

[Please cite as: Schwendimann, B. A., & Linn, M. C. (2015). Comparing two forms of concept map critique activities to facilitate knowledge integration processes in evolution education. *Journal of Research in Science Teaching*, 4. doi:DOI 10.1002/tea.2124]

Comparing two forms of concept map critique activities to facilitate knowledge integration processes in evolution education

Beat A. Schwendimann and Marcia C. Linn

Graduate School of Education, University of California, Berkeley, CA 94720

Abstract: Concept map activities often lack a subsequent revision step that facilitates knowledge integration. This study compares two collaborative critique activities using a Knowledge Integration Map (KIM), a form of concept map. Four classes of high school biology students (n=81) using an online inquiry-based learning unit on evolution were assigned to one of two conditions. Student dyads in one condition compared their concept maps against an expert map while dyads in the other condition conducted a peer-review. Analysis of the concept maps suggests that students in both conditions improved their understanding of evolution from pretest to posttest. However, the two conditions lead to different criteria: Students in the expert-map condition focused mostly on concept-focused criteria like concept classification while students in the peer-review condition used more link-focused criteria like link labels and missing connections. This paper suggests that both forms of KIM critique activities can be beneficial for constructing more coherent connections across different topics in evolution education. These results support the value of collaborative KIM critique activities and help clarify the forms of collaborative activities that are most likely to be effective to facilitate knowledge integration processes. **Keywords:** conceptual change, science education, misconceptions, inquiry, evolution, concept mapping

Corresponding author [Current affiliation]: Beat A. Schwendimann; CHILI – École Polytechnique Fédérale de Lausanne (EPFL); EPFL IC ISIM CHILI Station 20; 1015 Lausanne (Switzerland); beat.schwendimann@gmail.com

This study compares and distinguishes two forms of collaborative concept map critique activities to support integrating ideas related to evolution. The theory of evolution has been found difficult to understand because it incorporates a wide range of ideas from different levels (Bahar et al., 1999; Tsui & Treagust, 2003; Novick et al., 2014) and multiple interacting levels (Wilensky & Resnick, 1999; Duncan & Reiser, 2007; Hmelo-Silver et al., 2007). Many students hold a rich repertoire of alternative ideas relevant to evolution that are often disconnected and in conflict with scientific ideas (Southerland 2001; Sinatra 2003; Shtulman 2006).

Traditional instruction is often not sufficient to improve students' understanding of evolution (Taber, 2013). Bishop and Anderson (1990) observed that a majority of students holds ideas that differ from accepted evolutionary theory even after instruction. Many students leave school with a fragmented understanding of biology that does not allow them to connect their ideas to their everyday lives (Mintzes, Wandersee, & Novak, 2000). Effective evolution instruction needs to support students to distinguish between normative and alternative ideas, connect new and existing ideas, and apply ideas in different contexts. Building connections between ideas can be challenging because the relationships among them are often not intuitively obvious to the learner. The knowledge integration framework (Linn & Eylon, 2011) describes conditions that make it possible for a new idea to be understood, linked to related ideas and used to interpret existing ideas.

This study uses concept maps as tools to facilitate knowledge integration processes. To visualize, generate, and distinguish existing and missing connections between ideas, concept maps are potentially powerful learning tools (Hamza & Wickman, 2013) but results have been mixed (Ruiz-Primo, 2000). Concept map creation is a generative activity but is often used as one-shot summative assessment (Trowbridge & Wandersee, 1998). This study explores extending concept map generation activities by comparing two forms of subsequent collaborative critique and revision activities designed to facilitate students distinguishing alternative ideas. This study took advantage of a novel form of biology-specific concept map called a Knowledge Integration Map (KIM) that incorporates research on knowledge integration and on concept mapping (Schwendimann, 2009a, 2009b). Embedded in an inquiry-based online learning activity, student dyads used expert or peer-generated KIMs to critique and revise their concept maps.

This comparison study investigates how collaboratively generating and then either critiquing peer- or expert-generated KIMs, embedded in a dynamic technology-enhanced inquiry environment, affects students' integration of ideas related to evolution. This study aims to answer the research questions:

- 1) What are the affordances of collaborative KIM peer or expert critique activities? Different definitions for the term "affordance" have been developed (Bonderup Dohn, 2009). In this study, the term affordance refers to features of visualizations or activities that enable or constrain the possibilities for learners.
- 2) Is peer or expert critique more likely to promote knowledge integration?

Integrating evolution ideas through collaborative concept mapping

This study uses knowledge integration (Linn, Eylon, & Davis, 2004) as its operational framework because its design guidelines emphasize eliciting connections between existing and new ideas and developing criteria to distinguish among alternative ideas. Knowledge integration research demonstrated that students can benefit from instruction designed following the knowledge integration pattern: Eliciting existing ideas and connections, adding new ideas to the existing repertoire of ideas,

developing criteria to distinguish ideas, distinguishing and sorting out alternative ideas, connecting ideas, and applying ideas in multiple contexts. KIMs have been developed to support knowledge integration processes through making connections between ideas visible (see Table 1). A visual representation of the connections between ideas can elicit existing or missing relations between ideas (Shavelson, Ruiz-Primo, & Wiley, 2005). As each idea is only represented by one node, all related ideas and their relationships are quickly accessible. The clustering of related ideas and elicitation of connections in concept maps promises to facilitate knowledge integration processes.

Generating artifacts, such as concept maps, can promote conceptual learning (van Amelsvoort, Andriessen, & Kanselaar, 2005). The ‘generation effect’ (Osborne & Wittrock, 1983) has been well-documented in a variety of different settings. For example, Chi (2000) found that generating explanations of a text or diagram, whether for oneself or for others, can be more effective for learning than receiving explanations. Zhang and Linn (2011) found that students generating diagrams after using a dynamic science visualization integrated more ideas than students who only used the dynamic visualization. By generating artifacts, students articulate and represent their knowledge in new forms, apply their representations to solve scientific problems, realize gaps in their knowledge, and reorganize ideas. Generating concept maps for knowledge integration requires revisiting prior ideas. Since KIMs constrain learners to create only one single connection between two ideas, they motivate students to decide which relation they want to generate. Negotiating can encourage students to revisit their ideas and critically reflect upon the relationships (Berland & Reiser, 2009).

Table 1
Concept mapping for knowledge integration

Knowledge Integration Process	KIM Activity
Eliciting existing ideas	KIM generation activities can elicit existing alternative ideas.
Adding new ideas and connecting to existing ideas in repertoire	New ideas can be added to existing propositions in the KIM. If several alternative relationships between two ideas are possible, students have to decide which one to use in the map. If applicable, students decide which ideas to add to the map.
Distinguishing/ Critiquing ideas	After adding new ideas, ideas can be rearranged into new groups, and the KIM network structure might need revision to reflect the new ideas.
Sorting out Ideas/ Refining	Different sources of evidence can as reference to sort out ideas and further refine the KIM.
Applying ideas	KIMs can be used as resources to generate explanations of scientific phenomena.

KIMs developed for this study aim to support the generation and revision of ideas related to evolution from different levels. The drawing area on the KIM worksheet was divided into the biology-specific levels ‘DNA’, ‘cell’, and ‘organism/population’, which refer to micro, meso, and macro-level ideas (see Figure 1). The ‘cell’ level aimed to serve as a meso-level bridge between the genetic (micro) and the phenotype (macro) levels. As a simplification, the macro-level included both ‘organism’ and ‘population’ to refer to one or multiple organisms. Learners were instructed to place ideas into the corresponding level and construct connections within and across levels. Detailed characteristics of Knowledge Integration Maps (KIMs) are described in table S1: Characteristics of Knowledge Integration Maps.

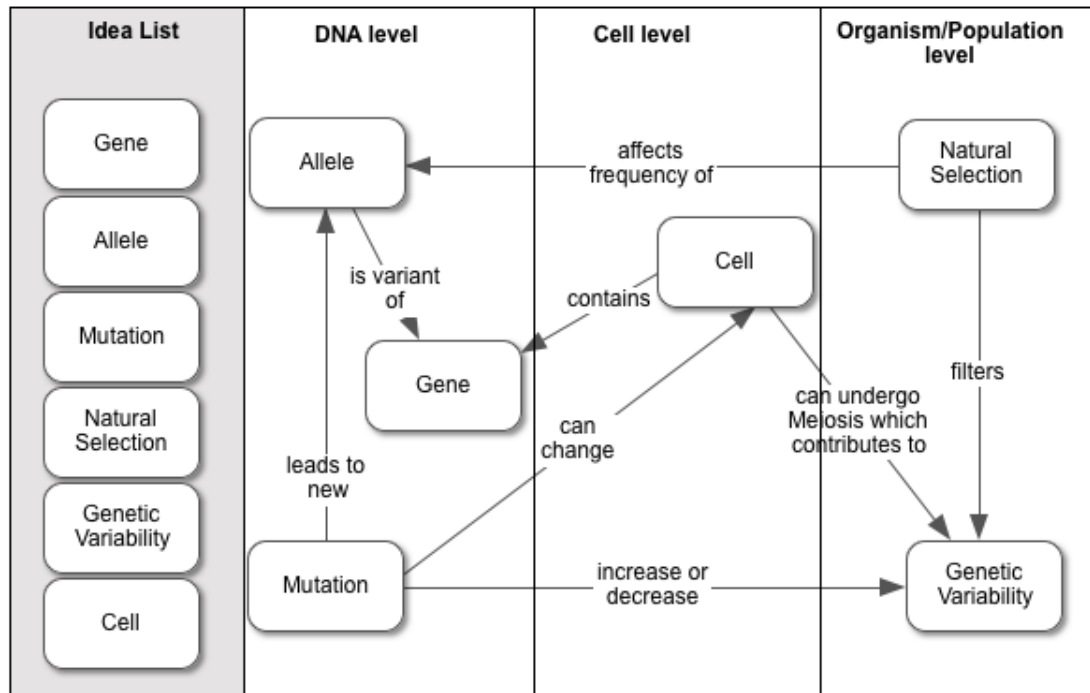


Figure 1. Knowledge Integration Map (KIM).

