Quality Science Simulations for Formative and Summative Assessment

Edys S. Quellmalz, Jodi Davenport, Michael J. Timms, Barbara C. Buckley
WestEd

Introduction

Multiple forces are converging to propel science assessment into the 21st century. The requirements of the No Child Left Behind Act (NCLB) for science testing at the elementary, middle, and secondary levels are focusing educators on the suitability of available assessments for measuring what students should know and be able to do in science. States must provide evidence that students are meeting challenging science standards, yet the content and structure of many large-scale science accountability tests are widely considered to be too limited to measure students' understanding of complex science ideas or their abilities to conduct scientific inquiry. Many see the powerful capabilities of technology as holding the key to transforming current science assessment practice (Quellmalz & Moody, 2004; Quellmalz & Haertel, 2004). Technology can expand what gets tested and how it is tested.

New technologies allow representation of domains, systems, and data in new, more powerful ways that are affecting the practice of science, mathematics, and engineering. Technologies such as dynamic models of ecosystems or molecular structures help scientists visualize and communicate complex interactions. Representational technologies have expanded the phenomena that can be investigated and the nature of argumentation and acceptable evidence in many fields (Holland, 1995). This move from static to dynamic models has profoundly changed the nature of inquiry among professional scientists and the way that science can be taught. Simulations and models are heralded as particularly powerful for dynamically representing the spatial, causal, and temporal processes in science systems and for permitting active, virtual investigations of phenomena that are too big or small, fast or slow, or dangerous to be conducted in hands-on labs (Gobert et al., 2004). New science curricula are integrating technologies into activities that engage students in constructing mental models of systems in the natural world and inquiry strategies for understanding and investigating those systems (White & Frederiksen, 1998; Clark & Linn, 2003; Songer, 2004; Horwitz et al., 2007).

Propelled by these trends, technology-based science tests are rapidly appearing in state, national, and international testing programs. For example, the *Science Framework for the 2009 National Assessment of Educational Progress (NAEP)* calls for the design of computer-based science investigation tasks, and the 2006 and 2009 cycles of the Programme for International Student Assessment (PISA) are piloting computer-based forms (NGB, 2006; Koomen, 2006). Moreover, technology-enhanced, formative classroom assessments mimicking the skills and types of items in high-stakes tests are becoming increasingly marketed and will affect the focus of science instruction.

Unfortunately, the press to implement more complex computer-based science tests has outpaced research addressing crucial questions about the validity, comparability, and complementarities of changing science assessment task designs. Also, while most assessment

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

developers are highly experienced in creating static items with no interactivity, they are likely unaware of the field of learning research that has studied how students learn and perform best in interactive multimedia formats. As assessment tasks and items become more interactive, developers need to take account of this body of cognitive research, as it offers much that is relevant to the development of complex technology-based assessments.

This presentation will describe research and development projects in the SimScientists program at WestEd that are studying how the features of science simulations can be designed to serve as powerful formative assessment resources. We lay out design principles drawn from cognitive and multimedia research that may inform the designs of complex science assessments and the methods we are employing to study them. We then describe simulation based science assessments being developed for middle school science as end-of-unit summative benchmark assessments, and curriculum-embedded assessments intended to benefit learning.

Background

The state-of-practice for the design of innovative, computer-based assessment item formats is embryonic, primarily occurring in large-scale testing programs, and typically not explicitly referenced to cognitive research or principles of multimedia design. Instead, to create assessment tasks and items aligned with specified science standards, test developers usually rely on the advice of content experts and teachers and refer to representations of science phenomena and graphics that appear in existing paper-based assessments and curricula. Since most current science instructional materials and tests are in print form, relatively few examples of dynamic representations of science principles and systems are available as exemplars.

In an attempt to provide a general framework for the PISA 2006 Computer-Based Assessment of Science, Koomen (2006) described a continuum that influenced the types of items and tasks developed. Different types of computer-based testing were located along a spectrum from conventional to transformational. At the conventional end of the spectrum were located "computer-fixed" forms that contained the same items as those in their corresponding paper tests (Paek, 2005). The term "transformational," proposed by Boston (2005), referred to dynamic test designs at the other end of the spectrum that exploit the capabilities of technology. Between the ends of the spectrum were tests characterized as "transitional."

