CRESST REPORT 821

EVALUATION OF CALIPERS II: USING SIMULATIONS TO ASSESS COMPLEX LEARNING- SITE VISIT FINDINGS

SEPTEMBER, 2012

Deborah La Torre Matrundola
Sandy Chang
Joan Herman

National Center for Research
on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

Evaluation of Calipers II: Using Simulations to Assess Complex Learning
Site Visit Findings

CRESST Report 821

Deborah La Torre Matrundola, Sandy Chang, and Joan Herman
CRESST/University of California, Los Angeles

September 2012

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by WestEd, subcontract #5448 S08-069, pursuant to WestEd's grant with the National Science Foundation, Grant #DRL-0733345, CFDA #84.369.

To cite from this report, please use the following as your APA reference: La Torre Matrundola, D., Chang, S., & Herman, J. (2012). *Evaluation of Calipers II: Using simulations to assess complex learning- site visit findings* (CRESST Report 821). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

EVALUATION OF CALIPERS II: USING SIMULATIONS TO ASSESS COMPLEX SCIENCE LEARNING

Deborah La Torre Matrundola, Sandy Chang, and Joan Herman
CRESST/University of California, Los Angeles

Abstract

The purpose of these case studies was to examine the ways technology and professional development supported the use of the SimScientists assessment systems. Qualitative research methodology was used to provide narrative descriptions of six classes implementing simulation-based assessments for either the topic of Ecosystems or Atoms and Molecules. Results revealed both strengths and weaknesses concerning technology support for the assessments, as well as technology and professional development support of the teachers. Furthermore, recommendations are provided concerning potential improvements to the assessments, reflection activities, and professional development.

Introduction

SimScientists' simulation-based assessments are intended to support rich learning of major concepts and principles in science and to promote inquiry skills. The program provides suites of assessment activities, with each focused on a major topic in middle school science and aligned with national and state standards for content and inquiry. Benchmark assessments are designed to test end-of-unit achievement of the selected topic. Sets of embedded assessments are designed to be used during the course of the instructional unit. The embedded science assessments are intended to function as formative resources by providing immediate feedback contingent on an individual student's responses, offering graduated levels of coaching in real time, and providing diagnostic information to guide offline reflection and extension activities. Technology-based, each benchmark and embedded assessment is contextualized in a real life scenario, engaging students in dynamic, interactive tasks. Reflection activities draw on embedded assessment results to differentiate subsequent instruction and deepen transfer student knowledge to new contexts. These activities are teacher directed, based on SimScientists' plans and materials. They utilize small group collaboration and engage students in communication and presentation of science ideas.

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) was contracted by WestEd to conduct site visits to the classrooms of selected teachers. This report is designed to provide narrative descriptions of the implementation of the assessment systems. Specifically the following questions were addressed:

- How was the SimScientists' technology-based assessment system implemented in the case study classrooms?
- To what extent were SimScientists' reflection lessons implemented as planned in support of science learning goals?
- What were teachers' reactions to the professional development and other supports for assessment practice?

Methodology

Qualitative research methodology was used to address the implementation evaluation questions. This included classroom observations and interviews with a convenience sample of six teachers who were involved in a larger pilot of the SimScientists assessments on the topics of (1) Ecosystems and (2) Atoms and Molecules. Site visits were conducted to observe the SimScientists activities in action and interview teachers about their reactions to the program. An examination of the professional development materials and activities used to prepare teachers for using the assessment system was also conducted.

Sample

The study sample included six teachers in three schools in Nevada: three sixth grade teachers who were implementing the Ecosystems assessment system and three eighth grade teachers who were implementing Atoms and Molecules. All Atoms and Molecules teachers were observed a total of seven times over the course of the study and Ecosystems teachers were observed five times. The exception involved T100 who had a last minute scheduling change for one of her assessments. The following table shows the specific components observed for each teacher.

Table 1

Components Observed for the Six Teachers in the Study Sample

	Ecosystems			Atoms & Molecules		
Component	T100	T101	T200	T300	T301	T302
Assessments						
Embedded 1	X	X	X	X	X	X
Embedded 2	X	X	X	X	X	X
Embedded 3	n/a	n/a	n/a	X	X	X
Benchmark		X	X	X	X	X
Reflection Activities						
RA1	X	X	X	X	X	X
RA2	X	X	X	X	X	X
RA3	n/a	n/a	n/a	X	X	X

Instrumentation and Procedures

Instruments were adapted by WestEd from protocols originally developed by CRESST and WestEd for use in the evaluation of an Enhanced Assessment Grant in which simulation-based assessment sites were implemented (see Herman, Dai, Htut, Martinez, & Rivera, 2011). The following describes the protocols used during the assessment observations, reflection activity observations, and teacher interviews.

Classroom observations. Separate observation protocols were used to examine the implementation of the online assessments and of the reflection activities. Each protocol included a section for the observers to record general observations on key themes, as well as structured checklists to record classroom organization and interactions at five-minute intervals. More specifically, each protocol had checklists for teacher role, the quality of the interactions among the students when working in pairs or groups, and student engagement. The protocol for the assessments also included a checklist for recording technology and other incidents while the protocol for the reflection activities also included class organization and classroom interaction.

Each observation took place over one class session lasting approximately 50 minutes. Observations were conducted from mid-March to late May during 2012. Three CRESST researchers were trained on the protocols and activities by WestEd prior to the site visits. In order to develop reliability, multiple researchers participated in each observation during the first round of site visits. Researchers then met to debrief and come to consensus.

Teacher interviews. Structured interviews were conducted with each teacher following her completion of the assessments. Questions focused on the comfort and challenges faced by the teachers and students, perceived strengths and weaknesses of the assessments in comparison to traditional paper-pencil assessments, and usefulness of the assessment reports. In addition, teachers were asked about their professional development experiences.

Analysis

Completed observation checklists were entered into SPSS. Observation checklists for the assessments and reflection activities were analyzed separately using descriptive statistics. First, overall frequencies for each theme were calculated in order to provide a sense of how the different teachers implemented the activities. Second, results were analyzed by five-minute intervals in order to provide a sense of how the SimScientists activities unfolded.

Interviews were taped using digital audio recorders, transcribed, and analyzed using Microsoft Word. Data were categorized according to the constructs identified in the research questions. Results were then compiled to identify overall themes as well as emergent themes by group (i.e., Ecosystems teachers and Atoms and Molecules teachers).

The next two sections summarize results for the three research questions. The concluding section will synthesize results across data sources and provide implications for next steps.

Implementation of the SimScientists' Assessments

One of the goals of the study was to determine the ways in which SimScientists' embedded and benchmarks assessments for Ecosystems and for Atoms and Molecules were implemented in case study classrooms. This section presents observation results for the assessments followed by interview results for the six case study teachers.

Observation Results for Delivery of the Assessments

This section presents observation results concerning the delivery of the assessments by the six case-study teachers. For ease of interpretation and because results were similar, observation data from the embedded and benchmark results are combined.

Teacher's role during the assessments. During the simulation-based assessments, the teachers tended to maintain some level of interaction with their students (see Table 2). In most cases, this involved the teachers assisting their students as needed or in a systematic manner where they went around to every student in the computer lab. These roles were most predominant for T101 and least predominant for T302, who tended to take a hands-off approach to the assessment administrations. When examining the results further, the sixth grade Ecosystems teachers were the least likely and the eighth grade Atoms and Molecules teachers were the most likely to be observed taking a hands-off approach.

Table 2

Teacher Role by Teacher and Focus (Assessments)

Teacher ID	*N*	Addressing whole class	Assisting systematically	Assisting as needed	Monitoring administration	Teacher not involved
Ecosystems						
T100	3	12.2%	11.1%	37.6%	23.3%	15.9%
T101	2	5.0%	8.3%	76.7%	10.0%	0.0%
T200	3	14.8%	16.7%	38.9%	29.6%	0.0%
Total	8	11.4%	12.5%	47.8%	22.3%	6.0%
Atoms & Molecules						
T300	4	9.8%	10.0%	59.7%	20.6%	0.0%
T301	4	2.5%	7.0%	37.5%	35.5%	17.5%
T302	4	2.8%	12.5%	28.1%	8.7%	47.9%
Total	12	5.0%	9.8%	41.8%	21.6%	21.8%
SimScientists Total	20	7.6%	10.9%	44.2%	21.9%	15.5%

When examining the overall SimScientists results by segment, teacher interaction with the students tended to lessen over time (see Figure 1). During the beginning of the assessments, teachers often provided instructions to their whole class. Once their introductions were complete, teachers tended to spend the next few segments helping their students get started on the assessments. Although students were expected to remember their usernames and passwords, many of them had trouble remembering these. In some cases, teachers developed techniques to help students overcome this difficulty, such as putting the login information on cutout strips of paper or using a naming convention for the usernames and passwords and putting this on the overhead at the start of each class. In some instances teachers also resorted to resetting passwords for certain students. Not surprisingly, as larger numbers of students finished their assessments (see Figure 3 on student engagement), teachers were more likely to monitor their class by walking around, standing observantly near the front of the computer lab, or checking student progress on the teacher interface.