To facilitate knowledge integration processes, KIMs extend the generation process by adding a collaborative critique and revision activity. Critiquing is the process of creating a set of criteria, applying criteria to compare one's own or others' alternative ideas against each other, reflecting on how those ideas apply to different ideas, and selecting supported ideas based on different resources (Shen 2010). Critique activities require students to use or develop criteria to reflect, revise their work, and self-monitor their learning progress which can foster the development of metacognitive skills for lifelong autonomous learning (Chi 2000). Prompts to critique a sample student explanation can support the development of scientific explanations (Donnelly, Linn, & Ludvigsen, 2014). Developing critical thinking is pivotal to integrating scientific ideas because it allows students to distinguish alternative ideas in their repertoire (Linn 2008). Asking students to critique and revise has been found to support the development of more coherent and generative criteria (Slotta & Linn, 2000; Lehrer & Schauble, 2004).

Collaborative critique activities involve the social processes of reaching agreement, which is critical in shaping one's ideas (Enyedy 2005; Clark & Sampson, 2008). In science education, collaboratively critiquing ideas requires learners to argue, negotiate, and make informed decisions (Berland & Reiser, 2009). Finding common ground can be a driving force for critique. To reach such common ground, students need to pose questions, make revisions, accept propositions, defend against criticism, and improve their criteria (Shen 2007). diSessa (2002) observed that students are able to develop their own criteria to critique representations.

However, students have often few opportunities to critique (Clark, 2000; Shen 2010). Students need genuine opportunities to develop criteria to distinguish valid alternative ideas based on different resources and scrutinize the reliability of sources (Davis & Kirkpatrick, 2002; Cuthbert & Slotta, 2004). Previous findings suggest that a critique activity following the generation of concept maps can provide a genuine opportunity to revise ideas (Schwendimann, 2007). Building on these findings, this study distinguishes two forms of collaborative KIM critique activities. Using expert-

generated artifacts as a reference for self-evaluation of student work is commonly found in traditional forms of instruction (Hmelo-Silver, Marathe, Surabhi, & Liu, 2007). Expert concept maps can serve as examples of good solutions, as ways to identify gaps in students' understanding, and as ways to reveal non-normative ideas. Cliburn (1990) noted that providing expert-generated concept maps can support integrative understanding. However, Novak (1980) observed that studying expert maps in genetics instruction could be confusing to some students because expert-generated concept maps could be interpreted as the single correct solution. Additionally, critiquing one's own work can be more difficult than evaluating other people's work (Linn & Clancy, 1992).

Peer-generated concept maps can be used as an alternative to expert-generated maps. Peer-generated maps might be easier to compare to one's own than expert-generated maps because they use familiar language and build on similar prior knowledge (Keppell, Au, Ma, & Chan, 2006). Peer evaluation can be mutually beneficial for the giver and the receiver (Topping 2005). Peer critique can motivate students to improve their work and better understand what might need to be refined (Hoadley 2004). Comparing one's own ideas against those of a peer could help students to value their own ideas. On the downside, peer critique might introduce or reinforce non-normative ideas. Additionally, students might be either reluctant to critique their peers or be overly critical.

Methods

Curriculum Design

WISE Environment. The evolution unit developed for this study, *Space Colony - Genetic diversity and survival*, used the Web-based Inquiry Science Environment (WISE) to deliver scaffolded inquiry activities including the dynamic visualizations Dragon Genetics (Concord Consortium 2006) and Evolution Lab (Leif 2005), animations, pictures, texts, and reflection items. Dynamic visualizations can support knowledge integration processes through inquiry activities (Lee, Linn, Varma, & Liu, 2010; Zhang & Linn, 2013). To illustrate the interplay between genetic diversity and selection processes, the WISE *Space Colony* unit was driven by the story of a group of human colonists who are planning to colonize planets with different environmental conditions (see Figure S1: Screenshot of WISE Space Colony unit). The guiding questions of the unit were "what are the sources of genetic diversity?" and "under which circumstances is genetic diversity beneficial?" The unit included the ideas of reproduction (sexual/asexual), cell division (mitosis/ meiosis), mutation (somatic/ germ line mutations), recombination (crossing over, independent assortment of chromosomes, and random fusion of gametes), evolution (natural selection), genetics (DNA, chromosomes and alleles), proteins and enzymes, cells (somatic body/ sex cells), haploid/diploid cells, sex determination in humans, cloning and twins. The major sources of genetic diversity - mutation and recombination - were placed into the bigger context of evolution to facilitate students understanding the role of genetic diversity. Students were asked to collect information to make an informed decision to either send out a genetically diverse or genetically homogeneous group of colonists. After the initial individual pretest, all students completed the same KIM training activity guided by the teacher (more details about the KIM training activity below). Students worked collaboratively on the WISE unit in dyads sharing

one computer. Scaffolded inquiry activities using dynamic visualizations (Dragon Genetics and Evolution Lab) applied ‘predict-observe-explain’ patterns to facilitate students’ explorations of the relations between genetic diversity and selection mechanisms. In ‘Evolution Lab’, students explored the relations between genetic diversity and natural selection. The dynamic visualization presented a population composed of individuals with varying phenotypes. The students received a worksheet with instructions to run four experiments to investigate the effects of low/high mutation rate and low/high selection strength on the phenotype of the organisms. For each experiment, the students first made a prediction, then ran the simulation, and finally provided an explanation for their observations. In the ‘Dragon Genetics’ activity, students explored the relations between genotype and phenotype by manipulating the alleles (genotype) of organisms (dragons) in the parent generation. The physical appearance (phenotype) of the offspring illustrated the mechanisms of inheritance of autosomal dominant and recessive traits. Students were instructed to manipulate parental alleles to create offspring with certain given traits. The two simulation activities and subsequent questions emphasized connections between underlying genetic processes, cell division, and overarching mechanisms of evolution. Halfway through the unit, each dyad generated a KIM followed by either an expert or peer map critique activity (see Figure 2). Classes were randomly assigned to one of the two conditions (peer or expert critique) (see Table 3). More details about the KIM generation and critique activity follow below. After the critique activity, students in each dyad revised their own KIM. Students spent five hours (a single 50-minutes or a double period each day over one week) to complete the unit and finished the project by taking individual posttests. Throughout the unit, the teacher and a researcher provided support for technical and content-specific questions. Prior to implementation, the unit has been revised by teachers, students, and biology experts to evaluate the content, clarity, and alignment with the curriculum. The WISE unit has been revised based on their suggestions.

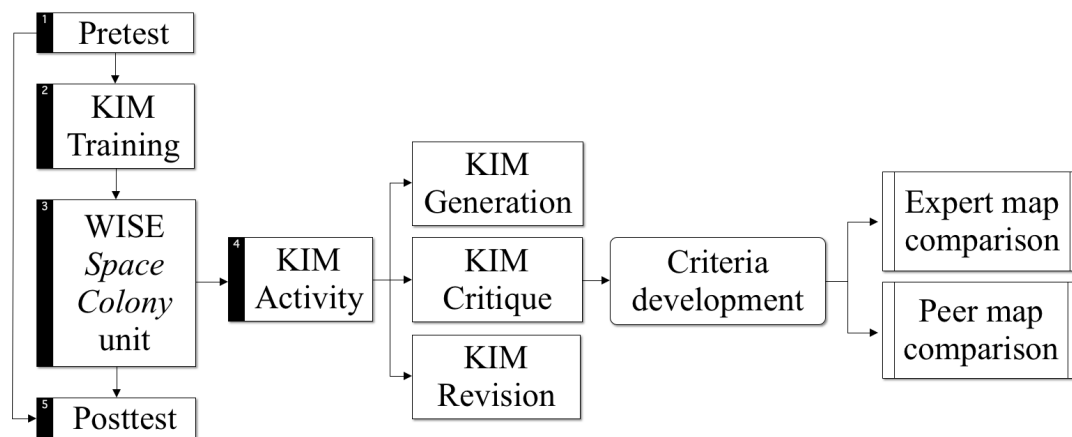


Figure 2. WISE Space Colony unit activity sequence.

KIM activities design

KIM Training. Shavelson (Shavelson, Lang, & Lewin, 1994) described the importance of an initial concept map training phase. The KIM training activities consisted of several tasks designed to introduce students to KIM generation, critique, and revision techniques. First, the teacher demonstrated the basic principles of

concept mapping, for example how to link two concepts with a labeled arrow. Student dyads then generated a paper-and-pencil practice concept map using a familiar context (“What does it take to have a pizza delivered to one’s home?”). After completion, dyads compared their maps against a worked-out example and discussed differences with the whole class. The teacher demonstrated how to compare and critique concept maps by pointing out different elements (such as idea placement, arrow direction, missing ideas, and link labels) and discussed them with the students. The KIM training activity took about thirty minutes.

