Calipers II: Using Science Simulations to Assess Complex Science Learning

Edys Quellmalz, Barbara Buckley, Mark Loveland (WestEd) and Michael Timms (ACER)

Presented at the DRK12 Principal Investigator Meeting, Arlington, VA, June 14, 2012

Funding provided by the National Science Foundation-Grant 0733345 and the US Department of Education-OESE Grant 09-2713-126
Calipers II Goals

1. Develop simulation-based assessment modules to *supplement* and *extend* the science knowledge and skills typically addressed in static print materials

2. Foster deep learning about science systems and use of inquiry practices

3. Document the effectiveness, technical quality, feasibility, and utility of simulation-based science environments for promoting and assessing science standards
Our Theoretical Foundation

Integrates research on

- **Model-based learning** (Gobert & Buckley, 2000)
  System Framework-components, interactions, and emergent system behavior
  The formation, use, evaluation and revision of mental models

- **Evidence-centered assessment design** (Mislevy et al, 2003)
  A systematic assessment development process that links targets, tasks & data

- **Cognitive science**
  Guides design and use of representations & interactions in tasks
Principled Assessment Design Approach (Mislevy et al., 2003)

**Student Model**
- What complex of knowledge, skills, or other attributes should be assessed?

**Evidence Model**
- What behaviors or performances should reveal the relevant knowledge and skills described in the student model?

**Task Model**
- What tasks or situations should elicit the behaviors or performances described in the evidence model?
Calipers II Design Principles

- Framework of models of science systems and integrated science practices
- Authentic, problem-based inquiry
- Scaffolding
- Formative assessment
- Collaborative science practices
- Discourse for sense-making and scientific arguments
Multiple Modes of Representation
Active Inquiry

Organism box shows icons that appear and disappear

Graph shows size of population over time.

Data inspector shows population for one point on graph.

Table shows population at start, end, and point selected by data inspector.
Calipers II Assessment Suites

1. A simulation environment representing age-appropriate models of a science system

2. Problem-driven curriculum inquiry activities

3. Embedded formative assessments with an intelligent scoring system that will provide immediate feedback

4. A coaching system tailored to individual responses and providing multiple re-teaching scaffolds
SimScientists Suites

5. Offline self-assessment and reflection activities to build scientific discourse and collaboration skills

6. End-of-unit benchmark assessments of progress

7. Standards-based professional development
Components of the SimScientists Classroom Assessments

Embedded in Classroom Instruction

**Embedded Formative Assessments and Reflection Activities** (2 or 3)

- Online module with feedback and coaching
- Progress Report
- Follow up Classroom Reflection Activity

**Benchmark Summative Unit Assessments**

- Online assessment without feedback
- Teacher scores constructed responses
- LMS
- Proficiency Report
What’s in a Model of a Science System?

**Components** that have structures and rules of behavior

**Interactions** among components and their environment, as permitted by their structures and behaviors.

Complex **behaviors or properties of the system that emerge** from these interactions.
# Ecosystem Target Model

<table>
<thead>
<tr>
<th>System Model Levels</th>
<th>Model Level Descriptions</th>
<th>Content Targets by Model Level</th>
<th>Inquiry Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components</td>
<td>What are the components of the system and their rules of behavior?</td>
<td>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.</td>
<td>Use principles to identify role of organisms.</td>
</tr>
<tr>
<td>Interactions</td>
<td>How do the individual components interact?</td>
<td>Matter and energy flow through the ecosystem as individual organisms participate in feeding relationships within an ecosystem.</td>
<td>Observe interactions among organisms.</td>
</tr>
<tr>
<td>Emergent Behaviors</td>
<td>What is the overall behavior or property of the system that results from many interactions following specific rules?</td>
<td>Interactions among organisms and among organisms and the ecosystem’s nonliving features cause the populations of the different organisms to change over time.</td>
<td>predict observe explain investigate</td>
</tr>
</tbody>
</table>
# Atoms & Molecules Target System Model

