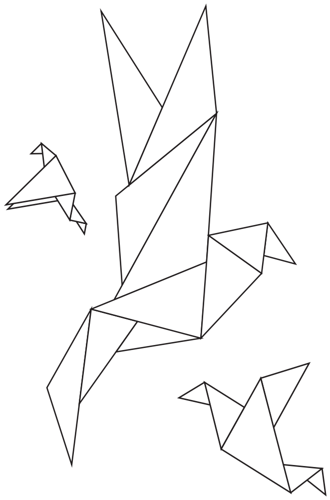


Acknowledgments

This material is based upon work supported by the National Science Foundation under grants DRL-1119670, ESI-0334199, and ESI-0455877. 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. The authors appreciate helpful comments from the Technology-Enhanced Learning in Science research group.

AUTHORS

Marcia C. Linn, Ph.D.
University of California, Berkeley

Jennifer Chiu, Ph.D.
University of Virginia

CORRESPONDENCE

Email
mclinn@berkeley.edu

COMBINING LEARNING AND ASSESSMENT TO IMPROVE SCIENCE EDUCATION

High-stakes tests take time away from valuable learning activities, narrow the focus of instruction, and imply that science involves memorizing details rather than understanding the natural world. Current tests lead precollege instructors to postpone science inquiry activities until after the last standardized test is completed—often during the last week of school. Students spend countless hours practicing and taking multiple-choice tests that have little educational value. Even college courses now devote class time to multiple choice clicker questions and often rely on similar items for course grades. Instead we need learning tests that help students understand science while at the same time measure progress.

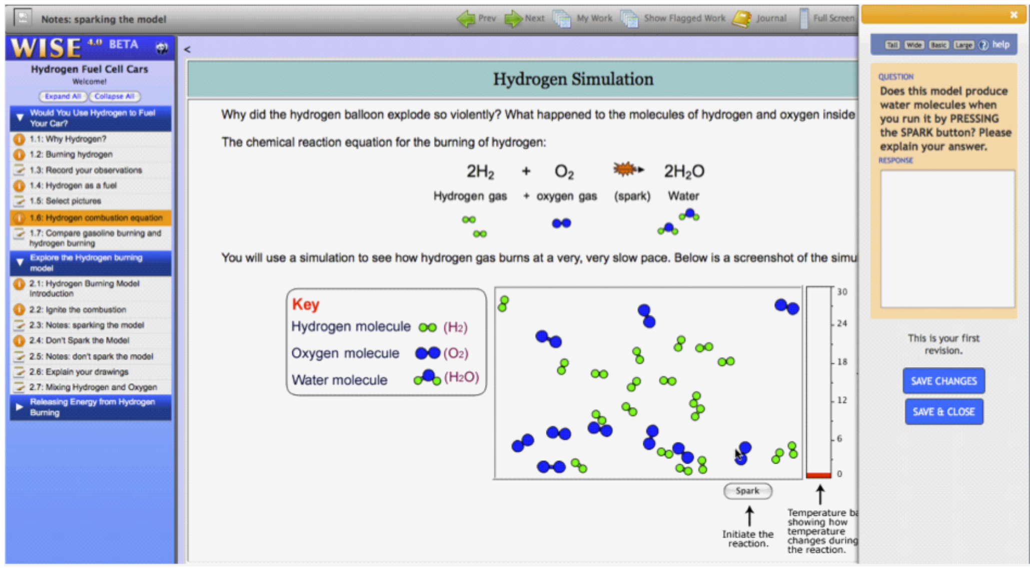
For example, an item on the California eighth grade science assessment asks: Which of the following best describes an atom?

a) protons and electrons grouped together in a random pattern
b) protons and electrons grouped together in an alternating pattern
c) a core of protons and neutrons surrounded by electrons
d) a core of electrons and neutrons surrounded by protons

These detail-oriented questions motivate teachers to stick to the textbook where students can access this information. Assignments ask students to memorize rather than encouraging them to understand the role of atoms and molecules in scientific processes such as recycling. Learning tests could ask students to design experiments to test their ideas about chemical reactions, to create concept maps to distinguish between energy transfer and energy transformation, or to construct an argument explaining how the chemicals in detergents can help clean up oil spills.

