specific* condition on the embedded *MySystem* and *Coal* items did not persist to the posttest. For the novel *Energy Story*, *KI* guidance benefitted participants' more than *specific* guidance.

RQ3: How do students' inquiry behaviors differ by condition and impact performance?

To measure autonomous learning, we looked at revisiting of the simulation. To simplify this analysis and avoid problems of attrition, we focused on students' revisits of the first simulation while revising the first *MySystem* item. As expected, those workgroups who chose to revisit the visualization produced significantly higher gains from initial to final diagram than those who did not revisit [revisited: $M = 1.1$, $SD = 1.4$; did not revisit: $M = 0.5$, $SD = 1.0$; $t(159) = 2.9$, $p = .004$, $d = 0.49$]. We also compared these sets of workgroups on gains from pretest to posttest (averaging together scores of all members of the workgroup, if pretest or posttest was completed individually). For this analysis, no significant differences between groups who chose to revisit the 1st simulation, or not, emerged on any of the posttest items.

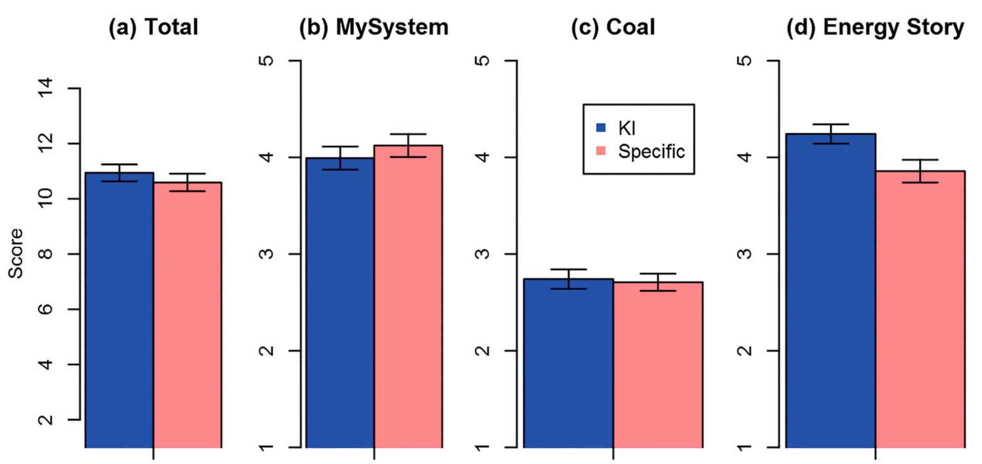


Figure 6. KI Scores for posttest total and individual items, by condition. Coal is a retention essay item from the unit, scored on a 5-point scale. *MySystem* is a retention item from the unit, scored on a 6-point scale. Energy Story is a novel essay item, scored on a 6-point scale.

