Promoting Student Learning through Automated Formative Guidance on Chemistry Drawings

Anna N. Rafferty, Computer Science Division, University of California, Berkeley, CA, USA
rafferty@cs.berkeley.edu
Libby Gerard, Graduate School of Education, University of California, Berkeley, CA, USA
libby.gerard@gmail.com
Kevin McElhaney, SRI Education, Menlo Park, CA, USA, kevin.mcelhaney@sri.com
Marcia C. Linn, Graduate School of Education, University of California, Berkeley, CA, USA
mclinn@berkeley.edu

Abstract: We investigated the effect of automated guidance on student-generated chemistry drawings in computer-based learning activities. Expert teachers provide guidance on generative tasks such as drawings or essays that encourages students to refine their understanding, often by gathering more evidence. We developed algorithms to score student drawings and designed guidance for each score level. The guidance was intended to promote coherent understanding. We compared computer-generated guidance to teacher guidance in two studies, conducted with over 300 students in secondary classrooms. The studies suggest that automated guidance is as effective as teacher guidance for improving student understanding. Teachers appreciated the assessment of class progress provided by the automated guidance. They reported that it took them several hours to grade their five classes of 30 to 40 students. Thus, automated guidance can reduce the time teachers spend evaluating student work, creating more time for planning lessons, facilitating inquiry, or guiding individual students.

Computer-assisted education has the potential to deliver timely guidance adapted to each student's individual ideas. Human tutors provide adaptive guidance by prompting learners to reconsider and revise their ideas, verify and elaborate on the correctness of ideas and consider ways to improve understanding (Merrill, Reiser, Ranney, & Trafton, 1992). Providing adaptive guidance is an important goal in designing computer tutors (e.g., Anderson, Boyle, Farrell, & Reiser, 1987; Anderson, Corbett, Koedinger, & Pelletier, 1995). While the majority of computer tutors provide formative guidance (Koedinger, Anderson, Hadley, & Mark, 1997; Slotta & Linn, 2009), it is often limited to student work on selection tasks (e.g. multiple-choice) or algebraic expressions. This paper explores the effect of automated, adaptive guidance on a generation task where students make drawings of chemical reactions as part of a web-based inquiry science unit.

Compared to the limited number of correct responses to selection tasks, generation tasks can adapt guidance to a wide range of student responses. Selection tasks often encourage students to recall facts rather than distinguish among ideas, and rarely provide opportunities for deep student inquiry (Shepard, 2000). Generative tasks, in contrast, elicit students' range of ideas and can encourage them to use evidence to sort out their ideas in order to create a coherent explanation. Mintzes, Wandersee, and Novak (2005) note that generative assessments can provide a fuller picture of students' conceptual understanding and drive students towards "making meaning'" rather than memorizing facts. Generative tasks can be difficult to evaluate due to the variety of responses and innumerable ways for students to express the correct answer. Previous research has found that due to the demands required in evaluating generative assessments, it is often challenging for teachers to provide detailed guidance to all students (Black & William, 1998; Ruiz-Primo & Furtak, 2007).

In this paper, we explore the effect of automated, adaptive guidance for student-generated drawings of chemical reactions as they interact with a web-based inquiry science unit. Drawings provide students with a way to express their understanding of atoms and molecules in chemical reactions (Chang, Quintana, & Krajcik, 2010). We support automated analysis of drawings by asking students to use a computer-based interface that features virtual atom stamps, rather than enabling open-ended drawings. This limits the degree to which student drawings can vary while still allowing for expression of multiple conceptual views. We designed an algorithm that diagnoses both normative and alternative chemistry conceptions in students' drawings, allowing us to align formative guidance with these conceptions. Our guidance design addresses the gap between students' ideas and the learning goal by prompting students to build on productive ideas they have, develop criteria for evaluating their own understanding, and revisit key concepts.

We explore the effectiveness of automated guidance through two classroom studies. The first investigates whether the automated guidance is as effective for promoting student learning as teacher-generated guidance. Comparing these two types of guidance allows us to identify key characteristics of effective teacher guidance, potentially informing future revisions to the automated guidance, and provides data for determining whether it is feasible to remove some of the demands on teachers by providing students with automated guidance. In this study, students receive automated guidance immediately, while they must wait until the next day to receive teacher-generated guidance, mirroring a typical classroom. The second study explores the role of immediacy by comparing delayed to immediate automated guidance. Together, these two studies address optimal ways to provide automated guidance on student drawings within the classroom and investigate how this guidance compares to that provided by teachers.