Emerging cyberlearning technologies can deliver and score learning tests continuously as students study complex science topics. Systems such as the Web-based Inquiry Science Environment (WISE, see WISE.Berkeley.edu) engage students in science units featuring learning tests, grade performance, guide students to refine their understanding, encourage students to monitor their progress, and diagnose class achievements for teachers (Figure 1).

Figure 1. A screenshot of the WISE: Web-based Inquiry Science Environment hydrogen cars project that includes visualizations of chemical reactions.



Learning tests in systems like WISE enable teachers to gather evidence about how their students learn. Teachers can use this information to identify places where students are struggling, provide feedback tailored to individuals or groups, and plan class discussions about topics that many students find difficult. When teachers use this kind of information to improve their practice, their students make substantial progress (Gerard, Varma, Corliss, & Linn, 2011).

Learning Test Goals

Learning tests combined with insights into how students learn have the potential to measure lifelong learning skills. Science courses need to produce lifelong learners who are capable of expanding their knowledge throughout their lives. Students need the ability to make sense of contemporary issues such as genetic engineering, global climate change, new cancer treatments, and alternative energy sources. Yet consistent with the emphasis on memorization, many adults claim they have forgotten any science they might have learned. To make the curriculum more relevant, science courses need to prepare students to use and refine their knowledge while improving the quality of their lives.

Research with thousands of students and hundreds of teachers shows that when students explore contemporary science issues like recycling, global climate change, and genetic inheritance using online units featuring scientific visualizations they learn more than students who study the same topics using the textbook (Linn, Lee, Tinker, Husic, & Chiu, 2006). Students who study these units learn to distinguish among alternative disease treatments, critique experiments about climate change, and reason about dilemmas such as designing an energy-efficient house. In addition, students prefer units with online visualizations to their textbook because visualizations (of phenomena such as chemical reactions) allow them to see how science works and test their ideas. By incorporating learning tests into online environments we can strengthen science learning and assess students at the same time.

"Learning tests combined with insights into how students learn have the potential to measure lifelong learning skills."

Promoting and assessing skills necessary for lifelong learning can prepare students to use science in their lives. We expect that when students learn to read they will use and expand their abilities every day. We prepare students to use mathematics regularly (although many complain that they have no need for calculus). We can change science courses so they prepare students to revisit their ideas and build more complex understanding. To accomplish this, we need to align curriculum, assessment, and professional development.

Teaching and Assessing Lifelong Learning

Teaching for lifelong learning is complicated because students come to science class with lots of intuitive, incomplete, contradictory, and idiosyncratic ideas. Research offers convincing evidence that adding new ideas in lectures, experiments, or visualizations is not sufficient to improve student understanding. Students need to integrate new ideas with existing knowledge to make progress in science. To develop useful and generative understanding students need to engage in the process of knowledge integration (Linn & Eylon, 2011).

"By incorporating learning tests into online environments we can strengthen science learning and assess students at the same time."

The knowledge integration framework, a constructivist perspective, emerged from an extensive longitudinal study to show that students need to not only comprehend new ideas but also to distinguish them from their existing ideas and to figure out how to incorporate them into a coherent account of the topic (Linn & Hsi, 2000). Knowledge integration has roots in studies showing that students maintain conceptual ecologies that include p-prims, analogies, epistemological beliefs, facets, facts, and intuitions.

Essentially, for any topic, students have developed multiple ideas along with evidence to support their existing views at home, in school, and in cultural activities. They may equate heat and temperature because they use the words interchangeably. They may argue that heat is a characteristic of a high temperature when discussing the weather. Furthermore, students tend to limit the applicability of their ideas to specific situations. Thus students may explain that objects in motion remain in motion in science class but come to rest on the playing field.