KIM Generation Activity. After completing the first half of the WISE unit, each student dyad generated a KIM using paper-and-pencil worksheet (see Figure S2: Knowledge Integration Map worksheet). Yin and colleagues (Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005) found that a forced choice design with a given list of expert-selected ideas to choose from allowed for better comparisons between maps than free choice of ideas. The worksheet provided students with a list of six ideas (gene, allele, mutation, cell, natural selection, and genetic diversity) but gave students free choice of idea placement, link selection, link direction, and link label. The number of given ideas has been kept small to reduce working time and limit complexity. First, students were instructed to sort out all provided ideas by placing each in the corresponding level (DNA; cell; organism/population). Idea placement can be seen as an indicator for students’ association of an idea with a certain level. Second, students connected ideas with labeled mono-directional arrows. Connections could be between ideas within the same level or across levels. Especially connections across levels can be seen as an indicator for a more integrated understanding (see Figure S2: Knowledge Integration Map worksheet). Prior studies using KIMs (Schwendimann, 2007, 2008, 2009a, 2009b, 2011) indicate that KIMs are sensitive to changes in learner’s understanding. Student dyads spent about twenty minutes on generating their KIM.

KIM Critique Activity. After generating their KIM, student dyads were instructed to collaboratively compare and critique KIMs. Student dyads in the ‘expert map’ condition conducted self-critique by comparing their own KIMs to an expert-generated KIM shown on a projector. The expert KIM represented a composite of maps generated by biology experts who were given the same six ideas. Students were told by the teacher that the benchmark map was generated by experts but also that this map did not represent a final solution but one of many possible solutions.

In ‘peer map’ condition classes, the teacher collected the KIM worksheets after the KIM generation activity and randomly distributed them to other dyads in the same class for peer review. Each KIM worksheet received a unique identification number to keep the authors anonymous and to return the worksheets back to the authors after the other dyad added their review (see Table 2).

To reduce task complexity and time requirements, students in both conditions were instructed to prioritize their critique by identifying the most salient element (for example, a missing connection, a misplaced idea, or an incorrect link label) in their own map (expert map condition) or in the peer map (peer map condition) that they considered most different from the comparison map (expert map or own map respectively). Students were asked to mark the chosen element on the worksheet and to provide an explanation for their decision. A meta-study by Falchikov and Goldfinch (2000) found that student-generated criteria work better for peer review activities than using a set of given criteria. Students in both conditions were required

to develop their own criteria, make decisions on how to change the selected element, and provide an explanation for their decision.

Table 2
Comparison of conditions

Groups	Training	Objects to Critique	Critique reference	Object to Revise	Criteria used to critique
Expert map comparison group	Training worksheet	Own Maps	Expert Maps	Own maps	Generated by dyads
Peer review group	Training worksheet	Peer Maps	Own Maps	Own maps	Generated by dyads

KIM Revision Activity. Student dyads were instructed to respond to the critique (self-critique in the ‘expert map’ condition or critique received from another dyad in the ‘peer map’ condition) by suggesting a possible revision in their own KIM and provide supporting resources for their revision (for example by referring to sections in the WISE unit or a textbook). Students were asked to decide if they agreed with the suggestion and provide a rationale for their decision (see Figure S2).

Participants

Two experienced science teachers in the same US public high school implemented the WISE unit *Space Colony*, each with two classes. All students (n=81) were in 9th and 10th grade and came from a variety of ethnic and socioeconomic backgrounds. Only students who completed the pretest, posttest, and the concept mapping activity were included in the study (see Table 3). Completion rates in both conditions were comparable (72% for the expert map condition and 70% for the peer review condition). Students did not complete the unit for various reasons (including illness, overlapping school projects, and athletic competitions). No student opted out of the study. The high school had an enrollment of around 2000 students and was located in the urban fringe of a large city. School-wide, 12% of students received free or reduced price meals. Four percent of students were classified as English Learners; 67% were White, 16% were Hispanic or Latino, 10% were Asian, and 3% were Black. All students were familiar with technology-enhanced learning environments and a few had used another WISE unit before. The unit was implemented after students had learned about genetics, but before the unit on evolution. One class from each teacher was assigned to one of two conditions (expert map or peer map) (see Table 3). On the first day, the first class of the day of the first teacher was assigned to the expert condition and the second class of the day to the peer condition. For the second teacher, the order for assigning conditions was reversed. Within each class, the teacher randomly assigned students to dyads by drawing popsicle sticks with students’ names on them.

Table 3
Participants and conditions

Teacher	Total # of students	Students included in study (m/f)	Pretest average	Condition
Teacher A	30	21 (10m, 10f)	9.9	Expert-Map
	29	17 (9m, 8f)	9.8	Peer-Map
Teacher B	27	20 (10m, 10f)	10.1	Expert-Map
	28	23 (17m, 6f)	10.4	Peer-Map

Data Sources and Analysis

Pre/posttest Design and Scoring. This study used a pretest-posttest design to measure students' prior knowledge and track changes in students' understanding of evolution ideas. The teacher administered paper-based pretests to individual students on the first day of the project and posttests immediately after finishing the project. After the posttest, semi-structured interviews were conducted with selected students, which will be reported separately.

Knowledge integration (Liu, Lee, & Linn, 2010) provided the design rationale for both the WISE unit and the assessment items. Knowledge integration items are constructed to measure qualitative and quantitative changes in the connectedness between ideas (described as knowledge integration). They consist of open response items where students construct an explanation. Constructed-response items provide students with opportunities to express their own ideas and explain science phenomena in their own terms (McCarthy 2005), and elaborate on their ideas using supporting resources, an activity that is similar to scientific reasoning and critical thinking in real-life science inquiries. The items tap into commonly found alternative ideas and aim to track conceptual changes

The six knowledge integration constructed-response items focused on distinct ideas related to evolution (see Table 4 for an overview and the supplementary material for the full instrument). The items distinguish the quality and the numbers of connections that students construct among ideas.

Table 4

Pre/posttest knowledge integration items

Item number	Main ideas	Item framing
Question 1	Sex determination and meiosis	Explain the predictability of a baby's gender.
Question 2	Difference between mitosis and meiosis	Explain why gametes produced through mitosis instead of meiosis would lead to genetic diseases in humans.
Question 3	Genetic diversity and natural selection	Explain why higher genetic diversity can increase a population's chances to survive a new disease.
Question 4	Natural selection, genetic drift, and inheritance	Interpret a real life scenario in which the thermostat of an aquarium fails and all but a few fish die. Explain the survival of these fish.
Question 5	Genetic diversity and mutations	Explain changes in the genetic diversity of the fish population.
Question 6	Genetic diversity and natural selection	Provide an explanation regarding the survival of these fish if the thermostat should fail again.

The items were scored according to a six-scale knowledge integration rubric (0-5) to measure changes in students' abilities to connect ideas (Liu, Lee, Hofstetter, & Linn, 2008) (see Table 5). A higher score indicates a higher number of connections between scientifically relevant ideas and was interpreted as more integrated knowledge of scientific ideas. Explanations including a single idea or a mixture of correct and incorrect elements were coded as 'partial'. 'Basic' explanations needed to connect two normative ideas with a correct link whereas 'full' explanations" consisted of three

or more correctly connected normative ideas. Each explanation item of the pre- and posttest was weighted equally. The sum of all six items was calculated as a total pre- and posttest score (maximum knowledge integration score 30 (6x5). Previous research indicates that the knowledge integration scoring rubric provides a sensitive measure for the development of students' ideas in science (Linn, Lee, Tinker, Husic, & Chiu, 2006).

Each knowledge integration item was independently scored by two raters: For the pretest, the ICC coefficient was 0.8022, analysis of variance $F = 1.8524$; and for the posttest, the ICC coefficient was 0.8288, analysis of variance $F = 2.2742$. For both tests, there was no significant difference between raters, and the level of concordance was high. The mean of the two ratings was used for the analysis.

Table 5

Pre/posttest knowledge integration rubric

Knowledge integration	Sample essay question: <i>“Why do grapes (with seeds) that reproduce sexually have a greater chance to survive a new disease than (seedless) grapes that reproduce asexually?”</i>
No answer (0) [Blank]	None
Off task (1) [Irrelevant to the question]	I don't know
Incorrect (2) [No link]	Because they have seeds to fight off the disease
Partial (3) [Single normative idea]	The seedless grapes are all genetically identical.
Basic (4) [One link between two normative ideas]	Seedless grapes are all the same, but grapes with seeds have better chance and traits to survive.
Complex (5) [Two or more links between normative ideas]	Grapes that reproduce sexually have more genetic diversity in their gene pool that allows for a greater array of organisms from which natural selection may choose – allowing some to survive.

Validity and reliability studies of knowledge integration items indicate that these items have satisfactory psychometric properties in terms of reliability, item fit, and differential item functioning (Liu et al., 2008; Liu, Lee, & Linn, 2010; Liu, Lee, & Linn, 2011). For the present study, item validity was determined by asking content experts, science teachers, and psychologists to evaluate the scientific significance of the items and their alignment with the curriculum and editing the items based on their suggestions. In addition, pilot studies demonstrated that the items distinguished between students that the researchers and teachers rated as having low versus high levels of understanding of the unit. As a measure of the reliability of measurements, the Intraclass Correlation Coefficient (ICC) was calculated (reported in the next section).

Mixed-effect linear models with robust standard errors were chosen to take clustering due to nested data (students within classes) into account. Robust standard errors were calculated using the Huber-White sandwich method. Effect sizes between the means of the posttest scores were calculated across the two conditions to indicate the size of the observed effect.

KIM generation analysis. In KIMs, ideas are represented by connected arrows (directionality of connection), labels (nature of connection), idea placement in designated levels (categorizing ideas into DNA, cell, or organism/population levels), and cross-links across levels (indication of understanding connections between levels). There are many ways to analyze KIMs (Schwendimann, 2014a). This study used a multi-level analysis strategy to make sense of KIMs (Schwendimann, 2014b), which consisted of propositional, placement, and cross-link analysis.