Designing Guidance for Inquiry

Guidance for generative assessments can help students improve their understanding and recognize gaps or inconsistencies in their ideas (Hattie and Timperley, 2007). It can promote learning by encouraging students to reconsider their ideas, building on their prior knowledge (Azevedo & Bernard, 1995). Generative assessments can be used to help teachers recognize students' level of understanding and adapt instruction. Ruiz-Primo and Furtak (2007) found that teachers' guidance on generative activities in an inquiry investigation was related to their students' science learning, suggesting that this monitoring can indeed help teachers boost student learning.

Providing guidance on generative assessments during instruction is difficult. Teachers often lack time to provide detailed guidance for all students on these assessments (Black & William, 1998; Ruiz-Primo & Furtak, 2007). Further, the specificity that effective guidance should exhibit remains unclear. For instance, generic guidance may prompt students to self-explain and generate their own insights (Butler & Winne, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Schmidt & Bjork, 1992) but may also allow non-normative ideas to persist (Koedinger & Aleven, 2007). Students with different levels of prior knowledge benefit from different levels of information specificity (Shute, 2008) and scaffolding (Razzaq & Heffernan, 2009). Due to the challenges of assessing and guiding student work on generative activities, selection tasks with one correct answer, such as multiple-choice questions, are the norm in science instruction. These activities are limited in their ability to capture the complexity of students' ideas.

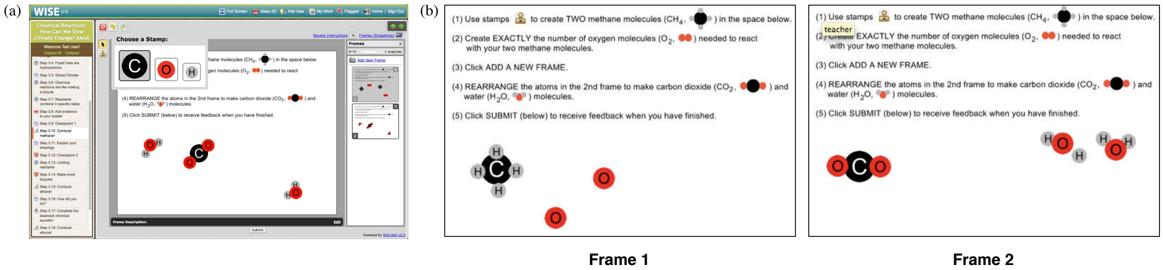
New technologies offer promise for implementing guidance strategies to support inquiry learning. In AutoTutor (Graesser et al., 2004), for instance, a computer avatar leads a tutorial dialogue as a student solves a challenging physics word problem. The avatar prompts for more information, elicits questions, identifies and corrects "bad answers," answers the learner's questions, and summarizes responses. Across different domains and comparison groups the average learning gain is approximately one letter grade. Automated guidance has also proved effective in helping students to develop concept maps in different middle school science domains and use evidence to strengthen the links among their ideas (Segedy, Kinnebrew, & Biswas, 2013). More recently, researchers have employed machine learning techniques to automatically recognize effective inquiry practices (Sao Pedro, Baker, Gobert, Montalvo, & Nakama, 2013).

Our work adds to this body of literature on automated guidance in inquiry science by examining the effectiveness of automated guidance for drawing tasks in which students pictorially represent scientific ideas. We compare automated guidance with teacher-generated guidance, allowing us to explore what types of guidance teachers provide to students. For the automated guidance, we designed guidance messages based on knowledge integration principles. Knowledge integration is based on constructivist ideas that focus on building on students' prior knowledge and helping them to connect new concepts with this knowledge, even if some of this prior knowledge is non-normative (e.g., Smith III, Disessa, & Roschelle, 1994). Knowledge integration guidance can assist students by prompting them to compare and contrast their views with evidence or scientific theories, or to add new ideas missing from their initial conception (Linn & Eylon, 2006). When guidance directly builds on students' own ideas, as articulated in their initial response to the activity, it can help students develop criteria for distinguishing between normative and non-normative ideas and push students to integrate ideas rather than holding separate, conflicting conceptions (Linn & Eylon, 2011).