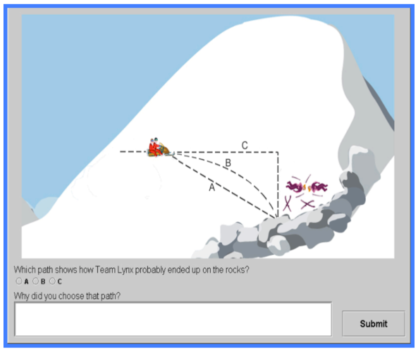
Foundations of 21st Century Science Assessments

The relationship between the designs of assessment tasks and items and the constructs they measure is being investigated in a SimScientists project at WestEd, Foundations of 21st Century Science Assessments (Foundations). We will be studying three task and item types that represent the lower, middle, and upper segments of the continuum, which we refer to as static, active, and interactive. In the *static* modality, learners are exposed to conventional items that could easily be administered in a paper and pencil context. In the *active modality*, learners can go beyond passive views of dynamic stimuli by controlling the pacing and direction of an animation. Animations become *interactive* simulations if learners can manipulate parameters as they generate hypotheses, test them, and see the outcomes, therefore taking advantage of technological capabilities well suited to conducting scientific inquiry. Figure 1 illustrates the three different types of tasks and items for testing knowledge of force and motion and use of

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

inquiry skills in the context of deploying a snowmobile on an incline plane (mountain) to rescue injured skiers.

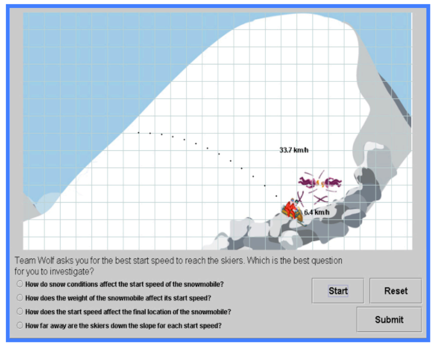
Static



Typical Task Features
- Fixed text and graphics
- Text-based selected and constructed-response formats

Level of Student Activity
Student looks at diagram and answers questions about force and motion.

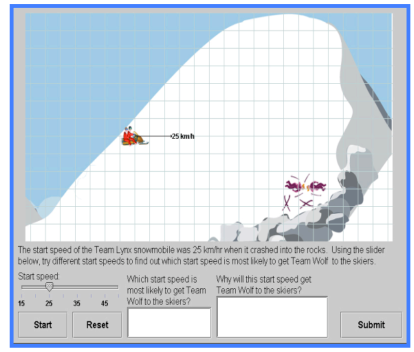
Active



Typical Task Features
- Animated phenomena, tables, and graphs
- Drag-and-drop response formats
- Text-based selected and constructed-response formats

Level of Student Activity
Student views and can replay animation of the snowmobile and answers questions about force and motion.

Interactive



Typical Task Features
- Simulations of phenomena
- Model creation and manipulation
- Drag-and-drop, hot-spots
- Interactive drawing and manipulation of variables
- Response log files of actions and responses

Level of Student Activity
Student manipulates the initial velocity of the snowmobile, observes the resulting location, then answers questions about force and motion. Student multiple attempts can be recorded and measured.

Figure 1. Examples of Static, Active, and Interactive Assessment Tasks and Items

In the Foundations project we are reviewing the relevant research literature and existing items in order to develop task design principles to inform the creation of assessments that meaningfully tap science inquiry and reasoning skills. Although the broad continuum of interactivity discussed above is useful for thinking about general differences among item types, cognitive research and multimedia research can help to articulate the task design features considered most effective as external representations of science problems presented to students and the types of student responses that will make student learning visible. Preliminary results from the Foundations project analyses of extant assessments suggest task design principles that specify *when* the different modalities (e.g. static, active and interactive) should be used for assessment items and *how* items should be developed to maximize construct validity.

Guidelines for when to use static, active or interactive items. A student's working memory is limited in capacity, so the modality of an item can contribute to a student's cognitive load (e.g. Mayer & Moreno, 2003). As greater interactivity requires more of a student's cognitive processing resources, a rule of thumb in the multimedia learning literature is to use only as much interactivity as required by the task demands (Clark, 2005). For assessments, the task design

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

principle of *alignment* suggests that the modality of the task should align with the task demands of the item in order to minimize non-essential cognitive processing. Thus, unless the task is assessing a construct that should make use of the affordances of technology (e.g., dynamic temporal, spatial, or causal changes), the item should be implemented in the *static* modality. The *active* modality affords items that include dynamic animations and should be used with tasks that require students to make observations from temporal data or extract patterns from dynamic displays. In a review of animation and interactivity principles in multimedia learning, Betrancourt (2005) cautioned that multimedia representations have evolved from sequential static text and picture frames to increasingly sophisticated visualizations and that often students may learn more effectively from static displays. Animations are considered particularly useful for providing visualizations of dynamic phenomena that are not easily observable in real space and time scales (e.g., plate tectonics, circulatory system). The *interactive* modality affords items that include simulations and should be used with tasks that require students to generate data and reason from outcomes. *Interactive* items can give students contingent responses, immediate feedback on their actions, and allow for a wide range of student responses.

