CRESST REPORT 791

Joan Herman

Yunyun Dai

Aye Mon Htut

Marcela Martinez

Nichole Rivera

EVALUATION OF THE ENHANCED ASSESSMENT GRANTS (EAGS) SIMSCIENTISTS PROGRAM: SITE VISIT FINDINGS

MAY 2011

The National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Sciences
UCLA | University of California, Los Angeles

Evaluation of the Enhanced Assessment Grants (EAGs) SimScientists Program: SITE VISIT FINDINGS

CRESST Report 791

Joan Herman, Yunyun Dai, Aye Mon Htut,
Marcela Martinez, and Nichole Rivera
CRESST/University of California, Los Angeles

May, 2011

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 #5756 s08-149, pursuant to WestEd's grant with the Nevada Department of Education (NDE), Grant Number 09-2713-126, which is pursuant to Federal Award Number S368A080005 between NDE and U.S. Department of Education, CFDA #84.369.

To cite from this report, please use the following as your APA reference:
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, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

EXECUTIVE SUMMARY

This evaluation report addresses the implementation, utility, and feasibility of SimScientists' simulation-based assessments for middle school science classrooms, with particular attention to the use of accommodations available in the program. Data were collected from a convenience sample of five schools and eight teachers across three states participating in a larger pilot of SimScientists. During site visits, CRESST researchers conducted classroom observations and teacher interviews addressing the major components of the SimScientists program in two topics: (1) Ecosystem and (2) Force and Motion. SimScientists components included embedded, formative assessments; reflection activities designed to deepen student understanding of key ideas and processes; and benchmark assessments to gauge student learning at the end of the unit. Site visit findings were confirmed using teacher survey results from the larger pilot on teachers' use and satisfaction.

Question 1: How Effective was SimScientists' Implementation?

Ecosystem was implemented in 6th and 7th grade classrooms; it consisted of two embedded assessments, two reflection activities, and a benchmark assessment. Force and Motion was implemented in 8th grade classrooms; it consisted of three embedded assessments, three reflection activities, and a benchmark assessment. The time range for students to complete embedded assessments was 35 to 45 minutes, and the range for reflection activities was 40 to 57 minutes. Both observations and interviews indicated that most reflection activities could not be completed in a single class period.

Both teachers and students generally believed that the SimScientists program was beneficial to learning. Observations showed that students were actively engaged most of the time during assessments. Teachers found students more focused because they were working individually on the assessments and were able to work at their own pace. Interviews with teachers confirmed that students performed better in simulation assessments compared to traditional assessments because the simulations provided visuals and interactions. 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.

Question 2: Was SimScientists Feasible and Useful in Middle School Classrooms?

A key feasibility issue was the availability of computers in the classroom to run the SimScientists program. Our observations indicated that there were enough computers in most

of the classrooms, although teachers explained that the computers were limited and not always accessible. Teachers had to reserve the computers early in the school year because many other teachers reserved the computers for their classroom activities. One school had enough computers, but not all the computers were SimScientists-compatible. Teachers found solutions to the problem of not having enough computers: they sent students to other rooms, computer labs, or borrowed laptops from other teachers. Most students finished the assessments by the end of the class period even when there were not enough computers for all the students in class; some of the students finished early enough that two students were able to complete the assessment in one class period with the same computer. Advanced planning and scheduling would be essential if teachers were to implement SimScientists on a regular or ad-hoc basis.

Teacher interviews confirmed that classrooms generally had enough computers, but if teachers were to use SimScientists several times a year, they would have to coordinate more often with other teachers. Therefore, some teachers might not be able to implement the program several times a year.

Most computers allowed students to easily user the SimScientists program. However, some networks were often slow, and we witnessed a lot of "reloading" or login problems. In general, this feasibility issue was not serious enough to prevent completion of the assessments. Most of the students, except for a few, liked the graphic and interactive element of the assessments, and teachers gave positive feedback for the technical quality of the simulation-based assessments as a whole. Teachers and students were pleased that the program provided immediate results. Teachers specifically liked that they could monitor students while they were taking the assessments.

Apart from the logistical difficulties, teachers agreed that the assessments would be useful in measuring their individual state standards. Teachers suggested that each question should be matched to a specific state standard so they can easily see what standards most of their students did not understand. Currently, the assessments cover general science concepts and are not directly aligned to a specific unit or state standard.

Question 3: How Effective were SimScientists' Accommodations?

There was not enough evidence to show that English language learners (ELLs) and students with disabilities benefitted from accommodations provided by the SimScientists program. Although the accommodations were available and although teachers were aware of them, we seldom observed them in use. Teachers said that the accommodations could have been helpful to a few students in prior years, but their current students did not really need them. Many teachers felt that ELL students would have benefitted more from a glossary or a vocabulary section.

Recommendations

While teachers and students alike were very positive about their experiences with SimScientists, we offer several recommendations for improvement—in particular, recommendations for refining the program's embedded and benchmark assessments and for increasing the feasibility of the reflection activities.

TABLE OF CONTENTS

EVALUATION OF THE ENHANCED ASSESSMENT GRANTS (EAG) SIMSCIENTISTS PROGRAM: SITE VISIT FINDINGS

Joan Herman, Yunyun Dai, Aye Mon Htut,
Marcela Martinez, and Nichole Rivera
CRESST/University of California, Los Angeles

Abstract

This evaluation report addresses the implementation, utility, and feasibility of SimScientists' simulation-based assessments for middle school science classrooms, with particular attention to the use of accommodations available in the program. Data were collected from a convenience sample of five schools and eight teachers across three states participating in a larger pilot of SimScientists. During site visits, CRESST researchers conducted classroom observations and teacher interviews addressing the major components of the SimScientists program in two topics: (1) Ecosystem and (2) Force and Motion. SimScientists components included embedded, formative assessments; reflection activities designed to deepen student understanding of key ideas and processes; and benchmark assessments to gauge student learning at the end of the unit. Site visit findings were confirmed using teacher survey results from the larger pilot on teachers' use and satisfaction.

INTRODUCTION

SimScientists' simulation-based assessments are intended to support rich learning of major concepts and principles in science and to promote inquiry skill. The program provides suites of assessment activities, 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 shorter, 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. A special feature of the SimScientists assessments is their capacity to automate accommodations for English language learners (ELLs) and students with disabilities who may need them. The accommodations include a text-to-speech option for students with limited English reading ability and text magnification for visually challenged students.

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 to help answer the following questions:

1. How effective was the SimScientists implementation?
2. Was SimScientists feasible and useful in middle school classrooms?
3. How effective were SimScientists' accommodations?

METHODOLOGY

To answer these questions, the study used a convenience sample of teachers in the states of Nevada, North Carolina, and Utah. Each teacher was involved in a large pilot of SimScientists assessments on the topics of (1) Ecosystem and (2) Force and Motion. Site visits were conducted to observe the SimScientists activities in action and to interview teachers about their reactions to the program. A group interview with state coordinators also was conducted to gather feedback.

Sample

The study sample included eight teachers in five schools across the three states: two schools and three teachers in Utah, two schools and three teachers in Nevada, and one school and two teachers in North Carolina.

The sample was chosen to ensure representation of the three participating states, the topic areas (1) Ecosystem and (2) Force & Motion, and the SimScientists embedded and benchmark assessments and reflection activities. The Ecosystem topic consisted of five components: Embedded Assessment 1 (E1); Reflection Activity 1 (RA1), E2, RA2; and a benchmark assessment (B). Force and Motion consisted of seven components: embedded assessment 1 (E1), reflection activity 1 (RA1), E2, RA2, E3, RA3, and a benchmark assessment (B). Table 1 shows the distribution of SimScientists components and topics observed.

Table 1

Components Observed for the Eight Teachers in the Study Sample

Component	T100 ECO	T101 ECO	T102 ECO	T200 ECO	T201 ECO	T202 ECO	T300 FM	T301 FM
Eco E1		X				X		
Eco R1		X				X		
Eco E2		X	X	X	X			
Eco R2	X	X	X	X	X			
Eco B	X			X	X			
FM E1								
FM R1								
FM E2								
FM R2								
FM E3							X	X
FM R3							X	X
FM B							X	X

Note. B = Benchmark Assessment; E1 = Embedded Assessment 1; E2 = Embedded Assessment 2; E3 = Embedded Assessment 3; Eco = Ecosystem; FM = Force and Motion; R1 = Reflection Activity 1; R2 = Reflection Activity 2; R3 = Reflection Activity 3.

Observations were conducted for each teacher's classes that were using the program. There were two unique situations within our sample. First, one school's students with disabilities were primarily gifted students despite their disability classification. We did not count these students as students with disabilities because we were interested in students who needed more assistance than the average student. Second, one class consisted entirely of ELLs. The ELL students had a teacher's aide who helped them with all class work, including the SimScientists program.

Instrumentation

Study instrumentation included observation and interview protocols jointly developed by CRESST and WestEd and a teacher survey developed by WestEd.

Observations

Two observation protocols were developed: one for the embedded and benchmark assessments and the second for the reflection activities. Both protocols asked observers to record identifying and descriptive information about the class and its composition, including start and end times. At five-minute intervals, the observers recorded the nature of:

- Classroom organization (explain briefly)
- Pair balance (explain briefly)
- Teacher role (explain briefly)
- Student dialogue (explain briefly)
- Technical incidents (e.g., log-on difficulties, computer failure)
- Student interactions for students not using accommodations (explain briefly)
- Student interactions for students using the accommodations (explain briefly)

After their classroom visits, observers added field notes and documented any use of accommodations. For the assessment protocol, observers described any use of the program's reporting function. For the reflection activities protocol, they recorded the extent to which teachers followed the implementation design, whether they were able to complete all activities, and any deviations noted.