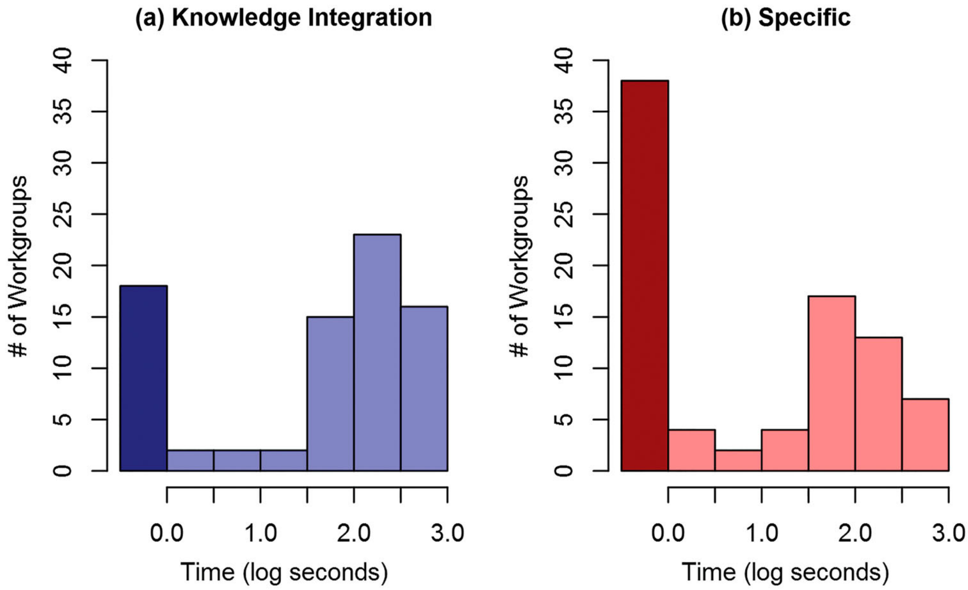


Figure 7. Histograms of time spent revisiting first simulation between initial and final construction of *MySystem* 1, by condition. For each condition, darker bar on left shows the number of workgroups who did not revisit the visualization. All other durations are log-transformed (base 10), such that 1 represents 10 seconds, 2 represents 100 seconds, and 3 represents 1000 seconds.

We tested whether the significant effect of revisiting on *MySystem 1* gains may have been moderated by condition. As Figure 7 displays, revisiting patterns differed by condition. Specifically, the number of workgroups who did not revisit the 1st simulation was greater in the *specific* condition than the *KI* condition [*KI*: did revisit = 60, did not revisit = 18; *specific*: did revisit = 47, did not revisit = 38; $\chi^2(1) = 7.5$, $p < .01$]. Yet, disaggregated by condition, the effect of revisiting on *MySystem 1* gains was similar for both conditions [*KI*: revisited: $M = 1.0$, SD = 1.5; did not revisit: $M = 0.3$, SD = 0.8; $t(76) = 2.1$, $p = .04$, $d = 0.57$; *specific*: revisited: $M = 1.2$, SD = 1.3; did not revisit: $M = 0.6$, SD = 1.1; $t(81) = 2.3$, $p = .02$, $d = 0.51$]. Once again, no significant effects, by condition, emerged for posttest items.

In addition to the choice to revisit the visualization or not, the quality of revisits likely impacted knowledge gains. While we do not have direct measures of actions taken while interacting with the simulation, previous research with complex climate simulations (Svihla & Linn, 2012) suggests that a great deal of time and effort is required to make sense of these complex visualizations. From this standpoint, more time spent on the simulation may be associated with stronger performance on related assessment items. To address this, we used log-transformed time data to normalize the data and highlight differences in duration magnitude.

To investigate the role of revisit duration, we analyzed Pearson correlations between total time spent revisiting (for those who did revisit at least one time) and various outcome measures. In contrast to the differences found comparing those who did revisit vs. those who did not, no significant association between *MySystem 1* gain scores and time revisiting emerged [$r = .10$, $t(105) = 1.0$, $p > .1$]. Likewise, no significant association between time revisiting and posttest gains on *MySystem* emerged [$r = .16$, $t(102) = 1.6$, $p > .1$]. However, in the case of the *Energy Story,* there was a significant association between time revisiting and posttest gains [$r = .23$, $t(99) = 2.3$, $p = .02$].

To test the effect of condition, we compared workgroups that chose to revisit in each condition. Workgroups in the *KI* condition spent significantly more time revisiting than workgroups in the *specific* condition [*KI*: $t(105) = 2.3$, $p < .05$, $d = 0.45$]. Yet, paralleling the general result, no significant effect of total time revisiting and *MySystem 1* gains emerged for either condition [*KI*: $r = .12$, $t(58) = 0.9$, $p > .1$; *specific*: $r = .11$, $t(45) = 0.7$, $p > .1$].

