Chapter 8
Simulations for Supporting and Assessing Science Literacy

Edys S. Quellmalz
WestEd, USA

Barbara C. Buckley
WestEd, USA

Matt D. Silberglitt
WestEd, USA

Mark T. Loveland
WestEd, USA

Daniel G. Brenner
WestEd, USA

ABSTRACT

Simulations have become core supports for learning in the digital age. For example, economists, mathematicians, and scientists employ simulations to model complex phenomena. Learners, too, are increasingly able to take advantage of simulations to understand complex systems. Simulations can display phenomena that are too large or small, fast or slow, or dangerous for direct classroom investigations. The affordances of simulations extend students' opportunities to engage in deep, extended problem solving. National and international studies are providing evidence that technologies are enriching curricula, tailoring learning environments, embedding assessment, and providing tools to connect students, teachers, and experts locally and globally. This chapter describes a portfolio of research and development that has examined and documented the roles that simulations can play in assessing and promoting learning, and has developed and validated sets of simulation-based assessments and instructional supplements designed for formative and summative assessment and customized instruction.

INTRODUCTION

Digital and networking technologies permeate school, work, personal, and civic activities. They are central, transformative tools for addressing goals and challenges in all walks of life. Conceptualizations of 21st century skills and new literacies go beyond traditional views of academic, disciplinary learning to emphasize the need to take advantage of the affordances of technologies to foster application of domain knowledge and competencies in real-world contexts, goals, and problems. Research in cognitive science about how people learn has long documented the importance of transferable knowledge and skills and how learning situated in one context must be explicitly

DOI: 10.4018/978-1-4666-9441-5.ch008

scaffolded to promote use in multiple contexts for new problems. Currently, research and development on the affordances of a vast, ever expanding array of digital and networking technologies are providing evidence of the power of technologies for transforming learning environments and the methods for monitoring and evaluating learning progress.

Technologies are revolutionizing the ways that learning can be both promoted and assessed. Interactive technologies such as computer-based learning environments and physical manipulatives enhanced by digital technologies provide teachers with powerful tools to structure and support learning, collaboration, progress monitoring, and formative and summative assessment. These digital tools enable new representations of topics that are difficult to teach and new approaches to individualized learning, that supports a wider range of learners' needs.

Large-scale national and international studies are providing evidence that technologies are truly changing and improving schools by enriching curricula, tailoring learning environments, offering opportunities for embedding assessment within instruction, and providing collaborative tools to connect students, teachers, and experts locally and globally (Quellmalz & Pellegrino, 2009; Quellmalz & Kozma, 2003; Law, Pelgrum, & Plomp, 2008).

In this chapter, we will describe projects in WestEd's Science, Technology, Engineering and Math (STEM) program that are capitalizing on the affordances of digital tools to deepen and extend the kinds of science learning highlighted in the *Framework for K–12 Science Education* and the *Next Generation Science Standards* (National Research Council [NRC], 2012a, 2012b). These projects draw upon a broad range of recent research to develop and evaluate interactive technologies for learning and assessment. This chapter will describe the principles extracted from work in the learning sciences, model-based reasoning, multimedia research, universal design for learning (UDL) and evidence-centered design (ECD) and employed in the design and development of these technology tools. We will summarize strategies for successful implementation of these new digital learning tools in current educational settings, as well as studies of the interventions' technical quality and impacts on learning. We will discuss how these interactive technologies support the development of learning progressions and multilevel, balanced assessment systems. We conclude the chapter with a discussion of additional lines of research and development.

This article is based upon work supported by the US Department of Education (Grant 09-2713-126), the National Science Foundation (Grants 0733345, 1108896, 1221614, and 1420386), and the Institute of Education Sciences, U.S. Department of Education (Grants R305A100069, R305A120047, R305A120390, and R305A130160). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Department of Education, the Institute of Education Sciences, or the National Science Foundation.

BACKGROUND

The research and development projects in WestEd's STEM program draw upon theory and findings from cognitive science and multimedia research and emphasize the schematic and strategic knowledge involved in systems thinking and the science practices related to inquiry-based problem-solving for real-world issues. The focus on real-world applications shifts attention from the inert retention of disconnected scientific domain knowledge to understanding the science relevant to environmental and social issues, making informed decisions, and communicating about the issues.

Focus on Significant Knowledge and Skills

In K–12 schooling, frameworks and standards recommend the knowledge and processes central within traditional academic domains and for 21st century skills. These documents lay out goals for *what* should be taught in K–12 education, recommending development of not just declarative and procedural knowledge, but integrated knowledge structures (schema), strategic use of knowledge, and transfer of knowledge to solve novel problems. Learning sciences research has documented that the mental models of experts can be represented as large, organized, interconnected knowledge structures, called schema, that are used in conjunction with domain-specific problem-solving routines (Bransford, Brown & Cocking, 2000). In the domain of science, models of science systems can serve as schema for organizing knowledge about dynamic system phenomena. Thus, formation of models and development of model-based reasoning is a foundational practice in science.

Moreover, the ever-widening horizons enabled by digital tools expand conceptualizations of literacy in science and other academic subjects to the larger context of "new literacies," a term that has emerged in recognition of the expanded ways that knowledge and information can be represented, accessed, processed, shared, and expressed. New literacies require expertise in the use of a range of digital media and information and communication technologies and exercised in academic and applied settings to collaborate, communicate, solve problems, and achieve goals (Quellmalz & Haertel, 2008).

Design of Learning Environments

Numerous national reports summarize key research findings about *how* to design effective learning environments and assessments for academic domains and 21st century skills (Branford, Cocking, & Glaser, 2000; Pellegrino & Hilton, 2013). For example, the reports *How People Learn* and *Applying Cognitive Science to Education* distill decades of learning research that informs strategies for supporting deep learning.

Representations of Science Phenomena

The use of physical, conceptual, and mathematical models has greatly benefitted scientific discovery. Models and simulations have profoundly changed the nature of inquiry in mathematics and science—for scientists, as well as for students (Nersessian, 2008). For example, economists, mathematicians, and scientists employ simulations to model alternative outcomes of complex systems.

Multimedia learning researchers have examined the effects of pictorial and verbal stimuli in static, animated, and dynamic formats, as well as the effects of active versus passive learning enabled by degrees of learner control (Clark & Mayer, 2011; Mayer, 2005; Lowe & Schnotz, 2008). Mayer's *Cambridge Handbook of Multimedia Learning* (2005) and Clark and Mayer's recently updated book, *eLearning and the Science of Instruction* summarize multimedia research and offer principles for multimedia design (Clark & Mayer 2011).

The majority of multimedia design principles address how to focus students' attention and minimize extraneous cognitive processing. Research addresses how to guide attention by making the most important information salient and omitting irrelevant representations (cf., Betrancourt, 2005; Clark & Mayer, 2011). Studies also recommend that complex simulations should be carefully focused to foster desired learner outcomes. Rather than realistically portraying every detail of systems, it is more important to ensure that the most relevant parts are easily discernible (cf., Lee, Plass, & Homer, 2006; van Merrienboer & Kester, 2005).

Extensive research has been conducted on external forms of *stimulus representation*s. Research on the perceptual correspondence of models to the natural systems they represent (e.g., cells, circuits, ecosystems) suggests features to consider in designing science learning environments. Research on models' physical similarity to natural systems and the ways in which system interrelationships are depicted through conventional physical and symbolic forms and signaled or highlighted can inform the design of science learning and assessment activities. The use of visual cues such as text consistency, color, and arrows can help students map between representations and gain deeper conceptual understandings, increasing the "readability" of dynamic visualizations (cf., Ainsworth, 2008; Kriz & Hegarty, 2007; Lowe & Schnotz, 2008). In a review of principles in multimedia learning, Betrancourt (2005) noted that multimedia representations have evolved from sequential static text and picture frames to increasingly sophisticated visualizations. Animations are considered particularly useful for providing visualizations of dynamic phenomena that are not easily observable in real space and time scales, cf., plate tectonics, circulatory system, animal movement (Betrancourt, 2005; Kühl, Scheiter, Gerjets, & Edelmann, 2011). Dynamic representations are well suited for portraying changes in temporal scale, spatial scale, and for depicting multiple viewpoints. For example, to represent changes in spatial scale, visual call-outs are frequently used for magnification. Cross-sectional views, cutaway views, and exploded views are used in both static and animated depictions of dynamic events. Color can cue key features of complex scenes, the ordering of events, and the categorization of structures so that learners can extract relevant information. Signaling in complex animations may include giving cues such as "there will be three steps" and directly instructing students to reason through the components of systems to increases comprehension (Hegarty 2004; Tversky et al., 2008). A growing body of research is developing principles for organizing and displaying information that will help focus learner attention (Ware, 2004).

User Control

User control refers to the degree of control the user can exert while interacting with representations. User control may allow students to pause, rewind, and replay dynamic visualizations, and manipulate features and sequences. Controlling the pace of presentation can increase the likelihood that students will learn from and understand the display (cf., Lowe & Schnotz, 2008; Schwartz & Heiser, 2006).

Digital media can also allow learners to explore, manipulate, and display the results of investigations of dynamic representations. Animations become *interactive* simulations if learners can manipulate parameters as they generate and test hypotheses, thereby taking advantage of technological capabilities suited to conducting scientific inquiry. Simulations can provide technology enhancements for science instruction by representing dynamic science systems "in action," making invisible phenomena observable and enabling manipulations of these models for active investigations of authentic problems (Gobert & Clement, 1999). For example, Rieber, Tzeng, and Tribble (2004) found that students given graphical feedback with short explanations during a simulation on laws of motion far outperformed those given only textual information. Plass, Homer, and Hayward (2009) found that manipulation of the content of a visualization, not just the timing and pacing, can improve learning outcomes compared to static materials.

Universal Design

Building on work by Rose and Meyer (2000), CAST (2008) developed a framework for Universal Design for Learning (UDL) recommending three kinds of flexibility: (1) representing information in multiple formats and media, (2) providing multiple pathways for students' action and expression, and (3) providing multiple ways to engage students' interest and motivation. Digital learning and assessment environments can present information in more than one modality (e.g. auditory and visual, static and dynamic), allow simultaneous presentation of multiple representations (e.g., scenes and graphs), and vary simple and complex versions of phenomena and models. Multiple pathways for expression may include interactivity, hints and worked examples, and multiple response formats (drawing, writing, dragging and dropping).

Universal Design for Computer-Based Testing (UD-CBT) further specified how digital technologies can create tests that more accurately assess students with a diverse range of physical, sensory, and cognitive abilities and challenges through the use of accommodations (Harns, Burling, Hanna, & Dolan, 2006; Burling et al., 2006). Accommodations are defined as changes in format, response, setting, timing, or scheduling that do not alter in any significant way the constructs the test measures or the comparability of scores (Phillips, 1993). UD-CBT has been found to level the playing field for English language learners (ELL) and students with disabilities (Wang, 2005; Case, Brooks, Wang, & Young, 2005). Tools already built into students' computers can allow multiple representations (text, video, audio); multiple media; highlighters, and zoom magnification (Twing & Dolan, 2008; Case, 2008).

Model-Based Learning

Researchers in model-based learning suggest that learners' mental models of science phenomena are formed, used, evaluated, and revised as they interact with phenomena in situ and with conceptual models, representations (including text), and simulations (Gobert & Buckley, 2000; Buckley, 2012; Clement & Rea-Ramirez, 2008). For example, cycles of model-based reasoning help learners build deeper conceptual understandings of core scientific principles and systems, interpret patterns in data, and formulate general models to explain phenomena (Stewart et al., 2005; Lehrer et al., 2001). A highly significant finding of cognitive research is that learners who internalize schema of complex system organization—structures, interactions, and emergent behaviors—can transfer this heuristic understanding across science systems (e.g., Goldstone, 2006; Goldstone & Wilensky, 2008).

Simulations for Science Learning

Numerous studies illustrate the benefits of simulations for science learning. Simulations can support the development of deeper understanding and better problem-solving skills in areas such as genetics, environmental science, and physics (Krajcik, Marx, Blumenfeld, Soloway, & Fishman, 2000; Schwartz & Heiser, 2006; Rieber et al., 2004; Buckley et al., 2004; Buckley et al., 2010). Students using simulations tend to rely more on conceptual approaches than on algorithmic approaches or rote facts during problem-solving (Stieff & Wilensky, 2003; White & Frederiksen, 1998), and can make causal connections among the levels of science systems (Hmelo-Silver, et al., 2008; Ioannidou, et al., 2010). Using dynamic, interactive simulations to make these connections explicit and salient benefits students' learning (Slotta & Chi, 2006).

Taking Science to School summarizes research-based recommendations for learning environments, suggesting that knowledge and skills be taught and tested in the context of larger investigations linked to driving questions, rather than teaching and testing individual ideas and skills separately (Duschl, Schweingruber, & Shouse, 2007). Learning theory holds that the environments in which students acquire and demonstrate knowledge should be situated in contexts of use (Simon, 1980; Collins, Brown, & Newman, 1989). Learning environments should involve active problem solving and reasoning. Cycles of feedback and scaffolding should be designed to promote and monitor learning progress. Cycles of feedback, revision, and reflection are aspects of metacognition critical for students to regulate their own learning (Pashler et al., 2007; White & Frederiksen, 1998).

Scientific literacy incorporates the goal that individuals can engage in science-related, real-world issues and ideas as reflective citizens. Interactive technologies can support the development of new literacies through affordances that help students develop collaboration and communication skills as they engage in deep, extended problem solving.

Evidence-Centered Design

Evidence-centered design (ECD) facilitates coherence of assessment and learning environments by linking the targeted knowledge and skills with evidence of proficiency, and with tasks and items to elicit that evidence (Messick, 1994; Mislevy, Amond, & Lucas, 2004; Mislevy & Haertel, 2007). The process begins by specifying a student model of the knowledge and skills to be addressed. Schematic, systems thinking about science phenomena should begin with explication of the kind of mental model that is to be constructed by the learner and for what purpose or application.

The ECD design process aligns the student model with an evidence model that specifies which student responses are evidence of targeted knowledge and skills, how student performances will be analyzed, and how they will be reported. The student and evidence models are then aligned with a task model that specifies features of tasks and questions intended to elicit student performances that provide evidence of the targeted knowledge and skills. The WestEd science projects used evidence-centered design to align the science content and practices addressed to scoring and reporting methods, and then to principled design of tasks that elicit evidence of understanding and use of the targeted science knowledge and skills.