Interviews

An interview questionnaire was developed in order to acquire teachers' reactions to the SimScientists program. Teachers commented on the following topics:

- Their individual comfort implementing the program
- Their students' level of comfort with the program (in particular, ELLs and students with disabilities)
- Strengths and weaknesses compared to traditional paper-pencil assessments
- Strengths and weaknesses of accommodations provided by the program

Teachers also provided recommendations on how to improve the assessments and reflection activities. In addition, teachers reported logistical factors that would prevent them from implementing the program on a regular basis.

Teacher Surveys

Using a WestEd survey, teachers provided feedback about SimScientists implementation, usage of embedded and benchmark assessment, report benefits, and program accommodations for ELLs and students with disabilities. There were seven survey forms, three regarding Ecosystem and four regarding Force and Motion.

Study Procedures

Prior to the site visits, four observers were trained to effectively use the protocols and practiced with them to achieve consistency. Each researcher observed different teachers.

Analysis

Observation data were prepared for quantitative analysis through input into Excel and SPSS. The initial descriptive analysis showed little difference between observations of embedded assessments and those of benchmark assessments. Thus, we combined the two in subsequent analysis and reporting.

In this report, we provide two types of descriptive analyses: first, the frequency of particular categories of activity across the entire observation period. Second, in order to provide a sense of how the SimScientists activities unfolded, we provide figures showing the frequency of activity in five-minute segments. While some classes lasted more than 45 minutes, we based the figures on the first 45 minutes of the classes for purposes of consistency. For most intervals, the figures represent 56 observation periods across eight teachers.

Interview data were transcribed and put into ATLAS.ti for qualitative analysis. Common themes and patterns emerged. In order to provide balance, we include contradictory statements made by a few or at least one teacher.

In the analysis of teacher survey data, we selected survey questions common across survey forms. These survey questions are related to the usefulness of the embedded reports, how closely teachers followed the activity guides, and how useful the embedded assessment and reflection activities were for students' learning.

We conducted a descriptive statistics analysis of survey questions related to the usefulness of SimScientists' features and activities. In addition, we used graphs to show teachers' responses about which accommodations options their students used in classroom activities and computer assessment.

Most teachers filled out all survey forms related to Ecosystem or Force and Motion. Some teachers filled out the same survey form twice, once for each unit. However, some teachers answered most of the survey questions in one survey but not the other. For these cases, we selected the one with fewer missing values.

The following sections summarize results separately from observations, interviews, and survey data. The concluding section synthesizes results across data sources to directly address study evaluation questions and provides implications for next steps.

OBSERVATION FINDINGS

Observation results are summarized for embedded and benchmark assessments separate from those for the reflection activities. We used Teacher IDs (TIDs) to protect the confidentiality of individual teachers and schools. The appendix provides a detailed summary of assessments, classroom demographics, and average time taken for each assessment.

Embedded and Benchmark Assessment

Table 2 reports computer availability in each classroom. Overall, an average of 37 computers were available in classrooms or computer labs, including the teachers' computers. The average number of computers in classrooms ranged from a low of 17.4 to a high of 58; however, on average, only 29 computers per classroom were either working and/or SimScientists-compatible.

Table 2

Computer Availability

TID	Average # of computers	Average # of computers working and/or compatible with SimScientists
T200(*N*=10)	37.0	34.7
T201(*N*=11)	37.0	34.6
T202(*N*=3)	33.0	33.0
T300(*N*=10)	17.4	17.4
T301(*N*=8)	58.0	20.4
T100(*N*=2)	33.0	33.0
T101(*N*=7)	35.9	34.1
T102(*N*=3)	36.0	34.0
Average (*N*=54)	37.4	29.2

Note. N = Number of periods observed; TID = Teacher ID.

Classroom organization. Classroom organization during the assessments consisted primarily of individual students working alone. Only one teacher had students working in pairs (see Table 3).

Table 3

Class Organization (Assessments)

TID	Individual students working alone	Pairs of students	Others*
T200(*N*=10)	85%	0%	15%
T201(*N*=11)	100%	0%	0%
T202(*N*=3)	0%	67%	33%**
T300(*N*=10)	100%	0%	0%
T301(*N*=10)	100%	0%	0%
T100(*N*=2)	95%	0%	5%
T101(*N*=7)	100%	0%	0%
T102(*N*=3)	100%	0%	0%
Average(*N*=56)	92%	4%	5%

Notes. *N* = Number of periods observed; TID = Teacher ID.
*Others: students were receiving passwords, or teacher was resetting passwords.
**In one class, the teacher asked students to work in teams but to use their own computers.

Figure 1 shows the distribution of classroom organization at five-minute intervals. The vast majority of students worked alone, except during the first 10 minutes of class, where, even then, over 75% of students worked alone.

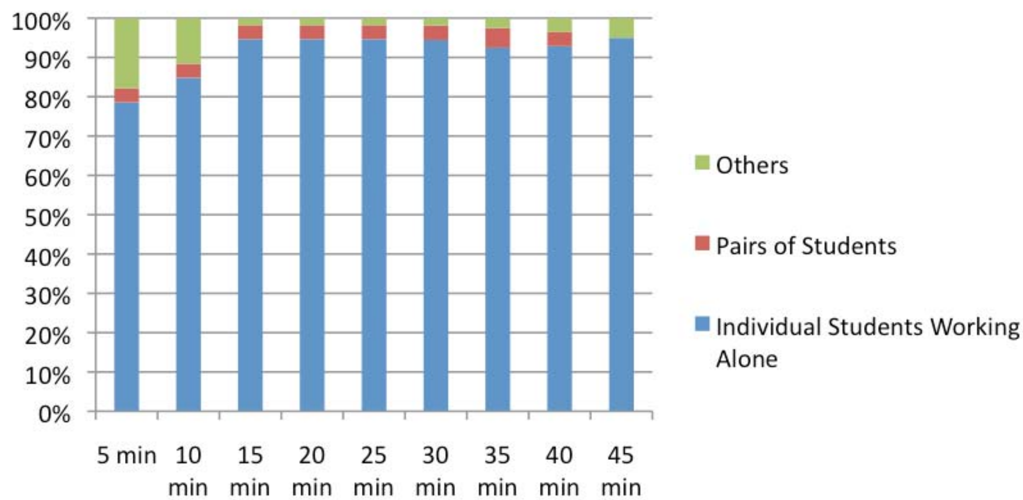

Figure 1. Class organization by interval (Assessment).
Note: The label "Others" means that the teachers were addressing the whole class, providing directions for the assessments, or giving out passwords.

Student interaction and engagement. Observations of students' interaction and engagement indicated that more than three quarters of the students in most classes were actively engaged during the assessments (see Table 4). There was only one class in which students were often not engaged and off-task: N202, whose teacher used two-person teams. The observer reported that the classroom management was relatively poor for that particular class. Otherwise, observers reported that students were both interested and engaged in the SimScientists system. Many students commented that they liked the system's interactive effect. Table 4 shows the levels of student engagement per teacher. Averages were calculated per teacher to get an estimate of how many students were low, moderately, and highly engaged.

Table 4

Student Interaction (Assessments)

TID	Low engagement	Moderate engagement	High engagement
T200(*N*=10)	9%	14%	77%
T201(*N*=11)	6%	15%	79%
T202(*N*=3)	25%	22%	53%
T300(*N*=10)	3%	29%	68%
T301(*N*=10)	4%	2%	94%
T100(*N*=2)	0%	6%	94%
T101(*N*=7)	2%	33%	64%
T102(*N*=3)	6%	0%	94%
Average(*N*=56)	6%	17%	77%

Note. *N* = Number of periods observed; TID = Teacher ID.

Observation data suggest students were less active at the beginning of the class and at the end of the assessments (see Figure 2). However, students appeared active or fully engaged when doing the assessments throughout the period.

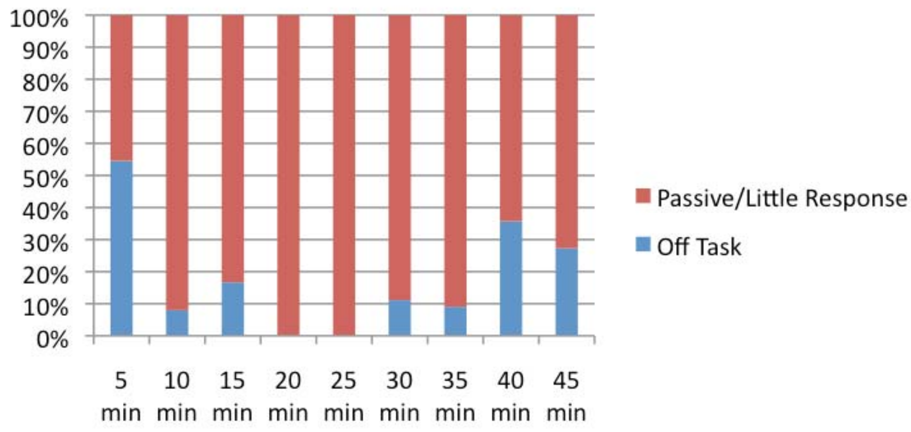

Figure 2. Student interaction (Assessments).