In contrast, for the pre-post *Energy Story,* we found an association between time spent revisiting and gains ($r = .23$), by condition. Specifically, in the *KI* condition, a (marginally) significant correlation emerged [$r = .26$, $t(55) = 2.0$, $p = .05$], whereas in the *specific* condition, no significant correlation emerged [$r = .11$, $t(42) = 0.7$, $p > .1$]. Similarly, those workgroups in the *KI* condition who revisited the first simulation achieved significantly higher posttest gains on the *Energy Story* than those in the *specific* condition [*KI*: $M = 1.4$, SD = 1.2; *specific*: $M = 0.7$, SD = 1.3; $t(99) = 2.9$, $p = .004$, $d = 0.59$]. In contrast, for those who did not revisit, there was no difference between conditions on *Energy Story* posttest gains [*KI*: $M = 1.3$, SD = 1.7; *specific*: $M = 1.0$, SD = 1.3; $t(50) = 0.7$, $p > .1$].

In addition to guidance condition, teacher-level differences in implementation may have impacted the value of revisiting the simulation. For example, in Teacher A_1's class, the teacher – who had taught this unit previously – displayed the simulation on a digital projector to discuss observations, thus reducing the need to revisit the simulation. This likely led to a low correlation between time spent revisiting the simulation and gains on the Energy Story pre–post item ($r = .13$).

On the other hand, in the other two classes in School A, the teachers – who were new to the curriculum – did not engage in any whole-class instruction, thereby maintaining the importance of autonomous inquiry. As such, in these classrooms, there was a strong relationship between simulation revisiting and Energy Story gains ($r = .56$). Additionally, students of Teacher A_1 demonstrated higher gains on the pre-post Energy Story than students of Teachers A_2 and A_3 [Teacher A_1: Mean = 2.0, SD = 1.2; Teacher A_2 and A_3: Mean = 1.4, SD = 1.3; $t(102) = 2.3$, $p = .02$, $d = 0.5$].

Likewise, teacher C was unfamiliar with the curriculum unit and did not engage in whole group discussion. As such, students in Teacher C's class gained less on Energy Story than Teacher A_1 [Teacher A_1: Mean = 2.0, SD = 1.2; Teacher C: Mean = 0.6, SD = 1.6; $t(125) = 5.0$, $p < .001$, $d = 0.9$]; although the socioeconomic obstacles in School C (e.g. limited computer literacy, language difficulties) likely influenced this result.

Discussion

The results indicate that the *specific* guidance facilitated stronger performance during the *global climate change* curriculum than *KI* guidance. A general effect of condition in the repeated-measures ANOVA of the embedded *MySystem* demonstrated an overall advantage for the *specific* condition, while an interaction with submission number suggests that this advantage accumulated over the unit. In the case of the written *coal* items, a condition by submission number interaction showed a trend toward significance. Taken together, these results indicate that *specific* guidance is more likely to produce immediate gains for students. On the other hand, posttest results did not favor the *specific* condition, and in the case of the transfer item, that is, *Energy Story*, results favored the *KI* condition.

These posttest results suggest that while the *specific* guidance helped students to add detail to their embedded *MySystem* steps, this guidance was less effective in promoting integrated understanding of general concepts, as measured by *Energy Story*. It may be the case that the *specific* guidance provided struggling students a straightforward hint that circumvented deeper processing of information. For example, if prompted to change the link between the Sun and Earth's surface from 'heat' to 'solar radiation', the participant could heed this guidance effectively without reflecting on the scientific foundation for this concept.

Conversely, *KI* guidance was more effective at promoting general concepts than specific representations within *MySystem*. This may be due to students' greater likelihood in the *KI* condition to revisit previous visualizations. Workgroups in the *KI* condition spent more time revisiting earlier visualizations than workgroups in the *specific* condition. Thus, students who revisited in the *specific* condition may have learned as much as those who revisited in the *KI* condition, but they simply chose to do so less frequently.