WESTED STEM RESEARCH AND DEVELOPMENT

Design Principles for Technology-Enhanced Interactive Learning Environments

WestEd technology-enhanced science projects address topics ranging from middle and high school to graduate courses. The projects address significant science and knowledge and practices aligned with national science frameworks and standards. The digital environments are designed according to the principles derived from learning research described above and distilled in *Taking Science to School* (Duschl, Schweingruber, & Shouse, 2007). The principles include actively engaging students in meaningful, real world problems, cycles of feedback, and scaffolding to promote learning. Findings from multimedia research inform the design of multiple, overlapping representations that cue attention to relevant features of the science phenomena and offer user control of a range of responses and expression. The projects use evidence-centered design (ECD) to structure the alignment of the science content and practices addressed (student models) with the types of instructional and assessment activities (task models) and the forms of evidence that are collected to document and summarize learning (evidence models) (Mislevy, Almond, & Lucas, 2004).

The SimScientists program (simscientists.org) developed suites of simulation-based assessments designed to promote and assess model-based learning in existing middle school science curricula. Each suite is composed of two or three curriculum-embedded modules that the teacher inserts into a unit. A summative simulation benchmark assessment is administered at the end of the unit. These interactive modules feature a simulation environment based on scientific principles for a model of a science system that is grade-appropriate and specifies core ideas to be applied during problem-driven inquiry activities. The modules are designed as supplements to ongoing curriculum units, to be implemented by the teacher at points in the curriculum sequence when key ideas have been introduced and the teacher judges that students can apply the concepts as they conduct the simulation-based investigations.

ChemVLab (chemvlab.org) is a collaboration between Carnegie Mellon University (CMU) and WestEd. This work is based on an existing Java user interface developed at CMU that simulates a chemistry stockroom and workbench for carrying out a wide array of investigations, along with a newly developed Flash-based user interface and programming interface between the Flash and Java components of the system that allow for delivery of structured tasks to students and assessment of their performance within the simulated environment. Designed for integration into high school chemistry lab courses, the activities improve upon the typical paper-based practice problems and provide students with practice that includes practical, simulated exposure to wet-lab work, data collection and interpretation, problem solving, and sense making. The system offers real-time customized feedback to guide student investigations and provides error correction in the application of chemistry concepts. Reports to students and teachers provide ongoing progress monitoring and allow teachers to adjust instruction based on gaps in students' knowledge and abilities (Davenport, Rafferty, Timms, Yaron, & Karabinos, 2012; Davenport, Rafferty, Yaron, Karabinos, & Timms, 2014).

The *Voyage to Galapagos* project *(VTG,* voyagetogalapagos.org*)* has created web-based software to help students "follow" the steps of Darwin through a simulation of the Galapagos Islands, guiding students' learning about natural selection and evolution. Students are encouraged to explore the islands, take pictures of iguanas, evaluate the animals' characteristics and behaviors, and use scientific methodology and analysis to "discover" evolution as they explore the virtual open environment of the Galapagos Islands. The program encourages students to follow the steps of good scientific inquiry, e.g., developing hypotheses, collecting and analyzing data, and drawing conclusions, while revealing basic principles of evolution theory to students.

Voyage to Galapagos is investigating the question: How much assistance is the right amount to provide to students as they learn with educational technology? To investigate this, *VTG* has been developed to provide middle school students with opportunities to do simulated field work, including data collection and analysis during investigation of three key biological principles: variation, function, and adaptation. The goal of the project is to find the right balance between minimum and full support, allowing students to make their own decisions and, at times, mistakes. Learning goals and tasks aligned with NGSS have been used to create an intelligent tutoring system to collect data about student actions, assign probabilities of students having made certain errors, and make decisions about error feedback and hints to provide students.

In the sections below, we use the evidence-centered design framework to describe the designs of the WestEd STEM simulation projects.

Student Models

The STEM technology-enhanced projects begin with specifications of the knowledge and skills to be fostered and assessed. National science frameworks and standards have been the major sources. For example, the College Board *Standards for Science Success,* the National Research Council *Framework for K–12 Science Education,* and the *Next Generation Science Standards (NGSS)* recommend deeper learning of the fundamental nature and behavior of science systems, along with the practices scientists use to study system dynamics (College Board, 2009; NRC, 2012a, 2012b). The projects then focus on science knowledge and practices particularly suited to dynamic, interactive modalities and that are difficult to promote and assess in static formats. The technology affordances permit visual representations of the structure, function and behaviors of systems "in action" that are typically too big, small, fast, slow, or dangerous for students to experience directly in classrooms. In addition, the technologies allow active investigations that support use of NGSS science and engineering practices. In the sections below we describe the sets of interrelated learning targets that serve as the student models of the projects.

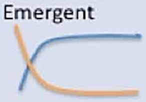
SimScientists Student Models

The overarching design of the SimScientists assessment and instructional modules integrates the frameworks of model-based learning and evidence-centered design (Buckley, 2012; Mislevy, Almond & Lucas, 2004). Incorporating the learning principles described above, design begins with specification of the science knowledge and practices to be addressed. The SimScientists computer-based modules are designed as supplements to ongoing curricula, therefore they selectively focus on integration of knowledge and application of science practices. The knowledge integration occurs within the organizational frame of an integrated science system model consisting of three tiers: 1) the system components, 2) interactions among components, and 3) the emergent system phenomena. The three-level science system model is intended to help learners form a schema of the organizational structure of all science systems (Bransford, Brown, & Cocking, 2000). The system model framework also serves as the target for the model-based reasoning promoted (Buckley, 2012). The projects reframe content standards identified by NGSS, the American Association for the Advancement of Science (AAAS), and the National Assessment of Educational Progress (NAEP) science in terms of multilevel science system models that explicate and integrate understanding of the system's components, their interactions, and behaviors that emerge from these interactions (Clement & Rea-Ramirez, 2008; Hmelo-Silver & Pfeffer, 2004; Grotzer, 2003; Perkins & Grotzer, 2000). The projects also reframe science practices in terms of the model-based reasoning needed for students to demonstrate and extend their understanding of the system models through investigations.

The first level of specification for the SimScientists student model is the System Target Model. As shown in Figure 1, SimScientists' ecosystems assessments and instructional modules focus on multiple levels of ecosystem organization, the interactions of components within levels and across levels, and the changes that emerge from those interactions over time. We characterize these levels as *components* and their roles, *interactions* between components, and *emergent* behavior that results from component-component interactions within communities over time. For the middle school grades, the ecosystem levels are represented in terms of food for energy and building blocks for growth and maintenance, organisms and their roles in dyad interactions (producers/consumers, predator/prey) and food webs (diagrams that represent the flow of matter and energy through ecosystems). The population changes that emerge from interactions among organisms and with abiotic factors in the environment are represented in models that include both the organisms and graphs of populations.

The model levels described above—components, interactions, and emergent behavior—are ubiquitous in science systems ranging in size from molecules to biospheres. The core ideas focus on understanding ecosystem components, interactions, and population behaviors and the science practices for studying ecosystems' dynamic phenomena.

Figure 1. Life science ecosystem target model

Model Level	Descriptions	Content Targets	Science Practices
Component 	What are the components of the system and their rules of behavior?	Every ecosystem has a similar pattern of organization with respect to the roles (producers, consumers, and decomposers) that organisms play in the movement of energy and matter through the system. (NGSS: LS2.A—Interdependent Relationships in Ecosystems)	Analyzing and Interpreting Data
Interaction 	How do the the individual components interact?	Matter and energy flow through the ecosystem as individual organisms participate in feeding relationships within an ecosystem. (NGSS: LS2.B—Cycles of Matter and Energy Transfer in Ecosystems)	Developing & Using Models; Analyzing and Interpreting Data
Emergent 	What is the overall behavior or property of the system that results from many interactions following specific rules?	Interactions among organisms and among organisms and the ecosystem's nonliving features cause the populations of the different organisms to change over time. (NGSS: LS2.C—Ecosystems Dynamics, Functioning and Resilience)	Planning and Carrying Out Investigations; Analyzing and Interpreting Data

ChemV Lab Student Model

The student model for ChemVLab focuses on conceptual understanding of chemistry. At the submicroscopic level, the student model integrates processes involving atoms and molecules to procedural knowledge such as quantitative problem solving. At the macroscopic level, the student model includes causal models for macroscopic processes based upon understanding the submicroscopic processes.

Voyage to Galapagos Student Model

In VTG, the student model specifies an understanding of evolution and natural selection at three levels:

Level 1 - Variation: Among species of animals, key trait variations are observed across populations.
Level 2 - Biological Function: Observed animal trait variations are tied to biological function.
Level 3 - Adaptation: Environmental factors have an impact on the observed biological functions in the animals.

The levels involve a conceptualization of increasingly complex ideas as students progress through the various levels of the software.

Table 1 describes learning goals and tasks aligned with the NGSS that have been used to create an intelligent tutoring system to collect data about student actions, assign probabilities of students having made certain errors, and make decisions about error feedback and hints to provide students.

Task Models

The STEM projects design tasks to elicit evidence that students understand core ideas and can use them in a range of practices to study science systems. Technology supports the design process by allowing development of re-usable templates for

Table 1. Alignment of levels within the VTG application to NGSS

	Crosscutting Concepts	Practices for K–12 Science Classrooms	Life Science Disciplinary Core Idea LS4 Biological Evolution: Unity and Diversity
Level 1	Patterns	4. Analyzing and Interpreting Data	
Level 2	Patterns	3. Planning and Carrying out Investigations	**LS4.B Natural Selection** How does genetic variation among organisms affect survival and reproduction?
	Cause and Effect	4. Analyzing and Interpreting Data	
		6. Constructing Explanations	
Level 3	Patterns	4. Analyzing and Interpreting Data	**LS4.C Adaptation** How does the environment influence populations of organisms over multiple generations?
		6. Constructing Explanations	

task types for investigating science phenomena. The templates specify key features of representations of system phenomena that are appropriate for the grade level. Multimedia research provides techniques for directing attention to relevant parts of the representations of the science phenomena. The templates also specify the types of responses that students are asked to make. Typically, the templates specify sets of tasks that students will complete as they use science practices to address real world problems.

Problems posed for investigation represent iconic problems addressed by scientists studying science phenomena such as observing components of a system, studying interactions, and conducting studies to predict and explain emergent system behaviors.

Models for task types deliberately incorporate design principles from learning research that include, among other features, multiple linked representations of system interactions and dynamic phenomena that are difficult to observe and manipulate in classrooms because of the phenomena's interactions at multiple scales, temporal dynamics, causal mechanisms. Based on recommendations from learning research, learners participate in active inquiry by designing, conducting, and interpreting iterative investigations and explaining conclusions. Scaffolding in the form of feedback and customized coaching guides and reinforces the learning.

SimScientists Task Models

In the SimScientists program, the conceptual framework guiding research and development is grounded in the belief that learners develop understanding and mental models of dynamic phenomena through a variety of routes that depend on the learner's starting point and interactions with phenomena and representations. These phenomena arise from complex systems of interacting components, which themselves may be complex systems. For example, learning about ecosystems might begin with a simple partial mental model of the ecosystem such as the idea that living creatures have survival needs—food, shelter, ability to avoid predators, etc. The first incomplete mental model of an ecosystem may be one of the organism—what it eats and who eats it. This simple mental model can become more complete and complex when learners consider the competing needs of populations of organisms over time, perhaps by conducting investigations with simulations. So for the development of a model of ecosystems, a learning trajectory could begin with tasks requiring identification of component organisms, adding understanding of their interactions before proceeding to a more complete model of the ecosystem emergent phenomena of changing population levels over time.

Science practices that focus on developing and using models, conducting investigations, and interpreting data are particularly relevant to helping students develop, test, and evaluate their mental models of science systems. Simulations of diverse types can enable students to conduct investigations with complex systems and system models. In SimScientists, a progression of tasks both develops and elicits students' conceptual understandings of the system model and associated science practices. Cognition and multimedia learning research guide the design of the representations of the system components, interactions, and emergent phenomena, in addition to ways that cueing and learner control guide student interactions with the simulations.

The SimScientists modules include two major types of assessment tasks. Curriculum-embedded modules are designed to foster integration of core ideas and their use in investigations. Each is designed to require one period. As recommended by learning research, the modules present real world problems that are recurring significant problems addressed by scientists in an area. Domain analyses provide one source for problems addressed by scientists. From these, problems are selected that focus on complex systems, in order to help counteract the fragmented understandings occurring among science learners. Examination of research papers published by scientists ensures accuracy and informs development of simulations of these complex systems.

The assessment tasks present real world problems, require use of core ideas, and focus on the investigations and reasoning of scientists as students create, observe, evaluate, and revise their models of phenomena. For example, students identify components and interactions, make predictions, design experiments, interpret data, evaluate their predictions, and explain the results and their reasoning, all key science practices.

The embedded modules further incorporate principles derived from learning research by providing opportunities for formative assessment during the sequence of investigations. The simulations provide individualized feedback as students perform a task or respond to questions. The feedback is accompanied by graduated coaching in the form of increasingly more information and, finally, a worked example. For example, within a unit on ecosystems, the teacher inserts the first embedded module after students have learned about different types of organisms in an ecosystem. The module engages students in helping to develop material for an interpretive center to describe a mountain lake ecosystem to visitors, beginning with an animation of organisms in the lake. At the component level of the ecosystem model, students observe what the organisms eat and identify their roles as consumers or producers. At the interaction level of the ecosystem model, students are asked to draw a food web that depicts the flow of energy and matter as organisms interact. The simulation uses affordances of the technology to provide immediate feedback about whether the arrow drawn connects the correct organisms and is in a direction showing the flow of energy and matter from the source. As shown in Figure 2, feedback highlights an incorrect arrow and includes coaching for the student to observe the animation of organisms eating in order to draw the arrow from the food source. If incorrect arrows remain, the following screen would show a correctly drawn worked example and require the student to draw the arrows correctly. This process is formative because the system evaluates a student response, provides feedback on its appropriateness, and offers additional instruction.

Figure 3 presents a formative embedded assessment of investigation practices and science knowledge of ecosystem emergent behaviors represented by changing population levels. The feedback addresses students' predictions about population level change over time, and is accompanied by coaching to analyze the graph in order to match observations to predictions.