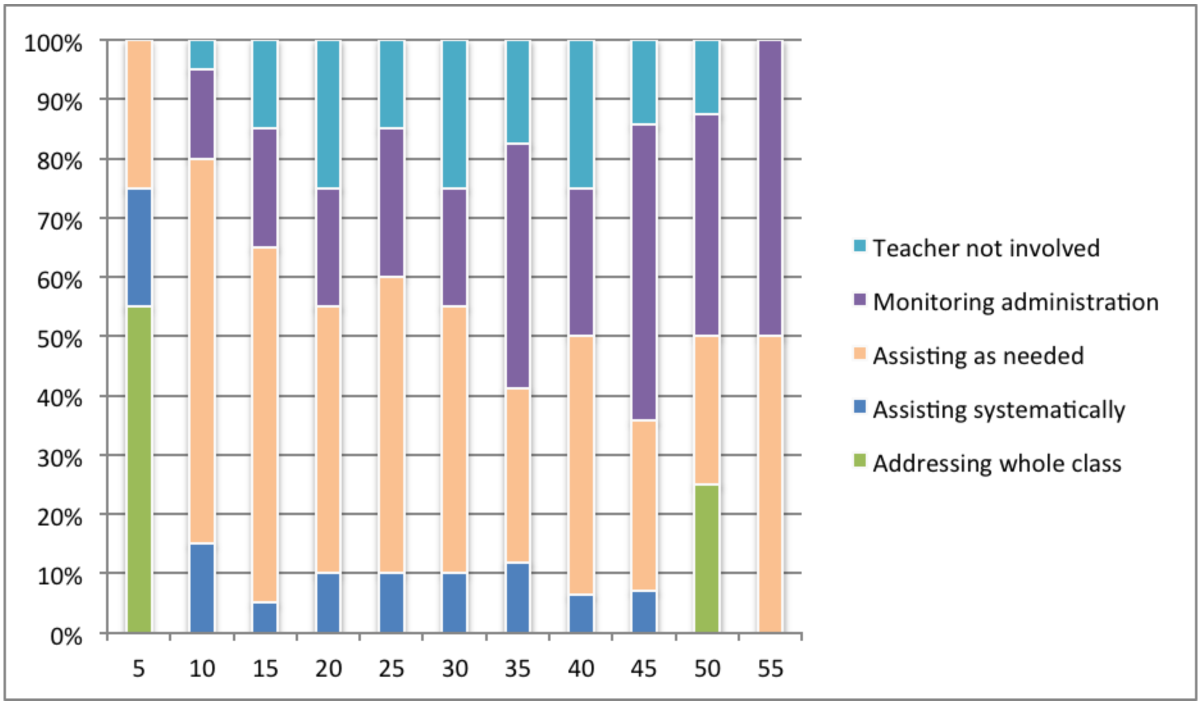

Figure 1. Teacher role by five-minute segment (assessments).

Technology. All of the assessments observed took place in large, dedicated computer labs at the schools. In these labs, Windows computers were laid out along three or four of the walls, with student desks in the center. During each of the administrations, individual computers were available for each of the students, as well as the teachers. As shown in Table 3, the average number of students present during the administrations tended to be in the low to mid-twenties for the Ecosystems teachers and in the mid-twenties to low thirties for the Atoms & Molecules teachers.

Table 3

Classroom Composition by Teacher and Focus (Assessments)

Teacher ID	*n*	Average male students per period	Average female students per period	Average total students per period
Ecosystems				
T100	3	10.33 (0.58)	14.33 (2.31)	24.67 (2.89)
T101	2	16.50 (2.12)	10.50 (6.36)	27.00 (4.24)
T200	3	11.33 (0.58)	11.00 (0.00)	22.33 (0.58)
Total	8	12.25 (2.82)	12.13 (3.27)	24.38 (2.97)
Atoms & Molecules				
T300	4	9.00 (0.82)	21.25 (1.26)	30.25 (1.71)
T301	4	14.75 (0.50)	11.75 (0.50)	26.50 (1.00)
T302	4	13.25 (1.50)	15.75 (2.22)	29.00 (2.94)
Total	12	12.33 (2.71)	16.25 (4.29)	28.58 (2.47)
SimScientists Total	20	12.30 (2.68)	14.60 (4.35)	26.90 (3.35)

When examining results by teacher, technology incidents occurred less than one-third of the time during the assessment observations (see Table 4). The exception involved T101 whose class had login failures about half of the time. Other minor issues observed included waiting for pages to load or having to reload an assessment. It should also be noted that network failures, power failures, and computers crashing/freezing were not observed.

Table 4

Technology Incidents by Teacher and Focus (Assessments)

Teacher ID	*n*	None	Login failed	Waiting for pages to load	Reloading assessment	Other
Ecosystems						
T100	3	76.7%	23.3%	0.0%	0.0%	0.0%
T101	2	46.7%	53.3%	0.0%	0.0%	0.0%
T200	3	77.8%	16.7%	0.0%	0.0%	5.6%
Total	8	69.6%	28.3%	0.0%	0.0%	2.1%
Atoms & Molecules						
T300	4	67.1%	30.4%	2.5%	0.0%	0.0%
T301	4	70.7%	22.0%	0.0%	2.3%	5.0%
T302	4	76.7%	23.3%	0.0%	0.0%	0.0%
Total	12	71.5%	25.3%	0.8%	0.8%	1.7%
SimScientists Total	20	70.7%	26.5%	0.5%	0.5%	1.8%

Note. Network failure, power failure, and computer crash/freeze were excluded from the table since no incidents were observed.

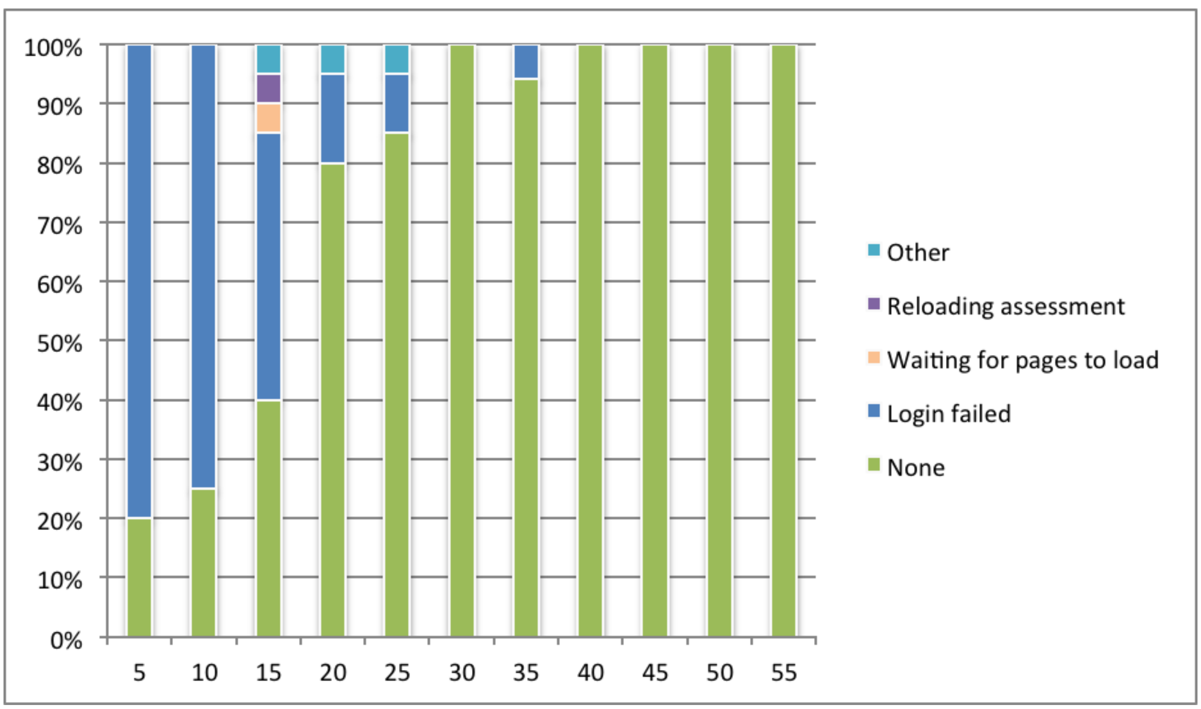

Figure 2. Technology incidents by five-minute segment (assessments).

As previously mentioned, most of the technology incidents occurred early in class when the students were logging in and beginning their assessments (see Figure 2). For example, while incidents were observed over 60% of the time during the first three segments, by the fourth segment they were observed only 20% of the time. Furthermore, the only incidents observed during the second half of the observations occurred when students had trouble logging on after arriving late to class.