Eight of the 56 assessments that we observed had students with disabilities or ELL students; however, they did not necessarily use the accomodations provided by SimScientists. Table 5 shows student engagement during those periods in which ELLs or students with disabilities were in attendance. Please note that as long as there were students with disabilities or ELL students, this section of the observation sheet was recorded, even if the students did not use SimScientist accommodations. Apart from one class (T301), the majority of the students requiring accommodations were highly engaged (see Table 5 and Figure 3).

Table 5

Student Interaction (Accommodations)

TID	Low engagement	Moderate engagement	High engagement
T200(*N*=1)	0%	0%	100%
T202(*N*=2)	17%	0%	83%
T300(*N*=1)	0%	0%	100%
T301(*N*=1)	33%	17%	50%
T101(*N*=2)	0%	20%	80%
T102(*N*=1)	0%	0%	100%
Average(*N*=8)	8%	7%	85%

Note. *N* = Number of periods TID = Teacher ID.

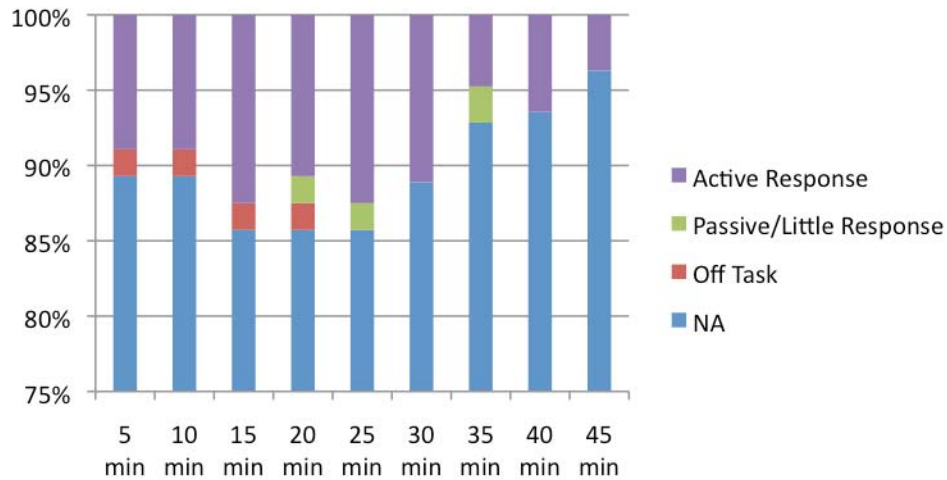

Figure 3. Student interaction (accommodated students only)

Only one teacher, who had 3 periods, assigned students to work in pairs. Observations indicated that on average, the paired students were in co-constructed meaning mode half of the time. A third of the time, they were disengaged and the rest of the time unbalanced.

Teacher's role. Teachers were primarily involved during the beginning of the assessments when they provided instructions and passwords to students. Although students were expected to remember their individual passwords, we consistently observed that teachers had to spend the first few minutes of class passing out or resetting passwords. While a repeating problem, it did not affect the overall implementaion of the program. Some students asked for assistance in logging in or loading the program, but the assessments were self-explanatory, and most students were able to figure things out on their own. The majority of teachers spent their time assisting students as needed during the assessments, while some teachers also monitored the performances of the class by walking around as well as checking students' status on their computer (see Table 6). Observers informally noticed that teachers' assistance given to students was mainly procedural.

Table 6

Teacher's role (assessment)

TID	Teacher not involved	Demonstration to whole class	Assisting students systematically	Assisting students as needed	Monitoring
T200(*N*=10)	2%	18%	0%	60%	20%
T201(*N*=11)	1%	0%	0%	99%	0%
T202(*N*=3)	0%	13%	0%	87%	0%
T300(*N*=9)	0%	19%	7%	29%	45%
T301(*N*=10)	38%	6%	0%	55%	0%
T100(*N*=2)	11%	5%	0%	16%	68%
T101(*N*=7)	14%	0%	0%	38%	48%
T102(*N*=3)	0%	3%	6%	31%	60%
Average(*N*=55)	10%	9%	1%	57%	23%

Note. *N* = Number of periods observed; TID = Teacher ID.

Figure 5 shows a similar pattern as Table 6: Teachers began the assessment lessons by demonstrating to the whole class. Afterwards, their direct involvement dropped substantially. There were very few instances, as little as 1% on average, where teachers continued to assist students systematically. In about 10% of the cases, teachers were not involved during the middle and at the end of classes. Because most of the students completed the assessments before the end of the assigned time, teachers asked them to work on other tasks or assignments. One teacher asked students to log into the BrainPop website where students could learn other scientific topics similar to those offered by SimScientists. Teachers were sometimes observed working on other administrative duties or checking their e-mail; such occurrences were reported as teachers "not [being] involved."

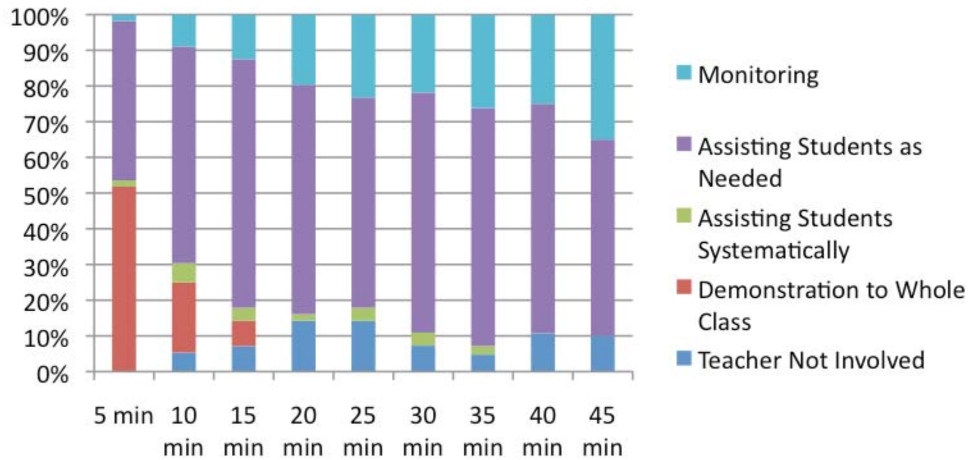

Figure 5. Teacher's role (Assessment).

Report use. Table 7 shows that students' use of reports varied by teachers. Three teachers directly encouraged all their classes to access and use the reports. For example, one teacher required that students look at their report and talk to her about it after completing their embedded assessments. A few other teachers explained that the report function was available and that students could use them to see how well they did. Since the unit of analysis is classes, observers recorded that the reports were "used" even if just one student viewed the report.

Table 7

Report Use for Embedded Assessments by Teacher

	How classes used the report							
TID	Not accessed	Not used	Review report	Discussed within pair	Discussed with other peers	Discussed with teacher	Total classes	%used
T200	0	0	0	3*	3*	0	3	100%
T201	0	0	0	0	0	5	5	100%
T202	1	0	0	0	1	0	2	50%
T300	4	1	0	0	0	0	5	0%
T301	0	2	1	0	0	2	5	60%
T101	1	4	1	1	0	0	7	29%
T102	0	0	0	0	0	3	3	100%
Total	6	7	2	1	9	5	30	57%

Note. TID = Teacher ID.
*Three classes were double counted, as the students discussed both within their pair and with other peers

Our observations indicated that in almost half of the classes, no students accessed or used the reports. Observers noted that even in classes where reports were accessed or used, only a small portion of the students did so. Even in classes where the the teacher directly encouraged their use, less than half of the students used the reports. In some classrooms, even if teachers did not mention viewing the reports, some students would use them. When an observer asked one teacher about the lack of report use, the teacher responded that they were just kids and wanted to be over with the task.

Technical problems. As mentioned earlier, there was a high number of login failures at the beginning of the class (Table 8). No power failure or network failures were witnessed during our observations. Waiting for pages to load was also another common problem.

Table 8

Frequency of Incidents by intervals

Incident type	5 min	10 min	15 min	20 min	25 min	30 min	35 min	40 min	45 min
Login fail	25	8	2	0	0	0	0	0	0
Waiting for pages to load	5	8	3	0	0	0	0	0	0
Reloading assessment	0	1	3	0	0	1	1	2	1
Computer crash/freeze	0	1	1	1	0	0	0	0	0
Network failure	0	0	0	0	0	0	0	0	0
Power failure	0	0	0	0	0	0	0	0	0
Other*	2	3	2	1	0	0	0	0	0
Total	32	21	11	2	0	1	1	2	1

Note. *Examples of “other” are error messages, cannot access database, keyboard stops working, and other hardware failures.

Problems occurred more on Apple computers than on PCs. There were also interface problems with Macintosh computers.

Most of the problems occurred during the first half of the class. In the second half, the only problem that occurred was that it took very long to reload the assessment. More loading problems occurred with benchmark assessments than with embedded assessments. The benchmark assessments took longer to complete, and with longer loading time problems, some classes did not complete the assessments on time. This occurred in less than 5% of all the classes.

Reflection Activities

We observed eight teachers during reflection activities, each teacher having three to seven class periods. The length of each period varied depending on the school and time of the day. Table 9 provides the average observed reflection activity time. Some classes were extended periods.

Table 9

Summary of Reflection Activities and Classroom Composition (total=34)

TID	Number of periods	Average time taken for the activity (in minute)	Grade	Average students per period	Average male students per period	Average female students per period	Activity	Topic
T200	5	51.4	6	32.2	14.2	18.0	RA2	Eco
T201	6	47.5	6	30.0	15.0	12.3	RA2	Eco
T202	3	40.3	7	28.0	15.0	13.0	RA1	Eco
T300	5	48.2	7	21.4	9.4	12.0	RA3	FM
T301	5	46.0	7	17.8	11.0	8.8	RA3	FM
T100	3	56.7	8	30.3	16.0	14.3	RA2	Eco
T101	7	39.6	8	30.6	14.5	16.0	RA1 & RA2	Eco

Note. Eco = Ecosystem; FM = Force and Motion; TID = Teacher ID.