Figure 4 presents a task asking students to build a model of the circulatory system by drag-

Figure 2. Mountain Lake embedded module, Draw Foodweb task

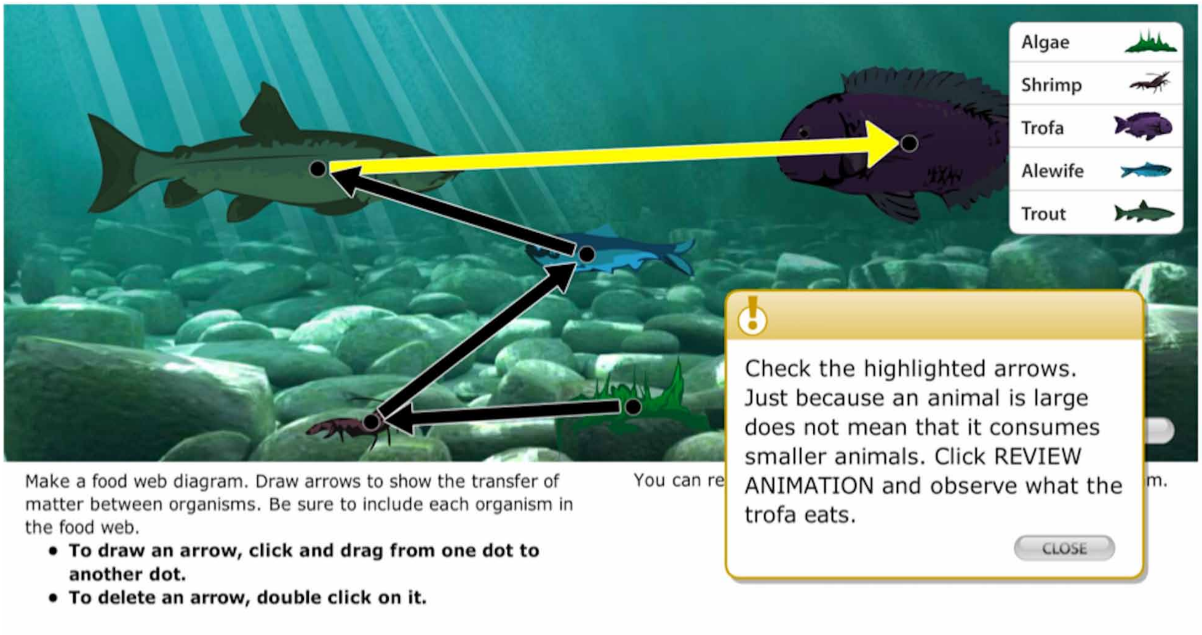


Figure 3. Mountain Lake embedded assessment, Predict Population task for science investigation practices of analyzing ecosystem population level emergent behaviors

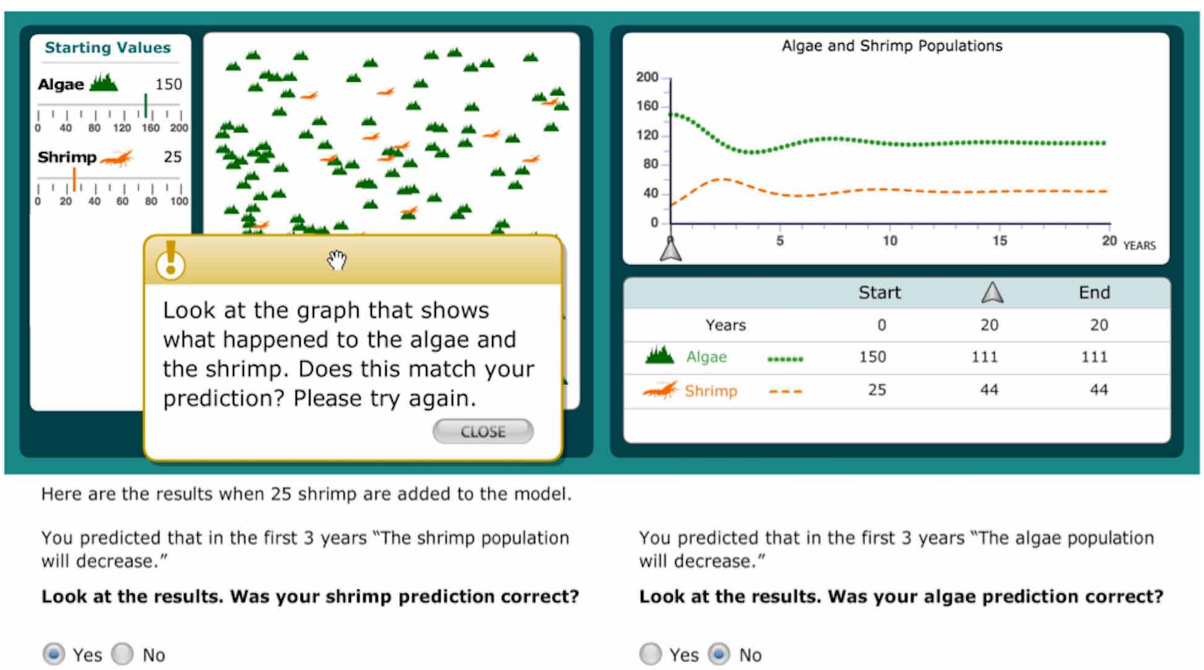
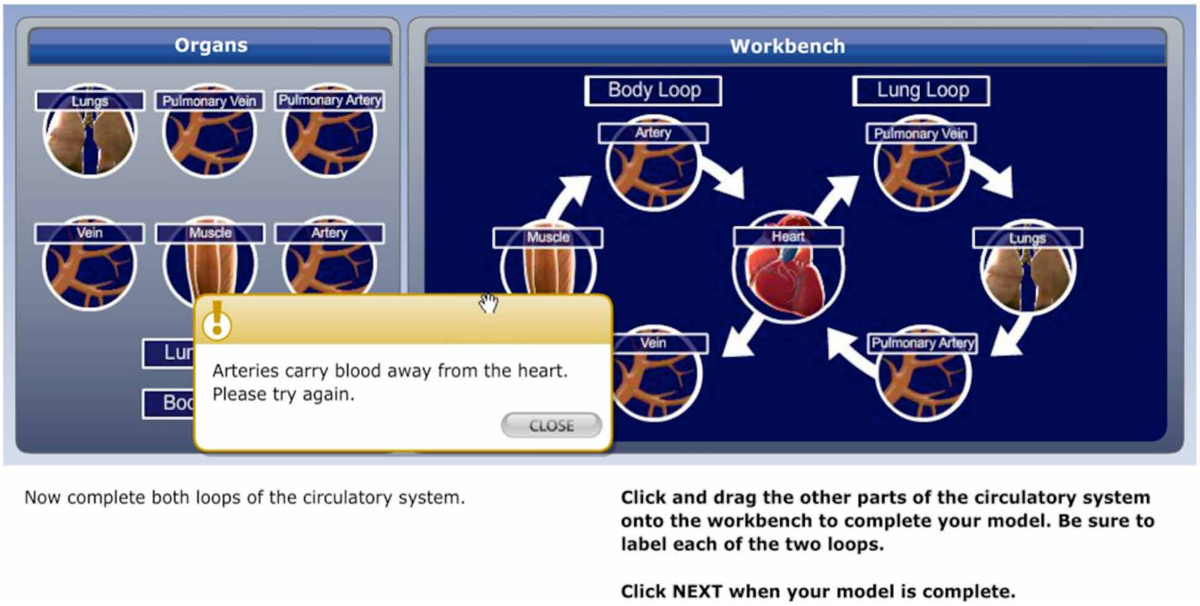


Figure 4. Human Body Systems embedded assessment, Build Circulatory System task



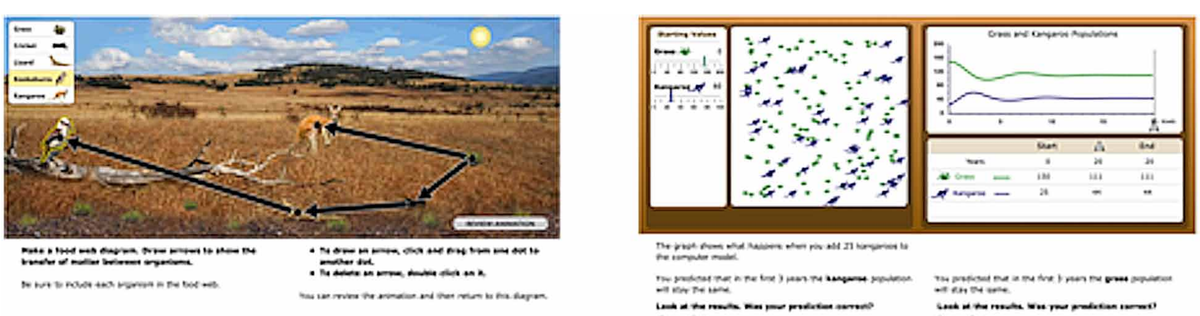
ging and dropping images of organs into the body and lung loops. The assessment provides feedback and coaching if the organ placements or sequences are incorrect.

Each curriculum-embedded simulation-based module intended for formative purposes is followed by an off-line reflection activity designed to adjust instruction based on progress reports that indicate which core ideas and practices need more attention. The reflection activities promote transfer of the core ideas and science practices to new settings.

The second major type of task model in the SimScientists assessment modules is a simulation-based benchmark assessment administered at the end of the unit. These assessments generate summative reports of student proficiencies on the targeted core ideas and practices. The tasks are parallel to those in the embedded modules, but do not provide feedback and coaching. Again, students' abilities to apply core ideas to a new ecosystem are assessed.

Figure 5 shows a benchmark assessment set in an Australian grassland. The real world problem is

Figure 5. Screen shots of foodweb and population investigation tasks in a SimScientists ecosystem benchmark assessment



to restore the grassland after a wildfire. Students must observe the eating behaviors of grassland organisms to identifying system components and construct a food web depicting interactions, and then manipulate the numbers of organisms re-introduced into the grassland ecosystem in order to restore a balanced ecosystem (the emergent phenomena). The two screens are sampled from the food web and population dynamics task sets, illustrating how task model templates can be re-used to create parallel tasks set in new contexts.

In addition, to broaden participation across a diverse range of students, the SimScientists assessments provide the three most common accommodations allowed in state testing programs—text-to-speech, screen magnification, and segmentation that supports re-entry into tasks when extended time is needed.

ChemVLab Task Models

An overarching goal of the *ChemVLab* project is to contextualize the procedural knowledge used in chemistry to solve problems and conduct investigations (Davenport et al., 2014). To this end, a series of activities has been developed to address topics common to high school chemistry curricula. The activities are all designed around a common approach: students investigate phenomena at the atomic and macroscale levels and solve problems using the properties of atoms and molecules to make predictions and to explain observations of properties of bulk matter. Through this approach, students gain a deeper understanding of the utility of the chemists problem-solving "toolbox" for reasoning about the world around them, rather than simply committing to memory a disconnected set of algorithms. Task models designed around concepts in chemistry include atomic, molecular, and bulk features, and a set of investigation "tools" that allow chemists to use observations at one scale for making inferences at another scale. These tools are specific to the domain and may be specific to the concept addressed in the task. While the ChemVLab portfolio of activities does not cover all of chemistry, the selected problems are common to high school chemistry curricula and address typical applications, while transforming these applications from discrete procedures to sets of contextualized, interrelated tasks. For example, in the Acid-Base activity (Figure 6), students are introduced to the mathematics that relate the concentrations of ions in these solutions to the primary logarithmic scale used to characterize their unique properties at the macroscale, the pH scale. The tasks for the student involve mixing acidic and basic solutions in order to change the concentrations of ions and the related pH in a way that reveals the nature of the logarithmic relationship between the two properties. Subsequent tasks then ask students to use their understanding of this relationship to predict the properties of given mixtures, and to explain macroscale phenomena as the result of interactions within systems of interacting ions.

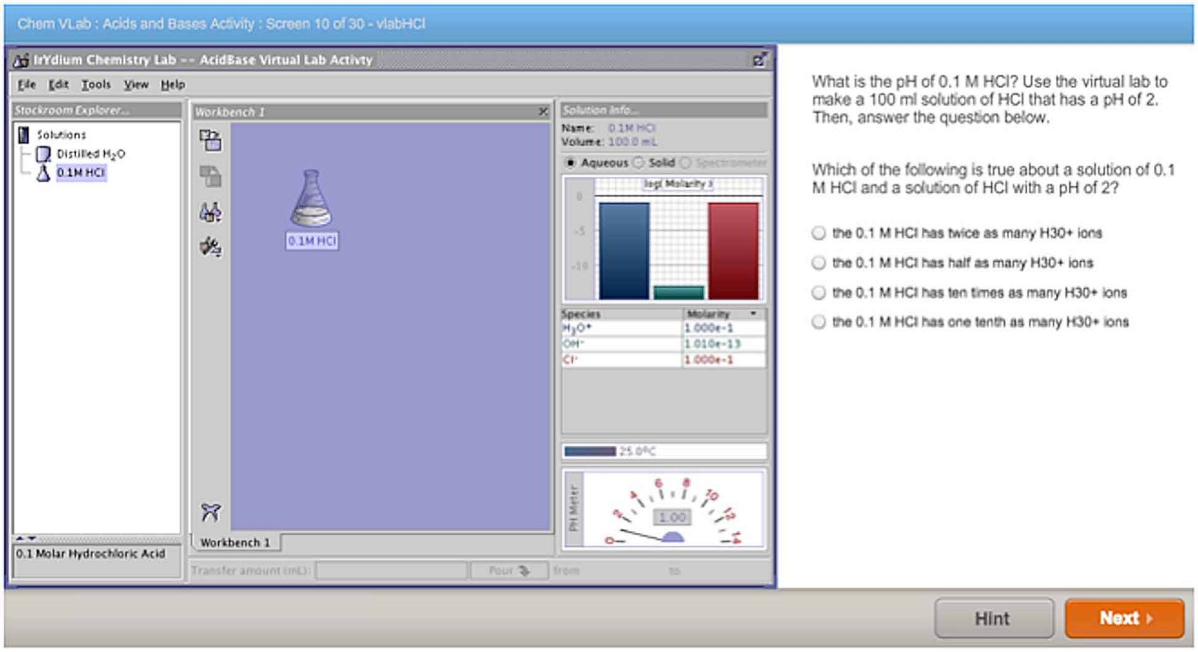
Voyage to Galapagos Task Models

The *VTG* software encourages the student to follow the steps of good scientific inquiry, e.g., developing hypotheses, analyzing data, drawing conclusions, and reveals the basic principles of evolution to the students. The open learning environment provides latitude for variability of student actions—and student errors—allows for a wide variety of assistance, and the ability to either intervene after those actions are taken with help—or not.

In VTG, students progress through a series of levels in which they complete a series of tasks using a set of virtual tools provided in the application. The task model in the VTG program is organized around the three main phases, or levels of the application:

Level 1: *Variation.* Students are tasked to explore the islands in the Galapagos Archipelago in search of evidence of trait variation among

Figure 6. The task that introduces the relationship between concentration and pH



iguanas found there. They use a camera to photograph a representative sample of iguanas. Back in a virtual lab, they then measure specific animal traits (body length, tail width, claw length, snout length, and color) with a Schemat-o-meter, classify the variation of the traits (e.g., for claw length, very long, long, neutral, short, and very short), and classify the variation of traits with a Schemat-o-meter and use their data to analyze geographic distribution of variant populations.