Student engagement and interaction. Over the course of the observations, over two-thirds of the students were actively engaged with the assessments (see Table 5). To be considered active, students needed to work actively with the simulations, respond to the questions in a purposeful manner, and/or discuss the assessment with their neighboring classmates. Most of the remaining students passively attended to the assessments. Some of the behaviors coded as passive included students failing to use the interactivity unless required to move forward in the assessment, students answering questions without reading the directions, and/or students randomly picking multiple-choice answers. When looking at the engagement results by teacher, more variation was found for the Atoms and Molecules assessments than for the Ecosystems assessments. For example, even though T300 and T302 had similar class sizes, nearly 25% more of the students were highly engaged in the latter teacher's class. In contrast, active engagement across Ecosystems teachers was more consistent and differed by 4.1% or less.

Table 5

Student Engagement by Teacher and Focus (Assessments)

Teacher ID	*n*	Active engagement	Passive engagement	Off task
Ecosystems				
T100	3	63.4%	33.7%	2.9%
T101	2	67.5%	29.7%	2.8%
T200	3	66.0%	34.0%	0.0%
Total	8	65.4%	32.8%	1.8%
Atoms & Molecules				
T300	4	59.9%	33.2%	6.9%
T301	4	60.4%	32.0%	7.6%
T302	4	84.0%	13.9%	2.1%
Total	12	68.1%	25.9%	6.1%
SimScientists Total	20	67.0%	29.0%	4.0%

As was observed with the technology incidents, passive engagement was predominant during the first segment of the observations (see Figure 3). It then dropped to less than one-third during the second segment and tended to stay at 10% or less once large percentages of students started completing their assessment. This change in engagement corresponded with students logging onto SimScientists, as well as the teachers giving students permission to start their assessments. Furthermore, similar to the results by teacher, less than 10% of the students were considered off task during any of the observation segments.

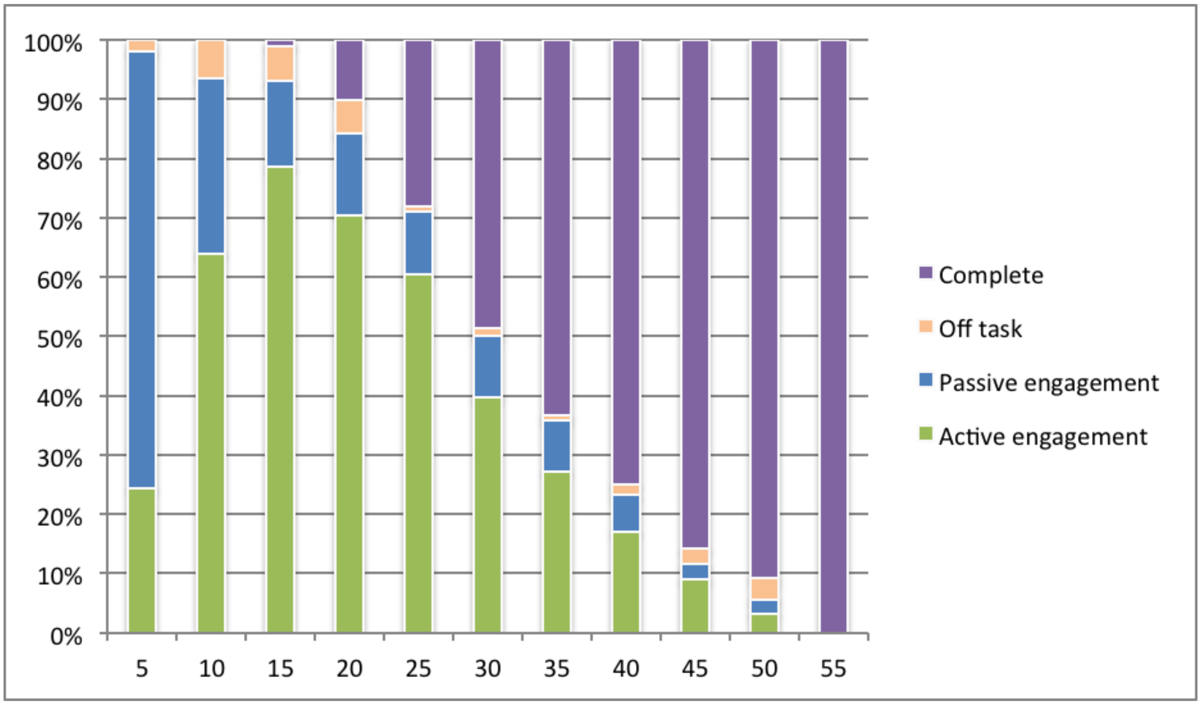

Figure 3. Student engagement by five-minute segment (assessments).

During the observations, there were some instances when one or more students failed to complete their assessment before the school bell rang (see Table 6). More specifically, this occurred during three of the Ecosystems observations and six of the Atoms and Molecules observations. The most common reason noted for this involved time, although other issues such as students having trouble logging onto SimScientists or students having second language issues were observed. In the latter case, students were encouraged by teachers to use Google translate or consult with their classmates, which affected both those who were struggling and those who were helping. The teachers with the most difficulty having their students finish on time included T101 and T301. In contrast, all of the students working with T200 were able to finish during their class period. Since all three of these teachers had good classroom management skills, it is unclear why this occurred.

Table 6

Incomplete Assessments by Teacher and Focus – Count of Sessions (Assessments)

Teacher ID	*n*	Sessions with student(s) incomplete	If incomplete, why		
			Not enough time	Technical issues	Other
Ecosystems					
T100	3	1	0	0	1
T101	2	2	1	1	2
T200	3	0	--	--	--
Total	8	3	1	1	3
Atoms & Molecules					
T300	4	2	2	1	1
T301	4	3	3	0	2
T302	4	1	1	0	0
Total	12	6	6	1	3
SimScientists Total	20	9	7	2	6

Note. Reasons incomplete are not mutually exclusive.

Assessment reports. Students' use of the embedded assessment progress reports varied depending upon their teacher (see Table 7). This behavior was most predominant during the sessions when T200 and T302 systematically moved around the computer lab, reviewing the reports with their students as they completed the assessment. In all other cases, students were given the option to review the report if they chose. With this autonomy, some students purposefully reviewed their reports on their own, while others glanced at theirs or did not open theirs at all. It is also interesting to note that observers did not witness any incidences of students reviewing their report with their peers.

Table 7

Student Report Use by Teacher and Focus - Count of Sessions (Assessments)

				If reviewed, how	
Teacher ID	*n*	Not accessed or used	Reviewed report	Alone	With teacher
Ecosystems					
T100	3	3	2	2	0
T101	2	2	0	--	--
T200	3	2	2	0	2
Total	8	7	4	2	2
Atoms & Molecules					
T300	4	4	2	2	0
T301	4	4	3	3	0
T302	4	4	4	3	1
Total	12	12	9	8	1
SimScientists Total	20	19	13	10	3

Note. Report use not mutually exclusive.

Interview Results

This section presents teachers' comments about the design and development of the assessments, their experience delivering the assessments, and their experience with scoring and interpreting the results.

Strengths. With respect to the design and development of the SimScientists assessment system, teachers reported that, in general, they and their students were comfortable interacting with the assessments. The visual aspects of the simulations, especially the "dynamic" animations versus "static" pictures found in other tests, helped convey concepts to students. Atoms and Molecule teachers noted that having a visual of "something they can't see" (T100) worked well because "it's really hard for the students to kind of grasp on to understand these concepts" (T302). One teacher (T302) particularly liked the narrative aspect of the simulations.

Teachers also appreciated the hands-on aspect of the simulations, such as the ability to manipulate data and the variety of response modes that were included. One teacher (T100) identified the graphing in the simulations as a strong feature. For the response modes, several teachers pointed out that although response modes were somewhat similar to those found on paper and pencil tests, having students respond on a computer allowed them to be “more interactive” (T101). Furthermore, some teachers felt that the hands-on features found in SimScientists allowed for higher cognitive levels of thinking compared to paper and pencil tests. Teacher T101 said, “I think by far the simulation-based assessments…are at a much higher cognitive level than at least the paper and pencil tests I tend to give because they require manipulation of data.”

Another design feature that seemed helpful to the students was the immediate feedback. One teacher (T300) noticed that she “saw a lot of activity” with students utilizing the feedback boxes. All teachers agreed that one major benefit of the computer simulations was increased engagement compared to paper and pencil tests.

Weaknesses. Teachers described some design issues that took away from student engagement and performance with the assessments. All three Atoms and Molecules teachers (T300, T301, and T302) shared similar complaints about the design of the simulations, whereas the Ecosystems teachers (T100, T101, and T200) had fewer complaints.