Each proposition was evaluated using a six-point KIM knowledge integration rubric (Schwendimann, 2008) by distinguishing between the link label and the link arrow direction (see Table S2: KIM knowledge integration rubric). A higher score was interpreted as an indicator for more complex integration of ideas. Each proposition was weighted equally. The student maps were compared against the expert concept map to establish proposition validity. All KIM propositions were coded independently by two researchers using the KIM knowledge integration rubric (Cronbach's alpha 0.85). The mean of the two ratings was used to calculate the total proposition score.

The KIM placement score indicates how many ideas have been correctly placed in the corresponding level (DNA, cell, organism/population).

The 'cross-link score' is the sum of the proposition scores of valid cross-links. Cross-links indicate connections between different levels.

A 'total knowledge integration concept mapping score' before and after the revision was calculated as a composite of three sub-scores (total proposition score, idea placement score, and cross-link score). To describe changes in KIMs before and after the revision, a linear mixed effect analysis using robust standard errors was conducted.

KIM critique analysis. A nine-scale concept map critique rubric was developed to categorize the different forms of critique (see Table 6: KIM critique rubric). The rubric distinguished which element in the KIM student dyads considered most different from their reference map.

Table 6
KIM critique rubric

Kind of critique	Description	Example
None	No critique given	
Off-Topic	Comment unrelated to biology or concept mapping	<i>I am tired.</i>
General Remark	General critical comment without giving specific feedback	<i>Make more links between your ideas.</i>
Critique of idea placement	Critique that an idea is placed in the wrong level (DNA/Cell/Organism/Population level)	<i>Mutation' should be in DNA-level.</i>
Critique of missing idea	Critique that one or more of the given ideas have not been used	<i>You forgot to add 'mutation.</i>
Critique of link direction	Critique of the direction of an existing link (while keeping the same label)	<i>Your arrow should go in the other direction.</i>
Critique of missing link	Critique that an important link has not been created.	<i>You missed to connect mutation and allele.</i>
Critique of missing label	Critique that one or more links have not been labeled (blank line).	<i>You should add a label for the link mutation and allele.</i>
Critique of existing label	Critique of the label of an existing label (while keeping the same direction)	<i>Connection between allele and mutation should be "leads to" instead of "includes".</i>

Results

Research Question 1

To answer the first research question, “What are the affordances of collaborative KIM peer or expert critique activities?” pretest-posttest gains as well as KIM responses were analyzed. A paired t-test analysis was conducted to ensure that prior knowledge as measured in the pretest did not significantly differ between the classes of the two different teachers [paired t-test $t(80) = -0.67, p > 0.05$ (two-tailed)]. This indicates that students in all four classes had comparable levels of prior knowledge.

Pretest-posttest gains. Student learning of ideas related to evolution was analyzed by comparing pretest and posttest scores. To determine the effects of the two conditions, a linear mixed effect analysis with robust standard errors was conducted using the mean posttest score as the dependent variable, the mean pretest score as the independent variable, the condition (peer review or expert KIM) as the grouping variable, and the classes as random variables. Mixed effect model analysis indicates a regression coefficient of the mean posttest score variable of 0.40 (SE = 0.11), $p < 0.001$ which shows that students (in both conditions) gained in their understanding of ideas related to evolution from pretest to posttest after using the WISE *Space Colony* unit (see Figure 3). The effect size was large (Cohen’s $d = 0.52$ (SD pretest = 2.78, SD posttest = 3.17)). There was no main effect for condition (see Table 7 and 8).

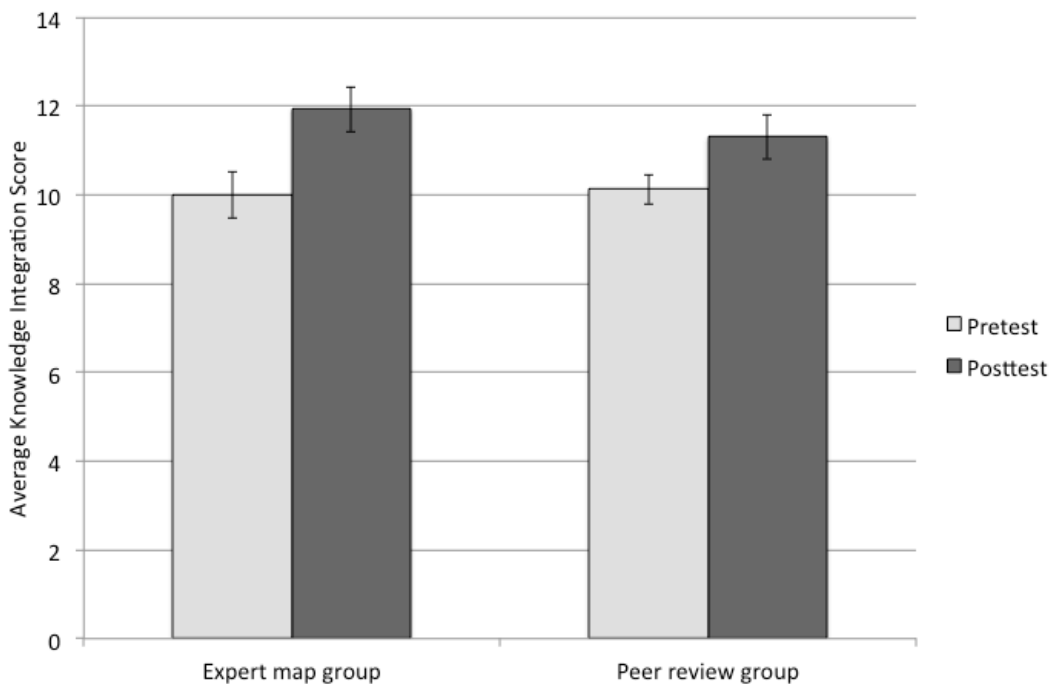


Figure 3. Pretest-posttest knowledge integration score gains (by condition).

Table 7

Overview table of descriptive statistics (knowledge integration items): All students

Knowledge	Pretest	Posttest	Linear mixed effect	Effect size
-----------	---------	----------	---------------------	-------------

integration items							analysis			(d)	
	n	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Coeff	SE	p-value	Cohen's d
Overall	81	5	24	10.06 (2.78)	2	18	11.62 (3.18)	0.40	0.11	0.000 ***	0.52

*p<0.05; **p<0.01; ***p<0.001

Table 8

Overview table of descriptive statistics (knowledge integration items): By condition

Knowledge integration items	n	Pretest			Posttest			Paired t-test	
		Min	Max	Mean (SD)	Min	Max	Mean (SD)	SD	p-value
Peer review	40	7	14	10.125 (2.10)	2	17	11.3 (3.13)	2.4069	0.0209 *
Expert map	41	5	24	10.00 (3.34)	5	18	11.93 (3.24)	3.3998	0.001 ***

*p<0.05; **p<0.01; ***p<0.001

To illustrate a student trajectory, consider this student response to the question “why sexually reproducing grape vines have a greater chance to survive a new disease than asexually reproduced vines that produce seedless grapes?” On the pretest the student responded “*Grapes with seeds that produce sexually are stronger and immune to new diseases because they are all natural*”. The pretest response uses the vitalist’ alternative ideas “stronger” and “immune” and implies that all grapes are identical (see table S3 for examples of the range of students’ alternative ideas of this item). In the posttest, the same student wrote, “*The seed grapes have a greater genetic chance to survive a new disease because they aren’t exactly alike. There is no genetic diversity in asexual reproduction.*” After the WISE unit, the student used the idea that organisms in sexually reproducing populations differ genetically from one another as well as the idea that sexual reproduction leads to greater genetic diversity than asexual reproduction. Such changes from pretest to posttest indicate improvements in integrating ideas about evolution.

KIM analysis. This study extended concept map activities by a collaborative critique and revision phase. Affordances will be discussed for each phase (generation, critique, and revision).

KIM generation. The KIM worksheet was divided into biology-specific levels to distinguish ideas (DNA, cell, organism/population) and make connections between levels visible. KIM proposition analysis indicates that students in both conditions successfully generated connections both within (see left box in Figure 4) and across levels (cross-links) (see center and right box in Figure 4). For the KIM cross-link analysis, a paired-samples t-test was conducted to compare the number of cross-links before and after the revision activity. There was a significant difference in the scores for cross-links pre-revision (M=27.37, SD=10.98) and post-revision (M = 27.92, SD = 11.07); $t(80) = -4.13$, $p = 0.0001$. Students were constrained to change only one element in the map but nevertheless significantly more cross-links were added in the revision. Cross-level links are especially desirable because they can be interpreted as “creative leaps” on the part of the learner (Novak 2006) and support reasoning across ontologically different levels (Duncan, 2007). Cross-links represent

connections across levels and they can be interpreted as indicators of more integrated knowledge of evolution that links micro, meso, and macro level ideas.