Level 2: *Biological Function.* The students then return to the island to find evidence of iguana functions (e.g., eating, swimming, foraging for food) by viewing videos found on some of the paths they have explored. They then are asked to hypothesize about the biological function of iguana trait variations (e.g., long claws are better for climbing rocks). After returning to the lab, they are provided with a Trait Tester, an instrument with which they can test animals for relative performance.

Level 3: *Adaptation.* Students are asked to review the island path steps where they found their iguanas and associate an environment with each sample animal. After examining the environments where animals with specific biological functions live, students hypothesize about selective pressures, use the Distribution Chart to plot where animals with different trait variations in order to draw conclusions about natural selection.

Within each level, there are 3 steps that correspond to:

- Sample collection and hypothesizing.
- Data testing and analysis.
- Synthesizing ideas.

Through the cyclical process of repeated exposure to employing these practice skills, students are given the ability to increase their proficiency in scientific inquiry.

Across the WestEd STEM projects, the task models use multimedia principles in the design of the representations of the science phenomena and the student interactions. Key features of the task models are the use of multiple representations, an array of technology-enabled cueing mechanisms, and a focus on active investigations that take advantage of the technology capabilities. In addition, the tasks collect learner responses for analysis of learning progress.

Evidence Models

A valuable affordance of computer-supported learning environments is their ability to capture, evaluate and summarize student responses to tasks and questions in problem-based modules. Each of the projects has designed an underlying database, a learning management system (LMS), to gather evidence of the targeted learning. In this section we describe the evidence models of the STEM projects.

SimScientists Evidence Models

The SimScientists embedded assessments generate progress reports based on the level of assistance students needed to complete the tasks. Typically, students have three opportunities to complete tasks and questions correctly. After each "try," students receive increasingly more coaching, with the last try a worked example. Each of the tasks and questions are aligned with knowledge and practice targets. The progress rubrics use the number of tries to classify student responses into the levels of "Needs Help," "Progressing", or "On Track" for each of the knowledge and practice targets. As shown in Figure 7, the progress report for an individual student describes performance for understanding core ideas related to the model levels of components, interactions, or system behavior. The progress report on the Predator/Prey simulation-based curriculum-embedded assessments documents progress on core ideas related to the emergent level of the system model: population dynamics. Progress reports also reflect students' application of science practices. The progress reports are provided to individual students. The teacher also receives reports of each individual's progress and class summaries.

The progress reports provide data for teachers to use formatively: to adjust instruction during an off-line reflection activity in the next class period. The reflection activities are designed to provide additional instruction and practice on core ideas and science practices on which progress reports indicate students need additional help or extension.

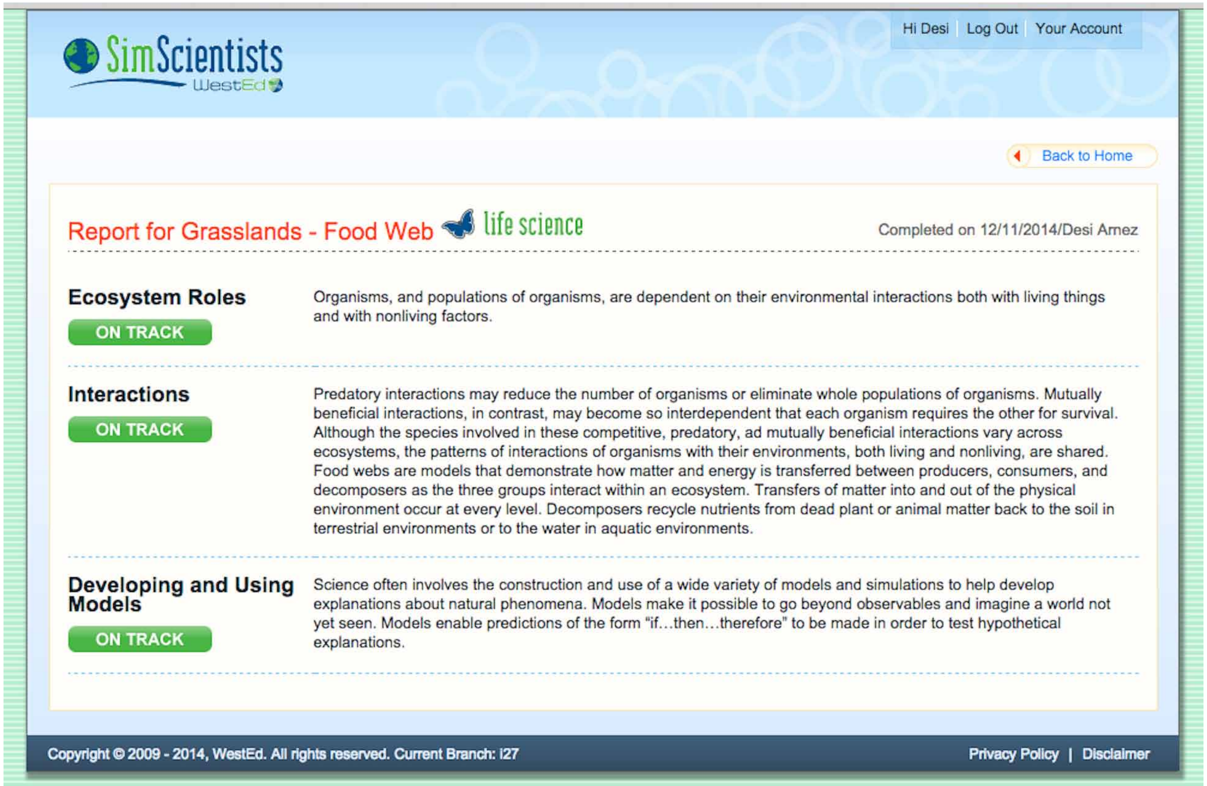
For simulation-based benchmark assessments administered at the end of units, the evidence model incorporates evaluations of student responses into a Bayesian Estimation Network (Bayes' Net) that then reports the proficiency levels for individual students and for the class on the NGSS core idea targets and science practices.

Figure 9 shows a class level report on proficiencies for core ideas within the three model levels (roles [components], interactions, populations [emergent] and for the science practices). A segment of the benchmark report for individual students is also shown.

ChemVLab Evidence Model

The *ChemVLab* project developed input variables aligned to the targets in the student model for each task model. The evidence model was then developed as a set of algorithms that use input variables to generate indicators of mastery. These algorithms are tailored to each input variable, and provide for multiple approaches to problem solving. For example, in a task that includes preparing a solution of a given concentration, input variables for the quantities of each substance are meaningless without a comparison included in the algorithm.

Figure 7. Student level progress report for a Life Science Ecosystem embedded assessment about food webs



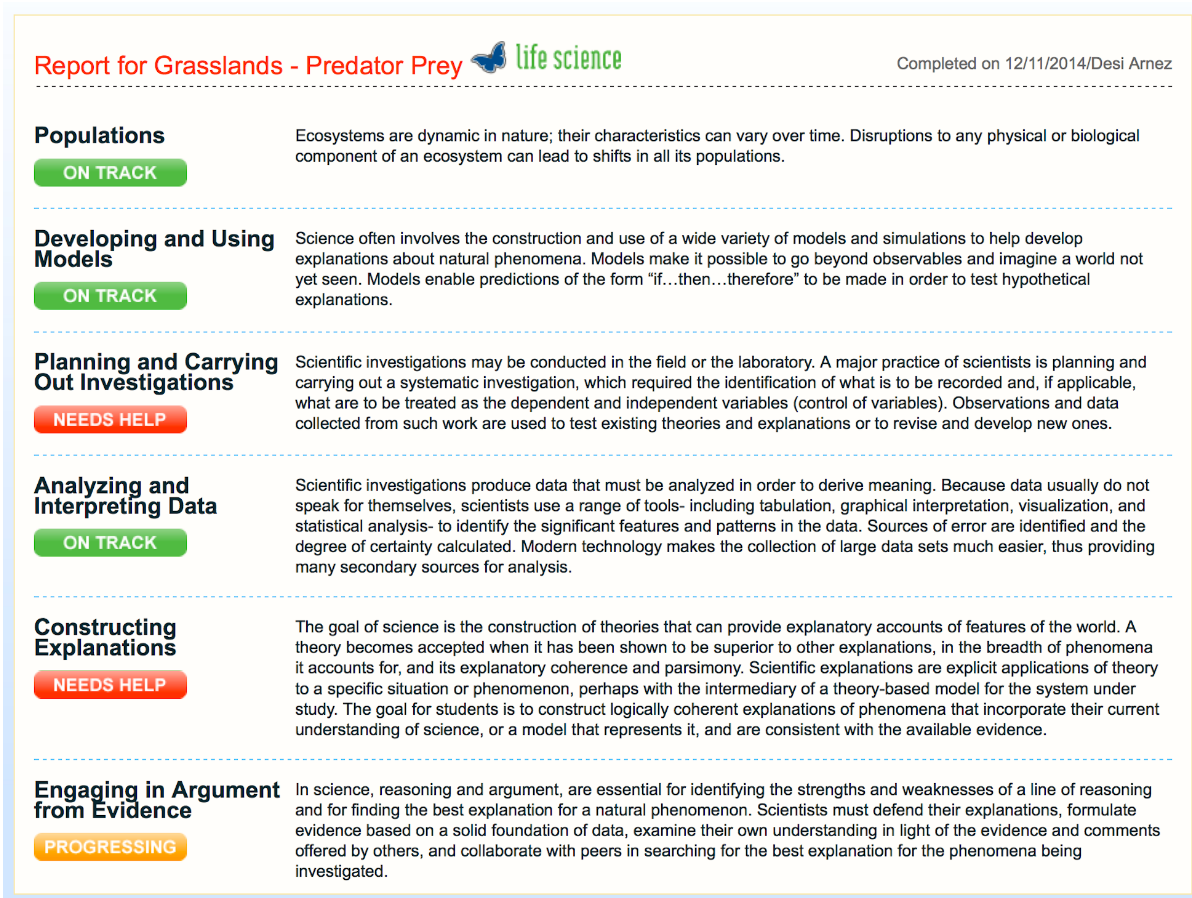
Individual indicators of mastery are aggregated in order to make inferences about students' abilities with respect to the targets.

Voyage to Galapagos Evidence Model

VTG uses Bayes' Nets to monitor when a student needs assistance in applying the relevant science practices for each task. When the probability that a student needs help reaches a threshold value, the assistance system switches on and can provide different levels of assistance. *VTG* has been used for studies in which five levels of assistance are being examined: (1) no support, (2) error flagging only, (3) error flagging and text feedback on errors, (4) error flagging, text feedback on errors, and hints, and (5) preemptive hints with error flagging, error feedback, and hints. The aim of the study is to learn which levels of assistance work best in an exploratory science learning environment.

For example in Level 1, students are asked to collect a sample of iguanas from the islands which shows the range of variation among the iguana populations. As they undertake the data collection task by exploring the islands, taking photos of iguanas that they see and saving them to their logbook, a Bayesian Network is used to collect data about student actions and assign probabilities of students having acted in such as way that suggests they are struggling with the task. The Bayes' Net contains a decision node that, when the probability exceeds a threshold value, turns on the assistance. The Bayes' Net has three top layers that range from the most general to most specific—the Knowledge, Skills, and Abilities (KSA) Layer, the Error Evaluation Layer, and the Error Diagnosis Layer. The specific nodes at each of these layers have associated error feedback and hints that are triggered when the nodes

Figure 8. Student level progress report for a Life Science Ecosystem embedded assessment about population dynamics



Report for Grasslands - Predator Prey life science

Completed on 12/11/2014/Desi Arnez

Populations
ON TRACK

Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations.

Developing and Using Models
ON TRACK

Science often involves the construction and use of a wide variety of models and simulations to help develop explanations about natural phenomena. Models make it possible to go beyond observables and imagine a world not yet seen. Models enable predictions of the form "if…then…therefore" to be made in order to test hypothetical explanations.

Planning and Carrying Out Investigations
NEEDS HELP

Scientific investigations may be conducted in the field or the laboratory. A major practice of scientists is planning and carrying out a systematic investigation, which required the identification of what is to be recorded and, if applicable, what are to be treated as the dependent and independent variables (control of variables). Observations and data collected from such work are used to test existing theories and explanations or to revise and develop new ones.

Analyzing and Interpreting Data
ON TRACK

Scientific investigations produce data that must be analyzed in order to derive meaning. Because data usually do not speak for themselves, scientists use a range of tools- including tabulation, graphical interpretation, visualization, and statistical analysis- to identify the significant features and patterns in the data. Sources of error are identified and the degree of certainty calculated. Modern technology makes the collection of large data sets much easier, thus providing many secondary sources for analysis.

Constructing Explanations
NEEDS HELP

The goal of science is the construction of theories that can provide explanatory accounts of features of the world. A theory becomes accepted when it has been shown to be superior to other explanations, in the breadth of phenomena it accounts for, and its explanatory coherence and parsimony. Scientific explanations are explicit applications of theory to a specific situation or phenomenon, perhaps with the intermediary of a theory-based model for the system under study. The goal for students is to construct logically coherent explanations of phenomena that incorporate their current understanding of science, or a model that represents it, and are consistent with the available evidence.