Curriculum and Drawing Activity

We focus our investigation of automated guidance on students' drawings of chemical reactions. The drawing tasks are part of an inquiry unit in the Web-Based Inquiry Science Environment (WISE), entitled *Chemical Reactions: How Can We Help Slow Climate Change?* (Chiu & Linn, 2011). *Climate Change* addresses difficulties students have with understanding chemical reactions as composed of discrete particles (Ben-Zvi, Eylon, & Silberstein, 1987) by highlighting the role of conservation of mass, ratios, and excess reactants in combustion reactions.

Past work has shown that learning multiple representations of chemical reactions and providing students with ways of visualizing the particles in reactions can help to strengthen understanding (Harrison & Treagust, 2000; Schank & Kozma, 2002). The drawing tasks ask students to draw the arrangement of atoms before and after a chemical reaction [Figure 1(a)]. One of the tasks focuses on the combustion of methane and the other ethane. The WISE Draw screen provides students with "stamps" for each atom; for instance, the methane reaction includes stamps for oxygen, carbon, and hydrogen. Students must choose how many of each atom to add to their drawing and arrange the atoms to reflect how they are grouped into molecules. The draw interface allows students to create multiple *frames*, one to show the atoms before the chemical reaction (reactants) and one to show the atoms after the chemical reaction (products) [Figure 1(b)]. After creating a new frame, students rearrange the atoms to show the products of the reaction. This activity presents chemical reactions as the rearrangement, rather than the creation or destruction, of matter. Students may still add or delete atoms in a way that reflects their conceptual misunderstandings about conservation of mass. The drawings enable students to articulate their ideas about chemical reactions, while constraining the representation to enable automatic evaluation. The drawings provide an opportunity to work with a different model of chemical reactions than the typical equation-based format, and students frequently demonstrate non-normative ideas in the activity. Our goal is to provide conceptual guidance targeting non-normative or missing ideas in the students' drawings.



(1) Use stamps to create TWO methane molecules (CH_4,) in the space below.

(2) Create EXACTLY the number of oxygen molecules (O_2,) needed to react with your two methane molecules.

(3) Click ADD A NEW FRAME.

(4) REARRANGE the atoms in the 2nd frame to make carbon dioxide (CO_2,) and water (H_2O,) molecules.

(5) Click SUBMIT (below) to receive feedback when you have finished.

(1) Use stamps to create TWO methane molecules (CH_4,) in the space below.

(2) Create EXACTLY the number of oxygen molecules (O_2,) needed to react with your two methane molecules.

(3) Click ADD A NEW FRAME.

(4) REARRANGE the atoms in the 2nd frame to make carbon dioxide (CO_2,) and water (H_2O,) molecules.

(5) Click SUBMIT (below) to receive feedback when you have finished.

Figure 1. Drawing and feedback interface. (a) The WISE Draw interface. Students use "stamps," the black, red, and gray circles, to represent atoms and show the molecules before and after a chemical reaction. (b) The frames of a student drawing. The scorer recognizes that CH_4, CO_2, and H_2O are allowed molecules, but that the student has not conserved mass and has placed two separate oxygen atoms rather than an oxygen molecule.

