Calipers II: Using Simulations to Assess Complex Science Learning

Edys Quellmalz, PI; Co-PIs: Barbara Buckley and Mark Loveland – WestEd; Michael Timms- Australian Center for Educational Research
Partners: George DeBoer (AAAS) and Joan Herman (CRESST)

GOALS

1. Design and develop simulation-based formative and benchmark assessments of core ideas and inquiry practices for physical, life, and earth science.
2. Develop formative assessment simulation modules with immediate, individualized feedback and graduated coaching followed by offline Reflection Activities.
3. Provide evidence of the technical quality, feasibility, and usability of the simulation-based assessments.
4. Study the effects of formative assessments on complex science learning and inquiry practices.
5. Align the Calipers II benchmark and formative, embedded assessments to national science standards and the AAAS item bank.

PRODUCTS

Simulation-based embedded assessments and benchmark assessments for middle school:

- Force & Motion
- Atoms & Molecules
- Ecosystems

Assessment designs for:

- Plate Tectonics
- Climate

RESEARCH AND EVALUATION QUESTIONS

1. What impact does use of Calipers II embedded assessments have on student learning?
2. To what extent does evidence collected from the embedded and benchmark assessments support inferences about a student's proficiency on each standard?
3. To what extent does use of feedback and coaching in the embedded assessments relate to performance on the benchmark and external posttest?
4. Are the simulation-based assessments feasible for implementation across a range of classrooms and technical infrastructures?
5. Do teachers and students consider the simulation-based assessments useful for monitoring learning and summarizing proficiency?
6. Do the professional development (PD) strategies support teachers in their selection, administration, interpretation and use of the classroom-embedded and benchmark assessments?

KEY FEATURES

Model-based learning

Evidence-Centered Assessment Design

Simulations of age appropriate science system models

- Multiple representations
- Active inquiry

Simulation-based, curriculum-embedded assessments for formative use

- Immediate, individualized feedback and graduated coaching
- Reflection activities for transfer, collaboration, discourse

Simulation-based unit benchmark assessment for summative proficiencies

CRESST EXTERNAL EVALUATION

Review progress, designs, instruments, findings

Conduct case studies of classroom implementation

- Classroom observations
- Cognitive labs
- Teacher interviews

simscientists.org

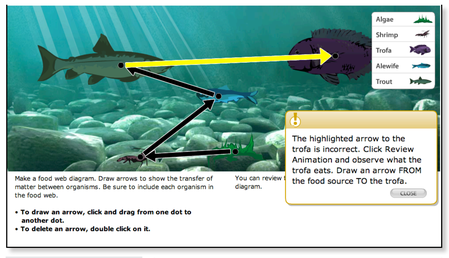
CALIPERS ASSESSMENT SYSTEM: CURRICULUM-EMBEDDED AND UNIT BENCHMARK ASSESSMENTS

Regular Instruction → Embedded Assessment + Reflection Activity → Regular Instruction → Embedded Assessment + Reflection Activity → Regular Instruction → Embedded Assessment + Reflection Activity → Regular Instruction → Benchmark Assessment

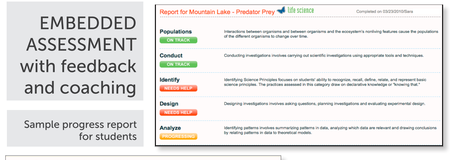
ECOSYSTEMS

System Model for Middle School Ecosystems

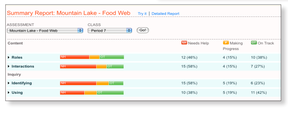
SYSTEM MODEL LEVELS	MODEL LEVEL DESCRIPTIONS	CONTENT TARGETS BY MODEL LEVEL	INQUIRY TARGETS
COMPONENTS	What are the components of the system and their rules of behavior?	Every ecosystem has a similar pattern of organization with respect to the roles (producers, consumers, and decomposers) that organisms play in the movement of energy and matter through the system.	Use principles to identify role of organisms.
INTERACTIONS	How do the individual components interact?	Matter and energy flow through the ecosystem as individual organisms participate in feeding relationships within an ecosystem.	Observe interactions among organisms.
EMERGENT BEHAVIORS	What is the overall behavior or property of the system that results from many interactions following specific rules?	Interactions among organisms and the ecosystem's nonliving features cause the populations of the different organisms to change over time.	predict observe explain investigate



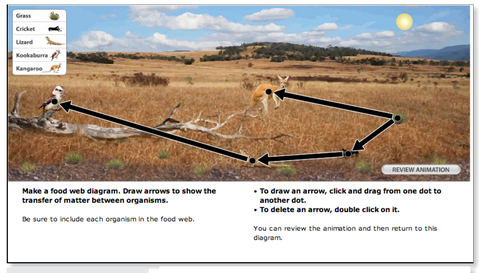
EMBEDDED ASSESSMENT with feedback and coaching