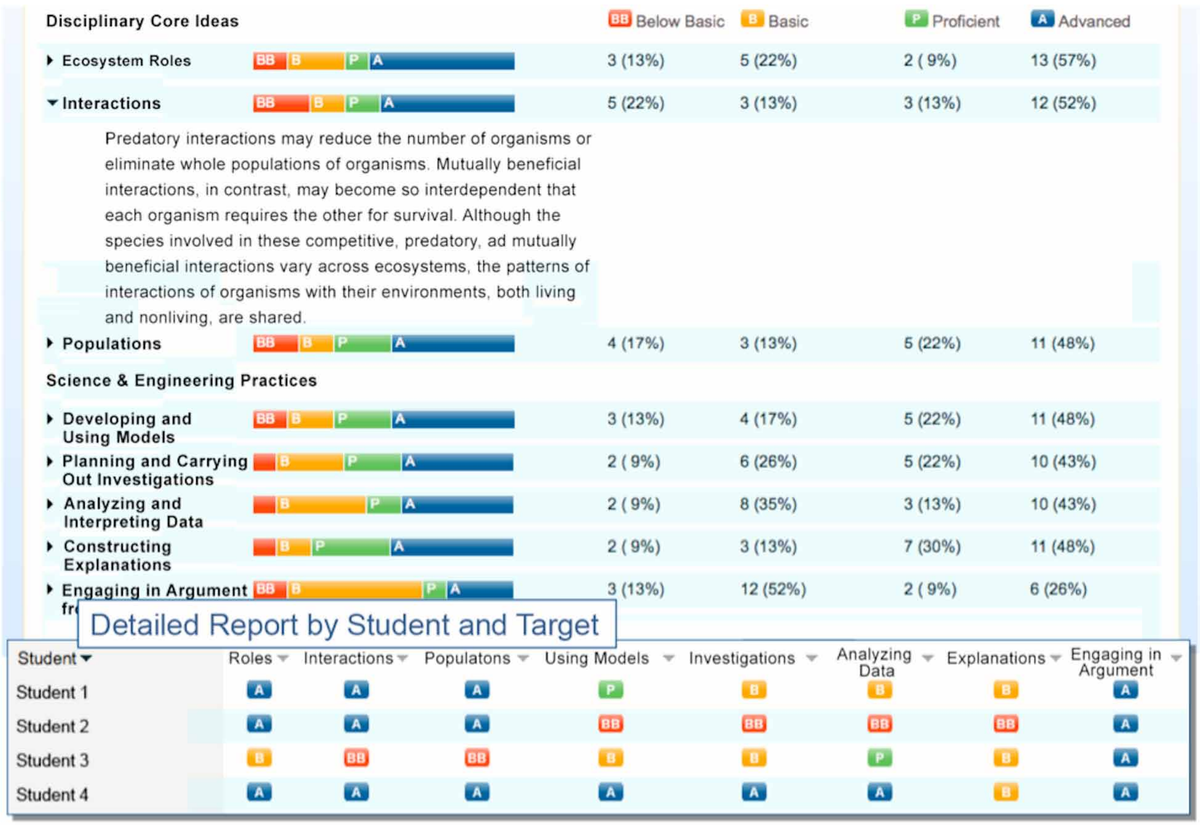
Engaging in Argument from Evidence
PROGRESSING

In science, reasoning and argument, are essential for identifying the strengths and weaknesses of a line of reasoning and for finding the best explanation for a natural phenomenon. Scientists must defend their explanations, formulate evidence based on a solid foundation of data, examine their own understanding in light of the evidence and comments offered by others, and collaborate with peers in searching for the best explanation for the phenomena being investigated.

Table 2. Spectrum of Assistance. The basis for the experimental design is a matrix that crosses Frequency of Intervention with Level of Support.

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">Frequency of Intervention
→</th></tr>
<tr><th>Never</th><th>When Struggling</th><th>Always</th></tr>
<tr><td rowspan="3">Level of Support
↓</td><td>Error Flagging</td><td rowspan="3">Condition 1
No support</td><td>Condition 2
Flagging errors when struggling</td><td>Skipped Condition</td></tr>
<tr><td>Error Flagging + Error Feedback</td><td>Condition 3
Flagging errors & providing feedback when struggling</td><td>Skipped Condition</td></tr>
<tr><td>Error Flagging + Error Feedback + Hints</td><td>Condition 4
Flagging errors & providing feedback and hints when struggling</td><td>Condition 5
Full support beginning with a preemptive hint is always provided</td></tr>
</table>

Figure 9. Class level report for Ecosystems benchmark assessment



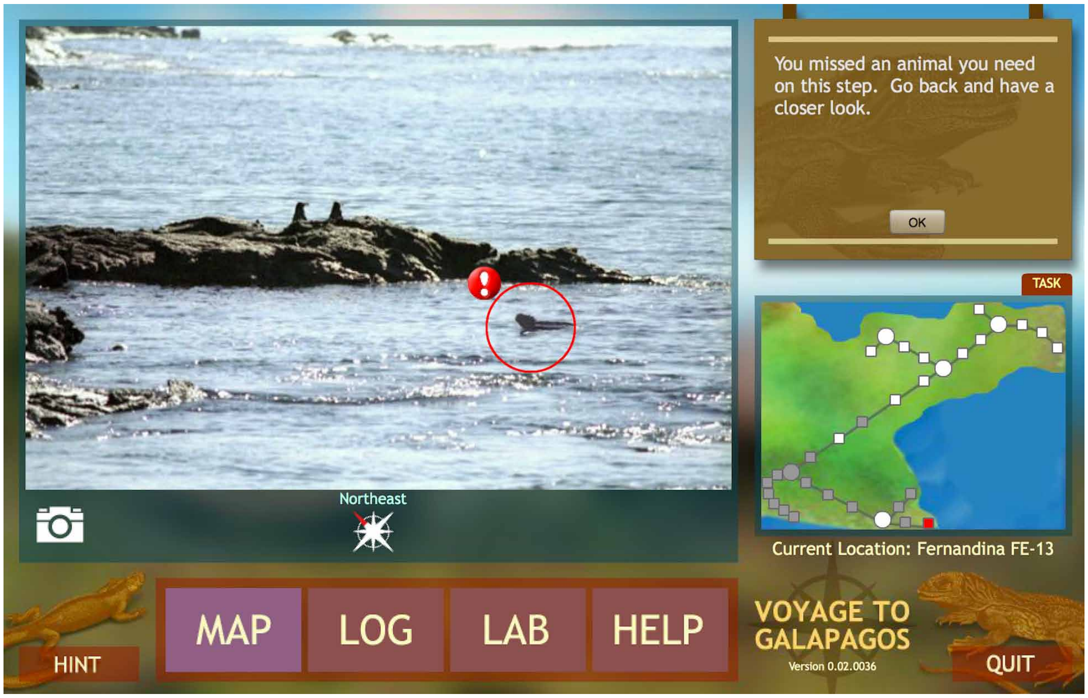
at the associated level reach a certain threshold. Whether a student receives the feedback or hints is configurable according to (a) what experimental condition they are in and (b), in the case of hints, whether they request help. With assistance having been configured this way, the conditions of assistance have been created that are the focus of the experimental design.

The designs of the WestEd STEM projects' student, task, and evidence models merge two critical affordances of simulations. Dynamic visualizations permit use of cueing and multiple representations of science phenomena that may not be directly seen or investigated. In addition, underlying databases can record learner actions and generate immediate reports, feedback, and customized scaffolding. The projects address deep learning in the form of integrated knowledge and processes ensconced within meaningful problem situations. The interactive capabilities of simulations support active learning and result in better measurement of inquiry skills than produced by static test formats (Quellmalz, et al., 2013)

Research and Development Methods Overview

The WestEd STEM technology-enhanced learning projects employ systematic, iterative design and development processes. Following the design phase, when learning outcomes are specified, tasks are designed, and evidence models are detailed, the projects seek expert reviews of the science content and assessment tasks. Cognitive labs are conducted with individual students to confirm intended construct validity and usability. Classroom

Figure 10. Assistance Provided in VTG. For condition 4 when assistance is available, errors are flagged and students are provided with feedback and hints when they are struggling.



tryouts then proceed from small-scale feasibility testing to pilot and field testing with progressively larger numbers of students and teachers. In this section we summarize the research and development methods for the WestEd STEM projects.

Implementation Studies

The STEM projects have been implemented in a range of classrooms representing the intended student populations. The sections below summarize data about the projects' use in classrooms.

SimScientists Implementation

In 2010, a large-scale implementation study was conducted to determine whether simulation-based assessments could be delivered in a wide range of settings (Quellmalz, Timms, Silberglitt, & Buckley, 2012). Over 5,000 students participated in the classrooms of 55 teachers in 39 different schools, from 28 school districts in 3 states. Table 3 shows that this sample represents a wide range of student backgrounds, including students with disabilities and English language learners.

Table 3. Total numbers of English language learners (ELL) and students with disabilities (SWD)

SWD*	ELL	FRL	Caucasian	Hispanic	African-American	Asian	Other**
12%	6%	34%	66%	13%	11%	4%	6%

*11% IEPs; >1% 504 plans
**multiracial, native American, Pacific Islanders, or unknown

The implementation study included two suites developed by the SimScientists program: Ecosystems and Force & Motion. Each teacher used one of the two suites. A total of 3,529 students completed the Ecosystems assessments and 1,936 students completed Force & Motion. Each suite included 2–3 simulation-based, curriculum-embedded assessments with feedback and assessment.

Teachers participated in a 1½-day professional development workshop prior to using the suites in their classrooms. In addition to familiarizing teachers with the assessments and reflection activities, this workshop focused on two key components of the implementation: curriculum integration and formative use of assessment data. The process of curriculum integration included several steps before, during, and after the PD workshop. This process was supported by facilitators who helped participants understand the prerequiste knowledge required for each embedded assessment so that teachers would schedule the assessments after the core ideas and practices had been addressed in the teachers' ongoing curricula. Teachers determined the lesson sequence and the precise timing of the embedded assessments. The purpose of curriculum integration was to ensure that embedded assessments would serve as appropriate "checks for understanding" as well as opportunities for integration of knowledge about the components, interactions, and emergent behavior of each science system and active investigation of the dynamic phenomena. Such knowledge integration and active inquiry remain uncommon in traditional modes of instruction.

The process of curriculum integration began with a teacher survey, completed prior to the PD workshops. In this survey, teachers indicated the number of days they planned to teach particular aspects of the topic, including science practices and concepts. Teachers were asked to bring their curricula to the workshops and use their states' standards to bring together alignments of their curricula and the modules, which had been aligned to each states' standards during design and development. The teachers then decided at what points in the unit to insert the embedded assessments.

After each of the embedded modules, teachers completed follow-up surveys to indicate how closely the implementation resembled their plan, and whether they used progress reports from embedded modules as formative evidence to adjust their instruction.

As with any software technology, there were myriad potential pitfalls on the way to implementing. To address the implementation challenges, the SimScientists team devised protocols for troubleshooting, programmed safeguards to protect against data loss, and provided real-time help to teachers using telephone and email help lines.

During the use of the simulation-based, curriculum-embedded modules, teachers were given options to have students work one-to-one with computers, work in teams of two or three, or a hybrid approach in which students each have their own computer, but work side-by-side to support each other in learning. After implementation of the last embedded assessment and reflection activity, teachers administered the benchmark assessment.

The benchmark assessment was designed as a summative assessment, which students completed independently and without the assistance of feedback and coaching. Data from complex interactions and problem solving patterns were interpreted using Bayes' Nets, and reports were generated that categorized students' performance on each assessment target in one of four levels: advanced, proficient, basic, or below basic.

In the implementation study, the fidelity of implementation was evaluated by the UCLA Center for Research and Evaluation of Standards and Students (CRESST). Following observation of the professional development sessions, the evaluation sampled classrooms to observe as students used the embedded and benchmark simulation-based assessments. Teachers completed surveys describing their ongoing curriculum unit and how they used the simulation-based assessments to monitor their students' progress and adjust their instruction.

A sample of teachers was interviewed about their perceptions of the feasibility and instructional utility of the simulations. In addition, completion rates documented in the Learning Management System corroborated that students were able to complete the simulation-based assessments in a class period. The CRESST evaluation documented that the simulation-based assessments could be implemented across a wide range of schools with diverse populations, science curricula and infrastructures (Quellmalz, et al., 2012). The evaluation findings suggested that participating in the SimScientists program was beneficial to learning and feasible and useful in middle school classrooms.

ChemVLab Implementation

The ChemVLab project has been implemented extensively. In one study, 13 teachers and 1334 secondary students used four *ChemVLab* activities. Students completed pre- and post tests as well as the modules. Teachers participated in a 3-hour professional development workshop and completed surveys during the implementation. Researchers conducted classroom observations and collected student demographic information (Davenport et al., 2014). Findings from this research are anticipated in a manuscript currently in draft.

Voyage to Galapagos Implementation

In the early development phase, the project conducted cognitive labs with 12 students to identify usability issues and establish initial construct validity. Initial classroom feasibility testing was then conducted with 7th grade classes, 161 students, in two schools. Data were collected in the LMS as students worked through the application. The interactions and communication between the flash-based application, the LMS database, and the Bayes' Net was demonstrated to operate effectively in providing real-time feedback and assistance to students. The cognitive labs and classroom observations indicated that students with greater assistance advanced further through the tasks.

Pilot studies were conducted in two schools with 258 7th grade students. The software was embedded within the normal classroom lessons and used as a curriculum supplement. Students completed pre- and post tests and had three class periods to use *VTG*. In addition to using the *VTG* software, teachers completed two 2-hour professional development sessions that provided guidance for embedding the software in their curriculum. They also participated in interviews following the implementation. Classroom observations and student demographic information along with LMS data and selected case studies were analyzed to validate and refine the Bayes' Nets that provide assistance for the different research conditions (Brenner, Timms, McLaren, Brown, Weihnacht, Grillo-Hill, et al., 2014).

Technical Quality

Evaluations of the technical quality of the WestEd STEM projects combine qualitative and quantitative methods. Specifications of significant content and skills document alignments to national standards and frameworks. External experts review alignments and grade-level appropriateness of task features. Think alouds and classroom trials provide data on reliability and validity. The sections below summarize the projects' technical quality studies.

SimScientists Technical Quality

The quality and validity of the SimScientists simulations have been documented for multiple topics, in multiple projects by employing established evaluation methodologies: alignment with national standards for science, expert review of scientific content and task and item quality by the

American Association for the Advancement of Science (AAAS), cognitive analyses of students thinking aloud, and analyses of teacher and student data gathered from classroom testing (AERA/APA/NCME, 2014; Pellegrino, 2002; Quellmalz, et al., 2012; Quellmalz, et al., 2005).

Technical quality of the SimScientists assessments was established by standard measures of reliability and by gathering evidence of validity from a variety of sources. Independent, expert reviews of task alignments with science standards, accuracy of science system models, and grade-level appropriateness established initial construct validity of the simulation-based tasks prior to programming.

Once programmed versions were developed, researchers administered the assessments to individuals, including both students and teachers, asking examinees to think aloud while completing the tasks. Recordings of the computer screen, together with audio, were reviewed by content experts for further evidence of validity, as well as usability of the interface. Tasks were subsequently revised as needed to improve their validity. To establish the validity of the classifications in the embedded reports, a one-way ANOVA was conducted using scores on the simulation-based benchmark.