Guidelines for how to use active or interactive items. Assessment developers are well-versed in the creation of static items. Thus the focus of the Foundations project is to develop task design principles for the creation of *active* and *interactive* items. First and foremost, all tasks should be designed to provide evidence of student understanding of a targeted construct. However, dynamic assessment items also require additional attention to the cognitive processing of the learner. Here we give some examples of the design principles we are extracting from the cognitive and multimedia learning literature.

Active items. *Active* items often include animations and dynamic displays. As research in the animation literature suggests that these displays may be difficult for students to process, dynamic items should be *aligned with task demands* and allow for *user-control* (de Jong, 2005; Schnotz & Rasch, 2008). External representations of life, physical, and earth science models are hypothesized to instantiate the relationships among system components and make scientific phenomena accessible, visual, and transportable (Gooding, 1990; Latour, 1990). Some key features of models include scalability and reproducibility and the potential for superimposing multiple representations and permitting manipulation of structures and patterns that otherwise might not be visible or even conceivable. When active items are used to assess student understanding of these scientific models, assessment designers should ensure that the task *requires* the use of an animation so that the dynamic display is not a source of extraneous processing. Further, tasks involving *active* displays should give students control of the pace of the animation. When degrees of *learner control and interactivity* are introduced as variables, other research suggests that spatial representations enable effective mental simulations and visualizations (Schwartz & Heiser, 2005).

Interactive items. Current technology allows for the creation of simulation environments where students have the ability to choose inputs to complex models of scientific principles and observe the outcomes of their choices. These environments afford the ability to assess rich constructs such as scientific reasoning and inquiry skills. However, open-ended simulation environments can also lead to floundering if tasks are not well structured or if the visual representations are not designed to minimize extraneous demands on cognitive processing.

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

Our review of the cognitive an multimedia learning literature suggests a number of task design principles to guide the development of interactive assessment items and tasks.

The *mapping* and *fidelity principles* guide the design of test items in order to minimize extraneous processing of the simulation itself. The *mapping principle* suggests that a clear mapping be made between multiple representations of a science principle. Visual cues, such as color, and consistent terminology can help students make connections between graphs, pictures, and alternative views in dynamic displays (Ainsworth, Bibby & Wood, 2002). The *fidelity principle* reflects the degree of realism that is appropriate for an interactive simulation. While it may be technically possible to replicate intricate details of scientific phenomena, research suggests that the fidelity and scope of a simulation should be appropriate to the instructional goal (Clark, 2005; Gott & Lesgold, 2000).

The *guided discovery* and *progressive* complexity principles guide the design of interactive test tasks. Open-ended simulation environments allow a wide variety of possibilities for student interaction. However, with out guidance, students may flounder and unstructured student actions may not be interpretable for assessment purposes. Thus the *guided discovery* principle recommends that students be given clear goals for the tasks to accomplish within the simulation environment. Ideal tasks give students the flexibility to use a range of strategies, but constrain responses to allow for the identification of patterns of behavior that give evidence for different levels of performance on the specified constructs. Further, guided discovery can offer help by activating relevant prior knowledge, providing structure for experimental design and prompting reflection on inquiry processes (Clark & Feldon, 2005; de Jong & van Joolingen, 1998). The *progressive complexity* principle also reflects the fact that interactive items with multiple possible inputs and representations can initially overwhelm a student and recommends that interactive tasks be sequenced such that students are gradually introduced to the full complexity of a simulation. Early tasks may require only partial functionality of a simulation, and later tasks may allow for the use of more features and interactions.

SimScientists Assessments

In one of WestEd's SimScientists projects, simulation-based curriculum-embedded and benchmark science assessments are being developed in an NSF-funded study, Calipers II: Using Simulations to Assess Complex Science Learning. In this section, we summarize the project and indicate how the design principles we have been identifying in the Foundations project are being applied in the design of the simulation-based science assessments.