Yet, there is some evidence that quality of interactions with previous visualizations differed by condition as well. More time on visualizations was associated with higher *Energy Story* posttest scores in the *KI* condition, while the same association did not emerge in the *specific* condition. Additionally, of those who did revisit the visualization, students in the *KI* condition outperformed students in the *specific* condition on the posttest *Energy Story*, whereas no difference emerged for those who did not revisit. We suspect that this advantage for *KI* guidance was due to its design that cued students to attend to important features of the visualization (in order to find the obscured feature of the guidance image). Highlighting key features is of central importance for making sense of complicated scientific visualizations (Lin & Atkinson, 2011).

While this is a complex pattern of results, they suggest that students' use of autonomous inquiry strategies, motivated by *KI* guidance, mediated learning outcomes. These inquiry behaviors not only include revisiting a previous simulation, but doing so purposefully to address a specific gap in understanding. While it is not clear why some participants in the *specific* condition chose to revisit the visualization, their revisiting had little influence on posttest performance. Thus, while providing students with the opportunity to interact multiple times with a complex visualization may only be beneficial if students are assigned a concrete task.

In general, these results fit with previous findings on knowledge integration and desirable difficulties. More challenging tasks, which inherently require deeper processing, facilitate better integration of ideas, but at the cost of short-term performance (Bjork & Linn, 2006). In this case, investigating a complex visualization supported students' knowledge integration and transfer, but required more deliberate efforts.

Limitations and future research

Laboratory and classroom studies require different procedures to account for vastly different goals and constraints, and therefore may result in different outcomes (Richland, Linn, & Bjork, 2007). For example, many laboratory studies find a benefit for the more difficult condition on later recall of the material being directly studied during training. However, in this study, no difference between conditions emerged on the directly studied *MySystem* and *Coal* items at posttest.

Procedurally, this study differs from some laboratory studies of desirable difficulties where participants are trained to obtain mastery on the target material to ensure that neither group began the posttest with an advantage (e.g. Vitale, Black, & Swart, 2014). In studies that use a training-to-mastery approach, performance differences are measured in terms of time to obtain mastery. In our case, however, participants in the *specific* guidance condition completed their final curriculum-embedded diagram more accurately, and therefore had greater potential to achieve a high score at posttest by recalling recent work. An alternative experimental paradigm, compelling students to achieve mastery on each *MySystem* item before proceeding, might achieve different results at posttest. Additionally, because the effects of desirable difficulties are often found after a delay, it may be the case that differences between the conditions would have emerged at a delayed posttest. Due to time constraints and classroom realities, these approaches were not feasible in the current study.

More generally, for many of our analyses, effects that were found ranged from small to moderate. By pairing students during the unit, embedded assessment analyses lost significant power. Furthermore, by asking workgroups to contribute common responses, we could not determine the unique contributions of each participant during the learning process. Thus, while there may be logistical and collaborative advantages to paired workgroups, this approach makes it difficult to track ideas generated in pairs to responses given individually.

Future research with qualitative methods is needed to shed light on the relationship between groups and individuals. For example, Visintainer and Linn (2015) interviewed students after completing a previous version of the climate change curriculum, and discovered a strong relationships between their ideas expressed in the responses given during the WISE curriculum and posttest interviews. Combining the quantitative design of this study and the qualitative design of the previous study would likely illuminate further interesting relationships between ideas generated through the course of study.

Furthermore, the generally positive relationship between inquiry behaviors and learning gains may be due to the characteristics of students (e.g. mindfulness) who chose to revisit, rather than experiences during revisiting. Both qualitative methods and event log data may reveal relationships between inquiry behaviors and learning. Alternatively, future research could experimentally manipulate (e.g. compel, restrict) access to previously encountered simulations for some students. Yet, in consideration of logistical and curricular needs of students, such a study may be more suitable to laboratory study.

Finally, while our analyses suggest that there were different patterns of behavior and learning across our various teachers and student populations, it is difficult to place the sources of these differences on any single characteristic of the classroom. For example, teacher A1 engaged in several curriculum-specific discussions, but also had generally more experience teaching than the other teachers, leading to high pre-post gains. Conversely, while teacher C's students were generally socioeconomically disadvantaged, the teacher had little prior experience working with online curriculum, leading to lower pre-post gains.