One issue for the Atoms and Molecules teachers involved the decreasing level of student engagement during progressive assessments. Teachers attributed students’ fatigue and loss of engagement to the length and repetitiveness of the simulations. One teacher (T300) noticed that students “were intrigued at first, but then that went away pretty fast.” Another teacher (T302) noted that the stories were:

> …Kind of the same thing each time, so maybe by the third time, the students got a little bored with it, the same types of questioning, the same types of boxes to type your answers…they all kind of look the same too, even though there’s a little bit different story line.

Another issue raised by the Atoms and Molecules teachers involved the difficulty of the assessments in relation to their students’ backgrounds. Two of the teachers (T300 and T302) noted that the “amount of reading” in the assessments was too high in comparison to the “reading level[s]” of their students. Another issue was the accessibility of the assessments for English language learners. One teacher (T301) stated that her English language learners had “a very hard time understanding English so then that’s a detriment to them in…how they process the information that’s on the assessment.” This teacher also noted that with the school’s student demographic, the students might not have the necessary background experiences to make the simulations comprehensible:

> You know, I've done activities in class, like roasting marshmallows, and…the majority of students have never roasted a marshmallow and so there's some simple experiences that they may not have that may be on those assessments…that may impede their performance. And I think some of the information that was in your guys' assessments...you know, using the robot to determine what gas was present in the sample and things of that nature, I think…they may not have much experience with that…

A couple of teachers noticed that a few students had difficulty with the interactive aspect of the simulations. That is, some students did not know which buttons to click in order to change variables or get the models to work. For example, one teacher (T101) saw that a few students did not notice all the interactivity, such as how the graph would light up as they moved a button. Another teacher (T301) saw some students struggling "to understand what they needed to manipulate, what variables need to be changed, or what the change in variables was actually producing on the screen."

Two Ecosystems teachers also noted that their students did not answer the short response questions in the simulations as thoroughly as they would have for a paper and pencil test. One teacher (T200) attributed the lack of thorough responses to students' desire "to go on to the next frame and see what was on the next frame."

Suggestions. The Atoms and Molecules teachers offered some thoughts as to how to improve the design and development of the computer simulations. One suggestion was to shorten the simulations. One teacher (T301) suggested reducing the number of questions and breaking the simulations down into smaller units. The other teachers (T300 and T302) suggested making the simulations appear "a little different" or have "something that was a little bit different from time to time," which "would have kept them more engaged" (T302). Lastly, one teacher (T300) questioned the need for the password system since students were monitored as they took the assessments.

Delivery of the assessments. Teachers encountered some issues with the delivery of the assessments. For the Atoms and Molecules teachers, it was difficult coordinating computer lab times with the scheduling of the SimScientists assessments. One teacher (T300) reported that she was not able to access the first embedded assessment, which caused "lots of ramifications" in terms of scheduling a computer lab. Another teacher (T301) discussed the inflexibility of scheduling the assessments, especially when there was a change in the school schedule. Teacher T301 added that it would have been easier with scheduling and coordination if the assessments were open for a longer period, such as having the assessment available for a week rather than just one day.

An issue raised by one teacher (T300) involved students not being able to go back to previous screens. Unlike a paper and pencil test where the students could flip back to previous pages, the SimScientists program did not allow students to be able to return to prior screens.

Lastly, there were a few logistical and technical issues encountered by the teachers and their students. Several teachers (T101, T300, T301, and T302) noted that some screens would freeze while students were taking the assessments. Students would have to exit the program and log back in, and for all but one teacher, the simulation would begin where the students left off. For one teacher (T301), the students had to start from the beginning of the assessment. Teachers also noted that students frequently forgot to use the correct browser (i.e., Internet Explorer). One pressing issue was that many students kept forgetting their passwords to log onto SimScientists, and resetting passwords took a lot of time.

Scoring of the assessments. For the benchmark assessments, teachers were required to score the open-ended student responses. Two Ecosystems teachers (T100 and T101) encountered usability problems when they failed to recognize that a checkmark at the top of the page indicated that all student responses were scored, and the lack of a checkmark meant that the scoring was incomplete. Furthermore, the scoring program allowed teachers to move to the next page despite not having every student response scored. When teachers wanted to access the score reports, they could not until all student responses were scored.

Interpretation of the benchmark reports. In general, teachers expressed that the benchmark reports gave them an idea of students' content and inquiry knowledge. One teacher (T100) was surprised at how well students did when asked about use of benchmark report to gauge scientific inquiry knowledge. Another teacher (T101) stated using the benchmark report to judge science inquiry and that the benchmark report "did a really good job" at helping her determine students' content knowledge. However, one teacher (T300) expressed some uncertainty to how the benchmark report compared to the tests she was used to giving and interpreting: "I'd like to go back and kind of retest using my tests on some of this material and do some comparisons of mine own because I'm just not sure."

When asked how their students performed on the benchmark assessment compared to teacher expectations, most teachers felt that students performed as expected. One teacher (T100) was excited that her students exceeded expectations, especially given how low her students were earlier in the year. However, another teacher (T200) was surprised that the accelerated students in one period performed worse than what she expected.

Lastly, teachers varied in their opinions on whether student performance on the SimScientists benchmark assessment would be similar to district and state assessments. The Atoms and Molecules teachers (T300, T301, and T302) thought that their students would perform comparably with the district and state assessments. One Ecosystems teacher (T100) noted that her students performed better on the benchmark assessment compared to the district assessment. In contrast, the other two Ecosystems teachers (T101 and T200) stated that the benchmark assessed areas that were not normally covered in their district assessments, notably inquiry, and therefore the two were not comparable.

Implementation of the SimScientists' Reflection Activities

The second goal of the study was to examine the ways that the assessment system supported teachers' use of formative assessment through reflection activities. This section presents an overview of the reflection activities, observation results for the reflection activities, followed by interview results on teachers reactions and use of SimScientists' reports for planning the reflection activities.

Overview of the Reflection Activities

Teachers who participated in SimScientists were provided with face-to-face training, detailed lesson plans, and materials for carrying out reflection activities following each embedded assessment, which were designed to build on students' current understanding and help them move to a deeper level of knowledge and inquiry. Each reflection activity included instructions on how to use the information provided in the student reports to assign students to appropriate follow-up activities. For the Atoms and Molecules lesson plans, the teachers were guided to assign specific activity components to their students based on whether they were classified in the A, B, C, and/or D groups. In contrast, the Ecosystems lesson plans had teachers use the report results to jigsaw students into different roles within each activity.

Each lesson plan was organized with an initial teacher review and introduction followed by at least one hands-on activity for students, group work where students create and present a poster, and a teacher-led wrap up. In addition, some of the lesson plans included a science demonstration for teachers to do in front of their class. In most, but not all cases, each lesson plan was designed to be carried out within one class session.

Observation Results for the Reflection Activities

Organization of the activities and roles of the teachers. During the reflection activities, the class organization used by the teachers varied depending upon the topic being taught (see Table 8). The sixth grade teachers using the Ecosystems unit made slightly greater use of introductions and student presentations than did their colleagues. In contrast, the eighth grade teachers using Atoms and Molecules were moderately more likely to have students work in pairs and/or trios and were much more likely to use whole class activities, such as teacher demonstrations and class discussion.

Additional differences were found when examining the results by teacher. Among the Atoms and Molecules teachers, T302 was fairly evenly split between having students work in pairs/trios or in small groups. In contrast, both T300 and T301 strongly favored having students work in pairs/trios. Similarly, T301 used a whole class organization nearly two-thirds of the time while the other two Atoms and Molecules teachers used this organization only about one-third of the time. Differences were less predominant among the Ecosystems teachers with the biggest difference involving T200 have students work on or do presentations over one-quarter of the time while none of the other teachers used this strategy during the reflection activity observations.

Table 8

Class Organization by Teacher and Focus (Reflection Activities)

Teacher ID	*n*	Teacher intro	Pairs or trios	Small group	Student presentations	Whole class	Students working alone
Ecosystems							
T100	2	31.7%	52.2%	30.0%	0.0%	21.1%	0.0%
T101	2	50.0%	50.0%	30.0%	0.0%	10.0%	0.0%
T200	2	33.3%	44.4%	27.8%	27.8%	5.6%	0.0%
Total	6	38.3%	48.9%	29.3%	9.3%	12.2%	0.0%
Atoms & Molecules							
T300	3	19.9%	68.1%	22.2%	0.0%	36.1%	14.8%
T301	3	30.0%	82.0%	33.3%	0.0%	62.4%	0.0%
T302	3	31.5%	44.4%	41.1%	0.0%	29.6%	0.0%
Total	9	27.1%	64.8%	32.2%	0.0%	42.7%	4.9%
SimScientists Total	15	31.6%	58.5%	31.0%	3.7%	30.5%	3.0%

Note: Items are not mutually exclusive.

