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 benchmarks and formative, embedded assessments to national science standards and the AAAS Item Bank.

PRODUCTS

Simulation-based assessments and benchmark assessments for middle school:
- Force & Motion
- Atoms & Molecules
- Ecosystems

Assessment design 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
Simulations-based, curriculum-embedded assessments for formative use
- Immediate, individualized feedback and graduated coaching
- Reflection activities for transfer, collaboration, discussion
Simulation-based unit benchmark assessment for summative use

CREST EXTENSIONAL EVALUATION

Screen capture of CALIPERS II System Model for Middle School Ecosystems

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 class was randomly assigned to treatment or control group.

Participants
- Currently enrolled
- Ecosystems: ~21 teachers, ~2,400 students
- Atoms & Molecules: ~10 teachers, ~500 students

Methods
- LMS data: actions & answers
- Technical quality: USM, IRT
- Construct validity: simulation, reflective activities
- Content validity: curriculum, standards
- Usability in interface: reflection activities, teacher interviews

Analysis
- All of the above
- Technical Quality
- IRT Benchmark IRT reliability = 0.76

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;
- 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
Translation across data sources to determine technical quality, usability, and feasibility
External evaluation including classroom case studies