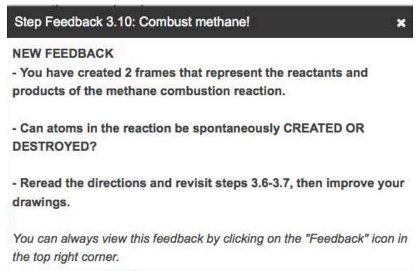
Evaluating Student Drawings

To evaluate student drawings, we created an algorithm that processes each drawing and assigns a score. Based on examination of 98 drawings from past students, half methane and half ethane, we identified common student ideas and grouped these ideas into conceptual categories, shown in Figure 2(a). Each category includes a different conceptual feature, such as conserving mass from the beginning to the end of the reaction or correctly representing the reactants. The concepts are organized into a hierarchy from more basic to more complex. We evaluated the accuracy of the algorithm on the development set of 98 drawings as well as on a test set of 200 additional drawings from past students. Both sets of drawings were scored by a trained human scorer, and the test set was not examined until after the algorithm was developed. For the development set, the algorithm's score matched the human's score for 96.9% of the drawings; for the test set, the scores matched in 91.5% of the drawings. This compares favorably with other systems for scoring student answer to generative assessments; for instance, the C-Rater system scores short answer responses, including responses to science prompts, and matched human scores about 84% of the time in two separate evaluations (Leacock & Chodorow, 2003). Note that while development focused on the methane and ethane tasks, the system can score other chemical reactions drawings. Information about the drawing task is provided as an XML input file specifying the correct molecules in each frame, allowing the scoring algorithm to be agnostic to the specific task being scored.

(a)

Criteria	0	1	2	3	4	5
Two frames		✓	✓	✓	✓	✓
Conserves atoms			✓	✓	✓	✓
Reactants correct				✓	✓	✓
Products correct					✓	✓
Groupings clear						✓
Rate in dev. set	11%	19%	16%	5%	3%	45%

(b)



Step Feedback 3.10: Combust methane!

NEW FEEDBACK

- You have created 2 frames that represent the reactants and products of the methane combustion reaction.

- Can atoms in the reaction be spontaneously CREATED OR DESTROYED?

- Reread the directions and revisit steps 3.6-3.7, then improve your drawings.

You can always view this feedback by clicking on the "Feedback" icon in the top right corner.

Figure 2. Drawing rubric and automated guidance window. (a) The automated scoring rubric. Drawings receive the highest score for which they exhibit all checked criteria. (b) An automated guidance window. Students are given personalized guidance, designed to promote knowledge integration, based on the automated scoring rubric. This guidance message corresponds to a score of 1, where the drawing has not conserved mass.

Creating Guidance from Scores

Given the scorer's ability to accurately evaluate student drawings, we can provide guidance based on the conceptual understanding that the student has. For each of the six possible scores, we designed a textual message to help students revise their drawing. The textual guidance was designed to promote knowledge integration by recognizing students' normative ideas and helping them to refine and revise their non-normative ideas (Linn & Eylon, 2011). Drawings that were scored as having some conceptual error (scores 0-4) all received textual feedback of a similar format. First, a correct feature of the drawing was recognized, anchoring the guidance with students' prior knowledge. For example, a student whose drawing received a score of 2 would be acknowledged for conserving mass, since this is the most complex conceptual feature exhibited by the drawing. The textual feedback then posed a question targeting the student's conceptual difficulty, such as identifying what molecules should be present in the reactant frame; this elicits student ideas about the topic of difficulty. Finally, the feedback directed students to a relevant step earlier in the unit, and encouraged them to review the material in that step and then revise their drawing. This promotes adding new ideas and distinguishing normative and non-normative ideas. Figure 2(b) shows the guidance for a score of 1.

Study 1: Comparing Student Learning with Teacher vs. Automated Guidance

We compare automated knowledge integration guidance to teacher-designed guidance on students' chemistry drawings. If computer-selected knowledge integration guidance could help students improve, the computer could save the teacher valuable time, which could be spent planning instruction and working with individual students. Computer-assigned guidance differs from teacher-generated guidance in several ways. The teacher-generated guidance can be customized based on the teacher's knowledge of individual students. For example, a teacher might respond to a conservation of mass error differently if the drawing was made by a student who typically struggles in science versus a student who typically excels. The teacher is also likely to be able to differentiate conceptual misunderstandings from errors due to use of the interface. The timing of the guidance also differs. Automated guidance is provided to students immediately, allowing them to revise their understanding before moving on in the unit. Previous studies suggest mixed results for immediate and delayed guidance (Shute, 2008). Given that immediacy is a unique affordance of computer-assigned guidance, we wanted to test its value. Overall, we hypothesize that the automated guidance will be as effective as teacher-generated guidance for promoting student learning given that the automated guidance targeted common conceptual errors and was designed to promote cohesive integration of ideas.

Methods

Participants

Eighth grade physical science students (N=263 completed both pre- and post-test, N[groups]=129; ages typically 13-14) from 10 classes in a public middle school participated in the study. Classes were taught by one of two teachers; each teacher taught five of the ten classes. Teachers were selected who have over three years experience teaching the WISE *Climate Change* unit and writing guidance for student work on the drawing tasks.