Standard psychometric analyses were conducted for the summative benchmark assessments. For the Ecosystems and Force & Motion benchmark assessments, which include a variety of dichotomous and polytomous items of various formats, Cronbach's alpha was 0.76 and 0.73, respectively. To establish the validity of the benchmark scores, correlations were measured between the simulation-based benchmark assessments and a set of traditional multiple-choice items aligned to the same assessment targets and administered to students in tandem with the benchmark assessments. Correlations were moderate (0.57 to 0.64), showing that the two types of assessments measured similar constructs, but the measures were not exactly the same. Further, correlations between the dimensions of science practice and content were lower within each benchmark (0.70 and 0.80) than within each set of post test items (0.85 and 0.92), suggesting that the simulation-based benchmark assessments were better for detecting differences between students' abilities in each dimension. (Quellmalz, et al., 2012).

ChemVLab Technical Quality

Analyses were conducted using data on student engagement and learning in the ChemVLab activities, including classroom observations, pre- and posttests, logs of students' interactions with the online activities, and interviews with teachers (Davenport et al., 2012, 2014). Classroom observations recorded that students stayed on task while using the virtual lab, and that discussions between students focused on the content of the activities. Students' scores improved between pre- and posttest administrations of a measure composed of released items from an American Chemical Society exam and researcher-developed items. Data mining of the log files from students interactions and problem solving processes revealed changes in student behavior over the course of each activity. Evidence included comparisons between parallel tasks, in which students needed fewer attempts to complete later tasks, and were less likely to pursue incorrect lines of investigation in the virtual lab, such as continuing to add water to a solution after the target concentration had been reached. During interviews, teachers indicated that the activities were feasible for classroom use and helpful to improve students' abilities. Analyses of the reliability and validity of the activities as assessments themselves is currently underway.

Voyage to Galapagos Technical Quality

The critical interactions and communication between the flash-based application, the LMS database, and the Bayes' Net were demonstrated to operate effectively in providing real-time feedback and assistance to students. Analyses were conducted on data gathered from multiple classroom feasibility studies in the 7th grade classes in two schools with 260 students. The coded data from the LMS, Bayes' Net, cognitive labs, and classroom observations revealed that the different experimental conditions could be distinguished and that students with greater assistance advanced further through VTG than those with less assistance. A randomized controlled study in classrooms of 12 teachers is underway to help understand how much guidance students need as they learn—and how to cater guidance to the prior knowledge level of students—and thus to be able to appropriately design software to best support student learning.

Impacts on Learning

SimScientists Impacts

To study whether the simulation-based curriculum embedded assessments, intended to provide formative assessment and adjusted instruction, had positive impacts on learning, in 2012 a cluster-randomized controlled study was conducted in the classrooms of 26 teachers, with 2,318 students. Each teacher's classes were randomly assigned to one of two conditions: the treatment condition, which included a suite of simulation-based assessments and off-line classroom reflection activities embedded into a teacher's regular instruction, a simulation-based benchmark assessment, and a traditional multiple-choice pre- and post test, or the control condition, which included the same number of days of instruction, with only the simulation-based benchmark assessment and the pre- and post tests. Effect sizes were determined using a two-level HLM with terms for the nesting of students within classes and classes within teachers. As shown in Table 4, based on ability estimates from posttests composed of traditional multiple-choice items, treatment effects (the effects of the embedded, formative assessments) were small but significant overall for the Ecosystems suite, and within each suite, for inquiry in Ecosystems and for content in Atoms & Molecules. Given that students only experienced the simulation-based embedded assessments for two or three times during multi-week units, the effects supported the promise of the active inquiry, individualized feedback and coaching in the simulation-based assessments and the additional reinforcement and adjusted instruction in the subsequent reflection activities for promoting progress, particularly on inquiry practices.

Table 5 shows that, based on ability estimates from the simulation-based benchmark assessments, treatment effects were small to moderate and statistically significant overall and for each dimension.

These data documented the benefit of formative use of the simulation-based embedded assessments. They also provide evidence that such effects are more likely to be detected by measures that employ similar formats, compared to more traditional tests. (Quellmalz, Timms, Buckley, Loveland & Silberglitt, 2012; Quellmalz, Silberglitt, Timms, Buckley, Loveland & Brenner, 2012)

Table 4. Treatment effects based on posttest ability estimates for ecosystems and atoms & molecules

Ecosystems	Treatment Effect Size	p value	Atoms & Molecules	Treatment Effect Size	p value
Overall	0.061	0.047	Overall	0.075	0.079
Content	0.057	0.082	Content	0.106*	0.019
Inquiry	0.092*	0.020	Inquiry	0.029	0.597

* Significant at $p < .05$

Table 5. Treatment effects based on benchmark ability estimates for ecosystems and atoms & molecules

Ecosystems	Treatment Effect Size	p value	Atoms & Molecules	Treatment Effect Size	p value
Overall	0.286*	< 0.0001	Overall	0.390*	< 0.001
Content	0.148*	0.0005	Content	0.327*	< 0.001
Inquiry	0.297*	< 0.0001	Inquiry	0.498*	< 0.001

* Significant at p < .05

English Learners and Students with Disabilities

To evaluate the benefits of simulation-based assessments for English language learners (ELLs) and students with disabilities (SWDs), performance of each focal group was compared to the general population on the simulation-based benchmark assessments and on posttests. Total numbers of each sample are listed for two topics, ecosystems and force & motion, in Table 6 below.

Analyses of performance on each assessment found significant differences between performances of each focal group and the general population. Although performances on all assessments were lower for the focal groups, gaps in performance were smaller on the simulation-based benchmark assessments than gaps on the multiple-choice posttests. Figure 11 compares the average percent correct on each assessment for ELLs and SWDs. Also included are comparisons of average scale scores on four administrations of the 8th grade NAEP science. (Quellmalz & Silberglitt, 2011)

These data provide evidence for the benefits of the simulations for assessment. These benefits include presentation formats with multiple representations and response formats that engage students in active investigations and on-screen manipulations. These modes provide alternatives to text.

Multilevel Assessment Systems

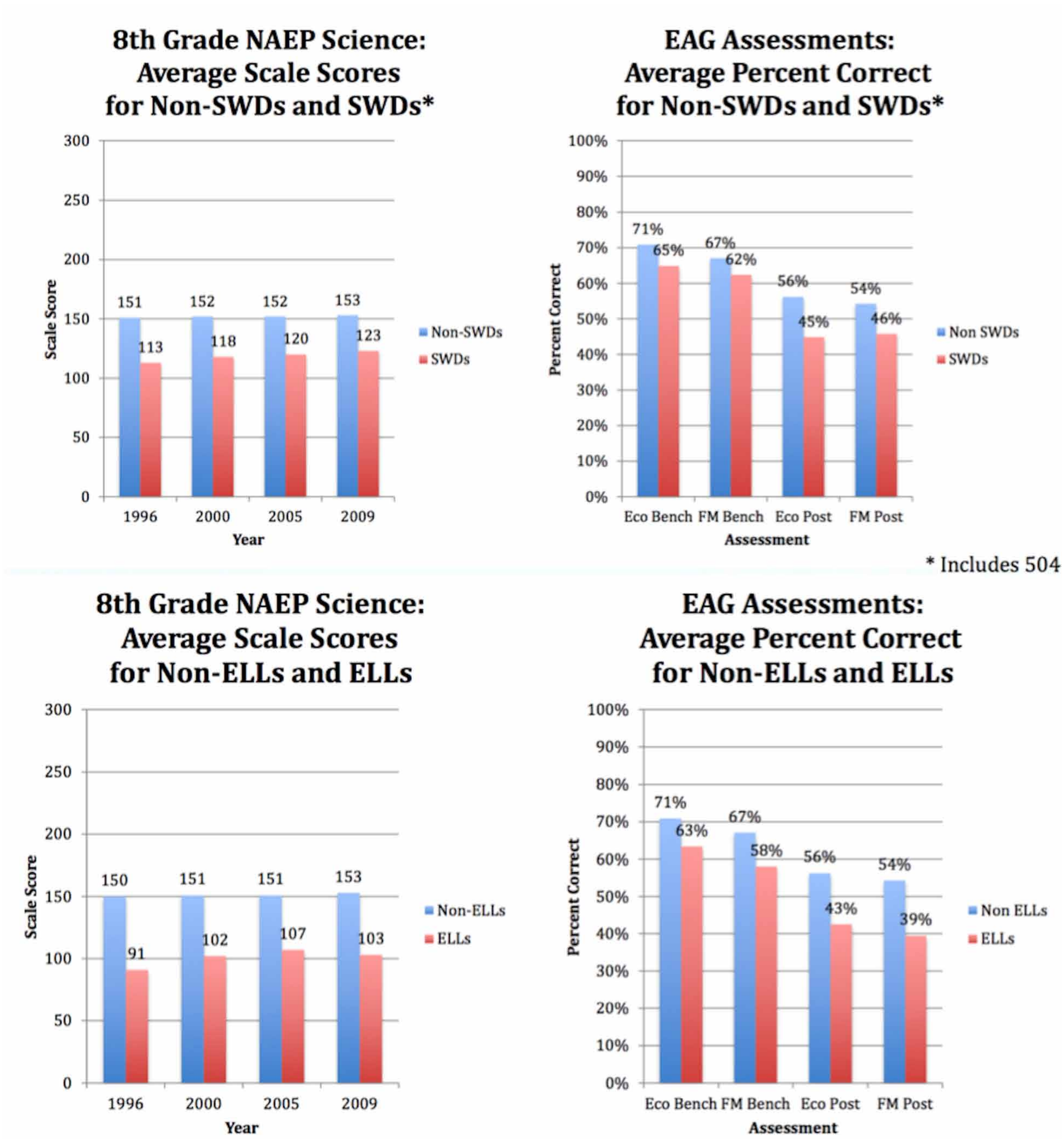
Science simulations are being included in national and international tests as means for measuring students' proficiciencies in science practices. Although large scale testing programs have limited time for administration, simulations designed for classroom use can offer opportunities for more extended investigations with individualized feedback and coaching. Reports can be used formatively by teachers for adjusting instruction during curriculum units.

WestEd SimScientists projects are developing science assessment system designs for formative and summative purposes at multiple levels of the educational system in which variants of templates for simulation environments can be used in classrooms during and at the end of units, and in district, state and national assessments—templates for observing phenomena at different scales, building models of a science system, and conducting investigations of the emergent behaviors resulting

Table 6. Total numbers of English language learners (ELLs) and students with disabilities (SWD)

Group	Ecosystems Posttest	Ecosystems Benchmark	Force & Motion Posttest	Force & Motion Benchmark
English learners	123	126	50	50
Students with disabilities	183	189	153	153

Figure 11. Gaps between ELL, SWD and the general population



from interactions among system components. By taking advantage of principled specifications for simulation-based assessments, coherent, vertically articulated science systems can be achieved.

In the large-scale implementation study described above, the primary goal was to determine the suitability of simulation-based assessments for a state science assessment system and to describe models for incorporating them (Quellmalz, et al., 2012). The simulation-based assessments consisted of two or three curriculum-embedded assessments with feedback and coaching to be used by the teacher as formative resouces to adjust instruction. An end-of-unit benchmark assessment, without feedback and coaching was designed to serve as a summative measure of the state of students' proficiencies on specified core ideas and science practices.

A six-state Design Panel reviewed the study findings supporting the technical quality, feasibility, and utility of the benchmark assessments and judged that the SimScientists simulation-based assessments could serve as credible components of a state science assessment system. Interviews of state representatives by UCLA's Center for Research on Evaluation of Students, Standards, and Testing (Herman, Dai, Htut, Martinez & Rivera, 2011) documented positive feedback overall. The state representatives reported that the SimScientists assessments worked well, and that teachers were willing to participate. The state representatives, impressed with teachers' reactions and the nature of the assessments and associated reflection activities, encouraged development and implementation in additional science topics and in subject areas beyond science, such as mathematics.

The six states on the Design Panel collaborated with WestEd to formulate two models for states to use simulation-based science assessments. The models aimed to describe how simulation-based assessments could become part of balanced state assessment systems, at the classroom, district, and state levels, with common designs that would make them mutually reinforcing (Pellegrino, et al, 2001; Quellmalz & Moody, 2004). The two models created combinations of simulation-based science assessments that would be *coherent* with each other, *comprehensive* in coverage of state science standards, and provide *continuity* of assessments through multiple forms and occasions.

The two models proposed included using classroom assessment proficiency data to augment state reports and use of a sample of simulation-based "signature tasks" parallel to those in the benchmarks to administer as part of state or district tests. Figure 12 presents a sample report that could be generated in the "Side-by-Side" model in which data at the state, district, and classroom levels are mutually aligned and complementary. District and classroom assessments can provide increasingly rich sources of information, allowing a fine-grained and more differentiated profile of a classroom, school, or district that includes aggregate information about students at each level of the system. In this "Side-by-Side" model, the unit benchmark assessments can function as multiple measures administered after science units during the school year, providing a continuity of in depth, topic-specific "interim" or "through-course" measures that are directly linked in time and substance to units on science systems.

Figure 13 portrays the "Signature Task" model in which states and districts draw upon specifications and rich simulation environments developed for classroom assessments to create a new, parallel set of tasks. These signature tasks could be administered in a matrix sampling design during the state or district testing to collect data on inquiry practices and integrated knowledge not fully measured by traditional item formats on the state test.

For example, the first task in each row shows a signature task for investigating the effect of forces on objects. On the state test, the object is a train. On the classroom assessment, the object is a fire truck. The masses, forces, and results of the investigations vary between the parallel tasks, but the simulation interface and the task structure are otherwise identical.