<table>
<thead>
<tr>
<th>Component</th>
<th>Atoms and Molecules</th>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>Speed – Spacing – Collisions</td>
<td>Skill</td>
</tr>
<tr>
<td>Emergent</td>
<td>Boiling &amp; Melting Point – States of Matter</td>
<td>Skill</td>
</tr>
</tbody>
</table>

**Component**
- Atoms and Molecules

**Atoms and Molecules**
- Nitrogen
- Water Vapor
- Argon

**Skill**
- Observe
- Analyze
- Measure & Investigate
### Examples System Models for Other Science Domains

<table>
<thead>
<tr>
<th>Model Levels</th>
<th>Force &amp; Motion</th>
<th>Atoms &amp; Molecules</th>
<th>Climate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components</td>
<td>Objects have mass. A force is a push or pull on an object. Forces have direction and magnitude. Friction is a force that acts opposite the direction of motion.</td>
<td>All matter is made of particles that are in constant motion. Particles have size, shape, and structures that influence their interactions.</td>
<td>Earth’s orbit and axial tilt strongly influence the amounts of solar energy received by different locations.</td>
</tr>
<tr>
<td>Interactions</td>
<td>Forces are pushes and pulls that can affect the motion of an object. The change in motion is dependent on the strength of the forces, the direction of the forces and the mass of the object.</td>
<td>When particles collide, larger structures may be created. The size, shape and energy of the particles determine how the particles pack together.</td>
<td>This produces differential heating that results in convection currents in the atmosphere and the ocean that move energy and water around the planet.</td>
</tr>
<tr>
<td>Emergent Behaviors</td>
<td>An object’s motion can be described by its position, direction of motion and speed. This motion may be represented graphically.</td>
<td>Matter has properties that emerge from the nature, arrangement and motion of its constituent particles.</td>
<td>Climates in a region are strongly influenced by latitude, patterns of air and water circulation, altitude, and proximity to large bodies of water.</td>
</tr>
</tbody>
</table>
Demo: Embedded Assessment
Formative Assessment Features

• Immediate, individualized feedback and coaching
• Progress reports for students and teachers
• Reflection activities that address students’ needs, promote transfer, and scientific discourse
• Timely information that teachers can use to tailor and adjust instruction
Can you do better than Dr. A? Design three trials so that both the shrimp and alewife populations survive for 20 years.

- Use the sliders to change the starting numbers of shrimp and alewife.
- Click RUN to see what happens.
- When you have saved 3 trials in which shrimp and alewife survive for 20 years, click NEXT.

In the highlighted trials, all three organisms did not survive for 20 years. Here are starting values that will allow all the organisms to survive 20 years:

- shrimp 15, alewife 15
- shrimp 40, alewife 15
- shrimp 50, alewife 20

Use one pair of starting values for each highlighted trial. Then click RUN.
# Progress Reports to Students (Formative)

**Report for Mountain Lake - Predator Prey**  
*life science*  
Completed on 03/23/2010/Sara

<table>
<thead>
<tr>
<th>Category</th>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Populations</td>
<td>ON TRACK</td>
<td>Interactions between organisms and between organisms and the ecosystem’s nonliving features cause the populations of the different organisms to change over time.</td>
</tr>
<tr>
<td>Conduct</td>
<td>ON TRACK</td>
<td>Conducting investigations involves carrying out scientific investigations using appropriate tools and techniques.</td>
</tr>
<tr>
<td>Identify</td>
<td>NEEDS HELP</td>
<td>Identifying Science Principles focuses on students’ ability to recognize, recall, define, relate, and represent basic science principles. The practices assessed in this category draw on declarative knowledge or “knowing that.”</td>
</tr>
<tr>
<td>Design</td>
<td>NEEDS HELP</td>
<td>Designing investigations involves asking questions, planning investigations and evaluating experimental design.</td>
</tr>
<tr>
<td>Analyze</td>
<td>PROGRESSING</td>
<td>Identifying patterns involves summarizing patterns in data, analyzing which data are relevant and drawing conclusions by relating patterns in data to theoretical models.</td>
</tr>
</tbody>
</table>
## Progress Reports to Teachers (Formative)