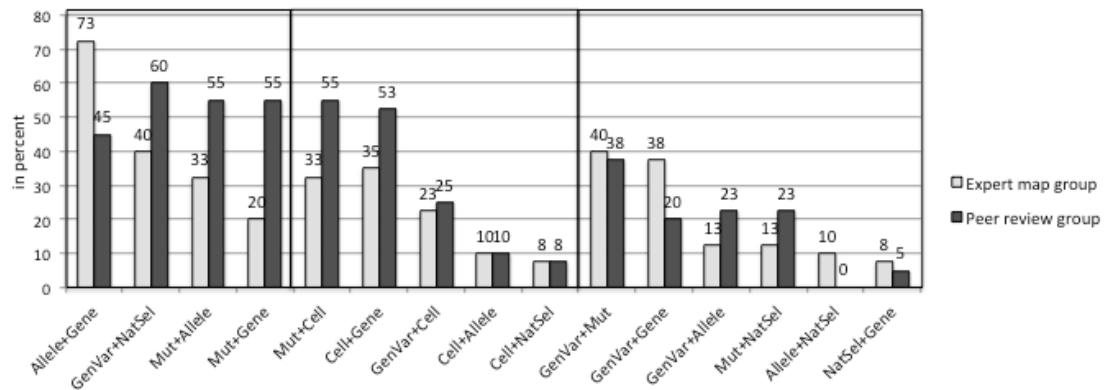


Figure 4. Student generated propositions (after revision) (in percent). Left box = within level connections; Center box = connections across one level; Right box = connections across two levels.

KIM Placement. Placement analysis of KIMs indicates students’ associations between levels (DNA (micro); cell (meso), and organism/population (macro) and particular ideas. Analysis of pre-revision KIMs suggests that the idea ‘cell’ was most often placed correctly (100% in expert map condition and 95% in peer map condition) (which is not surprising as the corresponding level was labeled ‘cell’). ‘Mutation’ (66% and 53%) was most often placed incorrectly and was frequently placed in the organism/population level. Some students’ explained “mutations change traits of organisms”. This indicates that these students perceived mutations to belong to the phenotypic level instead of a genetic level. Many dyads improved their initial idea placement in the revision activity (see below). Differences in students’ interpretations of ideas related to evolution emphasize the importance of using multiple measurements of understanding in concert to track changes, for example changes in elements of KIMs, students’ explanations of changes made in KIMs, and corresponding items in the pre/posttest.

KIM critique. Student dyads generated their own criteria for the KIM revision. Results indicate that students generated a broad variety of criteria to review different aspects of KIMs (see Figure 5). Students’ criteria have been coded using the KIM critique rubric (see Table 5).

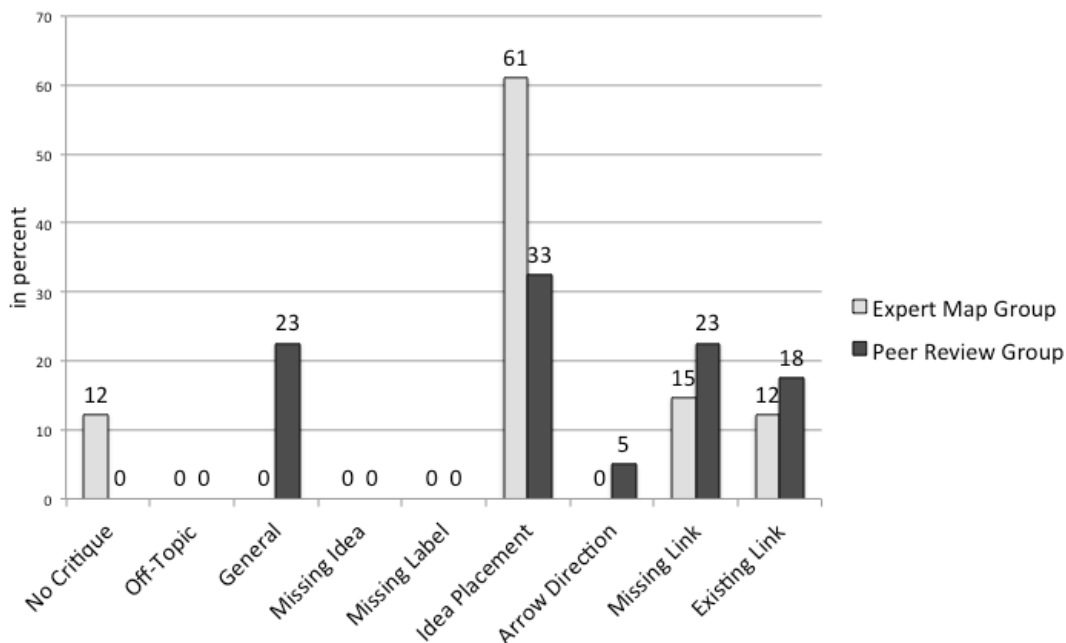


Figure 5. Knowledge Integration Map critique criteria by condition (in percent of total responses of students in each condition (n=81).

KIM revision. To complete the extended KIM activity (after generating and critiquing), student dyads were asked to decide how to revise their KIM and provide an explanation for their decision. The mixed linear effect model analysis used the mean KIM post-revision total improvement score as the dependent variable, the mean KIM pre-revision total improvement score as the independent variable, the condition (peer review or expert KIM) as the grouping variable, and the classes as random variables. Mixed linear effect model analysis with robust standard errors for the post-revision KIM total improvement score indicated a regression coefficient for the mean KIM post-revision total improvement score variable of 1.00 (SE = 0.01). $p < 0.001$, which suggests that students in both conditions improved their KIMs after the critique activity. On average, the single KIM revision improved the KIM score by 0.5 (on a 0-5 scale). Multiple regression analysis indicates that an improvement in the KIM score after the revision was positively associated with an estimated increase in the mean post-test score of 2.5; $p < 0.001$ (see Tables 7 and 8 for an overview of descriptive statistics for KIMs) (see Table S4: Quality of student dyads' revision suggestions).

Research Question 2

To answer the second research question, "Is peer or expert KIM critique more likely to promote knowledge integration?" each phase of the extended KIM activity was analyzed.

KIM generation. Student dyads in both conditions collaboratively generated their own KIMs from a given list of ideas. The KIM generation activity aimed to elicit

existing alternative ideas through idea placement and connections. Results indicate that generating KIMs can elicit alternative ideas as well as missing connections. Students in both conditions generated similar KIMs.

KIM critique. Analysis of dyad-generated criteria suggests that students in the expert and peer conditions differed in the ways they critiqued and revised KIMs (see Figure 5). Students in the expert map condition used mostly criteria regarding idea placements (61%), missing links (15%), and existing link labels (12%). Students in the peer map condition showed a different distribution: Only 33% critiqued idea placements, but 23% critiqued missing links, 18% existing link labels, and 5% link directions. No student critiqued a missing idea or a missing label. This might be explained by the explicit instructions to use all given ideas and label all connections.

All students in the peer condition provided some form of critique but 12% of dyads in the expert map condition did not provide critique. Students may have been more motivated to critique peers' work than to critique their own work.

Table 9

Categories of student-generated criteria

Non-relevant criteria = No Critique + Off Topic + General	Non-relevant criteria include missing, off-topic, and general comments.
Idea-focused criteria = Idea Placement + Missing Idea	Idea-focused criteria allow for a quick visual comparison between Knowledge Integration Maps without necessary conceptual reflection (for example "Is idea placed in corresponding level?"; "Is an idea from the given list missing?")
Link-focused criteria = Arrow Direction + Missing Link + Existing Label	Link-focused criteria provide conceptual feedback by identifying an important missing connection, pointing out that an arrow direction should be reversed, or suggesting the revision of an existing label.

For further analysis, student-generated criteria were grouped into the three categories *idea-focused criteria*, *link-focused criteria*, and *non-relevant criteria* (see Table 9). *Idea-focused criteria* evaluate the presence or placement of ideas whereas *link-focused criteria* identify missing links, the direction of a link, or the link label (see Figure 6). Students in the expert and peer condition differed in the prominence of criteria for KIM critique. The chi-square statistic is 16.293. This result is significant at $p < 0.05$. This suggests that students in the two conditions significantly differed in the kinds of criteria they used in the KIM critique activity. Z-scores were computed for raw scores of the criteria categories of each condition: *Idea-focused critique* ($z = 2.97$, $p = 0.001$) and *link-focused critique* ($z = 1.68$, $p < 0.05$).

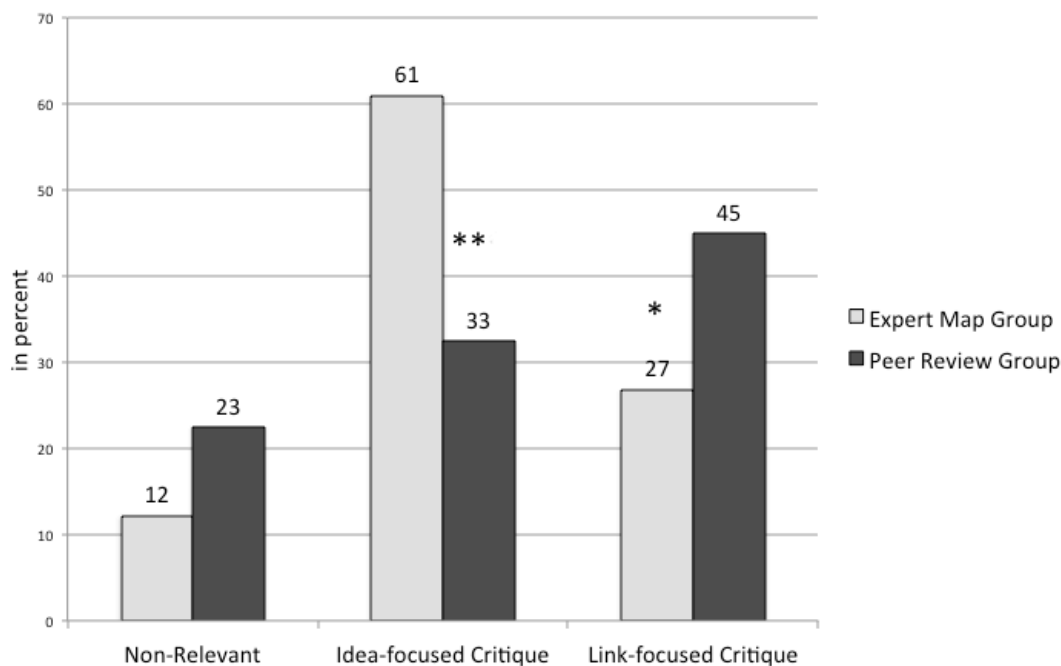


Figure 6. KIM criteria grouped by categories (in percent of total responses of each condition (n = 81). * = significant at $p < 0.05$; ** = significant at $p < 0.01$