Goals of the Calipers II project include: (1) develop simulation-based assessments that can be embedded in curriculum units for formative purposes and as benchmark assessments that can be administered at the end of a unit for summative information about proficiency on intended standards, (2) document the re-usable designs and processes employed, (3) provide evidence of the technical quality, feasibility, and utility of the new assessments, and (4) study the influence of formative assessments on complex science and inquiry learning.

The simulation-based assessments are designed to present dynamic, engaging interactive tasks of established technical quality that test complex science knowledge and inquiry skills that go well beyond the capabilities of print tests. Benchmark assessments are designed to test end-of

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

unit achievement of content and inquiry standards addressed in curriculum units on a middle school science topic such as human body systems or climate. Sets of shorter, embedded assessments are designed to be used during the instructional unit. The simulation-based embedded science assessments are intended to function as formative resources in three ways: (1) provide immediate feedback contingent on an individual student's performance, (2) offer graduated levels of coaching in real-time, and (3) provide diagnostic information to guide offline reflection and extension activities. The software and technical infrastructure of these simulation-based assessments can overcome many of the practical constraints that have limited the use and effectiveness of formative assessments. The SimScentists projects couple embedded assessment simulations with follow-on, off-line self-assessment and reflection activities. The embedded assessment and reflection activities incorporate the features of effective formative assessment: *frequent* use of standards-based classroom assessments; feedback that is *timely*, *individualized*, and *diagnostic*; online *supplementary instruction* that is individualized; and off-line *self-assessment* and reflection activities that help students confront misunderstandings, make new connections, and become more reflective, self-regulating learners (Herman et al., 2005).

The design principles shaping the assessments draw on recommendations from research on methods to promote science learning. They include: (1) a focus on integrated knowledge about the dynamic relationships among structures, behaviors, and mechanisms within models of science systems, (2) use of authentic, problem-driven inquiry practices, (3) scaffolding that generates immediate, individualized feedback, and levels of customized coaching, (4) metacognitive self-assessment and reflection, (5) scientific explanations and arguments, and (5) use of the affordances of simulations to provide multiple representations of dynamic, causal, and temporal phenomena and to offer multiple ways for students to interact and respond (Duschl et al., 2007).

Figure 1 presents a screen shot with an example of feedback and coaching provided for an assessment task in which students are asked to draw arrows in a food web to represent the flow of energy among organisms the students have previously observed. The technology enhances the feedback and coaching possible in multiple ways, such as allowing students to observe again the roles of organisms by re-running an animation of the organisms in the environment as they interact. The feedback and coaching system can highlight incorrect arrows a student has drawn between organisms, and provide more scaffolding by running an animation that highlights arrows being drawn correctly, and prompting the student to then draw the correct arrows before the student is allowed to proceed to the next screen. Multimedia principles employed in the animation include *task relevance*, since the assessment is designed to test students' ability to apply scientific principles to observations of a dynamic system. In addition, the task permits *user-control* allows students to search the animation and view at his/her own pace.