Table 10 shows the total instructional topics observed (34) and the observation frequency of the two topics, Ecosystem and Force and Motion.

Table 10

Science Topic Distribution

Topic	Frequency	Percent
Ecosystem	24	70.6
Force and Motion	10	29.4
Total	34	100.0

Completion rate. One of the goals of the evaluation was to see if the reflection activities were completed as intended. We found that only 41% of the classes that were observed completed the activities and 59% (n = 34) did not complete. Although they were close to completion, teachers consistently said that they wished they had more time. Even the teachers who completed the activity mentioned that they felt rushed and didn't have a chance to thoroughly go through the materials with the students to make sure they understood and to allow them to ask questions. Also, activity guidelines suggested that teachers review the materials and wrap up at the end of the classes. However, very few teachers were able to do so (see Table 11). Therefore, we examined the data to see where teachers spent most of their time and to investigate reasons for not completing the activities. We found that reflection activities took longer than assessments, stretching out in some cases to almost an hour in extended classes.

Table 11

Reflected Activity Completion Rate (by period)

TID	Completed	Not completed	Total	% Completed
T200	1	4	5	20%
T201	4	2	6	67%
T202	0	3	3	0%
T300	3	2	5	60%
T301	3	2	5	60%
T100	3	0	3	100%
T101	0	7	7	0%
Total	14	20	34	41%

Class organization. In reflection activities (RAs), students worked either in pairs, groups or as a whole class (see Table 12). Activities varied depending on the topic covered. Teachers followed the guidelines from the Quick Guide provided by WestEd. Only one teacher mentioned that she administered an activity before our observation, administering RA 1 across two days. She reported that this allowed her to do the activity better.

As teachers introduced reflection activities, students were very responsive, asking questions and answering teachers' questions enthusiastically. Throughout the activities, teachers became more comfortable as the day went by. Although teachers were prepared for seating arrangements and trasitioning into different tasks, many adjustments were necessary. For Force and Motion activities, especially for the nine-part story section, there was frequent regrouping and moving around before students settled into their respective groups. Consequently, students did not have enough time and rushed through their presentations. Apart from one class, all the classes spent about one quarter of the period presenting (see

Table 12). One teacher mentioned that she felt pressed for time and that the pace was not conducive for students to learn and grasp the materials, especially for developmental classes. Another teacher suggested that the guide provide more specific guidance—that is, start times and the amount of time to be spent on the first few activities. For example, the guide suggests that groups should present for five minutes; however, this was insufficient time when coupled with the Q&A session plus the classmate evaluations.

Table 12

Class Organization (RA)

TID	Individual students working alone	Pairs of students	Small groups (3+ students)	Whole class working on one activity	Student presentations	Teacher introducing activity
T200 (*N*=5)	0%	0%	37%	4%	41%	18%
T201 (*N*=6)	0%	4%	50%	2%	18%	25%
T202 (*N*=3)	0%	11%	35%	9%	22%	22%
T300 (*N*=5)	0%	0%	43%	2%	25%	31%
T301(*N*=5)	0%	6%	31%	21%	26%	16%
T100(*N*=3)	0%	0%	36%	5%	26%	32%
T101 (*N*=7)	0%	0%	44%	3%	20%	33%
Average (*N*=34)	0%	3%	41%	6%	25%	26%

Note. *N* = Number of periods observed; RA = Reflection activities; TID = Teacher ID.

As shown in Figure 6, teachers spent five to fifteen minutes in the beginning of the class explaining the activity, reviewing key points, and demonstrating to the whole class. The classes would work together as a whole while transitioning from one group activity to another. Sometimes, following the activity guide, more than one group would come together to discuss a presentation for the next activity.

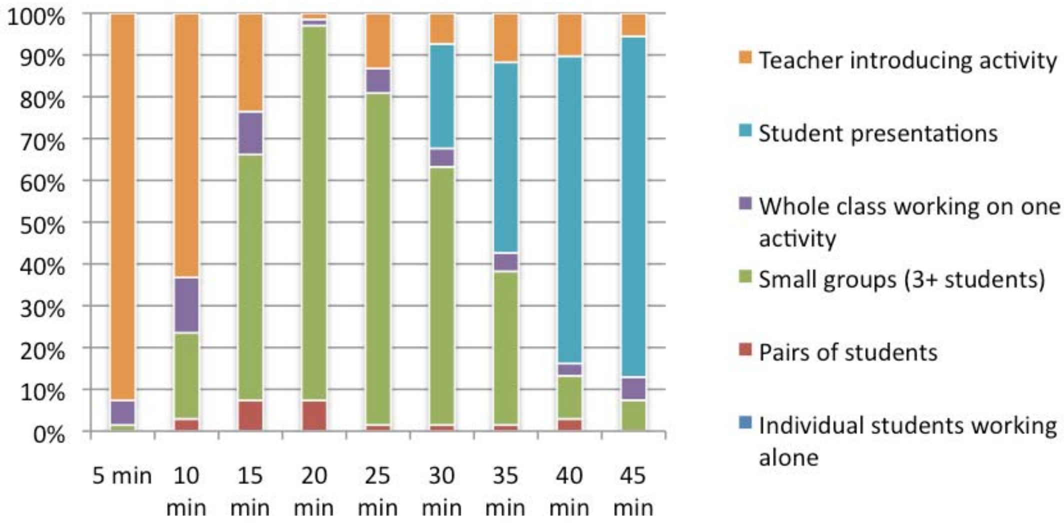

Figure 6. Class organization (RA). Note: "Teacher introducing activity" includes teachers giving students directions for the next tasks, reviewing materials, or wrapping up.

Virtually all student presentations started after the first 30 minutes. The activity guide suggests that each student presentation should be about five minutes per presentation; however, there was insufficient time for all students to complete their presentations. As shown in Figure 6, the largest percentage of student presentations occurred during the last five minutes of class, meaning that very few classes finished by the end of the period. After 30 minutes of the class time, only about 20% of the classes had started presenting, meaning that about 80% of the students had less than 15 minutes to present. Also, less than 10% of the observed classes had a chance to do wrap-ups at the end, as suggested by the reflection activity guide. There was no opportunity for the question and answer section at the end of the class.

Classroom interaction and student interaction. As shown in Table 13 and Figure 7, the majority of classes were led by students for more than half of the class period. This was in accordance with the program's guidelines. According to the observation sheet definition, "student-led" means that students dominate interactions. Discussions may be wide-ranging but on topic.

Table 13

Classroom Interaction (RA)

TID	Teacher-led	Student-led	Neither
T200 (*N*=5)	29%	44%	27%
T201 (*N*=6)	26%	74%	0%
T202 (*N*=3)	26%	67%	7%
T300 (*N*=5)	38%	52%	10%
T301(*N*=5)	39%	61%	0%
T100(*N*=3)	46%	54%	0%
T101 (*N*=7)	36%	58%	6%
Average (*N*=34)	34%	54%	12%

Note. N = Numbers of periods observed; TID = Teacher ID.

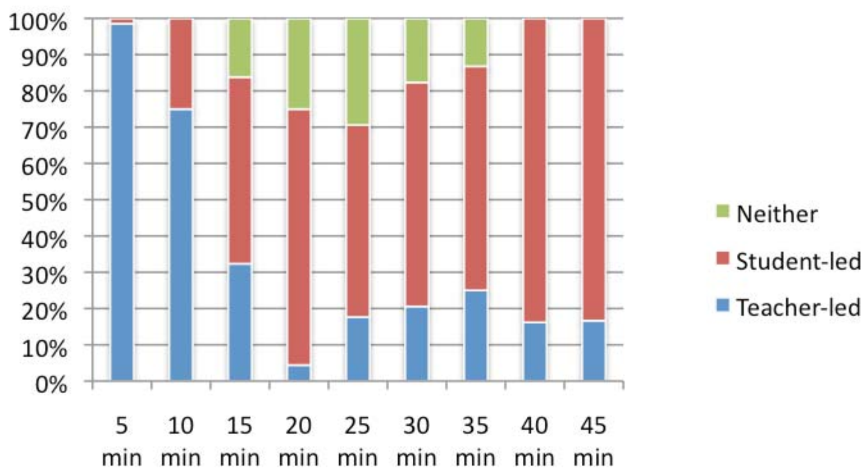

Figure 7. Classroom interaction (RA).

In most classes, students were co-constructing meaning and actively responding (see Table 14). However, we also witnessed some passive interactions among students, especially if one or a few students dominated the class group. Co-constructing meaning occurred most frequently during the middle of the class period after students were settled into their groups (see Figure 8).

Table 14

Student Interaction (RA)

TID	Passive/ little response	Active response	Co-construct meaning
T201 (*N*=6)	38%	38%	24%
T202 (*N*=3)	37%	4%	59%
T300 (*N*=5)	27%	34%	39%
T301(*N*=5)	4%	42%	53%
T100(*N*=3)	31%	48%	21%
T101 (*N*=7)	41%	39%	20%
Average	30%	36%	34%
Total periods	29		

Note. *N* = Number of periods observed; TID = Teacher ID.