These differences in implementation and experiences reflect the inherent complexities of classroom-based research. While expected, they raise important questions about generalizability. Namely, to what extent are the effects of our treatment due to specific classroom factors? Overall, we found that the *specific* condition afforded no clear, long-term advantages, relative to *KI*, across any sites. Thus, while the advantage for *KI* guidance may be moderated by particular classroom characteristics, we believe that this approach provides a stronger basis for ongoing curriculum design and research. In future studies, we will continue to explore the intersection of *KI* guidance and teachers' instructional strategies to delineate guidelines and practices that enable teachers to optimize their use of automated technologies in the classroom.

Conclusions

Inquiry activities provide students with the opportunity to engage in authentic practices used by scientists, while addressing complex content (Linn & Eylon, 2011). Yet, a number of researchers advocate directed instructional approaches due to their efficiency. For example, Adams, Mayer, MacNamara, Koenig, and Wainess (2012) suggest that neither narrative context nor discovery learning is likely to help students retain or transfer target scientific concepts.

In contrast, this study suggests that guided inquiry can be equally effective or more effective than specific guidance. Specifically, students asked to re-engage with a difficult visualization in a targeted manner were more likely to succeed on the novel *Energy Story* essay than those who were told how to add an idea or fix an error in their responses. Therefore, removing self-guided inquiry from educational activities is likely to leave students with shallow knowledge. At the same time, providing clear steps to direct student behavior is necessary to avoid inefficient strategies and confusion.

More generally, the pattern of results suggests that laboratory findings on desirable difficulties are applicable to the classroom and can foster knowledge integration. Yet, a desirable difficulties approach should be applied with caution. Difficult tasks have the potential to stimulate or overwhelm students. Knowledge integration and similar approaches to inquiry science provide guidelines for designing appropriately challenging activities, but require significant professional development to gain competence (Gerard, Varma, Corliss, & Linn, 2011).

Well-designed learning tasks should focus students on their own ideas with clear tasks (e.g. answering a question about a visualization), allowing them to focus on limited, concrete ideas. In regard to Koedinger and Aleven's (2007) 'assistance dilemma', we believe these results support a middle ground in which specific inquiry tasks are provided in lieu of more vague guidance that leaves students confused or more narrow guidance that short-circuits thoughtful activity. Effective guidance therefore requires instructional designers to clearly define common student difficulties, develop a set of targeted inquiry tasks to resolve these difficulties, and use automated guidance to set students on the right trajectory. Further research will be necessary to apply this and similar procedures to a wider range of subject areas and student populations. As self-guided, online curriculum becomes more prevalent in the classroom, we believe that this effort is necessary to ensure that all students are successful.

Note

1. Socioeconomically disadvantaged is defined as a student who is eligible for free or reduced-price lunch or a student for whom neither parent has received a high school diploma.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by National Science Foundation (NSF) DR K-12 grant 1119670. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF.

Notes on contributors

Jonathan Vitale is a postdoctoral researcher who investigates the role of graphs and models in science learning. Jonathan received his Ph.D. from Teachers College, Columbia University and was formerly a high school mathematics teacher in New York City.

Elizabeth McBride is a PhD candidate in Science and Mathematics Education at the University of California, Berkeley. She has earned as Masters of Engineering in Applied Climate Science and a Bachelors of Science in Engineering in Climate and Space Sciences, both from the University of Michigan. She is interested broadly in strengthening connections between science and engineering education at the secondary school level, technology in classrooms, and using data mining approaches to understand student learning.

Marcia C. Linn is Professor of Development and Cognition, specializing in science and technology in the Graduate School of Education, University of California, Berkeley. She is a member of the National Academy of Education and a Fellow of the American Association for the Advancement of Science (AAAS), the American Psychological Association, and the Association for Psychological Science. Her books include Computers, Teachers, Peers (2000), Internet Environments for Science Education (2004), Designing Coherent Science Education (2008), WISE Science (2009), and Science Teaching and Learning: Taking Advantage of Technology to Promote Knowledge Integration (2011).