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

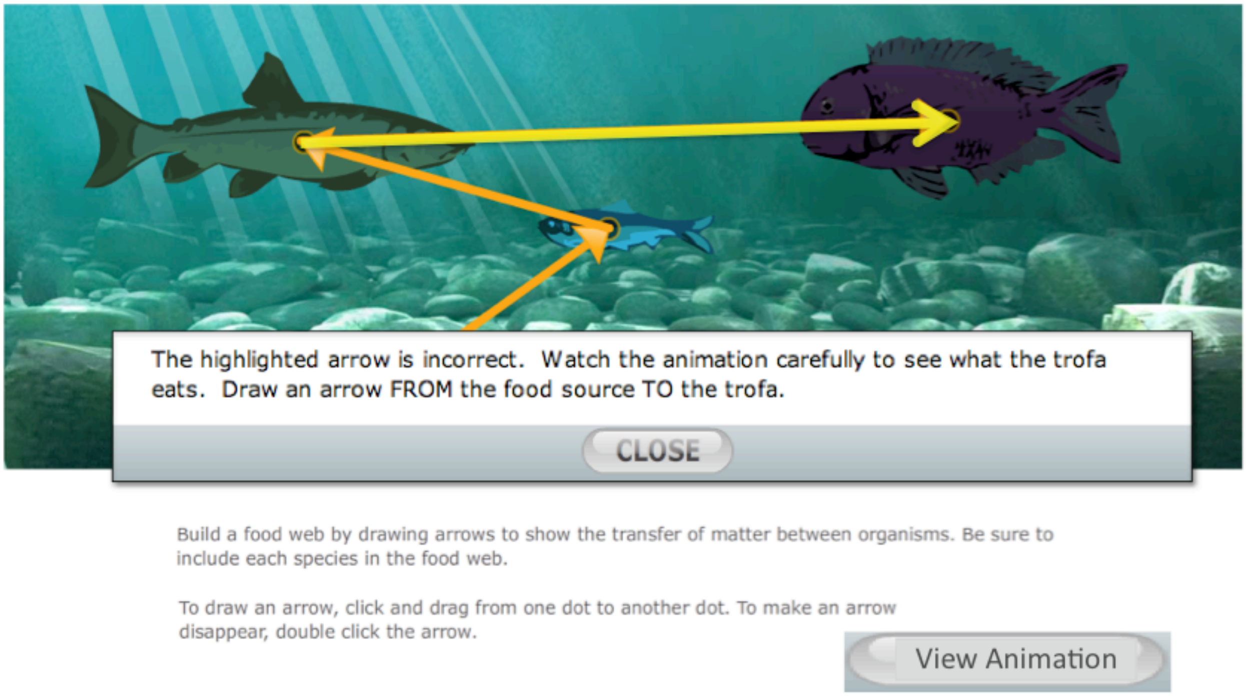


Figure 1. Sample of feedback on task in embedded assessment.

Figure 2 presents a screenshot of an ecosystem simulation in which students can conduct multiple investigations of the effects of varying the number of organisms on a model of the population of organisms in the system.

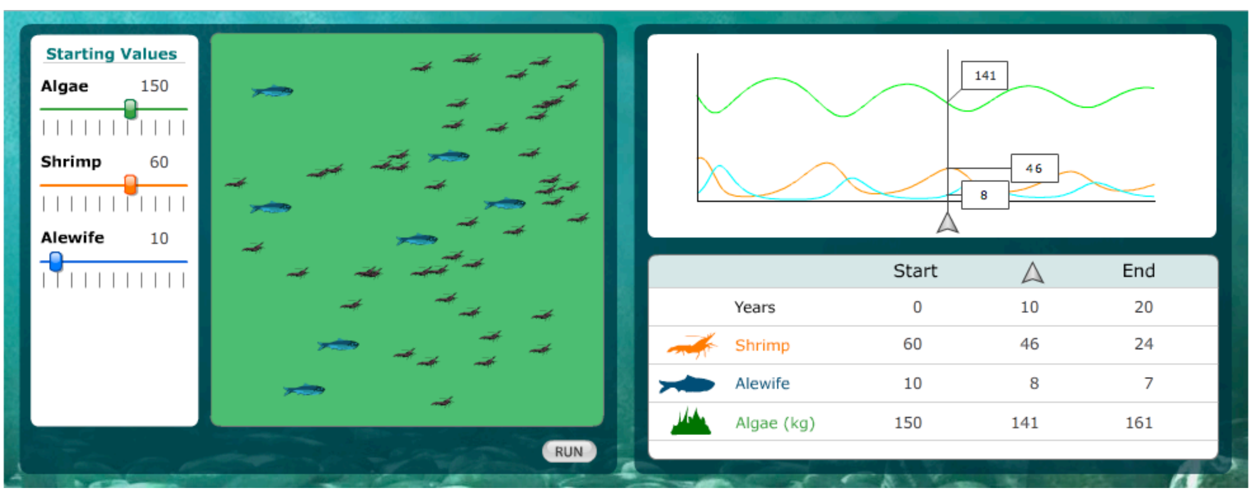


Figure 2. The population simulation for use in both embedded and benchmark assessment tasks.

The model of the population level of an ecosystem is simulated in three ways. A physical representation of changes in fish and algae populations is animated in the left box. On the right, a graph and table represent the changing population levels over time. The simulation allows

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

students to run multiple experiments. A graph inspector arrow allows students to read the graphic data generated from the simulation at different points in time to examine the relationship among the numbers of organisms. Students are asked to predict how the number of an organism will be affected initially and later when another organism is added. Feedback and coaching for these tasks in which students use the model to predict, observe, and test their predictions graduates from informing them that their prediction was incorrect and to try again, to prompting students to use the graph inspector, to highlighting the places on the graph and data in the table and pointing out the changing population numbers.