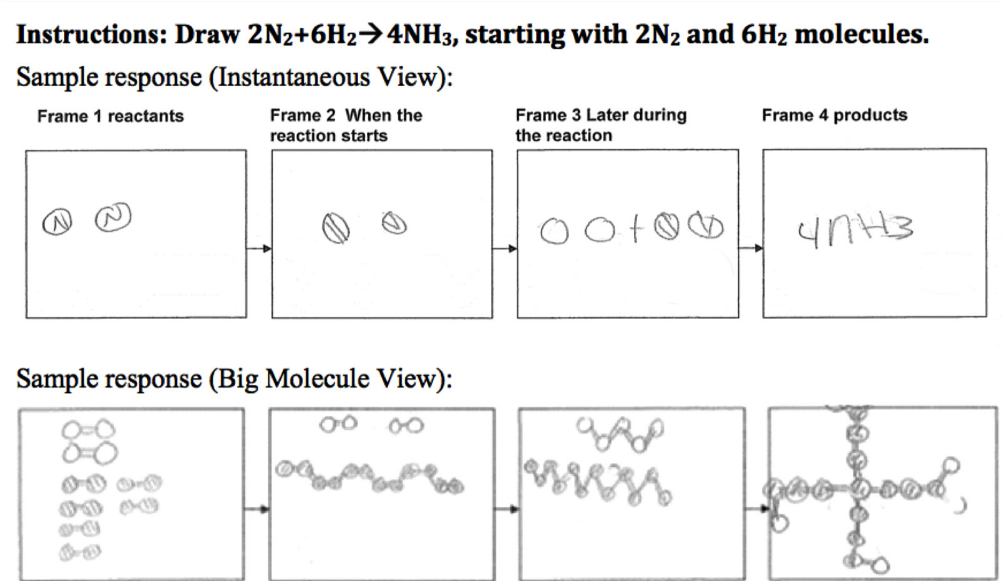
To gain more integrated understanding, students need to refine their repertoire of varied, often contradictory, and contextualized ideas. To help curriculum designers create knowledge integration based instruction, researchers have identified design principles (Kali, Linn, & Roseman, 2008). These principles have recently been synthesized in the knowledge integration pattern (Linn & Eylon, 2011). The pattern involves articulating existing ideas, adding new ideas, distinguishing new ideas from existing ideas, and building a coherent argument by reflecting on the evidence for the ideas in the repertoire. To reform science instruction so that it promotes lifelong learning we need curriculum materials that implement this pattern and learning tests that measure the integration of knowledge. We illustrate how this works for activities featuring visualizations, concept maps, and essays.

Visualizations and Assessment

Scientific visualizations can illustrate phenomena that are too fast, small, or vast to observe such as chemical reactions (See Figure 1). By themselves, visualizations are often deceptively clear—motivating students to report that they understand when, in fact, they lack deep insights (Chiu & Linn, in press). Instruction can overcome deceptive clarity by using the knowledge integration pattern (Linn & Eylon, 2011). Learning tests can increase the efficiency and effectiveness of the instruction.

Using the knowledge integration pattern to overcome deceptive clarity starts with asking for predictions to elicit existing ideas about the visualization topic. Students need to make predictions to generate their existing ideas. When they make predictions they are ready to compare these ideas to the ones introduced in the visualization. When making predictions students may report that chemical reactions involve breaking molecules into individual atoms and then recombining them in a new configuration based on their interpretation of symbolic equations (Figure 2). Consistent with the knowledge integration pattern, students use the visualization to add new ideas. Chiu and Linn (in press)

Figure 2. Examples of student drawings of chemical reactions



found that students often reported that they understood after viewing the visualization but assessments revealed that this was not the case.