As shown in Figure 4, teachers' use of some organizational features increased over time while their use of other organizational features decreased. For example, teachers were most likely to make introductions during the first ten minutes of the reflection activities, as specified in the lesson plans. In addition, whole class demonstrations or discussions lessened somewhat over time. While some of the teachers mainly addressed their whole class as specified in the lesson plans, others liked to bring up information on a regular basis. At the same time, teachers' use of student groups increased greatly over time, becoming the only organizational feature when observations lasted up to 50 minutes. It should also be noted that the one teacher who did use student presentations always started this activity after the 30-minute segment.

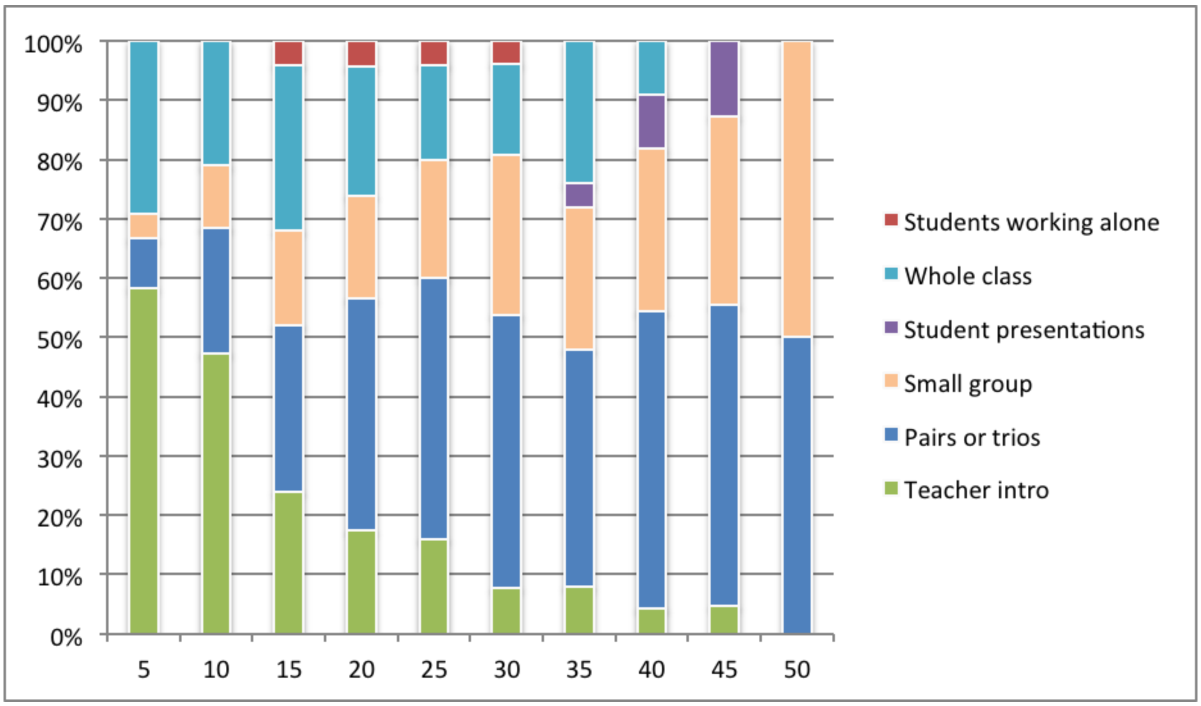

Figure 4. Class organization by five-minute segment (reflection activities).

During the reflection activities, the teachers' maintained high levels of interaction with their students (see Table 9). In general, over one-quarter of their time was spent presenting demonstrations to their whole class and about two-thirds of their time was spent assisting their students either systematically or as needed. Having a teacher monitoring their class was rare, with the exception of T300 who took on this role almost one-fifth of the time. In addition, only T302 was observed not being involved with her students while prepping activities or talking with her teaching assistant. It is also interesting to note that T101 and

T301 were primarily systematic with their assistance, while most of the other teachers primarily assisted their students as needed.

Table 9

Teacher Role by Teacher and Focus (Reflection Activities)

Teacher ID	*n*	Demo to whole class	Assisting systematically	Assisting as needed	Monitoring	Teacher not involved
Ecosystems						
T100	2	27.8%	25.6%	41.1%	5.6%	0.0%
T101	2	25.0%	60.0%	15.0%	0.0%	0.0%
T200	2	27.8%	11.1%	55.6%	5.6%	0.0%
Total	6	26.9%	32.2%	37.2%	3.7%	0.0%
Atoms & Molecules						
T300	3	27.3%	12.5%	43.5%	16.7%	0.0%
T301	3	24.3%	66.1%	9.5%	0.0%	0.0%
T302	3	25.9%	30.7%	28.5%	7.4%	7.4%
Total	9	25.9%	36.5%	27.2%	8.0%	2.5%
SimScientists Total	15	26.3%	34.8%	31.2%	6.3%	1.5%

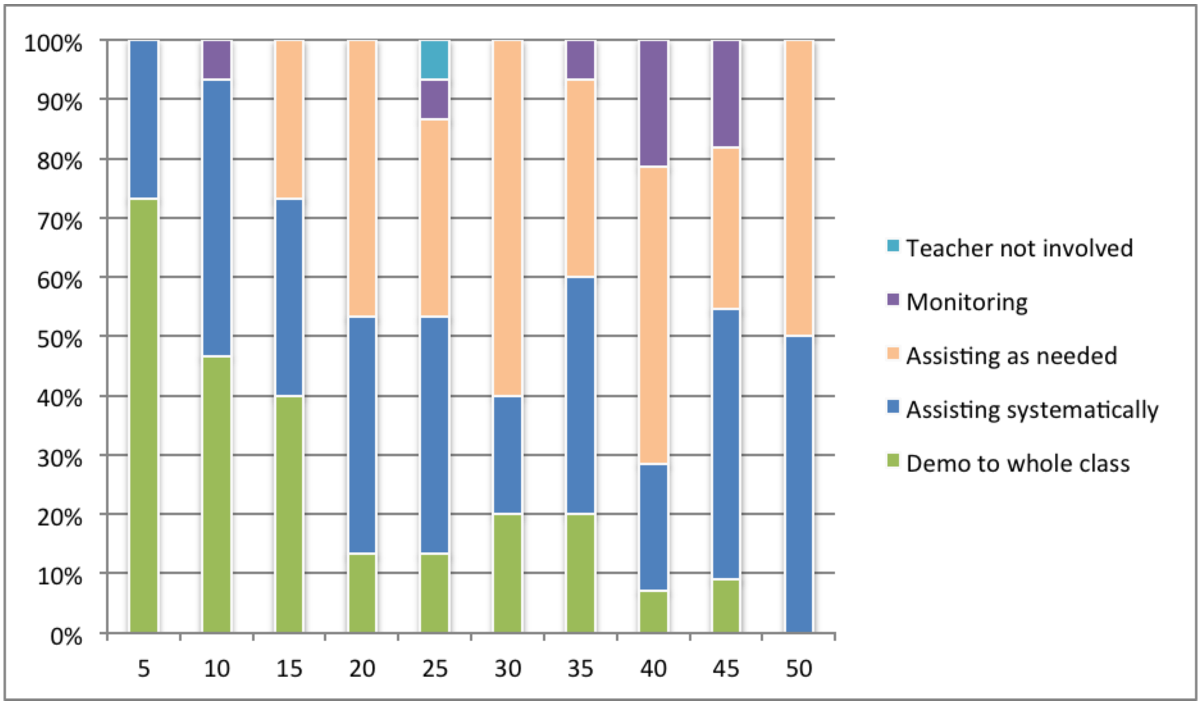

Figure 5. Teacher role by five-minute segment (reflection activities).

As with the organizational features, the role played by the teachers varied over time (see Figure 5). While demonstration to the whole class was predominant during the first five-minute segment, its use decreased greatly over the next few segments. At the same time, the teachers' focus on assisting their students increased from about one-quarter during the first segment to over three-quarters by the fourth five-minute segment. When teachers did take a less active role with their class by simply monitoring, this tended to take place during the later segments of the reflection activities while students were completing group work.

The reflection activities were designed to have both teacher-led and student-led components (see Table 10 and Figure 6). Therefore, it was not surprising to find both types of interaction common during the observations. Despite this, most of the teachers were more likely to utilize student-led interaction than teacher-led. The exception involved T101 who strongly favored teacher-led interaction. When examining the results across time, the teachers' emphasis on student-led interaction became even more apparent. In alignment with the class organization results, teacher-led interaction was predominant for the first 15 minutes of the reflection activities while student-led interaction was predominant for the remainder of class.