Multimedia principles employed include *mapping* of salient system components provided by color that unifies the multiple representations of organisms on the slider, table, dynamic portion and graph. The simulation task also applies the principle of *fidelity*, in that the simulation focuses on consumers and producers in a small, fictitious ecosystem. Also the set of simulation tasks designed for the population model employ the principle of *guided discovery* by providing specific tasks designed to assess particular constructs that increase in *progressive complexity,* starting with one aspect of simulation (just producer), then gradually introducing other organisms into the population dynamics model.

The Foundations of 21st Century Assessments project is in the first of its three years. The project will continue to carry out a detailed study of the stimulus features of static and dynamic science assessment problems presented to students and the nature of the responses technologies allow them to make. Our initial set of task design principles will inform the design of assessment items to investigate in static, active, and interactive forms. Once the items have been developed, we will study how well the different modalities elicit science constructs. The Foundations of 21stCentury Science Assessment project will offer recommendations for science task design structures that elicit valued constructs, for documenting the construct validity of tasks and items in the different task design structures, and for supporting the interpretations of tests along the technology spectrum.

References

Ainsworth, S.E., Bibby, P.A & Wood, D.J. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences.* 11(1), 25-62.

Betrancourt, M. (2005). The animation and interactivity principles. In R. E. Mayer (Ed.), *Handbook on multimedia learning*. New York: Cambridge University Press.

Boston, K. (2005, October). Strategy, technology and assessment. Keynote address at AR+t — Assessment, Reporting & Technology: System-wide Assessment and Reporting in the 21st Century, Roundtable Conference 2005, Melbourne, Australia

Clark, R. C. (2005). Multimedia learning in e-Courses. In Richard E. Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning* (pp. 589-616). Cambridge: Cambridge University Press.

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

Clark, R. E. & Feldon D.F (2005) The guided discovery principle in multimedia learning. In Richard E. Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning* (pp. 97-116). Cambridge: Cambridge University Press.

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. *Journal of the Learning Sciences, 12* (4), 451—493.

Duschl,R.A., Schweingruber, H.A., Shouse, A.W. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Committee on Science Learning, Kindergarten Through Eighth Grade. Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

de Jong, T. (2005) In Richard E. Mayer (Ed.) The guided discovery principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning* (pp. 215-228). Cambridge: Cambridge University Press.

de Jong, T, & van Joolingen, W. R. (1998). Scientfic discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68,* 179-201.

Gobert, J., Buckley, B. C., & Clarke, J. (2004, April). *Scaffolding model-based reasoning: Representations, information processing, & cognitive affordances*. Presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Gooding, D. (1990). *Experiment and the making of meaning*. London: Kluwer Academic Publishers.

Herman, J. L., Osmundson, E., Ayalya, C., Schneider, S., & Timms, M. (2005). *The nature and impact of teachers' formative assessment practices.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. New York: Addison-Wesley.

Horwitz, P., Gobert, J., Buckley, B. C., & Wilensky, U. (2007). *Modeling across the curriculum annual report to NSF:* Concord, MA: The Concord Consortium.

Koomen, M. (2006, April). *The development and implementation of a computer-based assessment of science literacy in PISA 2006.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice*. Cambridge, MA: MIT Press.

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist,* 38, 43–52.

National Governing Board (NGB). (2006). *Science framework for the 2009 National Assessment of Educational Progress.* Washington, DC: Author.

National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academy Press.

Paek, P. (2005). *Recent trends in comparability studies* (PEM Research Report 05-05). Iowa City, IA: Pearson Educational Measurement. Available at http://www.pearsonedmeasurement.com/research/research.htm.

Quellmalz, E. S., & Haertel, G. (2004, May). *Technology supports for state science assessment systems.* Paper commissioned by the National Research Council Committee on Test Design for K-12 Science Achievement.

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

Schnotz, W. & Rasch, T. (2007). Functions of animation in comprehension and learning. Richard Lowe and Wolfgang Schnotz (Eds.) *Learning with Animation: Research implication for design.* (92-114). Cambridge: Cambridge University Press.

Schwartz, D. & Heiser, J. (2006) Spatial representations and imagery in learning. In R. Keith Sawyer (Ed.) *The Cambridge Handbook of the Learning Sciences* (pp. 283-298). Cambridge: Cambridge University Press.

Songer, N. B. (2004). *Evidence of complex reasoning in technology and science: Notes from Inner City Detroit, Michigan, USA*. IPSI-2004 Pescara Conference, Italy.

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

This material is based upon work supported by the National Science Foundation under Grants DRL-0733345 and DRL-0814776 awarded to WestEd. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Contact: equellm@wested.org