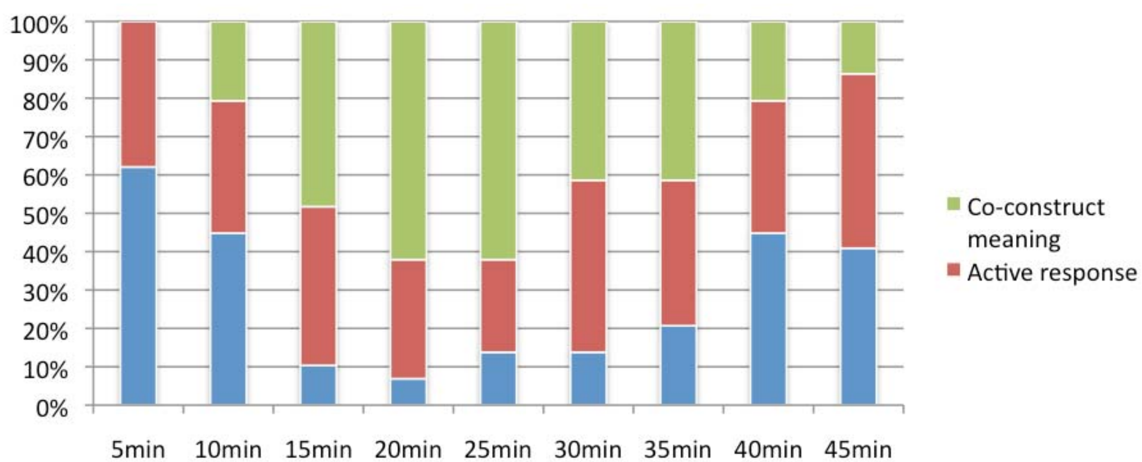

Figure 8. Student interaction (RA).

In one fifth to one half of the periods, students initiated dialogue with fellow students or the teacher and constructed their own meaning from the lesson activity. However, co-constructing meaning decreased towards the end of the classes when students rushed to finish their presentations. In almost half of the classes, students were very passive as time ran out (see Figure 8).

Table 15

Group Participation (RA)

TID	Disengaged	Unbalanced	Balanced
T201(*N*=6)	15%	51%	34%
T202(*N*=3)	10%	33%	56%
T300(*N*=5)	8%	17%	75%
T301(*N*=5)	0%	5%	95%
T101(*N*=7)	36%	46%	18%
Total	16%	32%	52%

Note. *N* = Number of periods observed; TID = Teacher ID.

For the students who worked in groups (Table 15), we observed an almost equal amount of balanced and unbalanced group participation. Students were very balanced in the middle of the class; however, they seemed unbalanced, disengaged, and distracted towards the end of class (Figure 9). Observers noticed that as the day went by, students were more tired and disengaged.

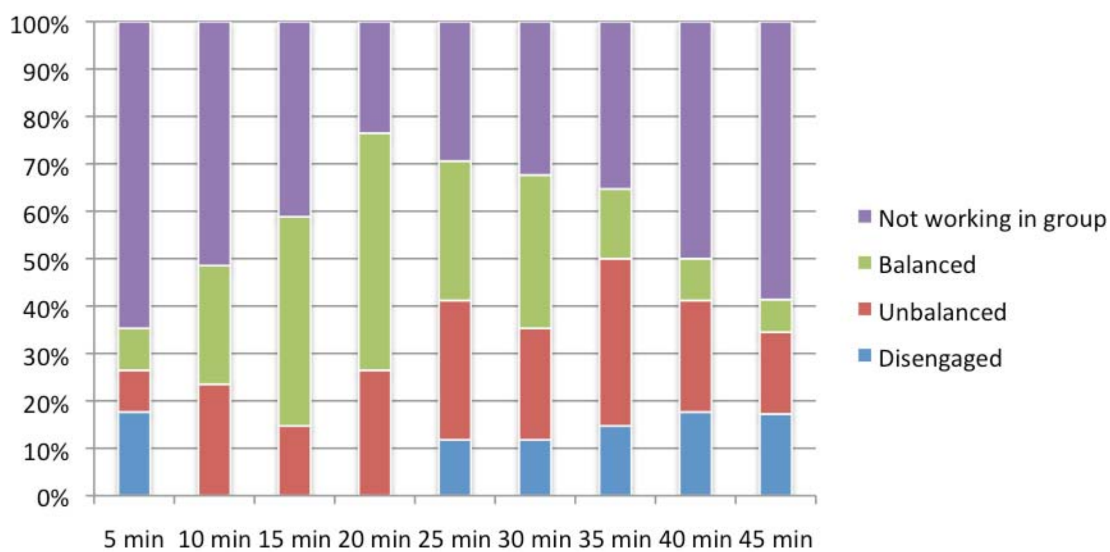

Figure 9. Group participation.

Teacher's role. During reflection activities, teachers primarily assisted students as needed (see Table 16 and Figure 10), ranging from 40% to 64% of the time. Teachers demonstrated to the whole class between 22% and 36% of the time, while during student presentations, teachers primarily monitored. As noted previously, there is very little evidence that teachers performed wrap-ups and reviews.

Most of the teachers were very well prepared and organized. For example, they planned for grouping. Nearly all teachers pre-printed the names of the students who should be sitting together and placed them on the desks before the students arrived. One teacher used a timer so that she could keep the activities on track; however, she was also among the teachers who did not have enough time.

Teachers spent a quarter to a third of the period explaining the activity, reviewing materials, and giving students instructions about grouping and arrangements. When students were in groups discussing their presentations, teachers answered their questions or walked around and monitored them. We saw two forms of teacher involvement during the presentations: facilitating students and helping them directly through feedback, or monitoring and evaluating. Some teachers tried to include the whole class by encouraging them to ask questions of the presenters, but time constraints limited this opportunity.

Table 16

Percentage Distribution of Teacher Role Categories by Teacher

TID	Teacher not involved	Demonstration to whole class	Assisting students systematically	Assisting students as needed	Monitoring
T200 (*N*=5)	0%	22%	12%	40%	26%
T201 (*N*=6)	3%	29%	6%	57%	5%
T202 (*N*=3)	0%	31%	0%	61%	7%
T300 (*N*=5)	0%	29%	11%	60%	0%
T301(*N*=5)	0%	36%	0%	64%	0%
T100(*N*=3)	0%	34%	0%	52%	13%
T101 (*N*=7)	6%	35%	7%	47%	5%
Average (*N*=34)	2%	31%	6%	54%	8%

Note. *N* = Number of periods observed; TID = Teacher ID.

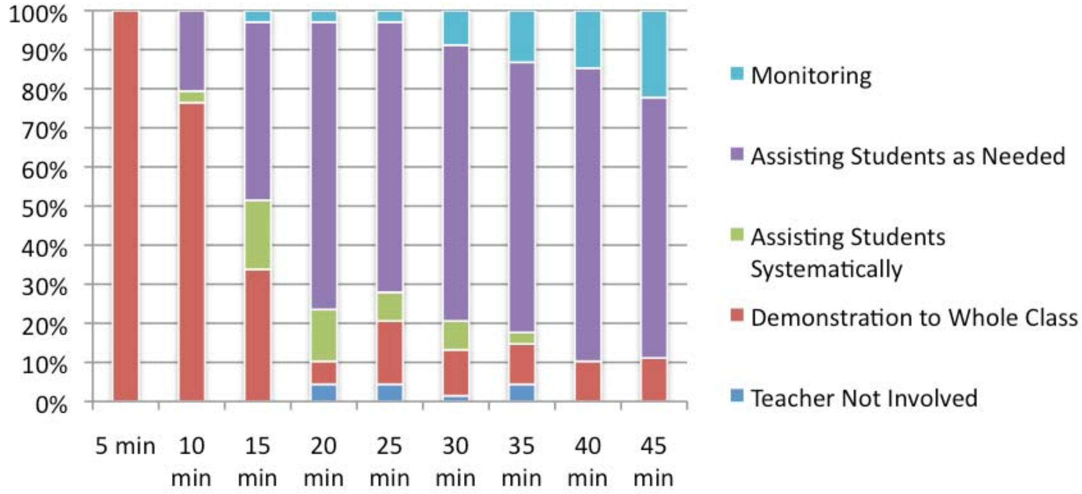

Figure 10. Teacher role (RA).

Accommodations. Table 17 summarizes the population of ELLs and students with disabilities in the classrooms observed. However, not all these students used accommodations.

Table 17

Summary of ELLs and Students with Disabilities by Teacher

TID	NumofELL	NumofSD	NumofSs
T200(*N*=10)	0	2	321
T201 (*N*=11)	0	3	334
T202(*N*=3)	3	9	90
T300(*N*=10)	4	0	185
T301(*N*=10)	0	12	205
T100(*N*=2)	0	2	61
T101(*N*=7)	0	13	209
T102(*N*=3)	1	4	92
Total	8	45	1497

Note. ELL = English language learner; *N* = Number of periods observed; SD = Students with disabilities; Ss = Students; TID = Teacher ID.

During assessments, screen magnification and text-to-speech were available for students who needed accommodations. Based on the observations, only two students used the text-to-speech accommodations. No students used screen magnification. ELLs and students with disabilities said that they liked the SimScientists program because they could work at their own pace and pause if they needed more time. In most cases, there were few differences between students who needed accommodations and those who did not. However, two classes were exceptions.

One of the five schools in the study had mainstreamed its ELLs and students with disabilities into regular classrooms. We observed two teachers and several classrooms in that school. There were only a total of five students who were considered students with disabilities, non-gifted. These disabilities included behavioral and mental cognitive problems versus physical disabilities.