Study Design and Administration of Feedback.

Students were randomly assigned by class period to receive either automated or teacher-generated guidance. Three class periods from each teacher were assigned automated guidance (AG), and the other two periods teacher guidance (TG). Students completed the *Climate Change* unit in the classroom as part of the curriculum.

Students in both conditions completed the same pre-test prior to *Climate Change.* The same items were administered as a post-test after *Climate Change.* Pre- and post-tests were completed individually.

Students worked through *Climate Change* in groups of one to three students. All students experienced the same activities in *Climate Change* except for the draw steps. The two draw steps occurred in a part of the unit focusing on combustion reactions and their contributions to climate change. In the draw steps, all students received the same instructions about the use of the WISE Draw interface and the chemical reaction to depict. Students in the automated condition were told to click the "Submit" button when they wished to receive feedback. When students clicked this button, they were warned that they only had two chances to receive feedback and to confirm that they wanted to proceed. After confirming, a pop-up box with the textual feedback appeared. Students could close the feedback or re-open it to view their existing feedback at any time. If students clicked the "Submit" button more than twice, they were told that they had used all of their opportunities to receive feedback, but that they could continue to revise their drawing if they wished.

In the teacher guidance condition, the researchers met with the teachers before the start of the unit to review previous student drawings on this item and discuss a possible scoring rubric and guidance approaches. After the unit started, the teachers reviewed student work at night after all students had completed the drawing steps, and wrote guidance. Students received the guidance at the start of class the next day when they logged into the WISE unit. After logging in, a pop-up told students they had received guidance from their teacher on the draw task and provided them a link to jump immediately to that step in the unit. The guidance interface on the draw steps was identical to the interface in the automated condition.

The two conditions differed in how many rounds of guidance were provided: students in the automated guidance could receive up to two guidance messages, while students in the teacher-generated condition could receive only one. Automated guidance can be provided multiple times without significantly lengthening the unit; providing multiple rounds of teacher guidance, given that teachers need to review student work after class, is infeasible given the number of days available for students to work on *Climate Change* and the location of the drawings in the activity. While this distinction means that we cannot draw conclusions about whether teacher-generated guidance would be more effective if multiple rounds were provided, we believe this design better reflects the way that guidance would actually be available to students and thus can provide evidence for the best way to incorporate guidance in classroom activities.

Data Sources and Scoring of Knowledge Assessments

Measures of learning included students' initial and revised drawing for each of the methane and ethane drawing tasks, and an item from the pre- and post-tests. The pre-post test item called for students to transfer ideas learned in the drawing tasks, critiquing a molecular drawing made by a hypothetical peer for a new chemical reaction formula. A knowledge integration rubric was used to score student responses on the drawings and the item on the pre- and post-tests. The five-point scoring rubric evaluated student response in terms of integrating ideas about conservation of mass, ratios and excess in chemical reactions. Unlike the automated scoring system, this rubric was not hierarchical. We also conducted teacher and student interviews during enactment, documented the teacher-generated guidance, and collected rich classroom observation notes.

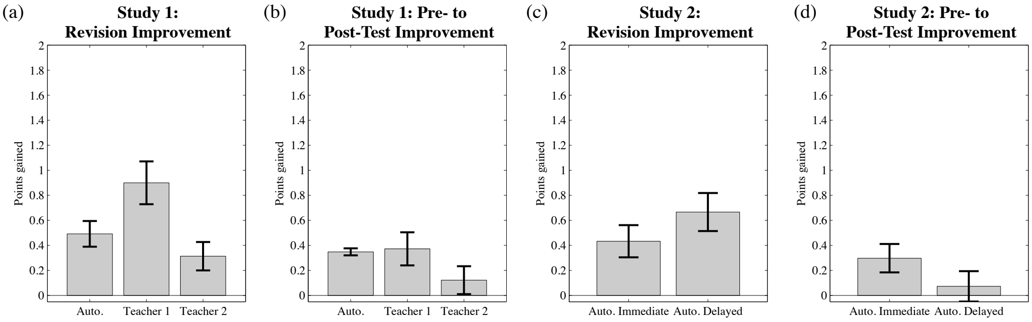


Figure 3. Improvements based on guidance type. Error bars show one standard error. (a) Revision improvement in Study 1. (b) Pre- to post-test improvement in Study 1. (c) Revision improvement in Study 2. (d) Pre- to post-test improvement in Study 2.

