Examining the Impacts of Annotation and Automated Guidance on Essay Revision and Science Learning

Libby Gerard, University of California, Berkeley, libby.gerard@gmail.com
Marcia C. Linn, University of California, Berkeley, mclinn@berkeley.edu
Jacquie Madhok, University of California, Berkeley, jjmadhok@gmail.com

Abstract: Automated guidance can facilitate student revision of explanations and arguments in online inquiry science units. We explore ways to design guidance for short essays that promotes meaningful revision rather than superficial changes. Specifically, we compare the affordances of annotation of a fictional essay to knowledge integration guidance on revision of science writing. 293 middle-school students were randomly assigned to condition. Students who annotated an essay made significantly greater pre to post test gains and were also better able to use automated guidance on a posttest item than students who only received knowledge integration guidance. These findings suggest ways to support students to revise science writing and build integrated understanding of science.

Keywords: technology, revision, writing, student learning, science, knowledge integration

"I like when there is writing because with typing it feels like I can explain more instead of…just doing multiple choice, because then its kind of like I want to explain it really bad but you can only put an answer." (7th grade study participant)

Introduction

Writing in science gives students the opportunity to defend claims and integrate multiple pieces of evidence gathered from a variety of contexts. A substantive literature demonstrates the value of writing and revising activities that prompt students to build connections between their initial ideas and new ideas (Fitzgerald, 1987). The centrality of writing and revision in scientific practice are reflected by the Next Generation Science Standards. Students are expected to construct explanations and arguments, and, to "identify flaws in their own arguments and modify and improve them in response to criticism" (NGSS, 2013). Writing and revision of arguments is integral to long-term science learning (Rivard, 1994). In spite of the importance of writing and revision, these activities are rare in the science classroom. This is due largely to the time required of teachers to provide individualized feedback for often at least 150 students. A NAEP report found that about half of Grade 11 students (52%) had never completed a written report of any kind in science (Mullis & Jenkins, 1988).

Advanced natural language processing tools can help science teachers integrate writing and revision activities by automatically scoring student written short essays about an inquiry topic. Individualized guidance can be assigned immediately based on the score to help students revise their essays (Liu et al, 2014; Roscoe & McNamara, 2012). Some essay writing tools (e.g. Criterion) that give diagnostic feedback on essay mechanics and organization have been widely deployed in classrooms to increase writing and revision opportunities. We explore ways to design guidance that strengthens the argument in the essay.

Students generally find revision in science unfamiliar and challenging. We characterize revision as involving revisiting or gathering new evidence, comparing initial ideas with new ideas, and reformulating connections. Class norms typically support completeness and correctness rather than refinement. In analyses of classroom argumentation Berland & Reiser (2011) found students rarely revised their ideas in light of challenges or questions posed. Studies consistently find that students most often make mechanical revisions to their essays when given feedback, rather than substantive changes to the content as suggested (Roscoe & McNamara, 2012). Less competent writers tend to make changes to spelling and grammar rather than to revise for meaning (Fitzgerald, 1987). This is particularly problematic for science because most pre-college students are novice science writers and have limited understanding of what makes a good science explanation (Sato & Linn, 2014).

Our prior work suggests that individualized, automated guidance for students' short essays can strengthen students' revisions and science learning when designed to promote knowledge integration (Linn & Eylon, 2011). Guidance designed according to the knowledge integration perspective prompts students to distinguish between their own scientific ideas and new ideas introduced in instruction. The knowledge integration guidance led to significantly more productive essay revisions, and subsequently more coherent and accurate science essays than did generic guidance (e.g. "add more evidence) or specific guidance (e.g. "Incorrect. Energy transforms from light energy into chemical energy) (Author et al., 2015).

Analyses also revealed difficulties students faced in integrating ideas during the essay revision process. Only a small percentage of the students integrated a new idea into their essay meaning that they edited their initial essay when revising to build a connection between their initial ideas and the new information they added. Rather the majority of students added a new idea without connecting it to their initial response; the new idea was "tacked-on". On a post test taken a week after the writing and revision activities, the students who had made integrated revisions demonstrated a significantly deeper understanding of the science concepts than those students who had added new but disconnected ideas in their revision (Gerard & Linn, 2015). This suggests that the practice of revision, when done in a way that builds and refines connections among ideas, can set in motion a successful learning strategy for the rest of the unit.