Another class that deserved special reporting had four ELL students, two Hispanic and two Asian. One of them had been speaking English for three years and the others less than two years. One ELL teacher devoted her time just to these four students. During embedded and benchmark assessments, these students were in a different room working with the ELL teacher. These students did not use any of the accommodations available from the program, but received a lot of help from the teacher. Their main difficulty was that they could not understand some of the words or descriptions; therefore, they used the teacher as a dictionary. The teacher mentioned that it would be helpful to have a glossary section in the program.

Two ELL students worked together, reading the text aloud and explaining it to each other. One of the remaining ELL students worked independently, and the last ELL student needed a lot of help from the teacher. Although they needed more time, all four ELLs did as well as other students on the embedded assessments. However, they struggled on the benchmark assessments because most of the questions were open-ended, requiring a written response. This school had slow computers with technical glitches; therefore, only one out of the four ELL students finished the assessment. A second student was close to finishing, completing about 80% of the task. The two others were only halfway through.

INTERVIEW FINDINGS

Teachers

Most teachers had favorable impressions of the SimScientists program and felt that it would be a beneficial addition to their lessons as a review or recap tool. In addition, they believed the program would help them assess their students’ level of understanding. Teachers provided feedback on program implementation, feasibility, and accommodations. This section first presents teachers’ comments about the embedded and benchmark assessment, then the reflection activities.

Embedded and Benchmark Assessment

Teachers felt confident that this generation of students was easily capable of adapting to computer programs like SimScientists because of their exposure to many computer programs and electronic devices. Almost all students have had prior access to the internet, video games, iPods, and other electronics.

Teachers were asked to compare the simulated computer assessment versus traditional paper-and-pencil assessments. Teachers agreed that the simulation assessment had greater benefits than a traditional test. One benefit was that it provided students with visuals and instant feedback. Another benefit was that the simulation assessment allowed students to manipulate graphs and inquire about science at their own pace. Teachers found the simulation assessment engaging and beneficial for all their students. As one teacher stated:

> I think that they [students] are way more engaged. When I told them that we were going next door to work on the computers again, they all seemed pretty excited to go next door and work on it. If I were to just give them a worksheet, there is no way that they would get this kind of excited to do it. So just the fact that we are on the computer using this kind of interactive assessment, whether it be formative or whatever, I still think it was more engaging them simply, "Here is a pencil, here is a paper, answer questions" [T102].

Many teachers also felt that, because their students have different learning styles, they benefited from the multiple ways of interacting with the assessment. Another teacher added:

> I love the idea that the computer gives them feedback where they choose a question and then if it's not quite right, it comes up and says, "Oh, you know, you did something wrong. Go and check your answers." So I like that kind of feedback when it doesn't just go to the next page and they don't know whether they did it right or not. And I like that it goes into steps with that feedback where the first feedback is the yellow box and it says, "Go back and check your answers," but it doesn't give them the answer, so I like that, that it's work in progress [T300].

Another major benefit mentioned by two teachers was that the simulation assessment allowed them to easily see which concepts his or her students struggled with, allowing future lessons to be tailored to those concepts.

Science content. Teachers were asked how useful the embedded and benchmark assessments were for integrating content knowledge into a science system model. Only a few teachers commented. Two teachers (T201, T102) said that the assessments covered science content well. One teacher stated:

> Yes, the science content is really being tested. Students are asked to conduct experiments, investigate, and draw conclusions and to use scientific skills. Students are not able to

> guess on the multiple choice questions because it probes them until they choose the right answer. Students are also taught about food webs in one biome and they are tested on another biome [T201].

Another teacher added, “Yes students learn how science works in general. Students learn how to go about the process of investigating and how to conduct an experiment. Also, the assessments cover the predator relationship similar to how an ecologist would do it” [T102]. However, one teacher (T100) felt that the assessments covered general science reasoning skills and not science content. Statements made by teachers were broad, rather than specific.

Improvements. Overall, teachers found the computer assessments potentially more beneficial compared to the paper-and-pencil assessments. However, some teachers offered suggestions for improvement, including better assessment password procedures, improved selection of students per assessment, a demo before the first embedded assessment, and more specific prompts when students choose the wrong answer.

One reoccurring problem was that students couldn’t remember their passwords to access the assessment, as reported by four teachers (T200, T201, T101, T300). Consequently, teachers had to print out each student’s password and hand it out each day. Two teachers suggested that student identification numbers should be used as their password because students have it memorized and the passwords are confidential. Another teacher suggested that the passwords should be student’s birthdates and initials.

Teachers had other useful suggestions. A “select all” feature for assigning students to the embedded/benchmark assessment would save teacher time rather than having to click on each individual’s name. Another teacher suggested a demo for students before they did the first embedded assessment in order for them to understand how to use the program. Another teacher suggested providing students with more prompts when they made mistakes in order to better guide them through the questions.

One teacher stated,

> The only weakness that I saw was that I thought the simulations were too short. I thought that they should have a little bit more depth, a few more steps to go into things, because I think that we saw that a lot of kids were having trouble in the food web. Especially a lot of kids were having a hard time getting the arrows to go in the right direction, and a page just to say, “Ok here is a shrimp and here is an alewife, and which way would you draw the arrow?”—it would get them going, “Ok that is how the arrows are supposed to go, changing it from matter to energy in the simulation” [T101].

Importantly, four teachers mentioned that some of their students struggled with the "matter changes into energy" question in the Ecosystem assessment. Most students did not understand that they were being asked to place the arrow in the direction of the energy flow. Some teachers suggested that students be given some explanation at the beginning of this question emphasizing that the question was asking about energy transfer. Other teachers suggested a prompt when students place the arrow the wrong way, emphasizing that the arrow should point where the energy is being transferred. Force and Motion teachers stated that many students struggled with the experiment questions where they were asked to leave some values the same and only change one value in order to conduct an experiment. They suggested that more prompts be added, explicitly telling students that all variables but one must be kept constant in order to conduct an experiment.

Teachers also suggested that the computer simulations provide students with a vocabulary sheet in order to clarify some words. Words found in the assessments differed from words in students' science curriculums, which differed by state. For example, some state curriculums used the word *biome* to describe a region occupied by a large community of plants and animals, but the SimScientists Ecology assessment refers to it as *ecosystem*. Some students had not been introduced to words such as *energy* or *mass* prior to the assessment. In general, a vocabulary list would help students from all states perform better on the assessments, while especially helping ELL students or students with weak reading skills. Another teacher suggested that the assessment match state standards by individual state so that teachers could see which state standards are being met by each individual student and the class as a whole.

Computer access. Teachers were interested in implementing the computer simulation assessment after every unit as a substitute for traditional paper and pencil assessments. However, only one teacher (T101) said that there were enough computers in his school to do the computer assessment after each unit. Three teachers (T100, T200, T201) would find it problematic if they had to implement the program several times a year. These teachers (representing two schools) did not have enough computers in their schools for students to do the assessment after each unit. In addition, two other teachers (T300, T301) who both taught at the same school said that they had enough computers but that the computers were protected with Citrix security software, which did not allow them to open the SimScientists program. Because only a few of their school computers did not have Citrix, it would be difficult for them to do SimScientists for each unit.

Technical problems. Only three teachers reported having technical problems with the computer assessments. One teacher (T200) reported having a difficult time launching the

SimScientists program. Students were logged out unintentionally by the program. Another teacher (T201) reported that the program had too many error messages throughout the assessment. Consequently, some students thought they were done when they were not. The third teacher (T102) reported that the Macintosh computers were more problematic than the PC computers. On the Macintosh computers, the assessment opened behind the main SimScientists screen. Consequently, both teacher and students, neither familiar with Macintosh computers, struggled opening the assessment. This same teacher also stated that his students would select an answer and sometimes it would not record their answer. Other times it would record the answer but when he pressed the "next" button, the answer would disappear and the student would be marked wrong. They were using old computers with a slow internet connection.

Student reports. Teachers only gave feedback about the embedded reports because they had insufficient time to score benchmark assessments prior to interviews. Four of five teachers said that the reports were particularly useful for following the progress of each student. These teachers used the reports to see where each individual student struggled. Only one teacher stated that the reports were useful for gauging which questions her students struggled with as a class. Another teacher stated the students reports were not useful because they did not provide her with a full summary report on how the class scored. The same teacher suggested that the report should also include scores for each classroom so that the teacher could tailor the next lesson to the class needs. Additionally, one teacher reported that he was not allowed to access the report once comments were entered and saved for each particular student. This same teacher recommended that the reports remain accessible throughout the process in order to revisit information from prior days. One teacher did not know if the reports were useful because he had not viewed the embedded reports at the time of the interview.

Only one teacher (T301) stated that the students looked at the report after the embedded assessment and that some of her students asked her how to interpret the report. Two other teachers (T100, T300) stated that their students did not look at the report after they finished the assessment; they just closed the program. No other teachers commented on students viewing the reports.

State standards. Many of the teachers said they wanted the assessments to directly match their state curriculum. They wanted to be able to identify which state standards were met per student and per classroom. They also wanted to be able to identify which state standards were being addressed per assessment question. As one teacher stated:

> I would like to see exactly what core standards for [our state] core curriculum they are missing. It said that they are; it did really well with intended learning outcomes like science knowledge stuff, but as far as core curriculum, with what sections they were missing, it didn't really have that. I know this is designed for students in different states, but it would be nice for me to say, "Ok this student didn't learn Section 2, Letter A" [T100].

Teachers understood that the plan was to implement the assessment in all states and that states had different curriculums. However, they thought it was important to be able to match SimScientists' questions to each state's standards, supporting alignment between standards, curriculum, and state tests.