Results

Overall, the automated guidance was as helpful for student learning as the teacher-generated guidance. Students made modest improvements from their initial drawings to their final drawings, increasing their scores by an average of 0.56 points (Cohen's *d*=.36). Students in both the AG and TG conditions showed similar amounts of improvement (*d*=.33 and *d*=.39, respectively). A repeated measures analysis of variance (ANOVA) with factors for condition and initial versus final drawing showed significant improvement after revising the drawing ($F(1,481)=50.36$, $p<.001$) but no significant interaction between condition and amount of improvement ($F(1,481) = 0.85$, $p > .3$).

Students also showed improvement from pre- to post-test on the transfer item in which they critiqued the drawing of a hypothetical peer. Students in both conditions showed similar improvements in performance on this item (AG: $t(154)=4.63$, $p < .001$, $d=.42$; TG: $t(107)=2.93$, $p < .01$, $d=.36$). A repeated measures ANOVA showed that there was no significant interaction between guidance type and amount of improvement.

Analysis of the teacher-generated guidance demonstrated substantial differences in the quality of guidance given by the two teachers. One teacher wrote substantially more detailed comments that focused on both chemistry concepts and features of the drawings, whereas the other teacher wrote more terse comments; both teachers mentioned that providing guidance to students took several hours in the evening.

Students of the teacher who wrote more conceptual comments made significantly greater improvements on their drawings. An analysis of variance on the amount of improvement in drawing scores from initial drawing to final revision, with a factor for guidance type (automated, Teacher 1, or Teacher 2) and a random factor for student group showed that guidance type had a significant effect on amount of improvement ($F(2,481) = 4.04$, $p < .025$). As Figure 3(a) shows, students who received more conceptual guidance (Teacher 1) improved more than students in other conditions did, and students who received automated guidance improved more than students who received terse guidance (Teacher 2). While this interaction was not significant for pre- to post-test improvement, the same trend held: students who received guidance from Teacher 1 improved an average of 0.37 points (*d*=.48), students in the automated condition improved 0.35 points (*d*=.42), and students who received guidance from Teacher 2 improved 0.12 points (*d*=.20; see Figure 3b).

The teacher who wrote more conceptual comments used a relatively small number of comments for all students, customizing these comments slightly on a case-by-case base. Each comment focused on a particular conceptual issue. For example, one comment was *"You have only made one frame to represent the products and reactants. Your first frame should be for the reactants. A second frame should be made for the products.*

Follow the directions on the top of the page." In contrast, the second teacher gave short comments that were often solely procedural, such as directing students to read the directions; this teacher commented that he had little time to review the drawings. Conceptual comments from this teacher tended to state a concept in isolation, such as *Conservation of mass?"* These comments may have been too terse to help students integrate these concepts into their revised drawings.

Student interviews point to both challenges and benefits of automated guidance in terms of helping students to monitor their learning. The automated guidance may provide students an alternative to relying on the teacher for answers as several students noted , "*[Teachers]'ll be more specific. They'll show you. They'll point out what's wrong*." while the automated guidance encouraged students to sort through their ideas on their own. While a number of students wished that the automated guidance would "*tell the person exactly what they did wrong*," one pair noted the benefits of less specific guidance: "*If you don't get a problem, the teacher may give the answer away…they will be like 'No, it's like this,' and they will do it for you. But you need to learn for yourself.*" On the other hand, the automated guidance was unable to provide the extended dialogue that teachers may facilitate to ensure students grapple with the concepts *and* reach an understanding. A combination of teacher and automated guidance may provide the best solution for promoting student learning.