In this study, we explore ways to help students learn to revise essays by integrating new ideas. We designed an essay annotation activity to help students learn how to use guidance to revise science writing. In the annotator activity, students identify gaps in a fictional student's essay and place pre-determined labels on the essay as hints for revision (Figure 1). The hints call for the fictional student to edit her essay to connect new ideas. We compare students' use of annotation and then automated guidance to revise short essays in an inquiry unit, to students' use of multiple rounds of automated guidance to revise their short essays.

Methods

This instructional comparison study investigates two approaches to guiding students essay revision and examines their impacts on science learning. Specifically, we compare the affordances of annotation of a fictional essay versus additional practice with automated guidance on revision of science writing.

Participants

Two-hundred and ninety-three students in three teachers' classrooms in one public middle school were randomly assigned within class periods to either the Annotation + revision condition, or, Two revisions condition. The school serves a moderately diverse student population (47% of students are an ethnicity other than Caucasian; 26% receive a free or reduced price lunch). Students worked in pairs in the unit and completed the pretest and posttest individually.

Curriculum and embedded assessments

We used the Web-based Inquiry Science Environment (WISE, http://wise.berkeley.edu) to randomly assign conditions to students within classes. Students studied the WISE Photosynthesis and Cell Respiration unit, which guides students to investigate energy flow in plants and animals. The unit incorporates dynamic visualizations and generative activities, including writing short essays, to help students gather evidence and integrate ideas. Students studied the WISE unit, led by their regular classroom teacher, for 6-8 class periods (50 minutes each) spread over 2 weeks.

We selected three short essays in the unit for annotation, guidance and revision. The selected essays call for students to integrate multiple pieces of evidence from the unit to explain energy transfer and transformation. Our analysis focuses on one of the three essays, called GreenRoof (Figure 1):

> *GreenRoof: Mary heard that growing plants on a roof could lower a house's energy usage. Mary does not understand why and how plants could help. Write an energy story to explain to Mary what happens to energy from the sun in the picture? How could growing plants on the roof reduce the house's energy usage?*

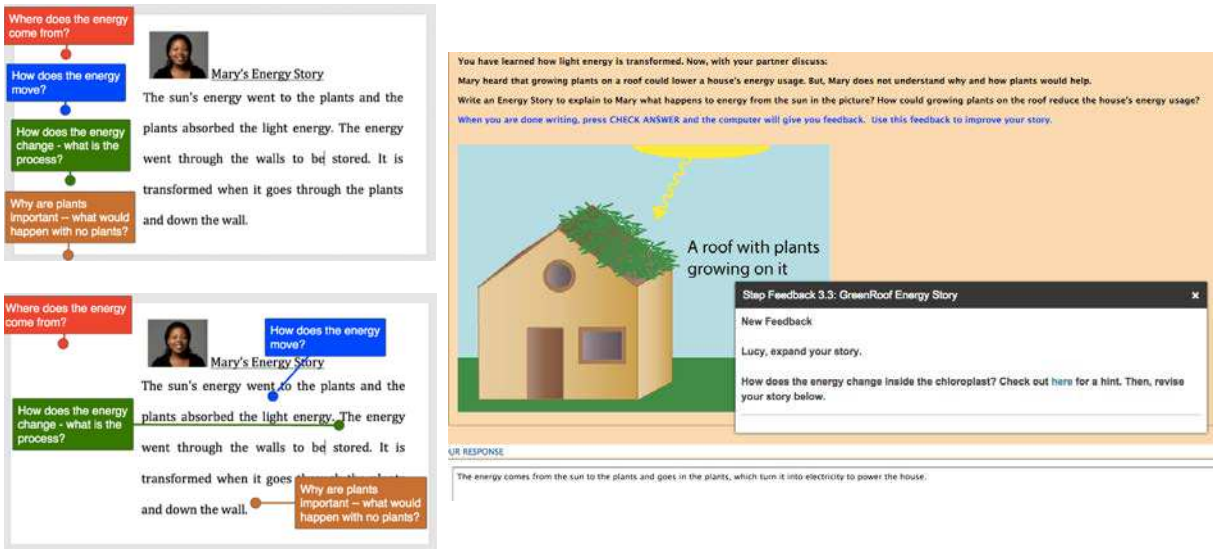
Each of the short essays included automated scoring by c-raterML™, a natural language processing tool developed by the Educational Testing Service (see Liu et al, 2014 for model development info). WISE assigned personalized guidance for the student's essay immediately based on the automated, craterML score (Table 1).