Following the knowledge integration pattern, helping students to distinguish the new ideas from their existing ideas (and test their understanding) can overcome deceptive clarity. Zhang and Linn (in press) showed that asking students to make drawings of a sequence of events is far more effective than exploring the visualization alone (see Figure 2). Students often run the visualization multiple times to complete their drawings (Chiu, 2010), revealing the value of the visualization.

Distinguishing ideas is helpful but students also need to consolidate their ideas. The final step of the knowledge integration pattern involves having students reflect on their investigations and create a coherent argument. Research shows the value of asking students to explain things like greenhouse gas accumulation in terms of chemical reactions (Chiu, 2010).

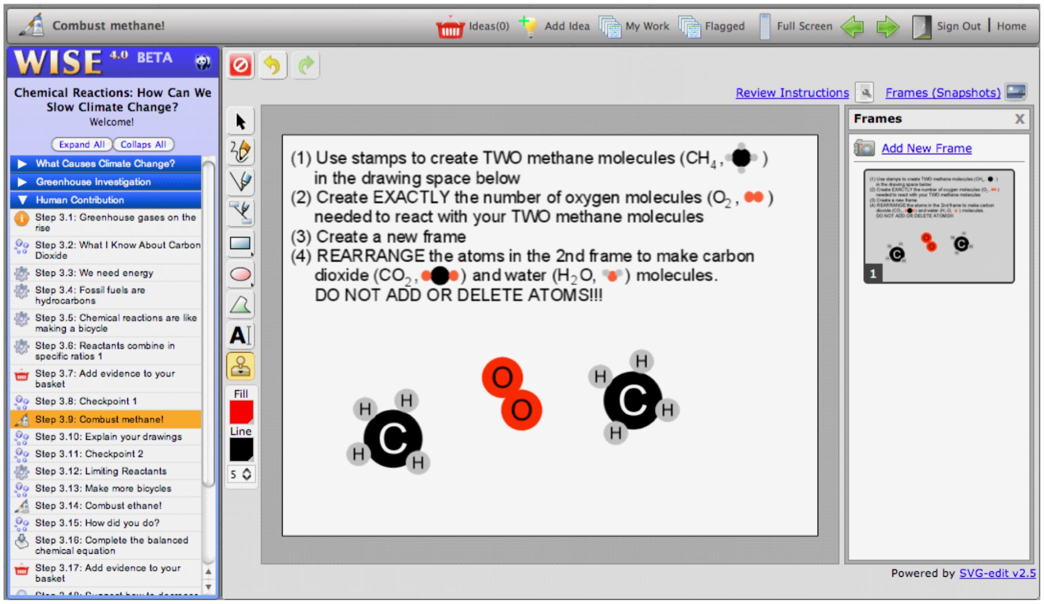
Several learning tests occur in chemical reactions. The drawings help students distinguish ideas and also assess their progress. Scoring the drawings, however, is time consuming. We have tried two ways to make scoring more automatic. Using WISE Draw we could analyze drawings (Figure 3). As discussed below, we can also score the essays students write when they explain phenomena using their understanding of chemical reactions.

Recently, Zhang (2011) created a learning test using a selection task. She identified 12 drawings that captured most of the variations generated by participating students when they were asked to create four drawings that capture the main events in the visualization. In the selection task, students selected among these drawings to illustrate four main events in the chemical reaction. There are over 12,000 possible sequences so it is unlikely that students will succeed by chance. She reported that students had difficulty selecting a valid sequence. The drawings in the selection task expanded the alternatives students considered. Zhang found that the selection task was just as effective as the drawing task for advancing student understanding but was also very easy to score automatically.

“The activity and the embedded learning test improved student understanding while also providing students and teachers with valid, automated scores to gauge their progress.”

In summary, designing instruction using the knowledge integration pattern can overcome the deceptive clarity of the chemical reactions visualization. In addition, designing a selection task by examining the drawings that students construct spontaneously resulted in a learning test that encouraged students to distinguish ideas. Furthermore, the activity and the embedded learning test improved student understanding while also

Figure 3. Example of how students could use WISEDraw to make their predictions.