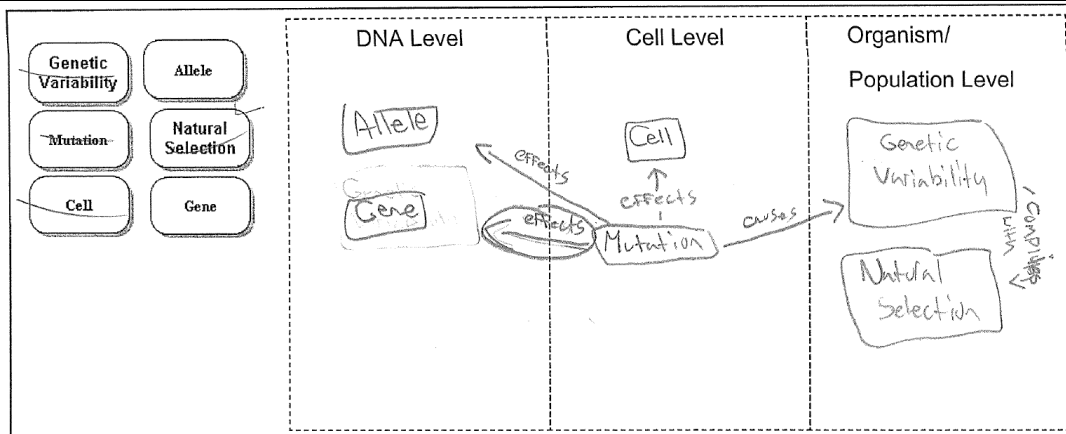
Students in the peer map condition generated more *link-focused criteria* (45%). Aligned with their criteria, students in this condition focused their revisions on improving propositions labels and directions. In the peer map condition example in table 7, the reviewing dyad commented “Genes make mutations”, which (incorrectly) suggests changing the (correct) existing link “Mutation effects gene”. This link-focused criterion refers to the connection between the ideas ‘gene’ and ‘mutation’. The author dyad rightfully disagreed with the reviewers’ suggestion and decided to keep their existing link (without providing further explanation).

Students in the expert map condition generated more *idea-focused criteria* (61%) that allow for quick comparisons with the expert map. Aligned with their criteria, most students in the expert map condition revised their idea placement. In the expert map condition example (see Table 10), the dyad commented on the connection between ‘genetic variability’ and ‘gene’ and decided to move ‘genetic variability’ to the ‘organism/population’ level “because it is more relevant to natural selection”. They used an idea-focused criterion to compare their map to the expert-generated map and identify the most saliently different element.

Table 10

Examples of KIM generation, critique, and comments (by condition)

Peer map dyad



Step 2) For Reviewers: Review

1. Analyze the given map.
2. Circle the one link you most disagree with.
3. Explain why you picked this link.

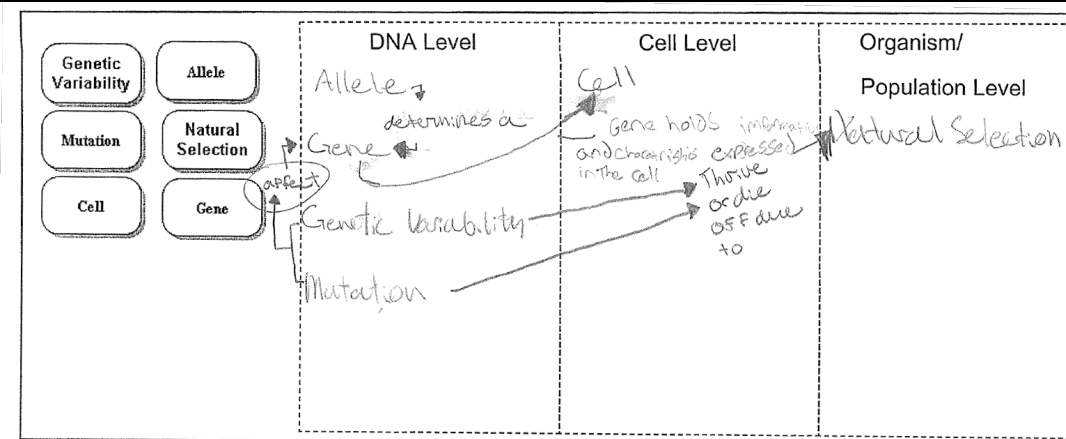
genes make mutations

Step 3) For Authors: Revision

- How would you change your map after the feedback?

I wouldn't.

Expert map dyad



Step 2) Master Map Review:

1. Compare your map to the Master map.
2. Circle the one link in your map that you consider most different from the Master map.
3. Explain why you picked this link.

We chose this because we considered the two to be synonymous, and because they ultimately affect the gene.

Step 3) Revision

- How would you change your map after the comparison?

We would move "genetic variability" to the Organism/population level because it is more relevant to natural selection.

KIM revision. In addition to differences in their criteria, students in the two conditions also differed in their decisions to revise their KIMs. In the expert map condition, 68% decided to make a specific revision. In accordance with the most frequent form of critique (idea placement critique) (see Table S5: Student dyad's revision decision after the critique activity), most students in the expert map condition decided to revise their idea placements (51%).

In the peer map condition, 46% indicated specific revisions. Several student dyads in the peer map condition (but none in the expert map condition) referred to sections in the WISE unit to support their ideas. This suggests that students in the expert map condition interpreted the expert map as a final solution that did not require revisiting other references, whereas students in the peer map condition referred to different resources to distinguish alternative ideas (their own and those of their peers).

It is interesting to note that 28% of students in the peer map condition, but no students in the expert map condition, decided not to revise their map and keep their original ideas. The dyads in the peer condition explicitly disagreed with the suggestions received from the reviewers, consistent with their decision not to revise. To illustrate, some responses were non-specific, for example “Well, you are wrong. At least that’s what I believe” offering no guidance about what to revise. Other responses were specific. For example, the reviewers suggested “You should move ‘genetic variability’ to the DNA level as it has something to do with genes” to which the authors responded “There is genetic variability within the population, so we believe it’s okay to leave it there.” For these students, findings suggest that it was not the feedback they received that led to improving their KIMs (because they disagreed with their peers’ comments), but the process of criteria generation, giving feedback, and self-critique. Findings suggest that giving specific feedback to and receiving (even flawed) feedback from anonymous peers who hold the same level of authority can provide a genuine opportunity to distinguish and sort out alternative ideas, facilitate critical reflection and use of supporting resources.

These findings suggest that extended KIM peer map activities align more strongly with knowledge integration principles than expert map activities. Knowledge integration emphasizes building connections between ideas. KIM peer map activities support the generation and revision of links between ideas through facilitating the use of link-focused criteria and responding to peers’ critique, which scaffolds students to revisit their own ideas and distinguish alternative ideas. Analyzing peer’s work and using peer’s critical feedback to revise one’s own work can facilitate knowledge integration processes. However, both collaborative extended KIM activities can support different knowledge integration processes of complex topics such as evolution. As learning tools, extended KIM activities can play an importance role in each phase of the knowledge integrating process through eliciting existing ideas and missing connections, adding new ideas and connections, revisiting prior ideas, distinguishing existing and new ideas and connections.

Discussion

This study compared two extended KIM activities, embedded in an inquiry-based online environment, designed to support knowledge integration processes. Results indicate that the combined effects of scaffolded inquiry activities and collaborative KIM activities contributed to a more integrated understanding of ideas related to evolution from pretest to posttest. This study extended traditional concept mapping in two ways. First, by introducing the concept mapping form KIM that facilitates sorting out ideas into different levels and visualizes cross-links between levels. Second, by extending the KIM generation activity by collaborative critique and revision phases. Findings from this study emphasize the importance of subsequent critique and revision phases for concept mapping activities and indicate that the combination of generating and critiquing KIMs can support the integration of ideas related to evolution across different levels.

Generating KIMs can elicit existing alternative ideas and connections, whereas subsequent critiquing and revising KIMs can foster distinguishing and sorting out alternative ideas. Generating and critiquing KIMs can support students to build and revise their own knowledge structures instead of rote learning. KIMs can serve as shared artifacts to initiate collaborative reviews and discussions among students. Different than essays, KIMs present each idea only once and show all relationships to or from that idea in one location. KIMs cluster ideas into domain-specific levels that can encourage critical reflection about relations between ideas within and across levels.

Students are often given too few opportunities to apply critique in science classrooms (Shen & Confrey, 2010). Critique activities can support the development of more coherent criteria (Lehrer & Schauble, 2004) to reflect, elaborate their ideas, revise their ideas, and self-monitor their learning progress, which supports the development of skills for lifelong autonomous learning (Chi, 2000; Linn, Davis, & Bell, 2004). Critiquing KIMs can elicit criteria to distinguish alternative ideas and provide students with genuine opportunities to negotiate and apply critique. Critique and revision activities can motivate students to revisit their initial ideas, distinguish alternative ideas, and revise connections, ideas, link labels, and idea placements.