Study 2: Timing of Guidance

In Study 1, the automated guidance was as effective as teacher-generated guidance in helping students revise their drawings and improve their post-test performance. Students received these two types of guidance at different timing intervals to reflect typical use – immediately from the computer and start of class the next day from the teacher. This variation in timing may have contributed to the effects of the automated and teacher-generated guidance on learning. Study 2 examines the question of whether the benefit of automated guidance is tied to its immediate timing. Previous studies have found mixed results concerning the effectiveness of immediate versus delayed guidance (for a review of the literature, see Shute, 2008), with evidence that particular learner and task characteristics may influence which method is more effective. Immediate guidance is often more effective, especially for struggling students and on more challenging tasks (Shute, 2008). For our tasks, we hypothesized that immediate guidance would be more engaging and motivating to students, and would allow them to improve their understanding prior to moving on to related tasks in the unit.

Methods

Participants

Ninth grade basic chemistry students (N=88 completed both pre- and post-tests, N[groups]=57; ages typically 14-15) from four classes in a public middle school participated. The same teacher taught all students.

Study Design and Administration of Feedback

Students were assigned to the immediate or delayed guidance conditions on a full-class basis. All students completed pre- and post-tests individually, and completed the WISE unit in groups of one to three students.

The immediate guidance condition was identical to the automated condition in Study 1. We provided guidance to students in the delayed condition the evening after they completed their initial drawings. When students returned the following day, they were informed that they had new guidance and viewed the guidance. In both cases, the comments were based on the score of their drawing, and the texts in the two conditions were identical. Students in the immediate guidance condition could submit their drawing up to two times; due to time constraints, students in the delayed condition received only a single round of automated guidance.

Data Sources and Scoring of Knowledge Assessments

We evaluated student drawings (initial and final, after revisions) and the pre- and post-test item using the same knowledge integration rubrics as in Study 1.

Results

Overall, the outcome measures showed similar learning regardless of guidance timing. Students in the immediate condition improved their drawings by an average of 0.43 points (*d*=.36) compared an average of 0.67 points improvement for students in the delayed condition (*d*=.49) [Figure 3(c)]. A repeated-measures ANOVA including factors for revision (initial versus final) and guidance condition, as well as an interaction between these two factors, showed a main effect of revision ($F(1,227)=25.5$, $p < .001$), but no significant effect of condition or of the interaction. On the post-test item, students showed small, reliable improvements from their pre-test scores, with an average improvement of 0.19 points (*d*=.27). A repeated measures ANOVA with factors for pre- versus post-test and feedback condition showed that both main effects were significant (pre- versus post test: $F(1,86) = 4.58$, $p < .05$; condition: $F(1,86) = 4.12$, $p < .05$). Closer examination revealed relatively little improvement for students in the delayed condition (*d*=.10) compared to students in the immediate condition (*d*=.43) [Figure 3(d)]; however, an analysis of covariance did not show a significant difference in gains when pre-test scores were included as a covariate ($p > 0.3$).

These results suggest that the benefit of automated guidance is not simply due to its immediacy, although further study is needed to determine if the two guidance types lead to differences in retention. While this study illustrates that automated guidance can also be effective when delayed, facilitating a variety of classroom implementations, immediate guidance is likely to be the most common approach. Immediacy is a unique affordance of computer-assigned guidance relative to teacher guidance. It can advance students during class enabling teachers to work with those who continue to struggle. Further, immediate guidance can be more easily integrated into activities and provides an intuitive appeal by helping students develop understanding of challenging material before moving on.

Discussion

Formative guidance can help students to distinguish between normative and non-normative ideas in generation tasks by building on their current knowledge. Our automated system accurately evaluated student responses and provided guidance on their chemistry drawings. The knowledge integration guidance focused students on comparing alternatives and integrating normative concepts to strengthen their own understanding. Of course, not all improvements on the pre- and post-test item were solely the result of the guidance. But, students' revisions suggest that they used the hints provided to revise their understanding. We compared providing automated guidance on demand or at a delay and found that these approaches were equally effective in helping students revise their understanding of chemical reactions. Overall, these studies show that student generated drawings with a wide range of responses are amenable to automatic evaluation. They also show that a small set of knowledge integration guidance options can encompass a wide range of student responses. Compared to guidance on selection items where students often just change the response to the correct one, guidance on these generation items motivated students to review their own work and identify revisions.