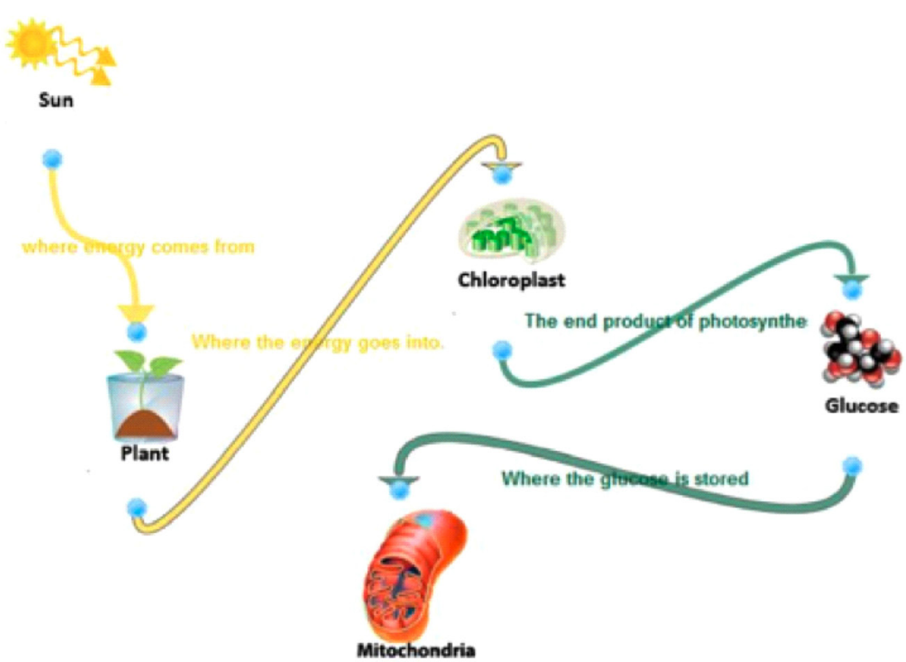
providing students and teachers with valid, automated scores to gauge their progress.

Teaching and Assessing with Concept Maps

Concept mapping activities can help students distinguish among their ideas just as we showed for drawings of chemical reactions (Linn & Eylon, 2011). MySystem, an open source WISE activity developed by the Concord Consortium supports a form of concept mapping. Students diagram connections and characterize the transformation of energy within systems (Figure 4). MySystem works best when embedded in the knowledge integration pattern.

To illustrate, in the photosynthesis unit learners use MySystem to explain to a new student how a rabbit gets and uses energy from the sun (Ryoo & Linn, 2010). Students first make predictions about energy transfer and transformation. They then interact

Figure 4. MySystem diagram. Students can select icons and label links to illustrate energy transfer and transformation.



with a visualization of photosynthesis to get new ideas. To distinguish their ideas they create a MySystem diagram. To sort out their ideas they write an energy story (see sophisticated example in Figure 5). Both the MySystem diagram and the energy story can serve as learning tests. WISE can compute an overall score capturing the coherence of a diagram as well as distinct scores for how well the diagram represents *energy source, direction of energy flow, modes of energy transfer,* and *thermodynamic properties*. It is possible to score MySystem diagrams while students are learning and to give students guidance to help them revise their ideas.

Essay Questions, Learning, and Assessment

Short and long essay questions require students to generate coherent arguments and explain complex phenomena. Research shows that they capture deep understanding in ways that multiple-choice items cannot (Lee, Liu, & Linn, 2010). Furthermore, studies show that asking students to write essays, even if they are not graded, can improve learning outcomes (Karpicke & Blunt, 2011). As learning tests, essay questions are important to help students consolidate their ideas.

Many teachers neglect essays because they do not have the time to grade them. It is common for middle school science classes to exceed 40 students and for teachers to have five or six sections of a class, yielding over 200 essay responses to each question

Figure 5. Photosynthesis science essay in response to: "Write a story about how the rabbit gets and uses energy from the sun."