Essay revision

For each of the three short essay steps students received personalized, automated guidance (Table 1). To assess the value of the annotation activity the two conditions were:

1. Annotation+revision: students first annotate the response of a fictitious student, then they get personalized automated guidance on their own response.
2. Two revisions: students get two rounds of personalized, automated guidance on their own response. We constrained the algorithm to assign unique guidance to the second revision even if the score did not change.

In the Annotation+revision condition, students were prompted to write their short essay, and then to annotate Mary's essay using pre-defined labels. Mary's essays included vague ideas as well as a mix of normative and non-normative ideas (Figure 1). The pre-defined annotation labels included questions about energy flow that were general enough to be used across all three short essays: (a) Where does energy come from? (b) How does energy change – what is the process? (c) How does the energy move? (d) Why are plants important – what would happen with no plants? (Figure 1). We used the same labels on each essay to model for students the questions they could ask themselves when they wrote and revised future essays. After annotating Mary's essay, students were prompted to submit their own essay for one round of automated guidance and revision.



Annotation activity **GreenRoof essay and auto guidance**

Figure 1. Revision activities.

In the Two revisions condition, students were prompted to write their short essay (Figure 1). They were then given two chances to submit their essay for automated guidance and to revise. All guidance was designed to help students move up one level in the knowledge integration rubric (Table 1).

Table 1: GreenRoof knowledge integration scoring rubric and personalized, automated guidance

Score, Criteria	Student Essay Examples	Automated Guidance Round 1
1 Off Task	IDK	{Student name}, think about the Sun's rays. Where do they go? How do they move? Check out Step 2.1 for a hint. Then, rewrite your story below.
2 Incomplete Scientifically non-normative OR irrelevant OR repeats the question	It helps reduce energy usage because you can get energy that you can use for your house through the plants.	{Student name}, expand your story. How does the energy change inside the chloroplast? Check out Step 3.2 for a hint. Then, revise your story below.
3 Partial Partial link of key ideas OR normative and non-normative ideas linked	The sun gives light energy to the plants so they can photosynthisize and the plant give off oxygen and energy to the world.	{Student name}, explain the connections between your ideas. How does the plant transform light energy to energy it can use? Check out Step 2.11 for a hint. Then, revise your story below.
4 Full 1 full link [ok if response also has non-normative ideas, as long as not connected to link]	One day Brent and Emilio … …observed the green roof. They noticed the sun's light energy was being absorbed by the plants on the roof.…They thought the plant turned the light energy to chemical energy.	{Student name}, good progress. Now expand your story. Why are the plants important - what would happen to sunlight that hits the side of the roof with NO plants? Check out Step 1.4 for a hint. Then, revise your story below.

		Step 1.4 for a hint. Then, revise your story below.
5 Complex 2 or more full links	Plants can lower energy usage when there on a roof because they absorb the sun and turn that energy into chemical energy. Without the plants on the roof it get hot since sunlight energy will turn into heat	{Student name}, great reasoning. Now, check over your story to make sure it describes the energy flow from the sun to both sides of the roof. Revise as needed.

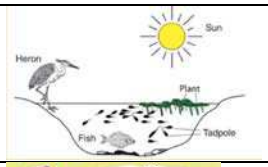
Data sources

We used a pretest, embedded assessments, and a posttest to capture student science learning and student ability to use guidance to revise. The pretest was taken one day prior to starting the unit. The posttest was taken approximately three days after completing the unit and one week after the last essay revision activity. The pre/post test included two short essays, Tadpole and Rabbit [Table 2].

To assess student ability to use automated guidance, for the Tadpole essay, students received one round of personalized, automated guidance and revised their response. The two essay activities were the same on the pretest and posttest.

We conducted student interviews and classroom observations to gather further information on how students used the writing and revision tools, and student views on writing and revision in science in general.

Table 2: Pre/Post Test Short Essays

Tadpole*		Explain how the tadpole in the picture gets and uses energy to grow. After you are done writing, press "Check Answer". You will have 1 chance to get feedback and revise your story.
Rabbit		There is a rabbit in the classroom. Mary wonders how the rabbit gets and uses energy from the sun. Write a story using scientific evidence to explain to Mary how the rabbit gets and uses energy from the sun.