### Summary Report: Mountain Lake - Food Web

**ASSESSMENT**: Mountain Lake - Food Web  **CLASS**: Period 7

<table>
<thead>
<tr>
<th>Content</th>
<th>NH Needs Help</th>
<th>P Making Progress</th>
<th>OT On Track</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roles</td>
<td>12 (46%)</td>
<td>4 (15%)</td>
<td>10 (38%)</td>
</tr>
<tr>
<td>Interactions</td>
<td>15 (58%)</td>
<td>4 (15%)</td>
<td>7 (27%)</td>
</tr>
<tr>
<td>Inquiry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifying</td>
<td>15 (58%)</td>
<td>5 (19%)</td>
<td>6 (23%)</td>
</tr>
<tr>
<td>Using</td>
<td>10 (38%)</td>
<td>5 (19%)</td>
<td>11 (42%)</td>
</tr>
</tbody>
</table>

**NH** = needs help  **P** = making progress  **OT** = on track
Grouping Recommendations for Classroom Reflection Activity (Formative)

<table>
<thead>
<tr>
<th>Student</th>
<th>Refl Gr.</th>
<th>Roles</th>
<th>Interactions</th>
<th>Identifying</th>
<th>Using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>C</td>
<td>P</td>
<td>NH</td>
<td>NH</td>
<td>OT</td>
</tr>
<tr>
<td>Student 1</td>
<td>C</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
</tr>
<tr>
<td>Student 3</td>
<td>A</td>
<td>OT</td>
<td>OT</td>
<td>OT</td>
<td>OT</td>
</tr>
<tr>
<td>Student 4</td>
<td>A</td>
<td>OT</td>
<td>OT</td>
<td>OT</td>
<td>OT</td>
</tr>
<tr>
<td>Student 5</td>
<td>C</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
</tr>
<tr>
<td>Student 6</td>
<td>C</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
<td>P</td>
</tr>
<tr>
<td>Student 7</td>
<td>C</td>
<td>P</td>
<td>NH</td>
<td>NH</td>
<td>P</td>
</tr>
<tr>
<td>Student 8</td>
<td>C</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
<td>NH</td>
</tr>
<tr>
<td>Student 9</td>
<td>C</td>
<td>NH</td>
<td>OT</td>
<td>NH</td>
<td>P</td>
</tr>
<tr>
<td>Student 10</td>
<td>B</td>
<td>OT</td>
<td>NH</td>
<td>OT</td>
<td>P</td>
</tr>
</tbody>
</table>

Group A students needed little help on either roles or interactions.
Group B students needed help with interactions, but not with roles.
Group C students needed help with understanding the roles of organisms in an ecosystem.
Classroom Reflection Activity (Adjusting Instruction)

- Formative use of assessment results
  - Students assigned to teams based on embedded results
- Transfer to different, more complex system
- Jigsaw structure
  - Allows differentiated instruction via tasks of varying difficulty
  - Promotes small and large group discourse and collaboration
- Guidance for teacher
  - Teacher review of key points in simulation
  - What to look for during group work and questions to pose in response
  - Posters and presentations
  - Evaluation of posters and presentations by students and teachers
Transfer to New, More Complex Ecosystem
Calipers II
Reflection Activities: Ecosystems
Ecosystem Benchmark Assessment: (Summative) Assess Transfer to New Ecosystem

Make a food web diagram. Draw arrows to show the transfer of matter between organisms.
Be sure to include each organism in the food web.

- To draw an arrow, click and drag from one dot to another dot.
- To delete an arrow, double click on it.