Overall, the automated knowledge integration guidance was as effective as teacher-generated guidance for promoting learning. The effect of automated guidance in the two teachers' classrooms suggests different roles that automated guidance may play to augment inquiry learning. In Study 1, we found that teachers differed in the type of guidance that they provided and that this lead to differences in student outcomes. For one teacher, the automated guidance was significantly more effective than their own written guidance, likely because this teacher had little time to write individualized conceptual guidance for each student drawing. In this classroom, the automated guidance could serve as a starting point for discussions between teacher and student by identifying conceptual issues that need refinement. For the teacher who had time at night to review students' work, automated guidance left him with less knowledge about his students' ideas, making it more challenging to plan instruction for the next day. In this classroom, automated guidance may work best as a tool to help the teacher plan instruction. The teacher could review the automated guidance and scores at night, and sort student work by auto-score to see categories of student ideas. This would enable the teacher to customize instruction and save valuable time reviewing responses one by one. As automated guidance becomes a part of computer-based curriculum, it is essential to work with teachers to customize use of the information.

Our work is consistent with a growing body of research suggesting that effective formative guidance can be provided for generative items within educational technologies. We are extending this work by exploring varied types of formative guidance. Textual guidance is typically provided by teachers because designing customized activities for each student is prohibitively time consuming. However, adaptive activities based on students' drawings could prove to be more engaging and lead to richer student insights than simple text messages. The automated scoring system provides a general tool for testing different types of adaptive guidance based on students' drawings. Our investigations provide support for the use of automated guidance in the classroom. These results also point to the potential of automated guidance and scoring for generative items in cases where teacher-generated guidance is infeasible, such as online courses with thousands of students. To design our scoring and guidance system, we relied on analysis of previous student work as well as educational principles, including the knowledge integration framework. This design process provides an example of how automated guidance can be created for new items and revised as additional student work is collected. Overall, our work demonstrates the effectiveness of automated guidance for student-created drawings and provides a new tool for exploring how best to deploy this guidance to help students and teachers.

References

Anderson, J., Boyle, C., Farrell, R., & Reiser, B. (1987). Cognitive principles in the design of computer tutors. *Modelling Cognition*, 93–133.

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

Azevedo, Roger, and Robert M. Bernard. "A meta-analysis of the effects of feedback in computer-based instruction." *Journal of Educational Computing Research* 13.2 (1995): 111-127.

Ben-Zvi, R., Eylon, B., & Silberstein, J. (1987). Students' visualization of a chemical reaction. *Education in Chemistry*, *24*(4), 117-120.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing about knowing* (p. 185-205). Cambridge, MA: The MIT Press.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Chang, H. Y., Quintana, C., & Krajcik, J. S. (2010). The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*, *94*(1), 73-94.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182.

Chiu, J., & Linn, M. (2011). Knowledge integration and wise engineering. *Journal of Pre-College Engineering Education Research (J-PEER), 1*(1), 1-14.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 180-192.

Harrison, A. G., & Treagust, D. F. (2000). Learning about atoms, molecules, and chemical bonds: A case study of multiple-model use in grade 11 chemistry. *Science Education, 84*(3), 352–381.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81–112.

Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*(1), 30–43.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264.

Leacock, C., & Chodorow, M. (2003). C-rater: Automated scoring of short-answer questions. *Computers and the Humanities*, *37*(4), 389-405.

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

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

Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences, 2*(3), 277–305.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). *Assessing science understanding: A human constructivist view*. Academic Press.

Razzaq, L. M., & Heffernan, N. T. (2009). To Tutor or Not to Tutor: That is the Question. In *Proceedings of Artifical Intelligence in Education,* 457-464.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84.

Sao Pedro, M. A., Baker, R. S. de, Gobert, J. D., Montalvo, O., & Nakama, A. (2013). Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction*, 23(1), 1–39.

Schank, P., & Kozma, R. (2002). Learning chemistry through the use of a representation-based knowledge building environment. *Journal of Computers in Mathematics and Science Teaching*, *21*(3), 253–279.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science*, *3*(4), 207-217.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2013). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development, 61*(1), 71–89.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29(4)*, 4-14

Slotta, J., & Linn, M. (2009). *WISE science: Web-based inquiry in the classroom.* Teachers College Press.

Smith III, J. P., Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115–163.

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations?. *Journal of Research in Science Teaching*, *48*(10), 1177-1198.

Acknowledgments

This work was supported by NSF grant number DRL-1119670 to Marcia C. Linn.