Accommodated learners/English language learners. Most teachers who were interviewed did not teach ELLs. Those that did had between one and four ELL students. Opinions about the utility of the text-to-speech accommodation among teachers were mixed. Some teachers reported that ELL students mostly benefited from having visuals and being able to interact with the assessment. One teacher (T200) said that ELL students benefited from text-to-speech because students develop aural comprehension before they develop reading comprehension. Another teacher (T101) stated the opposite, that text-to-speech was not very helpful because students who learn English as a second language have an easier time reading than listening to comprehend (this teacher learned English as a second language). A few teachers (T100, T102, and T300) suggested that adding a vocabulary list that defined difficult words throughout the assessment would specifically help ELL students with comprehension.

Accommodated learners/Students with disabilities. Six teachers had a few students with disabilities. They had different explanations as to how these students benefited from the SimScientists program. Some teachers felt that students with disabilities mostly benefited from having visuals and being able to interact with the assessment. This allowed students to gather more information than they would if they were just reading the question. Students with disabilities also benefited from the individualized feedback they received, according to teachers. One teacher stated that students with disabilities benefited from the fact that they were able to go at their own pace, which allowed them more time to think about the problem. One teacher said that one of her students with disabilities benefited mostly from text-to-speech because the student struggled with reading.

Accommodations. Most teachers stated that they did not use and did not need to use the magnification accommodation for any of their students. Their students did not have visual impairments where larger text would have been more helpful. Some teachers did say that in prior years they have had students where the magnification accommodation would have been very helpful.

Reflection Activities

Most teachers agreed that knowing students' level of comprehension (performance on the embedded assessments) before starting the reflection activities was a major benefit because they were able to tailor the activities according to students' comprehension levels. One teacher agreed with this, but she would have mixed A, B, and C students so that students who scored better could have helped students who scored lower. As one teacher explained:

> I don't always have the opportunity when we do Scantrons. I am going to have to personally go through, check each Scantron and be, "Ok, ok [number] 10 is being missed a lot. What is [number] 10?" Then look at that, and we are not always afforded the opportunity, if you will, to pull one group of kids and say you got this wrong, because what are the rest of the kids doing? The reflection activity was wonderful—it was already there for us. The computer put them in groups and…there was an obvious group that was not there. But when I kind of shuffled kids around a little bit to build a couple of groups, it still let me know which kids still really understood the concepts and that I would be able to look at one group of kids and be able to teach that group of kids in each activity. … I did it in each reflection activity. So for me that was wonderful as a teacher. I knew that that group of kids, that row of kids would have difficulty looking at the arrows, the direction of energy transfer, and I have that in front of me so that I can go back and say, "You and this group have another activity to do, and this group a different activity based on what you scored, and another group do this activity based on what you scored." [T201]

Another teacher said that the greatest benefit of the reflection activities was the format itself in that students worked in small groups and had the opportunity to present to the class. Her students normally do not have the opportunity to work in small groups or create and deliver presentations.

Improvements. Most teachers agreed that the reports were not useful for creating the reflection activity groups. Six of eight teachers stated that they struggled making the groups based on the embedded assessment scores. Their students scored similarly; many classes had a lot of A, B, or C students. Teachers often used their own knowledge and judgment of which students should be in a higher or lower group rather than the level suggested by the embedded assessment. One teacher (T202) had three classes that were already divided into low-, medium-, and high-level science comprehension groups. Therefore, many of her students from the same class scored similarly. Having students with similar embedded scores forced the teachers to gauge their own students' science knowledge, and many also used student behavior as a factor to choose which section they were assigned to. Having an uneven number of students per group and students with differing behaviors caused teachers to spend a lot of time creating groups, especially for those reflection activities that required teachers to create two sets of different small groups with the same students.

Many teachers stated that one day was not enough time to do a reflection activity for both the ecosystem reflection activities and the force and motion reflection activities. Students were rushed from one section of the activity to the next. There was no time to ask questions, to correct students, or to transition. Two teachers, one teaching Ecosystem and one teaching Force and Motion, said they struggled reading the reflection activity directions. Both teachers felt that the directions were too long and that having the same directions in different formats created more work for the teacher. Teachers felt they had to read each page in order to do the activity correctly, and many of the directions were repeated.

Some teachers expressed other concerns. Two teachers (T100, T202) felt that the reflection activities needed to be more challenging. Another teacher (T102) felt that the worksheet directions and questions were too vague and that the worksheet levels of difficulty for groups A, B, and C were too different from one another. Consequently, some students finished earlier than other students and had nothing to do. Teacher 102 also suggested that the embedded assessment should provide raw scores for teachers to determine if a student should be moved into a lower or higher reflection activity group. One teacher (T300) did not have a class of 27 students, so when her class was doing one of the reflection activities, she was not able to divide the class into nine groups as stated in the directions.

State Coordinators

Overall state coordinators provided positive feedback. They reported that the system worked well and teachers were willing to participate. There were no major complaints or negative feedback. Only one state coordinator reported witnessing any problems and these were logistics such as students forgetting passwords and teacher difficulties setting up and managing activities. Coordinator shared feedback from teachers that they were impressed with the software and activities and would welcome participate again. Impressed with teachers' reactions and the nature of the assessments and associated reflection activities, coordinators also were interested in knowing plans and topics for future development and likely topics to be developed. They also encouraged development and implementation in subject areas beyond science, such as mathematics.

SURVEY FINDINGS

In our analyses of EAG teacher survey data, we focused on the usefulness of the simulation and students' accommodations usage. Table 18 shows the exact number of teachers for each of the seven survey forms.

Table 18

Number of Teachers Who Completed the EAG Teacher Survey Data

Ecosystem survey form	# of teachers	Force and Motion survey form	# of teachers
ECO benchmark	30	FM benchmark	19
ECO embedded and reflection 1	33	FM embedded and reflection 1	22
ECO embedded and reflection 2	32	FM embedded and reflection 2	20
		FM embedded and reflection 3	20

Note. ECO = Ecosystem; FM = Force and Motion.

Usefulness of Simulation

One focus in our analysis of teacher survey data were teacher feedback on assessments and activities. Specifically in the survey forms, teachers were asked about the usefulness of the embedded reports, how closely they followed the activity guides and how useful the embedded assessment and reflection activities were for students' learning. From our initial analyses, we found that teacher ratings were quite similar across different units (i.e., Ecosystem and Force and Motion) and across different classroom activities and assessments. Thus, we present the average for the same survey questions across different units and different assessments in Table 19.

Table 19

Descriptive Statistics for Selected Survey Items

Survey questions	*N*	Mean	Std Dev
Embedded reports are useful	49	2.9	0.78
Usefulness of Reflection Activity (RA)			
RA supplements instructions	49	3.27	0.62
RA differentiates instructions	48	3.31	0.81
RA helps review content targets	51	3.21	0.46
RA helps review inquiry targets	48	3.12	0.55
Usefulness of Embedded Assessment (EA)			
EA helps understand content targets	51	3.19	0.51
EA helps understand inquiry targets	48	3.10	0.61
Logistics			
Teacher followed suggested RA grouping	50	3.06	0.62
Teacher followed design of RA	50	3.20	0.58

For each survey question, teachers were able to choose from responses (a) *not at all*, (b) *somewhat*, (c) *fairly well*, and (d) *completely*, coded 1, 2, 3, and 4 to represent the magnitude of teachers' agreement with each statement. As shown in Table 19, except for the questions about embedded reports, the average rating of each survey question is above 3.00. This indicates that, in general, teachers agree with the survey statements. In particular, teachers felt that the embedded assessment helped students understand both content and inquiry targets. In addition, they felt that reflection activities were helpful in reviewing content and inquiry targets. For logistics, most teachers stated that they followed the student grouping guides and the design of the reflection analysis.

Accommodations

We also analyzed accommodations usage during the benchmark and reflection assessments. Teachers were asked whether their students used any of the 10 accommodation options (see Figure A4 in the appendix) during the assessment or any of 13 accommodation options (see Figure A5 in the appendix) during science class and its reflection activity.

In Figure 11, we present Ecosystem accommodations used in the classroom setting. As shown, extended time (ET) and read aloud (RA) were used more often by students who needed accommodations. Figure 12 shows the percentage of teachers who indicated the usage of different kinds of accommodations in benchmark assessments and embedded assessments for Ecosystem. Specifically, ET, audio (AU), and headphone (HA) were the three mostly used accommodation options for students with disabilities and ELLs. The use of accommodations was similar for teaching Force and Motion and the related assessment (see Figures A4 and A5 in the appendix).