The sun creates energy through nuclear fusion, which moves through wave lengths in space until it reaches earth. Most is reflected by the atmosphere but a tiny bit makes it through, and an even smaller fraction is absorbed by the chloroplasts. They use it to rearrange CO_2 and H_2O into glucose and by that process it turns into chemical energy. Chemical energy is stored in glucose, and when the rabbit absorbs the plant, and thus the glucose, it gains the chemical energy.

(WISE units usually have 10 or more essay questions). Fortunately, new technologies make it possible to automatically score essays for knowledge integration (see rubric in Figure 6). For example, c-rater, a recent cyberlearning technology developed at ETS can score short essays (e.g., Sukkarieh & Pulman, 2005). C-rater evaluates essays based on a set of clear, distinct concepts. These concepts are developed using a 4-part scoring process: (1) model building, where researchers identify key concepts for the item; (2) natural language processing, where student and model responses are analyzed for linguistic features; (3) main points identification, where the linguistic features are used to identify the concepts in the student responses; and (4) scoring, where scores are assigned to responses based on main points (Sukkarieh & Blackmore, 2009). The accuracy of the scores depends on the linguistic complexity of the responses. Short science essays are good candidates for c-rater scoring because they have constrained vocabulary and syntax. C-rater can provide an overall score for each response, and distinct scores on how well the response addresses each key concept.

"… essays that ask students to create coherent arguments are excellent learning tests and essential to full implementation of the knowledge integration pattern."

Figure 6. Example Knowledge Integration Rubric used in the photosynthesis unit.

KI Score	Typical Responses
1 Off task	The sun helps animals survive by giving heat.
2 No Link, Partial Link	The sun helps animals survive by producing energy and helping some plants grow which animals eat.
3 Full Link	The sun helps animals survive by giving them food. Together, the sun and plants make food. Without this food, animals would not be able to survive for very long. Animals cannot make their own food because they are not full of chloroplasts and chlorophyll. Plants have BOTH of these things so they can get through photosynthesis, unlike animals.
4 Complex Link	The sun helps animals survive by the plants absorb the energy and perform photosynthesis. After that the energy is used for food in the plant and when an animal eats it the energy is transferred from the plant to the animal.

A proof-of-concept study using c-rater, produced reliable knowledge integration scores for student essays (Linn, Gerard, Matuk, & Liu, 2011). For an item in the WISE Photosynthesis unit, where students were asked to "Explain how the sun helps animals survive," the Kappa value between the c-rater score and human score was close to .70, higher than the Kappa value between two human raters who received a half-day of training on the knowledge integration rubric.

In summary, essays that ask students to create coherent arguments are excellent learning tests and essential to full implementation of the knowledge integration pattern. Methods for automated scoring of essays can empower teachers to use them more regularly. An open question is how best to use these scores to provide guidance for students.

"Questions that ask for explanations or critiques and that feature multiple right answers could encourage respondents to distinguish among ideas."

Improving Assessment in Lecture Classes

College courses may reinforce the image of science as requiring memorization by using clicker questions and machine-scorable tests that emphasize recall of information. Clickers are widespread. A quick search of publisher websites reveals that these devices are mainly used for recall questions. For example, an astronomy item asks:

The time for one cycle of lunar phases is:

a) about one day.

b) about 24.8 hours.

c) about one year.

d) the same as the time for one cycle of the moon relative to the stars.

e) the same as the time for one cycle of the moon relative to the sun.