* Students received one round of personalized, automated guidance and were prompted to revise.

Data analysis

We scored students' embedded and pre/post test responses for scientific accuracy and coherence using knowledge integration rubrics (see Table 1). The rubric rewarded students for making links among ideas (Liu, Lee & Linn, 2011). We scored students' initial and revised essays on the embedded items, and students initial and revised essays on the Tadpole pretest and posttest item. Two researchers blind coded all of the student essays. Researchers reached 97% agreement and worked out the disagreements until coming to a consensus.

In addition to scoring the data with knowledge integration rubrics, we also coded the students' embedded essays for the kind of revisions they made based on the automated guidance. We compared their initial essay to their revised essay and categorized the type of writing change (see Table 3). In this analysis, we coded the changes in the students' science writing not the scientific accuracy of the change.

Table 3: Revision Rubric

Score	**Category/Description**	**Example**	
		Initial Essay	**Revised Essay**
3	**Integrated**: Incorporated an idea into the middle of the essay. Edited the initial essay to formulate the connection to the new ideas.	The plants on the roof will take the energy to create photosynthesis and grow. Since the plants are on top of the roof it is closer to the sun taking energy from the house.	The plants on the roof will take the energy to create photosynthesis and grow. The plant uses carbon dioxide fro which we breathe out, water, and light energy to create glucose. Since the plants are on top of the roof it is closer to the sun taking energy from the house.

2	**Added Different**: Added a new idea to end of essay that is different than ideas in initial essay. Does not make any changes to initial essay. Or, deleted initial essay and wrote new essay with different ideas.	The plant uses the left over sunlight from what the house leaves behind	The suns energy goes to the plants then, plants absorb the light energy. The energy goes to the house. The plants change the energy into the chemical energy when it goes through the plants
1	**Added Similar**: Added a new idea to the end of the essay that expanded with more details, or repeated in different words, the ideas in the initial essay. Does not change initial essay.	The plants soak up the energy and provide light to the house. This will make it so that you will not have to use lights during the day.	The plants soak up the energy and provide light to the house. This will make it so that you will not use lights during the day. Also it will help the energy bill go down because it is using solar energy.
0	**No Change**		

Findings

Embedded assessment

In both conditions, students used the automated guidance tools to significantly improve their GreenRoof essay. Students assigned to annotation+revision outperformed students assigned to two revisions [N=152 pairs Annotate M=.46 SD=.72; Auto Guidance Two Rounds M=.28 SD=.56, t(149)=1.72, p=.09]. The difference between conditions was significant for students who began the writing activity with mid to high prior knowledge [Gain, High Prior, n=92 pairs, Annotate+revision M=.46 SD =.55; Two revisions M=.18 SD=.39, t(90)=2.9, p<.01]. We defined mid/high prior knowledge as students who scored a 3 or above on the knowledge integration rubric on their initial GreenRoof essay.

Students' revision approach differed significantly across the annotation+revision and two revisions conditions [Pearson chi2(3)= 14.11 Pr=.003]; Table 5]. The annotation+revision condition led significantly more students to add new and different ideas to their essay when they revised. Alternatively, students who received two rounds of automated guidance were more likely to add ideas that were similar to those they had already written. Further, the annotation condition led substantially more mid/high prior knowledge students to make a revision; 88% of high prior knowledge students in the annotation+revision condition compared to 78% of high prior knowledge students in the two revisions condition. Approximately 19% of students in both conditions integrated a new idea into their initial essay. There was no difference between conditions in the percentage of students who made integrated revisions. Mid/high prior knowledge students were more likely to integrate than low prior knowledge, consistent with the literature on revision (Fitzgerald, 1987).

Table 5: Percent of students making each revision type on embedded essay by condition and prior knowledge

	All Students		**High Prior**		**Low Prior**	
	Two revisions	Annotation + revision	Two revisions	Annotation + revision	Two revisions	Annotation + revision
No revision	24	16	22	12	27	20
Added Similar	39	30	40	37	36	23
Added Different	18	34	14	24	27	46
Integrated	19	20	24	27	9	11

Student interviews and classroom observations suggest that the annotation+revision condition helped students use guidance to critically examine their own essay. As one student reported, "*I think that the Mary's story [annotate activity]…helps you change yours [essay] because you take the feedback you are giving Mary and then you kind of apply it to your own*." Students used the prompts that were repeated in annotation labels (e.g. how does energy change?) as they reflected on how to revise their essay. Another student reported: "Since you have to move [the labels] around and see where it goes, it makes you think that if that was you doing the question and the teacher gave you that feedback…it kind of makes you think 'Oh ya, that should go there' and then it can help you when you have a question similar you can be like oh I remember that and I should add that in mine now."