Table 10

Classroom Interaction by Teacher and Focus (Reflection Activities)

Teacher ID	*n*	Teacher-led	Student-led
Ecosystems			
T100	2	47.8%	52.2%
T101	2	65.0%	35.0%
T200	2	33.3%	66.7%
Total	6	48.7%	51.3%
Atoms & Molecules			
T300	3	31.5%	68.5%
T301	3	43.7%	56.3%
T302	3	32.6%	67.4%
Total	9	35.9%	64.1%
SimScientists Total	15	41.0%	59.0%

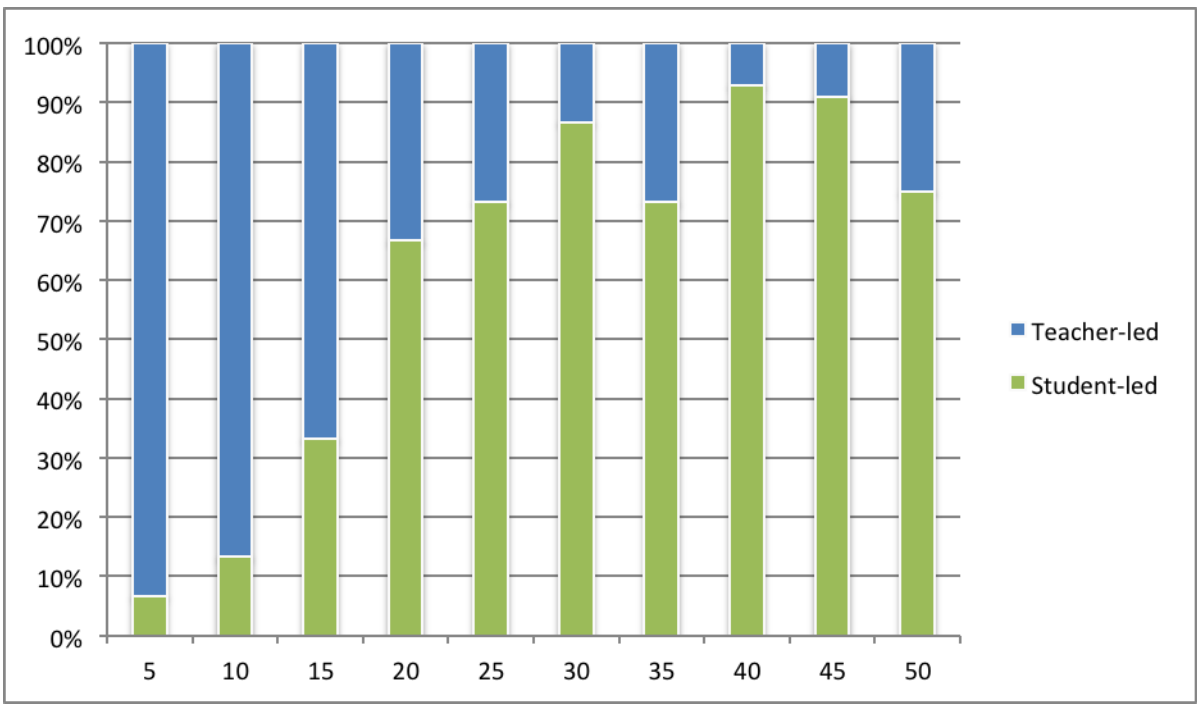

Figure 6. Classroom interaction by five-minute segment (reflection activities).

Fidelity to the lesson plans. One of the goals of the observers was to determine the fidelity that the teachers held to the reflection activity lesson plans (see Table 11). Surprisingly, none of the observers felt that the teachers completed any of the activities as designed. In most cases, a significant deviation was observed where the teacher purposefully changed one or more components of the lesson plan. Twice sessions were observed where the Ecosystems teacher ran out of time to finish the lesson. Furthermore, during three Atoms and Molecules sessions, the observers felt that the teacher both made a significant deviation and ran out of time. Based on the descriptive results, it was not surprising to find that the most common of these deviations involved teachers postponing the poster presentations to another class session. Among the Atoms and Molecules teachers, common deviations also involved changing activities from hands-on student experiments to teacher demonstrations or vice-versa.

Table 11

Fidelity to the Lesson Plans by Teacher and Focus (Reflection Activities)

			Reason sessions not completed as designed		
Teacher ID	*n*	Sessions completed as designed	Not enough time	Significant deviations	Both
Ecosystems					
T100	2	0	1	1	0
T101	2	0	0	2	0
T200	2	0	1	1	0
Total	6	0	2	4	0
Atoms & Molecules					
T300	3	0	0	1	2
T301	3	0	0	3	0
T302	3	0	0	2	1
Total	9	0	0	6	3
SimScientists Total	15	0	2	10	3

Student interaction and engagement. Reflection activities were normally conducted during the class sessions following the embedded assessments. Therefore, it was not surprising that most of the teachers had similar numbers of students present during the reflection activities as they did during the assessments (see Tables 12 and 3, respectively). More specifically, both types of observation had an average of just over 26 students present. Furthermore, the Atoms and Molecules sessions had an average of just over 28 students present. Interestingly, though, the average number of Ecosystems students went down slightly because T101 tended to have fewer students present on reflection activity days than during assessment days.

Table 12

Classroom Composition by Teacher and Focus (Reflection Activities)

Teacher ID	*N*	Average male students per period	Average female students per period	Average total students per period
Ecosystems				
T100	2	13.00 (1.41)	11.50 (2.12)	24.50 (3.54)
T101	2	15.00 (0.00)	6.00 (0.00)	21.00 (0.00)
T200	2	11.00 (0.00)	11.50 (0.71)	22.50 (0.71)
Total	6	13.00 (1.90)	9.67 (3.01)	22.67 (2.25)
Atoms & Molecules				
T300	3	10.00 (0.00)	19.33 (0.58)	29.33 (0.58)
T301	3	13.67 (0.58)	13.33 (2.31)	27.00 (1.73)
T302	3	13.00 (1.00)	16.67 (0.58)	29.67 (1.16)
Total	9	12.22 (1.79)	16.44 (2.88)	28.67 (1.66)
SimScientists Total	15	12.53 (1.81)	13.73 (4.45)	26.27 (3.56)

As was previously noted, students normally worked in pairs or small groups during the reflection activities. When students were doing group work, the students in the Ecosystems classes appeared to work together in a more balanced manner than did the students in the Atoms and Molecules classes. In other words, all members of the pair, trio, or group contributed equally or nearly equally. The exception involved T301 whose groups were balanced about two-thirds of the time. Interestingly, though, the Ecosystems classes were also more likely than were the Atoms and Molecules classes to have disengaged groups. This result may have to do with the structure of the activities and the ways in which the Ecosystems students were assigned to their group roles based on science competency (see Table 13).

Table 13

Group Participation by Teacher and Focus (Reflection Activities)

Teacher ID	*n*	Disengaged	Unbalanced	Balanced
Ecosystems				
T100	2	25.0%	0.0%	75.0%
T101	2	16.7%	0.0%	83.3%
T200	2	25.0%	0.0%	75.0%
Total	6	22.2%	0.0%	77.8%
Atoms & Molecules				
T300	3	0.0%	86.7%	13.3%
T301	3	0.0%	33.3%	66.7%
T302	3	7.4%	66.7%	25.9%
Total	9	2.5%	62.2%	35.3%
SimScientists Total	15	10.4%	37.3%	52.3%

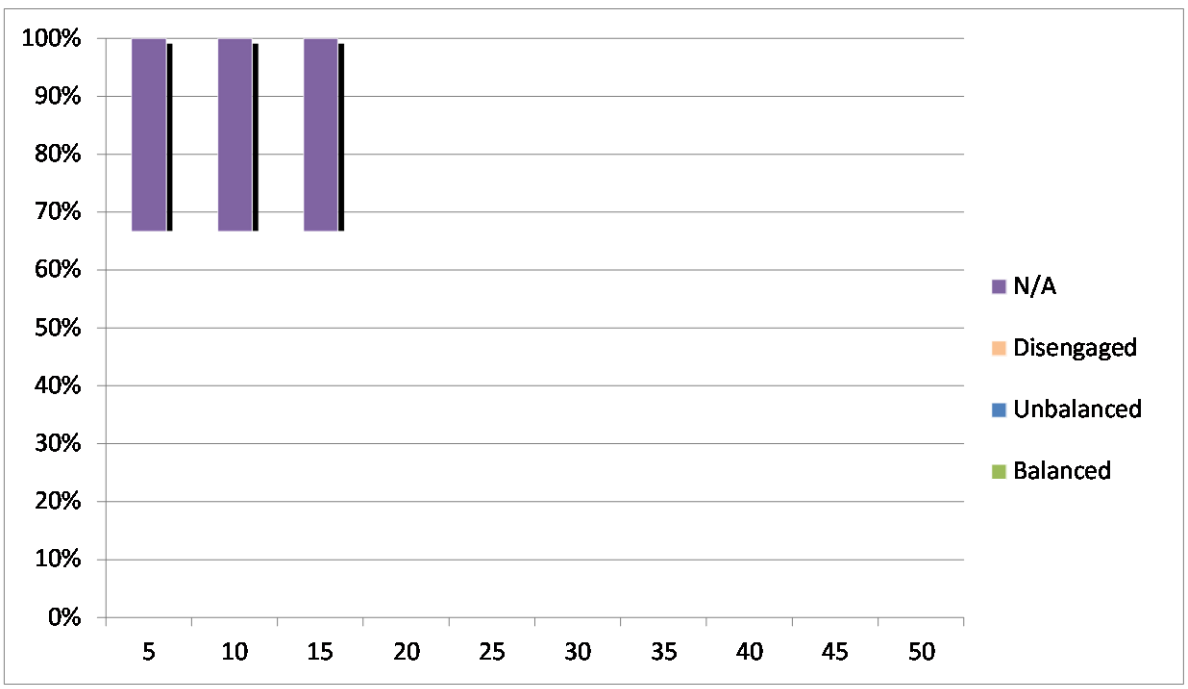

Figure 7. Group participation by five-minute segment (reflection activities).