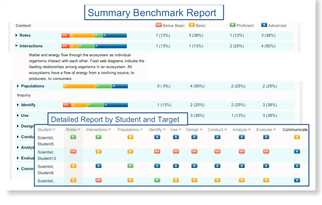
Sample progress report for students



Sample class summary report for teacher



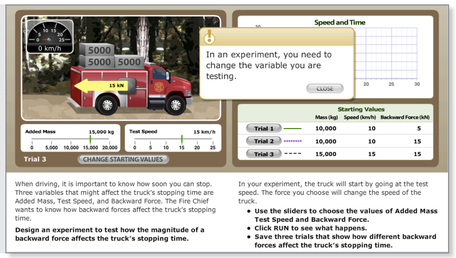
BENCHMARK ASSESSMENT: Draw food web in grasslands ecosystems



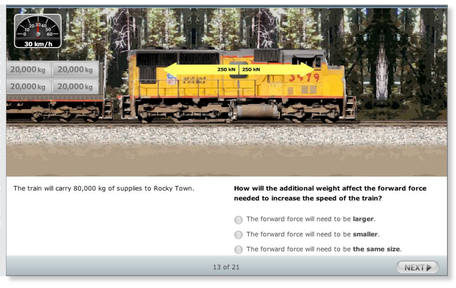
Sample benchmark report for teacher for class and by student

FORCE AND MOTION

EMBEDDED ASSESSMENT

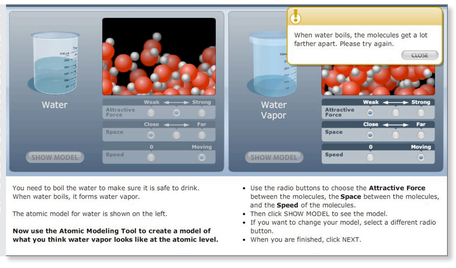


BENCHMARK ASSESSMENT

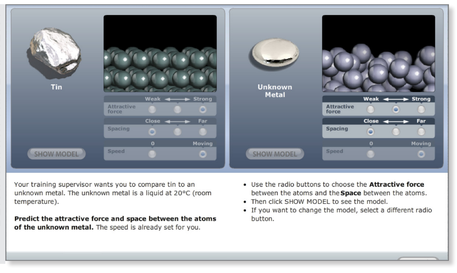


ATOMS AND MOLECULES

EMBEDDED ASSESSMENT



BENCHMARK ASSESSMENT



This material is based upon work supported by the National Science Foundation under grant # 0733345 awarded to WestEd and by the US Department of Education under OESE Grant# 09-2713-126 awarded to the Nevada Department of Education, Richard Vineyard, Principal Investigator. 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, the US Department of Education, or the Nevada Department of Education.

VALIDATION OF ASSESSMENT SUITES

External Review (AAAS)	Data collected	Use of Data	Results
	Ratings of • Alignment with standards • Scientific accuracy • Item quality • Grade level appropriateness	Documentation	Revisions as needed
Think-Alouds N	**Data collected**	**Use of Data**	**Results**
ECO 5 students 2 teachers F&M 5 students 2 teachers A&M 2 students 1 teacher	Audio of think-aloud Screen capture of actions Researcher notes	Usability of interface Accessibility of content Construct validity	Revisions as needed Tasks elicit targeted performances for 84% students.
Classroom Feasibility	**Data collected**	**Use of Data**	**Results**
ECO 125 students 1 teacher F&M 33 students 1 teacher	LMS data: actions & answers Cognitive labs Classroom observations Teacher surveys Teacher interviews	Usability in classrooms Participation patterns Engagement Instructional utility	Technical improvements (bandwidth, loading times) Revision of reflection activities Revision of teacher materials
Pilot Test	**Data collected**	**Use of Data**	**Results**
ECO 3529 students F&M 1036 students A&M 253 students	All of the above + External post test	All of the above + Technical Quality	ECO Benchmark IRT reliability = 0.76 F&M Benchmark IRT reliability = 0.73 A&M Benchmark IRT reliability = 0.82 77% of students were 'highly engaged'

FIELD TEST - IN PROGRESS

Impact of Formative Assessment

Randomized Control Trial

- Treatment includes simulation-based formative assessments.
- Control does not include simulation-based formative assessments.
- Each teacher's classes were randomly assigned to treatment or control group.

Participants currently enrolled

- Ecosystems: ~21 teachers, ~2,400 students
- Atoms & Molecules: ~10 teachers, ~1,500 students

Methods

- LMS data – students' answers and actions
- Multi-dimensional IRT analyses
- HLM analyses to determine effect size

Partial Sample Results for Ecosystems

- Participants: 763 middle school students, 5 teachers
- Treatment groups outperformed Control groups
  - learning gains = Post – Pre (effect size 0.19)
  - benchmark assessment (effect size 0.43)
- Holding pre test constant, mid-level students (pre test) outperformed low and high students on the benchmark (effect size 0.52).

CONCLUSIONS

Calipers II assessments are:

- feasible to implement on a large scale in a range of settings and technical infrastructures,
- useful for formative purposes to monitor progress and adjust instruction, and
- of sufficient technical quality to serve as credible components of multi-level state assessment systems.

NEXT STEPS: ANALYSIS AND DISSEMINATION

IRT analyses of assessment data

HLM analyses to determine effect size

Triangulation across data sources to determine technical quality, usability, and feasibility

External evaluation including classroom case studies