This model assures coherence of task types in different levels of the assessment system. The two models can provide a template for states to

Figure 12. Side-by-Side Model, showing how data reported from unit benchmark assessments can augment information from district and state science reports

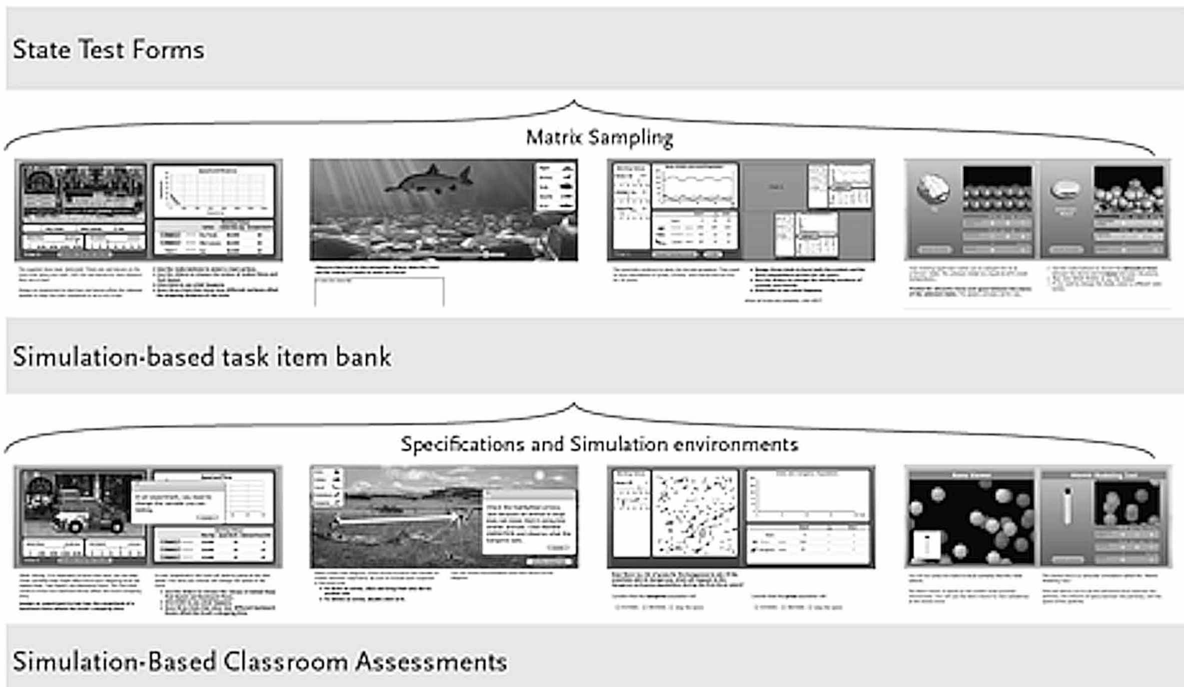
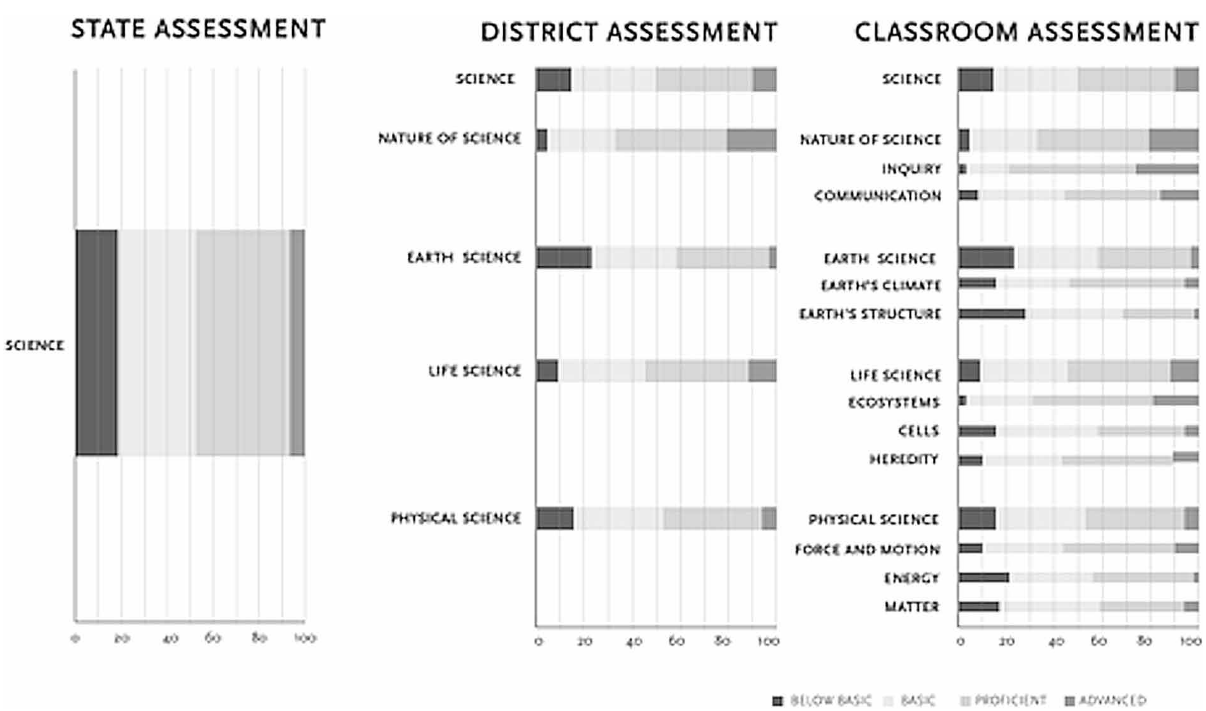


Figure 13. Signature Task model, showing how parallel tasks can be developed for state and classroom assessments



begin moving closer to the goal of a system for state science assessment that provides meaningful information drawn from a system of nested assessments collected across levels of the educational system.

In two recent SimScientists projects, a life science and a physical science strand of assessment suites are being developed for multiple units within a grade level. Each suite consists of simulation-based, curriculum-embedded assessments for formative use and end-of-unit benchmark assessments for summative evidence. Sets of simulation-based signature tasks are being developed from the template specifications used for the curriculum-embedded and benchmark assessments. End-of-year assessments are being developed for the life science and physical science strands that will consist of sets of simulation-based signature tasks.

These studies of the implementation, technical quality, and impacts on learning of the WestEd STEM projects provide evidence of the value of simulations for promoting and assessing science learning. Coupled with a principled approach to the design of simulation-based learning environments, the rigorous development and validation process can serve as a strong model for the design and empirical study of other technology-enhanced projects.

FUTURE DIRECTIONS

The WestEd STEM projects are conducting further research on the impacts of simulations on learning and assessment. Projects are also extending simulation designs into other genre of technology enhanced learning environments, including 3D simulations and games.

Research Directions

Learning Progressions

Two SimScientists projects are beginning to investigate the affordances of simulations for supporting and assessing the development of students' understanding of system models of natural and engineered phenomena studied throughout the school year. The *Model Progressions* project targets middle school students' understanding of genetics, evolution, and ecosystems as well as their ability to use genetics models to reason about evolution and ecosystems and the interactions among the three topics. The *SimScientists Crosscutting Concepts: Progressions in Earth Systems* aims to investigate learning trajectories for three crosscutting concepts (scale, cycles, systems) and will study development of these learning trajectories across three middle school Earth science topics (geosphere, climate, ecosystems).

Learning progressions should focus on foundational and generative ideas and practices of the domain, be grounded in research, possess an internal conceptual coherence, and be empirically testable (Corcoran, Mosher, & Rugat, 2009; Duncan, 2009). Like Songer and colleagues (2009), the SimScientists projects are investigating learning progressions from a disciplinary perspective. The focus is on foundational systems of a domain and science practices that enable learners or scientists to generate and test hypotheses. In model-based learning terms, learning progressions describe pathways by which learners' mental models of dynamic phenomena become more complex, accurate, and interconnected as they approximate the targeted system model. In classrooms the path is shaped by the curriculum for the year and the sequence in which topics are addressed.

Simulations provide learners with opportunities to interact with representations of phenomena. SimScientists modules scaffold learners' interactions with the simulations and provide an instructional pathway for the elaboration of learners' mental models as well as the development of science practices that further learners' ability to develop and use simulations to understand complex systems. SimScientists modules enable students to build, test and revise their models. Current projects are exploring how to help learners connect systems across topics in the life and physical sciences, for example, how at the emergent level, organisms' genetics lead to a variety of traits that interact and evolve within ecosystems. These projects are also exploring how evidence models can detect patterns of learner responses that might characterize learning progressions.

Development Directions

Simulations offer enormous potential for representing significant dynamic phenomena in science, social science, arts, and humanities. The technology can display and overlay phenomena that change in scale, time, and distance. In science, simulations can juxtapose microscopic and macroscopic representations, local, global, and galactic phenomena. In social science, simulations can slide back and forth in time and from place to place. In art and the humanities, simulations can embed visual arts into cultural and historical contexts, and fast-forward performances.

To date, the SimScientists simulation-based assessments embedded within a unit have had small, but significant impacts on science learning, importantly, the use of inquiry practices. The logistics of computer availability and teachers' pacing guides limited the number of periods that teachers could schedule access to computers during a unit. As the SimScientists projects develop strands of the simulation-based assessment suites for additional units, further research can seek evidence of potentially stronger impacts on learning, particularly improvement in inquiry practices, over multiple units across the school year.

In Touch With Molecules

In the *In Touch With Molecules* project (molecules.wested.org), collaborators at The Scripps Research Institute and WestEd are using physical models to represent biological structures and to simulate the functions that emerge from interactions among these structures, from individual nucleotides in DNA, to viral capsids composed of many hundreds of proteins and the genetic material they encapsulate.

This project builds upon the groundbreaking work of Dr. Arthur Olson, who leads the Molecular Graphics Laboratory at Scripps. In his lab, components of the physical models are created with 3D printers, embedded with magnets, and assembled into articulating models with conformational preferences. The lab has also developed augmented reality that merges the physical world with computer graphics, tracking interactions of the physical models through the camera in a mobile device and combining the images on the device's screen.

The *In Touch With Molecules* project is integrating model use into teaching and learning in a range of contexts, from 9th-grade general biology to graduate courses. Each activity scaffolds the initial interactions with the model, challenges students to make predictions about phenomena that can be simulated with the model, and then scaffolds the process of using the model to test these predictions. The activities also challenge students to make connections between the model and the actual molecules and processes it can be used to represent, recognizing the affordances. The goal of the learning is to be able to explain how interactions between the components of biological molecules give rise to more complex structures and associated functions.

For example, in the DNA model components give rise to structural properties, such as the helical shape of a single strand of DNA, and to interactions between structures, such as complementary base pairing that brings two strands together in a double helix. Magnets in the models simulate these interactions: the forces of attraction between the two strands. The simulated forces can be felt when the model is assembled, and again when pulling apart the two strands to simulate the "unzipping" of DNA that precedes replication and gene expression. Through a simulated process of replication, students can gain first-hand experience with the structural and functional consequences of complementarity between DNA strands and the semi-conservative nature of DNA replication.

Rather than simply constructing the final, double-stranded model as a puzzle to be assembled in an arbitrary way, the task asks students to consider how each component is added during the process of DNA replication. Table 7 below shows how the task model for DNA replication integrates model construction and use with aspects of the content, including the structure and function of DNA. This task scaffolds the simulation of an important bio-molecular process, and simultaneously prompts students to consider how function arises from structure.

The *In Touch with Molecules* project is developing evidence models of student understanding by exploring the data and generating hypotheses about how interactions in the activities can be interpreted. Evidence of conceptual understanding is gathered by documenting how students use the models to answer questions and test predictions as they simulated processes. For example, students produce video recordings of the replication process as simulated with the DNA model. In the future, tracking capabilities that augment the video recording could be used to capture students' interactions with the model. An evidence model could then be employed for interpreting interactions, providing feedback to students and teachers, and monitoring student progress in mastering the relevant targets.

SimScientists Games

The rapidly growing field of educational games is a particularly promising and logical direction for extending the design of simulation-based learning environments. Scientifically principled simulations can provide models and laboratories for investigating systems in the natural and designed world. System models can become digital environments for "serious" games that address the Next Generation Science Standards.

Games are seen as a promising strategy for immersing students in the excitement of doing science. Educational games can offer a sharp contrast to the prevailing activity structure in U.S. science classrooms characterized as "motivate, inform, and assess," treating science as a "final form" of solved problems and theories to be transmitted

Table 7. Task model for DNA replication

Step in Model Construction	Integrating Task
Add one nucleotide to begin the complementary strand.	How did you know which base to add? Explain how features of the model support your answer.
Add a second nucleotide to continue forming the complementary strand.	The hydrogen bonds in the new base pair form first, before the covalent bond in the backbone. Explain why the bonds form in this order. (Hint: what would happen if they didn't?)
Add a third nucleotide to the complementary strand.	Will the template strand and complementary strand of DNA be identical? Explain how features of the model support your answer.
Add a fourth nucleotide to the complementary strand.	Each strand of DNA is considered to be a polymer. What is the monomer of DNA? Explain how features of the model support your answer.
Add the fifth nucleotide to the complementary strand.	Bases can pair in other ways than by the base-pairing rules you learned in class. How do incorrect pairs affect the structure of DNA? Use the model to find at least two ways.

(Duschl, Schweingruber, & Shouse, 2007; Linn & Eylon, 2006). Such static transmission of science not only fails to promote deep science learning, but also squelches students' interest in the study of science. Reports by the NRC and others have summarized the potential of games to enhance motivation, conceptual understanding, and science process skills, but noted that much more research is needed on game design and learning impacts (Honey & Hilton, 2011; Martinez-Garza, Clark, & Nelson, 2013; Quellmalz et al., 2009).

Games are renowned for their appeal, but also for their dearth of focus on educationally significant, deep knowledge, strategic problem solving, and research-based mechanisms to promote or assess academic learning. To date, evidence of valued science learning is patchy, but studies are emerging about the benefits of games for science learning (Clark, Tanner-Smith, & Killingsworth, 2014; Honey & Hilton, 2011; Quellmalz, Timms, & Schneider, 2009). Research from cognitive learning, model-based reasoning, achievement motivation, and evidence-centered assessment design can be merged with the conventions of game design to produce activities that make learning effective and fun.

The SimScientists simulation-based supplementary curriculum-embedded assessments could be employed to conduct further research and development on how a new genre of *cognitively-principled* science learning games can promote, assess, reinforce, and extend deep science learning and also harness gameplay to motivate and engage. To foster learning, game features would include a focus on clear learning goals, compelling narrative quests, a balance of challenge and scaffolding with just-in-time feedback, hints, and explanation, adaptive problems, visual concrete and idealized representations, and user control (Clark, Nelson, Sengupta, & D'Angelo, 2009; Moreno & Mayer, 2005; Salen & Zimmerman, 2003; Squire, 2006). Within a game, students would take the role of empowered actors who must actively apply content knowledge and science practices to achieve a goal (Barab, Gresalfi, & Ingram-Goble, 2010). The games would provide adaptive levels of difficulty that challenge and engage students without interrupting the flow of play (Shute, Rieber, & Van Eck, 2011; Gee, 2007), and the scaffolding and engagement needed for students to engage in important science practices called for in the *NGSS* (Clark, et al, 2012; Kafai, Quintero, & Felton, 2010; Squire & Jan, 2006; Steinkuehler & Duncan, 2008). In addition to outcomes for science concepts and practices, games would promote and assess 21st century skills such as collaboration; the game platforms could allow massive, multiplayer games that promote collaborative problem solving.