Results suggest that the two forms of KIM critique activities can promote the generation and use of different kinds of criteria. The *expert map* condition received an expert-generated KIM as a trustworthy reference map. Studying an expert-generated map can model expert understanding, for example how ideas are connected across levels. Student dyads in the expert map condition had to critique their own work, which can be more challenging than evaluating other people's work (Linn & Clancy, 1992). Despite the teacher's introduction of the expert map as "just one of many possible solutions", results indicate that students interpreted the expert map as the one correct answer and might have limited opportunities for critique. Students might not have felt empowered to critique the reference map as experts generated it. Consequently, students in the expert map condition focused mostly on *idea-focused criteria* that allowed for a quick evaluation through visual comparison such as "idea placement".

This study identified the affordances of two different concept mapping conditions designed to promote knowledge integration in biology classrooms. Comparing student-generated work to expert-generated work (expert map condition) may have short-circuited reasoning by suggesting a solution. The alternative, students comparing their work to maps generated by peers was more generative. The *peer map* dyads compared their own maps against maps generated by their peers and provided critical feedback. Dyads in the peer map condition had to critique two KIMs (their own and their peers' map) and both generated and received critique. Findings indicate that students in the peer map condition developed and used more *link-focused criteria*, for example missing links, link direction, and link labels. Peer-generated KIMs might be easier to compare to one's own than to an expert generated KIM because they build on similar prior knowledge.

One initial concern for the peer review activity was that students might receive peer-generated work of varying quality and provide feedback that might reinforce non-normative ideas. Results suggest that some peer feedback consisted of non-normative ideas. However, students successfully distinguished alternative ideas, rightfully discarded non-normative suggestions, and expressed confidence in their own ideas. A bigger problem was that some dyads gave vague or unclear feedback. This suggests that it was not necessarily the content of the received peer feedback but

the knowledge integration processes of sorting out and distinguishing alternative ideas that led to improved KIMs. This aligns with findings by Chi (2000) who suggested that generating explanations could be more effective for learning than receiving explanations.

In the peer map condition, KIMs served as shared artifacts for collaboration within as well as in-between dyads. Several students in the peer map condition used the WISE unit as a resource to defend their ideas against peer feedback. In the peer map condition, students wrote for an audience (the other dyad in the same class), which can be motivating to provide explanations than the expert map condition who evaluated their own maps. Results suggest that comparing one's own ideas against those of peers can help students to recognize the value of their own ideas. Critical comparison of peer-generated KIMs can strengthen students' confidence in their own ideas, whereas comparing to an expert map might decrease such confidence.

In summary, findings from this study suggest that generating KIMs can elicit existing alternative ideas and help students visualize the connections between complex scientific ideas across different levels whereas critiquing and revising KIMs can facilitate revisiting and distinguishing alternative ideas and connections. Both forms of critical collaborative reflection of KIMs can facilitate knowledge integration processes and self-monitoring of students' learning progress. However, critiquing peer maps aligns more strongly with knowledge integration principles as the activity provides students with a genuine opportunity to engage in critique activities, negotiate, focus on links between ideas, and use different resources to support one's ideas.

Implications

What are the design considerations for setting up collaborative KIM critique activities? This study suggests that KIM activities should consist of training, generation, critique, and revision phases. An initial KIM training phase is necessary to familiarize students with the basic KIM generation and critique techniques. Learning how to generate, evaluate, and revise KIMs takes time and practice. Teacher-guided discussions should address different elements of KIMs (for example the presence or absence of ideas and connections, idea placements, link directions, and link labels) and how to conduct critical evaluations.

This study implemented an extended collaborative concept mapping form that represented a compromise between open and heavily constrained forms. Open-ended concept maps, in which students can choose their own ideas and links, might reflect the range of students' knowledge structures more accurately, but they can be more time-consuming and challenging to generate for students and more difficult to compare and evaluate for teachers and researchers. On the other hand, more constrained concept mapping forms can result in ceiling effects (Ruiz-Primo, 2000; Yin et al., 2005) and limit the representation of alternative ideas. KIMs aim to provide a balanced setting in several ways. This study suggests the value of using a forced choice design with a given list of expert-selected ideas to choose from combined with free choice of link labels, arrow direction, and idea placement.

To reduce time requirements and task complexity, only a relatively small number of expert-generated ideas should be provided to students. The properties of KIMs can foster collaborative activities by constraining learners to only one link between two ideas and idea placement in only one level. One unique characteristic of KIMs is the division into specific levels, which fosters clustering of related ideas in

close spatial proximity, and highlights cross-connections among ideas in different levels. Educators and researchers can create KIMs for different topics by identifying levels relevant to a specific domain.

This study emphasizes the importance of a review and revision step after the initial KIM generation phase to improve the quality of the map, to initiate critical discussions, and to strengthen knowledge integration. To reduce complexity and save time, students were constrained to selecting and revising only one salient element in the map. However, dyads still had to explore and discuss the whole map in order to decide upon which element they wanted to revise. Students should be encouraged to refer to resources in the online learning environment or other sources when providing feedback or evaluating received feedback. Evaluating alternative ideas can support students valuing their own ideas and distinguishing alternative ideas.

When using expert-generated KIMs for comparison activities, instruction should stress the point that there is no single correct solution for a KIM because even experts create many different maps when given the same concepts (Schwendimann, 2007). Expert maps should be presented as one of many possible good examples. Students could compare KIMs generated by different experts to illustrate that there are many different solutions. Peer critique should be conducted anonymously to reduce personal bias and reluctance to critique others.

Collaborative KIM peer critique activities can shift authority from the teacher to the student, encourage students to evaluate their own ideas, and reduce emphasis on right and wrong answers. As both forms of critique are important, this study suggests using peer and expert critique activities in concert to facilitate knowledge integration processes of complex scientific topics, such as evolution.

As embedded learning tools, KIM generation and critique can inform teachers and students of existing alternative ideas or missing connections. Particularly, cross-link scores can serve as KIM quality indicators to identify ideas that might require further revision and for which students need additional support. Teachers could provide additional support for students to facilitate adding missing ideas and links to their repertoire or distinguishing alternative ideas, for example by conducting scaffolded inquiry activities. Additionally, teachers could use KIMs as shared artifacts to lead class discussions to distinguish alternative ideas and connect ideas. In general, KIMs should be used in concert with a variety of learning activities, for example guided inquiry activities. Ideally, KIMs would be introduced early in students' academic careers rather than later, so they can integrate them into their developing study strategies.

Limitations and Future Research

There are several limitations to this study. The study is primarily exploratory and aimed to compare two different forms of concept maps to describe their affordances. Therefore, the results are restricted to description and comparison between two conditions rather than attribution to causal factors. The power of statistical tests generally depends on sample size, data collection, and other design aspects. In this study, the sample size and the nesting of data limits the power of the analysis. Data from this study are clustered within one school and classes of two teachers within that school. The nesting of data can result in less variability among individuals. Analysis methods need to take the clustering of data into account (for example by adjusting for the standard error of using multi-level modeling). Extensions of this study should be implemented in a wider range of randomly selected schools with more participants.

This study assigned conditions at the class level because participating teachers considered it problematic to implement two separate conditions within classes. Future iterations will assign conditions on the individual level by randomly assigning and distributing different conditions through the WISE platform. Additionally, using WISE instead of paper-based worksheets could reduce the number of omitted critiques by providing prompts and monitoring progress, such as by allowing students to proceed only after they provide a critique).

The WISE unit was embedded in an existing biology curriculum as a transition between the genetics and evolution units. Prior to the WISE curriculum, students already had an introduction into genetics, which might explain the relatively high pretest scores. Both critique activities (expert and peer map comparison) led to productive reflection and revision, and resulted in similar post-test performances.

Time constraints of participating teachers determined the length of the WISE unit. The similarities between conditions could be explained by the short duration of the critique activity and by the constraint to revise only one salient element in the KIM (implemented to reduce the complexity of the task and limit time requirements). Studies implemented in complex authentic classrooms have only limited control over conditions. Follow-up studies could explore the effects of extended KIM activities or multiple KIM critique-revision cycles over longer periods of time. Students could be instructed to revise more than one element.

The KIM used for this study divided the worksheet into three levels. Findings indicate that students interpreted the levels in different ways. A revised KIM could use a clearer distinction by keeping the organism and population levels separate or by using only two levels, for example genotype (micro) and phenotype (macro). A follow-up KIM study could provide a different set of concepts, for example by including “protein”. To further elicit students’ understanding of ideas and connections, KIMs could be extended to ask students to directly link to supporting resources in the WISE unit and explain their propositions. Although both conditions spent the same amount of time on the unit, the peer map activity took more time than the expert map comparison activity, mostly because collecting, shuffling, and distributing the paper-based KIM worksheets took additional time. Using computer-based KIMs instead of pen and paper could streamline the process of generating, anonymously distributing, and revising KIMs. As time in the science classroom is limited, future studies should explore timesaving forms of KIM critique activities for science education.

Acknowledgements

We thank participating teachers and students. We are grateful for the comments and suggestions by Prof. Andrew Martin and Dr. Łukasz Kidziński.

This material is based upon work supported by the National Science Foundation under grants DRL-0918743, and DRL-0334199. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Disclaimer

This paper is an edited and revised section of the dissertation work by Beat A. Schwendimann (2011). Mapping biological ideas: Concept maps as knowledge integration tools for evolution education. UC Berkeley, USA. Preliminary findings

have been presented at the 9th International Conference on Computer-Supported Collaborative Learning (CSCL), Hong Kong, China (Schwendimann 2011).

Conflict of Interest

The authors declare that there are no conflicts of interest.