As can be seen in Figure 7, balance among the student groups changed over the course of the class sessions. Balance started low and increased steadily until the middle of the reflection activities, after which it steadily decreased until the end of class. In contrast, the percentage of unbalanced groups was fairly steady from the 15-minute segment to the 45-minute segment. Furthermore, disengaged groups were normally observed during the second half of class when group work was more predominant. In some cases, such as the last five-minute segment, groups became disengaged after some, but not all of the members finished their work. This was particularly found during the Ecosystems activities where each student worked as part of a small team (2 to 3 students) as well as a larger group (4 to 7 students).

As with the assessments, observers kept track of overall student engagement throughout the reflection activities (see Table 14). Interestingly, while over two-thirds of the students were highly engaged during the assessments, only about half of the students were highly engaged during the reflection activities. The exception involved T301 who had over two-thirds of her students actively engaged and over one-quarter passively engaged. This result was not surprising considering the dynamic personality of the teacher and her practice of addressing the whole class on a regular basis. Among all of the teachers, only T100 and T300 had over 10% of their students off task.

Table 14

Student Engagement by Teacher and Focus (Reflection Activities)

Teacher ID	*n*	Active engagement	Passive engagement	Off task
Ecosystems				
T100	2	38.3%	50.1%	11.6%
T101	2	29.3%	61.9%	8.8%
T200	2	51.1%	43.0%	5.9%
Total	6	39.6%	51.6%	8.8%
Atoms & Molecules				
T300	3	49.0%	39.4%	11.6%
T301	3	71.9%	26.7%	1.4%
T302	3	43.6%	48.6%	7.8%
Total	9	54.8%	38.2%	6.9%
SimScientists Total	15	48.7%	43.6%	7.7%

As shown in Figure 8, student engagement tended to increase after the first ten minutes of the reflection activities. This trend corresponds with the change previously observed from teacher-led to student-led interaction at the 15-minute segment. Despite this, the percentage of students off task increased slightly after the first fifteen minutes.

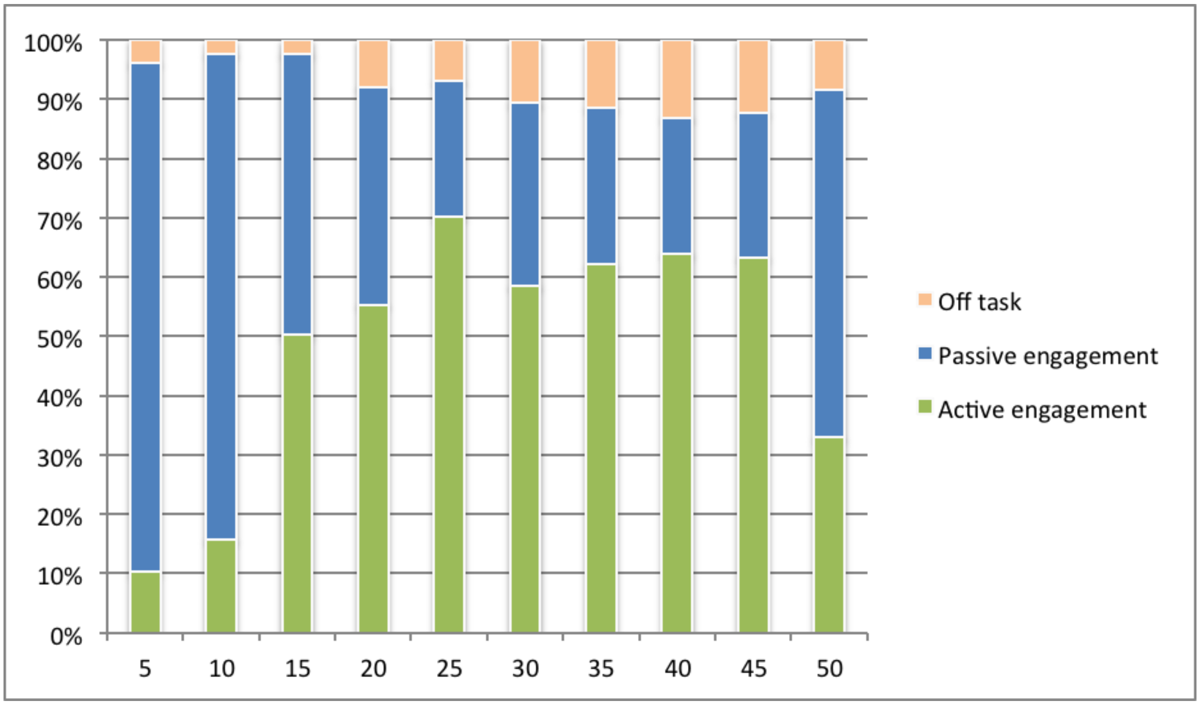

Figure 8. Student engagement by five-minute segment (reflection activities).

Interview Results

Teachers were asked their opinions about how useful the embedded assessment reports were for planning their reflection activities and differentiating instruction for their students.

Strengths. Overall, teachers felt that the student reports were easy to understand and useful. Some teachers appreciated that the reports identified students' competencies as a tool to differentiate the reflection activities. More specifically, one of the Atoms and Molecules teachers (T301) liked how the reports indicated which lesson plan to use while one of the Ecosystems teachers (T101) liked that it helped her in assigning students to the predictor, analyzer, and identifier roles during their group work. Some of the teachers also felt that the reports gave them information that they were not already aware of concerning some of their students. For example, one teacher (T100) realized that some of her normally high performing students were struggling with some of the content and/or inquiry targets, which was echoed in the quality of their poster presentations. While not providing specific examples, yet another teacher (T302) pointed out the benefit of having the reports provide ratings on specific targets rather than simply indicate a pass/fail for the entire assessment.

Weaknesses. Despite their generally positive comments about utilizing the assessment reports, most of the teachers also indicated some level of discomfort interpreting the student reports. One teacher's (T100) discomfort stemmed from the challenge of being sure that she was assigning her students to the correct roles when she implemented the Ecosystems reflection activities for the first time. Whereas another teacher (T200), who had already used Ecosystems during the previous year, struggled with how to assign her students since she primarily had group A students, whom the reflection activity guidelines stated should be put in the producers group.

Another issue that surfaced during the interviews involved the willingness of some teachers to trust the technology enough to rely on the assessment results over their own instincts about their students. One teacher (T300) who was very visual and tended to like a lot of structure was skeptical about how the reports ranked her students. While she disliked being unable to see how the embedded assessments were being scored, she was also concerned that the software could not make judgments about which components of the Atoms and Molecules reflection activities her students would find motivating. One of the Ecosystems teachers (T101) expressed a similar concern when she stated that she could not understand why four of her best students were scored low on the first embedded assessment. Yet another of the Atoms and Molecules teachers (T302) expressed concern that some of her students' rankings were simply based on luck, since she felt that some of them randomly clicked the buttons.

Professional Development Support for Teachers

During their interviews, most teachers discussed both strengths and weaknesses of the support systems provided to them when preparing for and implementing the SimScientists assessment systems. These supports included plans, guidance, and technology-based score reporting for implementing the reflective lessons discussed above. This section presents their comments about the face-to-face professional development.

Strengths. Teachers shared the value of the professional development. Three of these teachers (T101, T200, and T301) felt that some of these benefits stemmed from the WestEd trainer, who was described as well educated, knowledgeable, and supportive. The one teacher (T200) who had prior experience with SimScientists even stated that she appreciated how the trainer covered things differently than she had the previous year, increasing the benefit of attending the professional development again.

Teachers also expressed appreciation that they were directly exposed to the assessments and reflection activities rather than just being told about them. Two of the teachers (T100 and T302) pointed out that the hands-on aspect enabled them to experience how the assessments would work before exposing their students to them. One of these teachers (T302) expressed a similar sentiment concerning the reflection activities, highlighting how it helped her to "be better prepared to set it up for [her] students." Even the teacher who taught Ecosystems during the previous school year felt that getting hands-on experience with the reflection activities again was beneficial since she considered them challenging to implement.