Pre/post test

We report the results of each pre/post test essay separately since students received automated guidance on the Tadpole short essay, and did not receive automated guidance on the Rabbit essay. Students in both conditions significantly improved from pretest to posttest [Rabbit $t(282)=8.72$, $p<.0001$; Tadpole $t(276)=7.78$, $p<.0001$] [Figure 2]. There was no difference in gains between the two conditions on the Rabbit essay.

On the Tadpole short essay where all students received automated guidance, students in the annotation+revision condition made significantly greater pre to posttest gains than the students in the two revisions condition, Figure 2 [N=261 students, Annotate+revision M=.58 SD=1.02; Two revisions M=.29 SD=1.03, $t(259)=2.29$, $p<.05$].

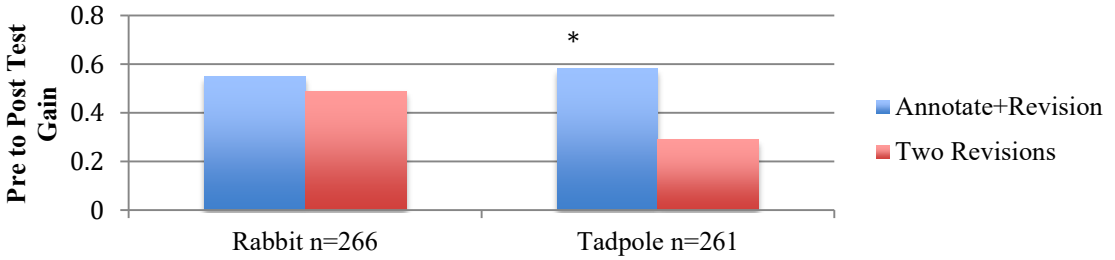


Figure 2. Student pre to post test gains on two short essays. The tadpole essay included automated, guidance.

Further, by posttest the students in the annotate+revision condition took better advantage of the automated guidance on the Tadpole essay than students in the two revisions condition. The analysis suggests that the annotation activity helped students learn how to use guidance to revise their writing and improve the coherence of their understanding.

Students who made an integrated change when revising their GreenRoof essay during the unit, made the greatest pre to post test gains on both short essays [Figure 3]. This strengthens our earlier findings demonstrating the benefits of integrating as a revision approach (Author, 2015). Further design work is needed to improve the percentage of students who integrate when revising.

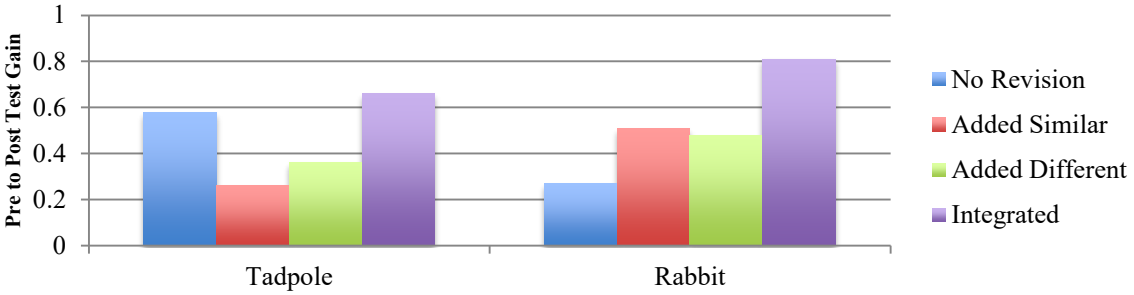


Figure 3. Pre to post test gain by type of revision made on an essay embedded in the unit.

Example of student revision using annotation and auto guidance

We selected a case of a student, Erica, whose work was representative of those in the annotate+revision condition. The case documents how Erica used annotation and automated guidance to strengthen her essays.