CONCLUSION

This chapter describes how research can inform the design of simulations that model science systems with the aim of promoting understanding of core ideas about systems in the natural and designed world along with the application of science and engineering practices to study and learn about these systems. The lines of research inform the design and linking of specified knowledge and skills, tasks for representing science phenomena and for eliciting observations of students' understanding of core ideas practices, and alignments of goals and tasks to elicit and evaluate evidence of learning, and to report it. Design principles derived from research and best practice inform designs of simulation-based environments to promote and assess science learning, along with research methods for evaluating the quality and validity of simulation projects. The findings from empirical work in schools demonstrate their technical quality and their impacts on learning. When guided by findings from learning research, technology-enhanced environments using simulations can fundamentally transform science education, and provide future directions for research and development.

REFERENCES

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). Dordrecht: Springer. doi:10.1007/978-1-4020-5267-5_9

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play using games to position person, content, and context. *Educational Researcher*, *39*(7), 525–536. doi:10.3102/0013189X10386593

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.019

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brenner, D. G., Timms, M., McLaren, B. M., Brown, D. H., Weihnacht, D., Grillo-Hill, A., . . . Li, L. (2014). *Exploring the Assistance Dilemma in a Simulated Inquiry Learning Environment for Evolution Theory*. Paper presented at the 2014 Annual Meeting of the American Education Research Association, Philadelphia, PA.

Buckley, B., Gobert, J., Kindfield, A., Horwitz, P., Tinker, R., Gerlits, B., & Willett, J. et al. (2004). Model-based teaching and learning with BioLogica™: What do they learn? How do they learn? How do we know? *Journal of Science Education and Technology*, *13*(1), 23–41. doi:10.1023/B:JOST.0000019636.06814.e3

Buckley, B. C. (2012). Model-based teaching. In M. Norbert (Ed.), *Encyclopedia of the sciences of learning* (pp. 2312–2315). New York: Springer.

Case, B. J. (2008). Accommodations to improve instruction and assessment. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability*. Washington, DC: Gallaudet Research Institute.

Case, B. J., Brooks, T., Wang, S., & Young, M. (2005). *Administration mode comparability study for Stanford diagnostic Reading and Mathematics tests*. San Antonio, TX: Harcourt Assessment, Inc.

CAST. (2008). *Universal design for learning guidelines version 1.0*. Wakefield, MA: Author.

Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009, October). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence.* Paper presented at Learning science: Computer games, simulations, and education: Workshop conducted from the National Academy of Sciences, Washington, DC.

Clark, D., Tanner-Smith, E., & Killingsworth, S. (2014). *Digital Games, Design and Learning: A Systematic Review and Meta-Analysis (Executive Summary).* Menlo Park, CA: SRI International.

Clark, D. B., Martinez-Garza, M. M., Biswas, G., Luecht, R. M., & Sengupta, P. (2012). Driving assessment of students' explanations in game dialog using computer-adaptive testing and hidden Markov modeling. In D. Ifenthaler, D. Eseryel, & G. Xun (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 173–199). Springer New York. doi:10.1007/978-1-4614-3546-4_10

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Wiley.com.

Clement, J., & Rea-Ramirez, M. A. (2008). *Model based learning and instruction in science*. Dordrecht: Springer. doi:10.1007/978-1-4020-6494-4

College Board. (2009). *Science: College Boards standards for college success*. Retrieved from http://professionals.collegeboard.com/profdownload/cbscs-science-standards-2009.pdf

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). London: Routledge.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform. CPRE Research Report# RR-63*. Philadelphia: Consortium for Policy Research in Education.

Davenport, J. L., Rafferty, A., Timms, M. J., Yaron, D., & Karabinos, M. (2012). ChemVLab+: Evaluating a Virtual Lab Tutor for High School Chemistry. *The Proceedings of the 2012 International Conference of the Learning Sciences*, (pp. 381–385). Academic Press.

Davenport, J. L., Rafferty, A., Yaron, D., Karabinos, M., & Timms, M. (April, 2014). *ChemVLab+: Simulation-based Lab activities to support chemistry learning*. Paper presented at the 2014 Annual Meeting of the American Educational Research Association, Philadelphia, PA.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, *46*(6), 606–609. doi:10.1002/tea.20316

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades k–8*. Washington, DC: The National Academies Press.

Gee, J. P. (2007). What video games have to teach us about learning and literacy (2nd ed.). Palgrave Macmillan.

Gobert, J. D., & Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education*, *22*(9), 891–894. doi:10.1080/095006900416839

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, *36*(1), 39–53. doi:10.1002/(SICI)1098-2736(199901)36:1<39::AID-TEA4>3.0.CO;2-I

Goldstone, R. L. (2006). The complex systems see-change in education. *Journal of the Learning Sciences*, *15*(1), 35–43. doi:10.1207/s15327809jls1501_5

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer through complex systems principles. *Journal of the Learning Sciences*, *17*(4), 465–516. doi:10.1080/10508400802394898

Grotzer, T. A. (2003). Learning to understand the forms of causality implicit in scientifically accepted explanations. *Studies in Science Education*, *39*(1), 1–74. doi:10.1080/03057260308560195

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction*, *14*(3), 343–351. doi:10.1016/j.learninstruc.2004.06.007

Herman, J., Dai, Y., Htut, A. M., Martinez, M., & Rivera, N. (2011). *Evaluation of the enhanced assessment grants (EAG) SimScientists program: Site visit findings*. Los Angeles: CRESST.

Hmelo-Silver, C. E., Jordan, R., Liu, L., Gray, S., Demeter, M., Rugaber, S., & Goel, A. et al. (2008). Focusing on Function: Thinking below the Surface of Complex Natural Systems. *Science Scope*, *31*(9), 27–35.

Honey, M. A., & Hilton, M. (Eds.). (2011). *Learning science through computer games and simulations*. Washington, DC: National Academies Press.

Ioannidou, A., Repenning, A., Webb, D., Keyser, D., Luhn, L., & Daetwyler, C. (2010). Mr. Vetro: A Collective Simulation for teaching health science. *International Journal of Computer-Supported Collaborative Learning*, *5*(2), 141–166. doi:10.1007/s11412-010-9082-8

Kafai, Y. M., Quintero, M., & Felton, D. (2010). Investigating the 'why' in Whypox: Casual and systematic explorations of a virtual epidemic. *Games and Culture*, *5*(1), 116–135. doi:10.1177/1555412009351265

Krajcik, J., Marx, R., Blumenfeld, P., Soloway, E., & Fishman, B. (2000, April). *Inquiry based science supported by technology: Achievement among urban middle school students*. Paper presented at the 2000 Annual Meeting of the American Educational Research Association, New Orleans, LA.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies*, *65*(11), 911–930. doi:10.1016/j.ijhcs.2007.06.005

Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior*, *27*(1), 29–35. doi:10.1016/j.chb.2010.05.008

Law, N., Pelgrum, W., & Plomp, T. (2008). *Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study*. New York: Springer. doi:10.1007/978-1-4020-8928-2

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology*, *98*(4), 902–913. doi:10.1037/0022-0663.98.4.902

Lehrer, R., Schauble, L., Strom, D., & Pligge, M. (2001). Similarity of form and substance: Modeling material kind. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 39–74). Mahwah, NJ: Lawrence Earlbaum Associates.

Linn, M. C., & Eylon, B. S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 511–544). New York, NY: Routledge.

Lowe, R., & Schnotz, W. (Eds.). (2008). *Learning with animation: Research implications for design*. New York, NY: Cambridge University Press.

Martinez-Garza, M. M., Clark, D., & Nelson, B. (2013). Advances in Assessment of Students' Intuitive Understanding of Physics through Gameplay Data. [IJGCMS]. *International Journal of Gaming and Computer-Mediated Simulations*, *5*(4), 1–16. doi:10.4018/ijgcms.2013100101

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.004

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13–23. doi:10.3102/0013189X023002013

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design. CSE Technical Report. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20. doi:10.1111/j.1745-3992.2006.00075.x

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*(1), 117–128. doi:10.1037/0022-0663.97.1.117

National Research Council (NRC). (2012a). *A framework for k–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

National Research Council (NRC). (2012b). *Next generation science standards*. Available at: http://www.nextgenscience.org/next-generation-science-standards

Nersessian, N. J. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/20072004.pdf

Pellegrino, J. W. (2002, February). Understanding how students learn and inferring what they know: Implications for the design of curriculum, instruction and assessment. In M. J. Smith (Ed.), *NSF k–12 mathematics and science curriculum and implementation centers conference proceedings* (pp. 76–92). Washington, DC: National Science Foundation and American Geological Institute.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press.

Perkins, D. N., & Grotzer, T. A. (2000). *Models and moves: Focusing on dimensions of causal complexity to achieve deeper scientific understanding*. Retrieved from http://files.eric.ed.gov/fulltext/ED441698.pdf

Phillips, S. E. (1993). *Legal implications of high-stakes assessment: What states should know*. Retrieved from http://files.eric.ed.gov/fulltext/ED370985.pdf

Plass, J. L., Homer, B. D., & Hayward, E. O. (2009). Design factors for educationally effective animations and simulations. *Journal of Computing in Higher Education*, *21*(1), 31–61. doi:10.1007/s12528-009-9011-x

Quellmalz, E., & Kozma, R. (2003). Designing assessments of learning with technology. *Assessment in Education: Principles, Policy & Practice*, *10*(3), 389–407. doi:10.1080/0969594032000148208

Quellmalz, E., Timms, M., & Buckley, B. (2005). *Using science simulations to support powerful formative assessments of complex science learning*. Paper from the annual meeting of the American Educational Research Association, San Diego, CA.

Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G., Jordan, K., Huang, K., & Buckley, B. (2013). Next-generation environments for assessing and promoting complex science learning. *Journal of Educational Psychology*, *105*(4), 1100–1114. doi:10.1037/a0032220

Quellmalz, E. S., & Haertel, G. D. (2008). Assessing new literacies in science and mathematics. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 941–972). Mahwah, NJ: Lawrence Erlbaum Associates.

Quellmalz, E. S., & Moody, M. (2004). *Models for multi-level state science assessment systems*. Paper commissioned by the National Research Council Committee on Test Design for K-12 Science Achievement.

Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, *323*(5910), 75–79. doi:10.1126/science.1168046 PMID:19119222

Quellmalz, E. S., Timms, M. J., Buckley, B. C., Davenport, J., Loveland, M., & Silberglitt, M. D. (2012). 21st century dynamic assessment. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age.

Quellmalz, E. S., Timms, M. J., & Schneider, S. A. (2009). *Assessment of student learning in science simulations and games*. Washington, DC: National Research Council.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*(3), 363–393. doi:10.1002/tea.21005

Quellmalz, S., & Timms, B. (2012). Multilevel assessments of science systems: Final report. Redwood City, CA: WestEd.

Quellmalz, Timms, Buckley, Loveland, & Silberglitt. (2012). *Calipers II: Using simulations to assess complex science learning: Final report*. Redwood City, CA: WestEd.

Quellmalz & Silberglitt (2011, February) *Integrating simulation-based science assessments into balanced state science assessment systems: Findings and implications*. Workshop from the 2011 Meeting of the Technical Issues in Large-Scale Assessment, State Collaboratives on Assessment and Student Standards, Atlanta, GA.

Rieber, L. P., Tzeng, S. C., & Tribble, K. (2004). Discovery learning, representation, and explanation within a computer-based simulation: Finding the right mix. *Learning and Instruction*, *14*(3), 307–323. doi:10.1016/j.learninstruc.2004.06.008

Rose, D., & Meyer, A. (2000). Universal Design for Learning. *Journal of Special Education Technology*, *15*(1), 67–70.

Salen, K., & Zimmerman, E. (2003, November). This is not a game: Play in cultural environments. In *DiGRA '03 – Proceedings of the 2003 DiGRA International Conference: Level Up*. Retrieved from http://www.digra.org/digital-library/forums/2-level-up

Schwartz, D. L., & Heiser, J. (2006). Spatial representations and imagery in learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. New York, NY: Cambridge University Press.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games... and... learning. In R. Reiser & J. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed.; pp. 321–332). Upper Saddle River, NJ: Pearson Education.

Slotta, J. D., & Chi, M. T. H. (2006). The impact of ontology training on conceptual change: Helping students understand the challenging topics in science. *Cognition and Instruction*, *24*(2), 261–289. doi:10.1207/s1532690xci2402_3

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610–631. doi:10.1002/tea.20313

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, *35*(8), 19–29. doi:10.3102/0013189X035008019

Squire, K. D., & Jan, M. (2007). Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*, *16*(1), 5–29. doi:10.1007/s10956-006-9037-z

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology*, *17*(6), 530–543. doi:10.1007/s10956-008-9120-8

Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn* (pp. 515–565). Washington, DC: National Research Council.

Stieff, M., & Wilensky, U. (2003). Connected chemistry—incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology*, *12*(3), 285–302. doi:10.1023/A:1025085023936

Tversky, B., Heiser, J., Mackenzie, R., Lozano, S., & Morrison, J. (2008). Enriching animations. In R. K. Lowe & W. Schnotz (Eds.), *Learning with animation: Research implications for design* (pp. 263–285). New York, NY: Cambridge University Press.

Van Merrienboer, J. J., & Kester, L. (2005). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 71–93). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.006

Wang, S. (2005). *Online or paper: Does delivery affect results? Administration mode comparability study for Stanford diagnostic Reading and Mathematics tests*. San Antonio, TX: Harcourt Assessment.

Ware, C. (2004). *Information visualization: Perception for design*. San Francisco: Morgan Kaufmann.

White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, *42*(1), 99–157. doi:10.1016/0004-3702(90)90095-H

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118. doi:10.1207/s1532690xci1601_2

KEY TERMS AND DEFINITIONS

Dynamic: Phenomena changing in time and scale.

Evidence-Centered Design: Specifications of assessment design in terms of knowledge and skills to be assessed (student model), tasks to elicit observations of the knowledge and skills (task model), and evaluations of student responses (evidence model).

Model-Based Learning: Framework characterizing learners' formation, use, evaluation, and revision of their mental models of phenomena as learners interact with phenomena in situ and with conceptual models, representations (including text), and simulations of phenomena.

Multilevel Assessment Systems: Coherent, articulated assessment systems from the classroom to district, to state to national levels based on common specifications of learning standards and task models.

Multimedia: Representations of phenomena and means of expression employing a variety of static, active, and interactive modalities such as pictures, graphics, text, animations, and simulations.

Representations: Static, active, and interactive renderings of phenomena.

SimScientists: Program of research and development projects at WestEd studying the capabilities of simulations for promoting and assessing science learning.

Universal Design for Learning: Methods for offering alternative means for representing information in multiple formats and media, providing multiple pathways for students' action and expression, and multiple ways to engage students' interest and motivation.