Weaknesses. Despite their generally positive opinion of the professional development, many of the teachers also pointed out issues that detracted from their ability to apply what they learned. The biggest issue involved receiving their training months before implementing the assessments and reflection activities. Although the teachers expressed understanding that this would have been a logistical decision, they felt strongly that their experience would have been better without the lag in time. One teacher (T101) was polite about the issue, noting that the information she learned would have "sunken in" better. Three other teachers were more direct (T100, T300, and T301), stating that they simply forgot many of the details of their training, including procedural issues. Another logistical issue brought up by one of the teachers (T101) involved the length of the training. She indicated that her group was released about four hours early one day, limiting the amount of time the trainer was able to model the implementation of SimScientists.

Suggestions. In addition to providing their opinions about the professional development, some of the teachers made suggestions for how WestEd might make the experience more beneficial. One of the Atoms and Molecules teachers (T300) would have liked more in-depth, systematic instructions on how to use the teacher interface known as the Learning Management System (LMS). This teacher also suggested that WestEd could provide a manual or set of job aides at the training including how to reset passwords and other tasks with the LMS. Another Atoms and Molecules teacher (T301) indicated that it would have been helpful if teachers were told how much creative freedom they had in adapting the reflection activities for their classes. For example, the teacher wanted to know whether it was okay to make the teacher demonstrations into hands-on student activities, which was the norm in her class. In addition, the Ecosystems teacher (T101) who stated that her training session was shortened suggested that the trainer include more breakout sessions during the professional development so that the teachers can practice implementing the assessment system.

Summary and Recommendations

In this section, we summarize study findings with regard to issues of technology support of the assessment systems, as well as the technology-based and face-to-face professional development support teachers of SimScientists received.

Technology Support for the Assessments

Design elements of the assessments. Overall teachers had favorable impressions of the design elements of the SimScientists assessments. Teachers particularly liked how much more engaged their students seemed to be when completing the simulations then when completing traditional paper and pencil assessments. They credited this added engagement to the visuals, animations, opportunities for students to manipulate data, and the varied response modes included. Despite this, Atoms and Molecules teachers felt that the assessments would be even more engaging for their eighth graders if they were shorter in length and had a little less repetition in the scenarios used. Furthermore, these teachers expressed some concern that the reading level and scenarios were not always appropriate or relatable for their predominantly low income and limited English proficient students.

Delivery of the assessments. Delivery of the assessments tended to run smoothly. Most students were actively engaged in using the simulations and answering the questions in a purposeful manner. Furthermore, most of the students who participated in the case-study classes were able to complete their assessment during regular class time. Some of the exceptions involved students who had greater problems remembering their logins for SimScientists, students who were not present for the entire class period and students who worked slowly because they were English language learners. In the last case, students were encouraged to use Google Translate or consult their classmates to understand the directions and questions. In addition, students were able to rely on their teachers, who tended to circulate around the computer labs giving assistance as needed.

Interpretation of the assessment reports. Many of the teachers reported that the assessment reports were useful in gauging their students' levels and performance. Some noted that this included being able to get a better understanding of their students' content knowledge and scientific inquiry skills. Despite this, most teachers felt that their students' results met rather than exceeded their expectations. Furthermore, teachers had mixed opinions concerning how the benchmark results would compare with their district and/or state science assessments[1]. While the Atoms and Molecules teachers felt that results would

[1] State science assessments are administered to eighth grade students, but not sixth grade students in Nevada.

be comparable, Ecosystems teachers had mixed opinions about whether results would be comparable and whether students performed better on SimScientists.

Professional Development and Other Sources of Support for Teacher Practice

Fidelity of the reflection activities. During the observations, teachers varied in their fidelity to the reflection activities as detailed in the lesson plans and professional development materials. For example, most of the teachers followed the lesson plans concerning student group work (e.g., chemistry experiments and food webs) and were careful to make sure the activities were primarily student-led. In contrast, teachers did not always provide introductions, reviews, and summaries. Furthermore, all but one teacher, who had prior experience teaching Ecosystems, chose to implement the poster presentations during the one class period. This was a deviation for all of the Atoms and Molecules reflection activities as well as the second Ecosystems reflection activity. The issue of fidelity was also brought up in the interviews when one teacher voiced that she would have liked to be told how much "creative freedom" she had to deviate from the lesson plans. For example, she wanted to know whether it was okay to change the teacher-led demonstrations included in some of the reflection activity lesson plans to student-led experiments.

Technology-based support for planning the reflection activities. One of the unique features of the SimScientists assessment systems involved the use of the student assessment reports to differentiate the reflection activities. With the Atoms and Molecules activities, teachers were able to use this information on student content and inquiry competencies to determine whether to implement the A and/or B versions of the lesson plans. Similarly, Ecosystems teachers were able to use student competencies to assign roles during the group work. While most of the teachers felt that the student reports were easy to understand, some expressed discomfort about whether they were properly using the information. Furthermore, some of the teachers questioned whether all of the information was accurate and whether it was more useful than their own instincts as a teacher. Interestingly, WestEd addressed this last issue in their professional development by letting teachers know that their judgment should outweigh the assessment-generated reports when preparing the reflection activities.

Professional Development

Teachers felt that the professional development they received had both strengths and weaknesses. For example, teachers felt positively about the trainer and really liked the hands-on aspects where they were able to experience some of the assessments and reflection activity components. The major weaknesses mentioned by the teachers focused on scheduling issues. More specifically, teachers stated that their trainings took place months before they started the assessments and in one case the length of the training was unexpectedly shortened.

Recommendations

Based on our findings, we make the following recommendations for SimScientists:

Embedded and benchmark assessments.

1. During the interviews and informally during the observations, some of the teachers expressed dislike of the scheduling system for the assessments. For example, some teachers had trouble coordinating both computer lab access and access the program developers provided to the assessments. This was especially difficult for teachers during state testing when the computer labs were in greater demand. In addition, some teachers felt that the scheduling limited their ability to maximize learning during both SimScientists and their regular class activities. Program developers should consider providing greater flexibility in scheduling when assessments are available and possibly consider keeping them available for more extended periods.
2. Although most of the teachers understood the reason for having logins, they were a major source of delay for individual students and entire classes at times. Program developers should consider simplifying the system for usernames and passwords so that teachers have an easier time setting them up and students have an easier time remembering them.
3. Even though the overall levels of engagement observed were similar between the two topics, Atoms and Molecules teachers felt strongly that their students' level of engagement decreased as they completed each progressive assessment. This decrease was attributed to the length of the assessments and the similarity of the scenarios. Program developers should consider shortening the Atoms and Molecules embedded assessments and/or putting more variation into the scenarios in order to decrease student-testing fatigue.
4. Each SimScientists assessment included a report indicating student competencies for the content and inquiry targets. These reports, along with the feedback pop-ups for the individual assessment pages were intended to provide students with valuable information about their own thinking or behavior in order to improve learning. Despite this, these features were not fully utilized during the assessments observed. For example, all or most students were observed viewing their reports during only two class sessions when the teacher systematically circulated around the computer lab for this purpose. Program developers should better train teachers concerning the purpose of the student feedback so that teachers can provide better support for their use.

5. Unlike the embedded assessments, where students evaluated their own open-ended questions as a learning tool, teachers were asked to score this question type for the benchmarks. Because of this, teachers were required to complete an online training to develop reliability prior to scoring these items. Despite this, some of the teachers who were new to SimScientists had usability issues when scoring their students' assessments. More specifically, some of the teachers reported unknowingly skipping items and then having trouble finding the missed items once they realized this. Program developers should consider making the checkmark indicator system more clear so that teachers are aware of their error before moving onto grading other questions.

Reflection activities.

1. While the lesson plans included guidelines for how to differentiate the activities based on student competencies revealed in the embedded assessments, some teachers felt limited by the process. Developers should consider providing greater opportunities and training on how teachers can differentiate the activities to meet their students' learning styles as well as their content and inquiry needs.

Professional development.

1. Although most teachers were generally positive about the professional development they received, many felt that the gap in time between their training and implementation detracted from the experience. Developers should revisit their scheduling of the face-to-face professional development and/or provide some form of interim follow up to help improve teacher retention of the material. For example, trainings could be videotaped and made accessible to teachers through the LMS for later review.

2. In addition to giving specific directions about how to complete the hands-on activities during the reflection activities, some teachers expressed a desire to have more in-depth training on how to use the LMS teacher interface. Furthermore, developers should consider providing a manual and/or job aides for teachers to use as they are trying to put their training into practice.

References

Herman, J., Dai, Y., Htut, A. M., Martinez, M., & Rivera, N. (2011). *Evaluation of the enhanced assessment grants (EAGS) SimScientists program: Site visit findings* (CRESST Report 791). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).