You can review the animation and then return to this diagram.

8 of 26
Summary Benchmark Report (Summative)

Matter and energy flow through the ecosystem as individual organisms interact with each other. Food web diagrams indicate the feeding relationships among organisms in an ecosystem. All ecosystems have a flow of energy from a nonliving source, to producers, to consumers.

**Detailed Report by Student and Target**

<table>
<thead>
<tr>
<th>Student</th>
<th>Roles</th>
<th>Interactions</th>
<th>Populations</th>
<th>Identify</th>
<th>Use</th>
<th>Design</th>
<th>Conduct</th>
<th>Analyze</th>
<th>Evaluate</th>
<th>Communicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simmons85, Sara85</td>
<td>BB</td>
<td>BB</td>
<td>A</td>
<td>BB</td>
<td>BB</td>
<td>BB</td>
<td>B</td>
<td>BB</td>
<td>BB</td>
<td>A</td>
</tr>
<tr>
<td>Simmons86, Sara86</td>
<td>BB</td>
<td>BB</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>BB</td>
<td>BB</td>
<td>A</td>
</tr>
<tr>
<td>Simmons87, Sara87</td>
<td>BB</td>
<td>BB</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>BB</td>
<td>BB</td>
<td>B</td>
</tr>
</tbody>
</table>
Calipers II Data and Findings

- AAAS review of alignment of content and inquiry targets with national and state standards
- Cognitive labs
- Classroom feasibility testing
- Pilot testing
- Field testing
  - 4 states, 28 districts, 58 teachers, 6,000 students
- Findings
  - Technical quality
  - Implementation evaluation
  - Effects of embedded on summative simulation benchmark and conventional posttest
<table>
<thead>
<tr>
<th>EAG State Design Panel Partners</th>
</tr>
</thead>
</table>

**Pilot states**
- Nevada (lead state)
- North Carolina
- Utah

**Advisory states**
- Massachusetts
- Connecticut
- Vermont
CRESST Evaluation Conclusions

- Teachers were able to implement as intended.
- Observations found that students were actively engaged.
- Teachers and students generally believed that the SimScientists program was beneficial to learning.
- Teachers found the automatically scored, immediate feedback—especially the reports generated by the questions—helpful to students. The instant reports allowed teachers to easily see which questions students had the most difficulty with so that they could tailor their lessons accordingly.
CRESST Evaluation Conclusions

• Teachers collectively agreed that the simulation-based assessments had greater benefits than traditional paper-and-pencil tests because of the instant feedback, interaction, and visuals.

• Teachers agreed that the assessments would be useful in measuring their individual state standards.
Analyses of Performance

- Compare performance on simulation-based assessments to traditional assessments
- Compare performance for all students, for English Learners, and for students with disabilities
Correlations Between Posttest and Benchmark Ability Estimates

<table>
<thead>
<tr>
<th>Topic</th>
<th>Content</th>
<th>Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystems</td>
<td>0.64</td>
<td>0.57</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>0.61</td>
<td>0.60</td>
</tr>
</tbody>
</table>
## Correlations Between Content and Inquiry Ability Estimates for Each Assessment

<table>
<thead>
<tr>
<th>Topic</th>
<th>Posttest</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystems</td>
<td>0.85</td>
<td>0.70</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>0.92</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Gaps in Total Performance Between English Learners and the General Population

<table>
<thead>
<tr>
<th>Topic</th>
<th>Posttest</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystems</td>
<td>24.0%</td>
<td>10.6%</td>
</tr>
<tr>
<td></td>
<td>(n=123)</td>
<td>(n=126)</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>27.4%</td>
<td>13.6%</td>
</tr>
<tr>
<td></td>
<td>(n=50)</td>
<td>(n=50)</td>
</tr>
</tbody>
</table>
# Gaps in Inquiry Skills Performance Between English Learners and the General Population