This question, like the one about atoms from the California assessment, focuses on science details. The use of questions like this helps explain findings that college students who completed astronomy courses were unable to illustrate the phases of the moon. Most students believed that the phases are caused by the moon passing through the earth's shadow, which occurs only during an eclipse rather by than explaining that half of the moon is illuminated by the sun and that the portion visible from the earth varies over time (Bell & Trundle, 2008). Research shows the value of embedding the clicker questions in

"Scientists spend little time memorizing. They spend more time conducting experiments and interpreting the results. They know the details relevant to their own work because they use the information every day to reason about dilemmas."

the knowledge integration pattern (Crouch, Fagen, Callan, & Mazur, 2004; Linn & Eylon, 2011; Lorenzo, Crouch, & Mazur, 2006).

Classes using clicker questions can be improved by implementing the knowledge integration pattern. This means substituting questions that ask for understanding for those that ask for recall. Questions that ask for explanations or critiques and that feature multiple right answers could encourage respondents to distinguish among ideas. For example, the selection question from the chemical reactions visualization would be good.

When instructors use clicker questions as part of a larger goal of knowledge integration they could elicit ideas by asking students to make predictions either individually or in small groups. For example, a prediction question might offer a set of moon representations and ask respondents to create a valid sequence of images. To add ideas about moon phases, studies show the advantages of visualizations (e.g., Bell & Trundle, 2008). A clicker question might initiate a distinguishing ideas activity by asking students to select among alternative drawings of the waxing moon. Students might then discuss their choices in their small groups using evidence from the visualization. Instructors could ask students to write a short essay comparing their predictions to the group solution to encourage students to build a coherent argument.

Teaching a topic like the phases of the moon for understanding takes more time than focusing on details and involves dealing with complex phenomena such as the relative position and motion of the earth, moon, and sun. New visualization technologies can make these topics accessible and intriguing. Combining visualization with judicious use of clickers or other class response systems by using the knowledge integration pattern could strengthen lectures. Incorporating these technologies into precollege and college courses could increase interest in science and satisfaction with science courses. C-rater can provide an overall score for each response, and distinct scores on how well the response addresses each key concept.

Conclusion

Transforming science education and developing lifelong learners is within our reach. Emerging technologies and instructional frameworks support the design of learning tests that enable students to develop deep understanding and teachers to become effective guides. Instead of focusing on the ideas that students add during instruction, these technologies can administer learning tests that measure how students distinguish among ideas and evaluate new and existing ideas while they learn. Learning tests can assess the coherence of students' understanding of a new topic. In addition to serving as learning opportunities for students to engage in knowledge integration processes, learning tests give teachers insight into students' progress.

The knowledge integration framework characterizes learners as developing a repertoire of ideas, adding new ideas from instruction, experience, or cultural interactions, distinguishing these ideas in varied contexts, making connections among ideas at multiple levels of analysis, and developing more and more nuanced criteria for evaluating ideas. This process culminates in an increasingly linked set of views about any phenomenon. This kind of scientific thinking is essential for lifelong learning. By focusing learners on using evidence to evaluate new and existing ideas, these activities encourage students to build a coherent understanding and to become aware of their own learning process.

Incorporating learning tests into science has important implications for educational policy. When No Child Left Behind legislation mandated annual testing in reading and mathematics schools often abandoned or neglected science instruction (Au, 2007). Many elementary schools dropped science in favor of increased emphasis on reading. Early reading programs increased emphasis on learning basic decoding skills. Students focused on learning to read but not on reading to learn science.

Now that science tests are included in evaluation of schools, the emphasis on detail-oriented questions deters students and instructors from emphasizing understanding and lifelong learning. This emphasis on details gives students a distorted picture of science and scientific careers. Scientists spend little time memorizing. They spend more time conducting experiments and interpreting the results. They know the details relevant to their own work because they use the information every day to reason about dilemmas. When they need a detail from another field, they are likely to look it up rather than depend on their memory.