Erica began the Photosynthesis/Cell Respiration project with a mix of ideas about energy transfer and energy forms. On the pretest Tadpole essay, she explained a tadpole "gets energy straight from the sun. The tadpoles use the energy to store and create kinetic energy." This response seems to draw on vocabulary that is not fully understood.

Erica gathered new ideas from visualizations of energy flow in photosynthesis. She connected these ideas with her intuitive views about energy in her initial GreenRoof essay [Figure 4]. She explained that plants get energy from the sun and give energy to the house. Erica then annotated Mary's GreenRoof essay placing the labels "where she wanted [Mary's essay] to explain more". Next, Erica received automated guidance on her essay. The guidance helped Erica see that her essay generally needed improvement, but she did not know what to change. Erica drew on her annotation experience to consider what to revise. "It [*the annotator] helped me use the [automated] feedback because I could go back and look at those questions [the annotator labels] I had put down and use it for myself*." Erica then revised her GreenRoof essay to add a new idea about photosynthesis, while still holding onto her non-normative ideas about the transfer of energy from plants to the house.

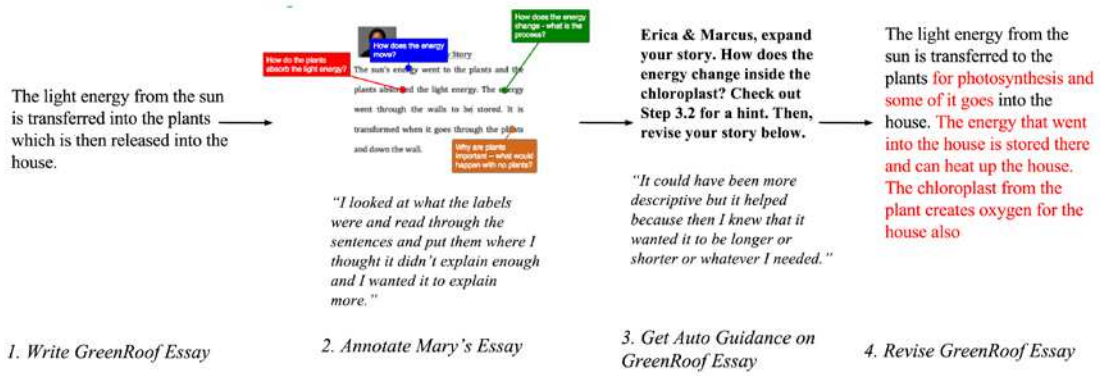


Figure 4. Erica's writing, annotation, and revision on the GreenRoof essay.

After the GreenRoof essay activity, Erica explored visualizations of energy flow in cellular respiration. She then wrote the EnergySun essay, the last essay writing activity in the unit [Figure 5]. Erica wrote a vague essay indicating she was still confused about energy transfer. Erica next annotated Mary's EnergySun essay "*placing the labels where it was hard to understand [what Mary wrote] and [she] wanted to know more*." These comments seem to reflect Erica's growing realization that she needs to find out more about energy transfer and transformation. Next, Erica received personalized, automated guidance on her essay. The guidance helped Erica understand what she needed to find out whereas at first she said, she "didn't know what to write." Erica revised her essay to integrate detailed links about energy transformation and storage in glucose.