<table>
<thead>
<tr>
<th>Topic</th>
<th>Posttest</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.6%</td>
<td>6.6%</td>
</tr>
<tr>
<td>Ecosystems</td>
<td>(n=123)</td>
<td>(n=126)</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>35.1%</td>
<td>10.9%</td>
</tr>
<tr>
<td></td>
<td>(n=50)</td>
<td>(n=50)</td>
</tr>
</tbody>
</table>
## Gaps in Total Performance Between students with Disabilities and the General Population

<table>
<thead>
<tr>
<th>Topic</th>
<th>Posttest</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystems</td>
<td>25.5%</td>
<td>5.6%</td>
</tr>
<tr>
<td></td>
<td>(n=183)</td>
<td>(n=189)</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>20.3%</td>
<td>6.2%</td>
</tr>
<tr>
<td></td>
<td>(n=153)</td>
<td>(n=153)</td>
</tr>
</tbody>
</table>
## Gaps in Inquiry Skills Performance Between Students with Disabilities and the General Population

<table>
<thead>
<tr>
<th>Topic</th>
<th>Posttest</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystems</td>
<td>20.2% (n=183)</td>
<td>8.4% (n=189)</td>
</tr>
<tr>
<td>Force and Motion</td>
<td>15.7% (n=153)</td>
<td>7.0% (n=153)</td>
</tr>
</tbody>
</table>
Calipers II Field Testing (in progress)

• To examine effects of the curriculum-embedded assessments on learning

• Randomized control and treatment groups
  – With and without curriculum-embedded assessments
  – Partial results - Ecosystems
    • 763 students, 5 teachers (full sample 21 T, ~2400 students)
    • Effect size on pre-post conventional AAAS items 0.19
    • Effect size simulation based benchmark 0.43
  – Full field test in progress for Ecosystems and Atoms and Molecules
Balanced, Multilevel Assessment System Models

- Reporting benchmark results alongside district and state data
- Matrix sampling of short “signature” tasks from different topics
Side-by-Side Model

**STATE ASSESSMENT**
- SCIENCE
- NATURE OF SCIENCE
- EARTH SCIENCE
- LIFE SCIENCE
- PHYSICAL SCIENCE

**DISTRICT ASSESSMENT**
- SCIENCE
- NATURE OF SCIENCE
- INQUIRY
- COMMUNICATION
- EARTH SCIENCE
- EARTH'S CLIMATE
- EARTH'S STRUCTURE
- LIFE SCIENCE
- ECOSYSTEMS
- CELLS
- HEREDITY
- PHYSICAL SCIENCE
- FORCE AND MOTION
- ENERGY
- MATTER

Legend:
- Below Basic
- Basic
- Proficient
- Advanced
Signature Task Model

State Test Forms

Matrix Sampling

Simulation-based task item bank

Specifications and Simulation environments

Simulation-Based Classroom Assessments
Current Findings

The Calipers II simulation-based assessments

- Measure constructs not tested well by static modalities
- Can provide separate measures of inquiry and content
- The curriculum-embedded assessments seem to have positive effects on student learning
- The summative benchmark assessments have sufficient technical quality to be components of a state science assessment reporting system
Challenges

• Scheduling computer access
• Convincing psychometricians
• Progress from prototypes to developing grade-band suites
• Replications of documentation of technical quality
• Delivery across multiple technical infrastructures
• Models for scaling
• Models for sustainability
Next Steps

• Study vertically aligned simulation based assessment suites of
  – Classroom assessments
    • curriculum embedded assessments (for formative purposes)
    • benchmark assessments (for summative purposes)
  – Accountability assessments
    • signature tasks (for summative purposes)
Potential Collaborations

- Use of developed Calipers assessments as one of your external outcome measures
- Development of assessments embedded within your programs
- Co-development of new simulation-based learning and assessments
Contact Information

Website:

www.simscientists.org

Email:

equellm@wested.org