Bibliography

- Bahar, M., Johnstone, A. H., & Hansell, M. H. (1999). Revisiting learning difficulties in biology 33 no2 84-6. *Journal of Biological Education*, 33.
- Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93(1), 26-55.
- Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, 27(5), 415-427
- Bonderup Dohn, N. (2009). Affordances revisited: Articulating a merleau-pontian view. *International Journal of Computer-Supported Collaborative Learning*, 4(2), 151-170
- Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In *Advances in instructional psychology: Educational design and cognitive science, vol. 5* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45(3), 293-321.
- Clark, D. B. (2000). Evaluating media-enhancement and source authority on the internet: The knowledge integration environment. *International Journal of Science Education*, 22(8), 859-872.
- Cliburn Jr, J. W. (1990). Concept maps to promote meaningful learning. *Journal of College Science Teaching*, 19(4), 212-17.
- Concord Consortium. (2006). *BioLogica*.
- Cuthbert, A., & Slotta, J. (2004). Fostering lifelong learning skills on the world wide web: Critiquing, questioning and searching for evidence. *International Journal of Science Education*, 27(7), 821-844.
- Davis, E. A., & Kirkpatrick, D. (2002). It's all the news: Critiquing evidence and claims. *Science Scope*, 25(5), 32-37.
- Derbentseva, N., Safayeni, F., & Canas, A. J. (2007). Concept maps: Experiments on dynamic thinking. *Journal of Research in Science Teaching*, 44(3), 448-465
- diSessa, A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change. Issues in theory and practice* (pp. 26-60). Kluwer Academic Publishers.
- diSessa, A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, 22, 293-331.
- Donnelly, D. F., Linn, M. C., & Ludvigsen, S. (2014). Impacts and characteristics of computer-based science inquiry learning environments for precollege students. *Review of Educational Research*, 84(4), 572-608
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching*, 44(7), 938-959

- Enyedy, N. (2005). Inventing mapping: Creating cultural forms to solve collective problems. *Cognition and Instruction*, 427-466.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Hamza, K. M., & Wickman, P. -O. (2013). Student engagement with artefacts and scientific ideas in a laboratory and a concept-mapping activity. *International Journal of Science Education*, 35(13), 2254–2277
- Hmelo-Silver, C. E., Marathe, Surabhi, & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert–novice understanding of complex systems. *Journal of the Learning Sciences*, 16(3), 307-331.
- Hoadley, C. (2004). Fostering collaboration offline and online: Learning from each other. In M. C. Linn, E. A. Davis, & P. L. Bell (Eds.), *Internet environments for science education* (pp. 145-174). Mahwah, NJ: Lawrence Erlbaum Associates.
- Keppell, M., Au, E., Ma, A., & Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment & Evaluation in Higher Education. Special Issue: Learning-Oriented Assessment: Principles and Practice.*, 31(4), 453-464.
- Lee, H. -S., Linn, M. C., Varma, K., & Liu, O. L. (2010). How do technology-enhanced inquiry science units impact classroom learning? *Journal of Research in Science Teaching*, 47(1), 71-90
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 635-679.
- Leif, S. (2005). *Evolution lab* [Computer Software].
- Linn, M. C. (2008). Teaching for conceptual change: Distinguish or extinguish ideas. In S. Vosniadou (Ed.), *International handbook of research on conceptual change*. New York: Routledge.
- Linn, M. C., & Clancy, M. J. (1992). The case for case studies of programming problems. *Communications of the ACM*, 35(3), 121-132.
- Linn, M. C., Davis, E. A., & Bell, P. (2004). *Internet environments for science education* (p. 412). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Linn, M. C., Eylon, B. -S., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 29-46). Mahwah, NJ: Lawrence Erlbaum Associates
- Linn, M. C., & Eylon, B. -S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration* (p. 336). Routledge.
- Linn, M. C., Lee, H. -S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313(5790), 1049-1050.
- Liu, O. L., Lee, H. S., & Linn, M. C. (2010). Multifaceted assessment of inquiry-based science learning. *Educational Assessment*, 15(2), 69-86.
- Liu, O. L., Lee, H. -S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, 13(1), 33–55.
- Liu, O. L., Lee, H. -S., & Linn, M. C. (2011). Measuring knowledge integration: Validation of four-year assessments. *Journal of Research in Science Teaching*, 48(9), 1079-1107
- McCarthy, C. B. (2005). Effects of thematic-based, hands-on science teaching versus a textbook approach for students with disabilities. *Journal of Research in Science Teaching*, 42(3), 245-263.

- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2000). *Assessing science understanding: A human constructivist view*. San Diego, CA: Educational psychology press. Academic Press
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76, 413-448
- Novak, J. D. (1980). Meaningful reception learning as a basis for rational thinking. *The Psychology of Teaching for Thinking and Creativity*.
- Novak, J. D. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization Journal*, 5(3), 175-184
- Novick, L. R., Schreiber, E. G., & Catley, K. M. (2014). Deconstructing evolution education: The relationship between micro- and macroevolution. *J Res Sci Teach*, 51(6), 759-788
- Osborne, R. J., & Wittrock, M. C. (1983). Learning science: A generative process. *Science Education*, 67(4), 489-508.
- Ruiz-Primo, M. A. (2000). On the use of concept maps as an assessment tool in science: What we have learned so far. *Revista Electrónica De Investigación Educativa*, 2(1), 30.
- Schwendimann, B. A. (2007). Integrating interactive genetics visualizations into high school biology. In *Annual meeting of the American Educational Research Association, Chicago, IL*
- Schwendimann, B. A. (2008). Scaffolding an interactive dynamic model to promote coherent connections in high school biology. *Annual meeting of the American Education Research Association (AERA)*. New York, NY
- Schwendimann, B. A. (2009a). Critiquing concepts maps for a more coherent understanding of biology. *Annual meeting of the National Association for Research in Science Teaching (NARST)*. Anaheim, CA
- Schwendimann, B. A. (2009b). Learning from a dynamic biology visualization through concept mapping and critique. In *Annual meeting of the American Educational Research Association, San Diego, CA*. San Diego, CA
- Schwendimann, B. A. (2011). Distinguishing evolution ideas through two different forms of collaborative critique-focused concept mapping activities. In *Proceedings of the 9th international conference on computer-supported collaborative learning (CSCL)*. Hong Kong, China: CSCL Conference
- Schwendimann, B. A. (2014a). Making sense of knowledge integration maps. In D. Ifenthaler & R. Hanewald (Eds.), *Digital knowledge maps in education: Technology enhanced support for teachers and learners*. New York: Springer.
- Schwendimann, B. A. (2014b). Multi-level analysis strategy to make sense of concept maps. In *Proceedings of the sixth international conference on concept mapping*. Santos (Brazil): International conference on concept mapping
- Shavelson, R. J., Lang, H., & Lewin, B. (1994). On concept maps as potential "authentic" assessments in science. *CSE Tech. Rep 388*.
- Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (2005). Windows into the mind. *Higher Education*, 49 no. 4, 413-430.
- Shen, J. (2007). Justifying alternative model: A case study on K-8 science teachers' understanding of frames of reference in astronomy. In *Annual meeting of the american educational research association, chicago, IL*. Chicago, IL.
- Shen, J. (2010). Nurturing students' critical knowledge using technology-enhanced scaffolding strategies in science education: A conceptual framework. *Journal of Science Education and Technology*, 19(1), 1-12.

- Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychol*, 52(2), 170-194
- Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, 40(5), 510-528
- Slotta, J. D., & Linn, M. C. (2000). How do students make sense of internet resources in the science classroom? In M. J. Jacobson & R. Kozma (Eds.), *Learning the sciences of the 21st century* (pp. 193-226). Hillsdale, NJ: Lawrence Erlbaum & Associates.
- Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? *Science Education*, 85(4), 328-348
- Taber, K. S. (2013). Representing evolution in science education: The challenge of teaching about natural selection. In B. Akpan (Ed.), *Science education: A global perspective* (pp. 65-91). Abuja, Nigeria
- Topping, K. J. (2005). Trends in peer learning. *Educational Psychology. Special Issue: Developments in Educational Psychology: How Far Have We Come in 25 Years?*, 25(6), 631-645.
- Trowbridge, J. E., & Wandersee, J. H. (1998). Theory-Driven graphic organizers. In *Teaching science for understanding: A human constructivist view*.
- Tsui, C. -Y., & Treagust, D. F. (2003). Genetics reasoning with multiple external representations. *Research in Science Education*, 33(1), 111-35.
- van Amelsvoort, M., Andriessen, J., & Kanselaar, G. (2005). Using representational tools to support historical reasoning in computer-supported collaborative learning. *Technology, Pedagogy and Education*, 14(1), 25-41.
- Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, 8(1), 3-19
- Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2), 166-184.
- 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
- Zhang, Z. H., & Linn, M. C. (2013). Learning from chemical visualizations: Comparing generation and selection. *International Journal of Science Education*, 35(13), 2174-2197

Supplementary information linked to the online version of the paper at Wiley-Blackwell

- Tables S1-5
- Figures S1-2
- Pretest/ posttest worksheet

List of figures

- *Figure 1.* Knowledge Integration Map (KIM).
- *Figure 2.* WISE *Space Colony* unit activity sequence.
- *Figure 3.* Pretest-posttest knowledge integration score gains (by condition).

- *Figure 4.* Student generated propositions (after revision) (in percent). Left box = within level connections; Center box = connections across one level; Right box = connections across two levels.
- *Figure 5.* Knowledge Integration Map critique criteria by condition (in percent of total responses of students in each condition (n=81).
- *Figure 6.* KIM criteria grouped by categories (in percent of total responses of each condition (n = 81). * = significant at $p < 0.05$; ** = significant at $p < 0.01$