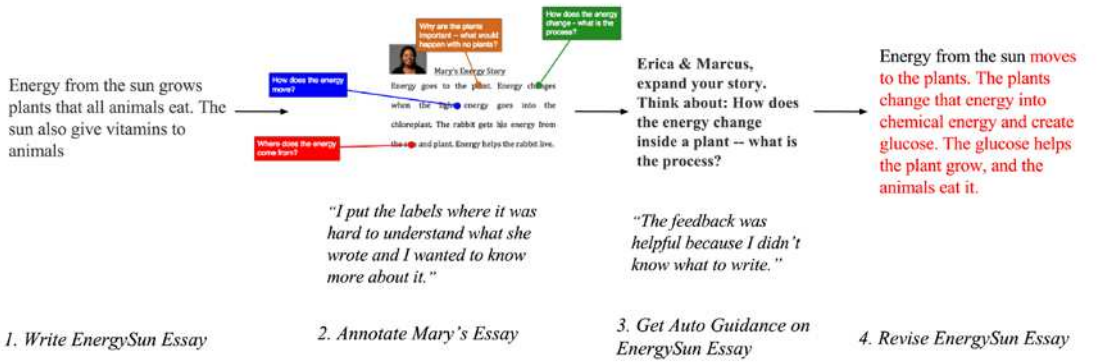


Figure 5. Erica's writing, annotation, and revision on the EnergySun essay

By the time of the posttest, Erica had formulated a coherent account of energy flow. She explained on the posttest how a tadpole gets and uses energy. "…The sun gives off radiation to the plant which absorbs it into its chloroplast. The chloroplast transforms it into chemical energy to make glucose which helps it grow. The tadpoles then eat the plants..." The annotate+revision condition helped Erica learn how to revise a science essay. She drew on the big questions about energy that were in the annotator labels she used to critique Mary's essay to inform her own writing and revision.

Conclusions and implications

This study investigated two different approaches to supporting students' use of automated guidance to revise and strengthen their science writing and understanding. These results demonstrate that automated knowledge integration guidance can help students refine their ideas about energy transfer and transformation in photosynthesis and cellular respiration. It is noteworthy that both conditions guided students to figure out how to revise their own essays rather than providing them with the answer.

In providing students a fictional peer's essay to annotate, we found that students transferred the hints they used to critique a peer's essay to their reflection on how to apply guidance to improve their own essay.

Students' reflective process may explain why students in the annotation+revision condition later showed greater gains on the posttest when revising an essay with automated guidance than those students in the two revisions condition. The annotation+revision condition helped students learn how to revise using guidance.

Students have difficulty revising scientific arguments and often tack ideas on to their essays rather than integrating them to increase the coherence of their responses (Gerard & Linn, 2015). We designed the annotation activities to model the process of editing within an essay, as opposed to only adding ideas to the end. A minority of students however in both conditions (19%) integrated new ideas when revising essays during the unit. The annotate condition increased the likelihood that students would add new and different ideas rather than elaborate on their existing ideas. Research shows that when students are asked to add new ideas they integrate their ideas more effectively than when they are asked to add similar ideas (Matuk & Linn, 2015). The case study reinforces these interpretations by showing that the annotate condition helped Erica add key ideas in energy transfer and transformation.

Students' propensity to add-on ideas, similar or different, rather than integrate new ideas may indicate the difficulty and the unfamiliarity that integrating ideas presents to students as they learn a new topic. Traditional school culture prioritizes adding a "right" idea over distinguishing and building connections among one's initial views and the new information presented. This finding also indicates areas for improvement in the annotation task design. Although the annotation hints (e.g. "how does energy change?") were intended to prompt the fictional student to both fill gaps in her essay *and* modify non-normative ideas, our student interviews suggest students perceived the hints as guidance suggesting the fictional student add more information about this idea. Future design iterations will aim to focus students' attention on both adding and modifying during the annotation activity.

In all, this study reveals how tools like annotation and automated guidance can help students to engage in meaningful writing and revision practices in the science classroom. Knowledge integration guidance as well as using annotations to model the process of using guidance can significantly strengthen essay revisions.

References

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

Gerard, L., Ryoo, K., McElhaney, K., Liu, L., Rafferty, A., & Linn, M.C. (2015, online first). Automated guidance for student inquiry. *Journal of Educational Psychology*

Gerard, L., & Linn, M.C. (2015). *Writing and revising in science*. Paper accepted for the annual meeting of the American Education Research Association, Washington D.C

Linn, M. C., & Eylon, B.-S. (2011). Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration. New York: Routledge

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

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

Matuk, C. & Linn, M. C. (2015). Examining the real and perceived impacts of a public idea repository on literacy and science inquiry. In *Proceedings of the 11th Intl Conference for Computer Supported Collaborative Learning*, Gothenburg, Sweden: International Society of the Learning Sciences.

Mullis, I.V. S. & Jenkins, L.B. (1988). The science report card: Elements of risk and recovery. Trends and achievement based on the 1986 national assessment (National Assessment of Educational Progress). Princeton, NJ: Educational Testing Service.

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

Rivard, L. O. P. (1994), A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching* (31), 969–983

Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, *105(4),* 1010-1025

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

Sato, E. & Linn, M.C. (2014). Designing critique to improve conceptual understanding. In Polman, J. L., et al. (Eds.). *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS) 2014*, Vol 1. Boulder, CO: International Society of the Learning Sciences.