Similarly, science instruction can encourage students to use reliable Internet sites to look up information rather than relying on a possibly faulty memory or being influenced by persuasive messages. For example, to answer the question from the California assessment, students could use a site such as wiki.answers.com and enter "describe an atom." This site returns the answer: "An atom consists of a nucleus and electrons. The nucleus consists of protons and neutrons crammed together. The electrons revolve around the nucleus in shells or orbits." Of course, learners need to know what information they are missing before they can look it up. Identifying gaps in knowledge is part of science reasoning and is emphasized in the knowledge integration pattern. The research from the Technology-Enhanced Learning in Science (TELS) center and the Center for Curriculum Materials in Science (CCMS) both funded by the National Science Foundation showed that students learn more when they explore science ideas than when they rely on typical textbooks (Kali et al., 2008).

For example, rather than memorizing the parts of an atom, students could learn and apply ideas about atoms in units that spur lifelong learning. While studying a unit on Hydrogen Fuel Cell Cars they could learn about atoms to investigate the tradeoffs between gasoline-powered and hydrogen-fuel-cell-powered cars and buses. They could study the chemical structures of materials and relate those to recycling policies.

In summary, we are abandoning lifelong science learning and hands-on experimentation so students can practice and take tests emphasizing details. We can reclaim some of this valuable classroom time by using online learning environments that incorporate learning tests to measure lifelong learning skills. Think about what would happen if scientists spent time memorizing new facts rather than investigating compelling problems. To retain our competitive advantage in science we need to restore a focus on lifelong learning to the classroom.

References

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Bell, R. L., & Trundle, K. C. (2008). The use of a computer simulation to promote scientific conceptions of moon phases. *Journal of Research in Science Teaching, 45*(3), 346-372.

Chiu, J. (2010). Supporting students' knowledge integration with technology-enhanced inquiry curricula (Doctoral dissertation). Retrieved from ProQuest Dissertation and Theses database. (UMI No. AAT 3413337).

Chiu, J., & Linn, M. C. (in press). The role of self-monitoring in learning chemistry with dynamic visualization. In A. Zohar & Y. J. Dori (Eds.), *Metacognition and science education: Trends in current research.* London, UK: Springer-Verlag.

Crouch, C., Fagen, A., Callan, J., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, *72*, 835-838.

Gerard, L. F., Varma, K., Corliss, S., & Linn, M. C. (2011). Professional development in technology-enhanced inquiry science. *Review of Educational Research*, *81*(3), 408-448.

Kali, Y., Linn, M. C., & Roseman, J. E. (Eds.). (2008). *Designing coherent science education.* New York, NY: Teachers College Press.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775.

Lee, H.-S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education, 24*(2), 115-136.

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

Linn, M. C., Gerard, L., Matuk, C., & Liu, L. (2011). "CLASS: Continuous learning and automated scoring in science." (National Science Foundation DRL-1119670). Berkeley, CA: University of California, Graduate School of Education.

Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners.* Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M.C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J.L. (2006). Teaching and assessing knowledge integration in science. *Science*, *313*, 1049-1050.

Lorenzo, M., Crouch, C., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*, 118-122.

Ryoo, K., & Linn, M. C. (2010). Students' progress in understanding energy concepts in photosynthesis using visualizations. In *Proceedings of the Ninth International Conference of the Learning Sciences.,* Chicago, IL: International Society of the Learning Sciences, Inc.

Sukkarieh, J. Z., & Blackmore, J. (2009). C-rater: Automatic content scoring for short constructed responses. In *Proceedings of the 22nd International Florida Artificial Intelligence Research Society (FLAIRS) Conference,* Sanibel Island, Florida.

Sukkarieh, J. Z., & Pulman S. (2005). Information extraction and machine learning: Automarking short free-text responses for science questions. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education,*Amsterdam, The Netherlands.

WISE; the Web-based Inquiry Science Environment. Retrieved October 10, 2011, from http://wise.berkeley.edu.

Zhang, Z. (2011). Promote chemistry learning with dynamic visualizations: Generation, selection, and critique (Doctoral Dissertation). Retreived from Proquest Dissertation and Theses Database. (UMI No. 3469523)

Zhang, Z. & Linn, M. C. (in press). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching.*