ORCiD

Jonathan M. Vitale http://orcid.org/0000-0002-0342-4875

References

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology*, *104* (1), 235–249.

Anderson, J. R., Boyle, C. F., Corbett, A., & Lewis, M. W. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modeling cognition* (pp. 93–134). New York, NY: Wiley.

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.

Anderson, J. R., & Schunn, C. (2000). Implications of the ACT-R learning theory: No magic bullets. *Advances in Instructional Psychology*, *5*, 1–34.

Appleton-Knapp, S. L., Bjork, R., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research*, *32*(2), 266–276.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). *Off-task behavior in the cognitive tutor classroom: When students 'game the system'*. In Proceedings of ACM CHI 2004: Computer-Human Interaction. Alagoas, Brazil (pp. 383–390).

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238.
Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: Introducing desirable difficulties. *APS Observer*, *19*(3), 29–39.
Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*(1), 21–32.
Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *79*(4), 474–482.
Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23–56). Washington, DC: National Academy Press.
DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*(6), 945–955.
Gerard, L. F., Varma, K., Corliss, S. B., & Linn, M. C. (2011). Professional development for technology-enhanced inquiry science. *Review of Educational Research*, *81*(3), 408–448.
Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, *22*(4), 521–563.
van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, *32*, 83–98.
van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*(2), 154–164.
Hansen, P. J. K. (2010). Knowledge about the greenhouse effect and the effects of the ozone layer among Norwegian pupils finishing compulsory education in 1989, 1993, and 2005 – what now? *International Journal of Science Education*, *32*(3), 397–419.
Hmelo-Silver, C. E., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *The Journal of the Learning Sciences*, *15*(1), 53–61.
Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154.
Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424.
Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.
Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.
Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.
Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*, *10*(3), 285–291.
Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers and Education*, *56*(3), 650–658.
Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, *87*(4), 517–538.
Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York, NY: Routledge.
Linn, M. C., Gerard, L., Ryoo, K., McElhaney, K., Liu, O. L., & Rafferty, A. N. (2014). Education technology. Computer-guided inquiry to improve science learning. *Science*, *344*(6180), 155–156.
Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Erlbaum, Mahwah, NJ.

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

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, *29*(4), 587–625.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14–19.

McElhaney, K. W., Chang, H.-Y., Chiu, J. L., & Linn, M. C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education*, *51*(1), 49–85.

McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction*, *5*(4), 381–413.

Phye, G., & Sanders, C. (1994). Advice and feedback: Elements of practice for problem solving. *Contemporary Educational Psychology*, *19*, 286–301.

Pridemore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology*, *20*(4), 444–450.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., & Duncan, R. G. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, *13*(3), 337–386.

Richland, L. E., Linn, M. C., & Bjork, R. A. (2007). Cognition and instruction: Bridging laboratory and classroom settings. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 1–60). New York, NY: Wiley.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*(2), 267–280.

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.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Singer, J., Marx, R. W., Krajcik, J., & Chambers, J. C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist*, *35*(3), 165.

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

Svihla, V., & Linn, M. C. (2012). A design-based approach to fostering understanding of global climate change. *International Journal of Science Education*, *34*(5), 651–676.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89.

Van Den Boom, G., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567.

Visintainer, T., & Linn, M. (2015). Sixth-grade students' progress in understanding the mechanisms of global climate change. *Journal of Science Education and Technology*, *24*(2–3), 287–310.

Vitale, J. M., Black, J. B., & Swart, M. I. (2014). Applying grounded coordination challenges to concrete learning materials: A study of number line estimation. *Journal of Educational Psychology*, *106*(2), 403–418.

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

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories – an embodied modeling approach. *Cognition and Instruction*, *24*(October 2014), 37–41.

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, *33*, 131–146.