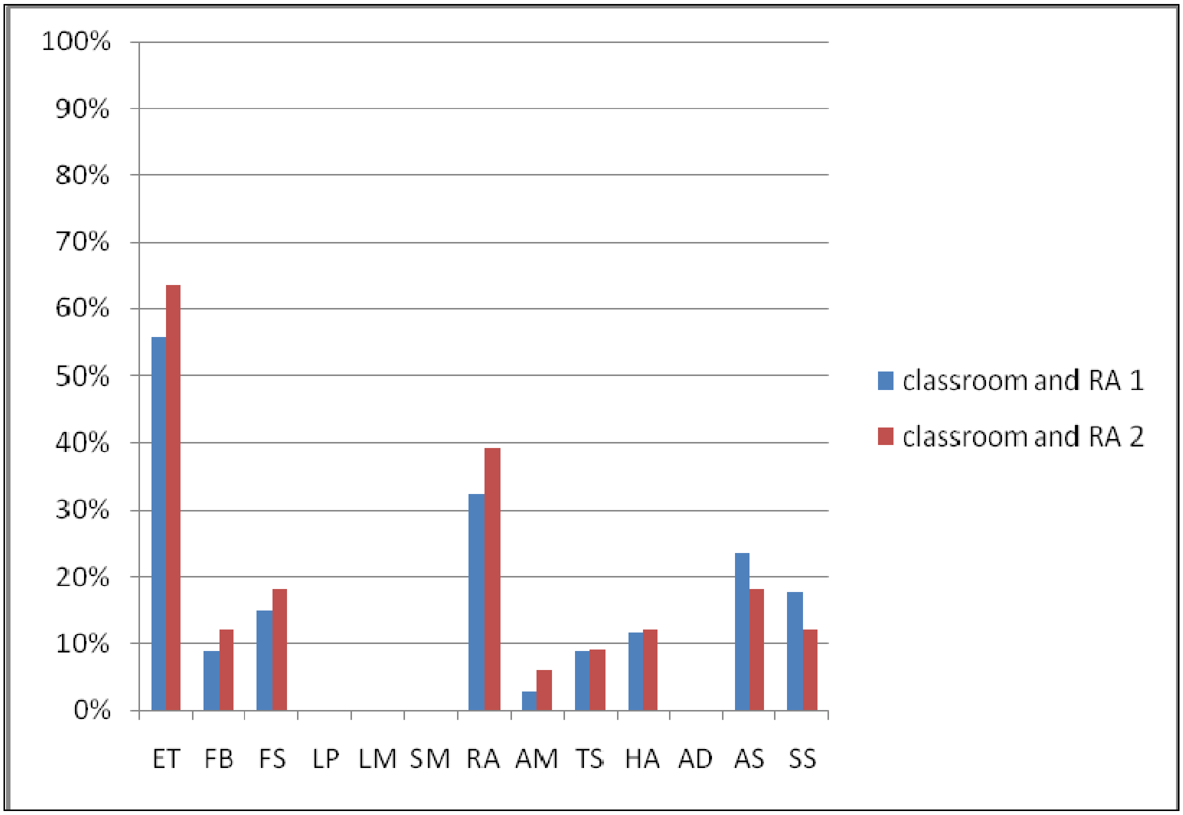

Figure 11. Accommodations during science classroom activities (including reflection activity) for Ecosystem. ET = Extended time; FB = Frequent breaks; FS = Flexible scheduling; LP = Large-print materials; LM = Larger monitor for computer; SM = Screen magnification for computer; RA = Read-aloud; AM = Audio materials; TS = Text-to-speech; HA = Headphones for audio materials; AD = Assistive devices for computer; AS = Alternate seating in classroom; SS = Separate setting.

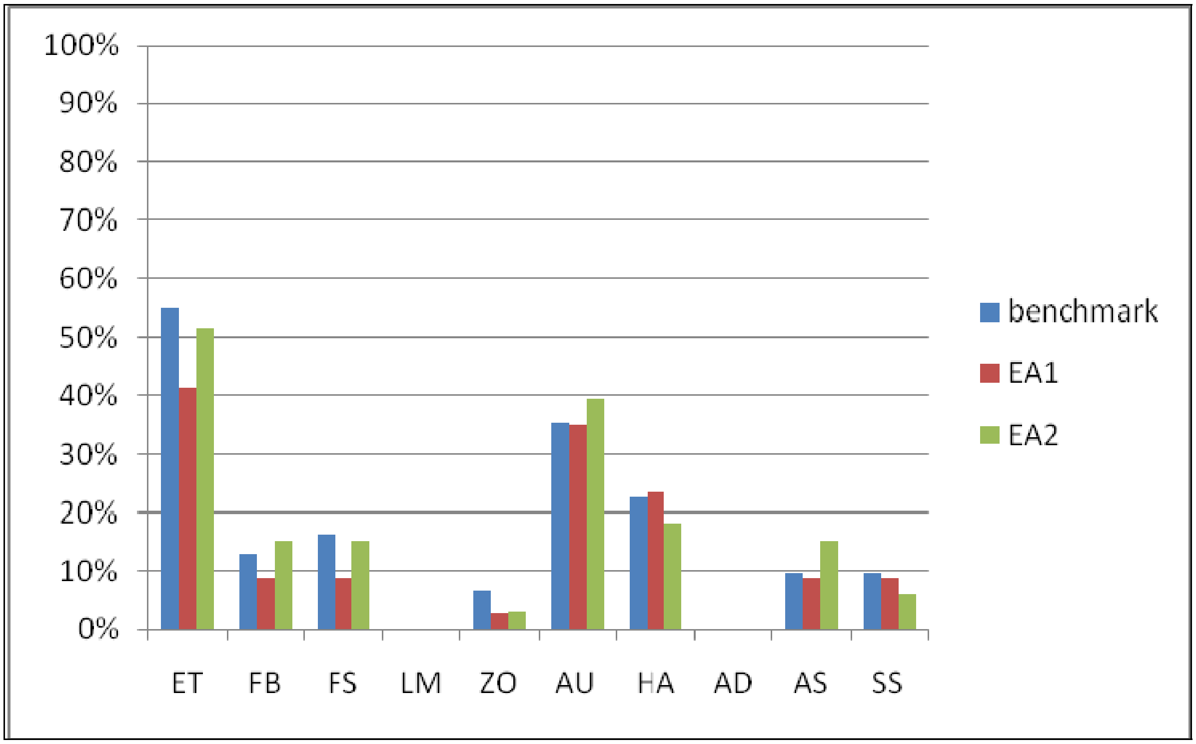

Figure 12. Accommodations during assessment for Ecosystem.
ET = Extended time; FB = Frequent breaks; FS = Flexible scheduling; LM = Larger monitor for computer; ZO = Zoom; AU = Audio; HA = Headphones for audio materials; AD = Assistive devices for computer; AS = Alternate seating in classroom; SS = Separate setting.

In addition to different accommodation options used, Table 20 presents basic statistics about the number of ELLs and students with disabilities who received accommodations during instruction or during the embedded assessment. It further shows how many students received accommodations in their Individualized Education Program (IEP) or 504 plan. During instruction, about 85% of the students receiving accommodation had an IEP or 504 plan. Among these students, there were very few ELL students. For students who used accommodation during the embedded assessments, about 87% of them had an IEP or 504 plan. In general, these findings drawn from teacher surveys support our classroom observations regarding accommodations usage by ELLs and students with disabilities.

Table 20

Average Number of Students Using Accommodations in Ecosystem.

Variable	*N*	Mean	Std Dev	Minimum	Maximum
Accommodations during Instruction					
Total	48	8.52	14.60	0	81.33
Student with IEP or 504	46	7.88	7.11	0	25.67
ELL Student with IEP or 504	41	1.27	2.55	0	11.5
ELL Student without IEP or 504	36	1.88	5.36	0	25.5
Accommodations during Embedded Assessment					
Total	46	8.40	14.40	0	81.33
Student with IEP or 504	44	7.39	7.22	0	25.67
ELL Student with IEP or 504	41	1.11	2.21	0	11.5
ELL Student without IEP or 504	38	1.27	4.21	0	25.5

SUMMARY AND RECOMMENDATIONS

In this section, we summarize study findings with regard to issues of program implementation and feasibility, teacher roles, and the nature of student interactions during SimScientists. Each section separates findings for the SimScientists' assessments from those of the reflection activities.

Program Implementation and Feasibility

Assessments

Overall, both teachers and students responded favorably to the SimScientists computer simulation assesments, both embedded formative and end-of-unit benchmark assessments. Teachers rated nearly all of the questions on their surveys 3 or higher on a 4-point scale (see Table 19). Observations provided evidence that students were active and engaged during the assessments and that teachers gave positive feedback when interviewed. Teachers collectively agreed that the simulation assessments had greater benefits than traditional paper-and-pencil tests because of the simulation's instant feedback, interaction, and visuals. Logistically, most teachers stated that they needed computers to be more easily accessible in order to implement the computer assessment several times in the year. Observations and interviews, both with teachers and state coordinators, suggested that teachers and students were highly satisfied with SimScientists and able to implement the assessments effectively.

Reflection Activities

While teachers were positive about the reflection activities, full implementation of the reflection activities proved less feasible. Teachers were able to complete the activity in less than half the reflection activities observed. Teachers reported in surveys that the activities helped to supplement instruction, plus to review content and inquiry targets. Teachers also stated they followed the design of the reflection activities, and observations supported that finding. In addition, students were engaged when they discussed their assigned science topics in their small groups, and teachers found the presentations a good opportunity to have students teach the class.

Several teachers thought that the activities were too long and too difficult to complete in one day. Among their suggestions for improving the activities was to extend them over two days or remove some components. Nonetheless, many teachers agreed that the reflection activities were a good addition to the assessment and it helped students to better understand the science topics. Overall feedback from state coordinators suggested that teachers were positive about the activities and willing to participate again.

Teacher Role

Assessments

According to the observations and as expected, the teacher's role was more passive during assessments compared to the reflection activities. For the assessments, teachers explained the procedures and assigned computers to the students in the beginning of the session. After that, teachers mainly helped students as needed or simply monitored student engagement.

Activities

The reflection activities required teachers to do intensive preparation and be well organized. Because the tasks were tightly scheduled within activities, teachers had to make sure the tasks proceeded on schedule. Some teachers grouped students, arranged tables and printed out activity sheets prior to the lesson. Even with the preparations, most of the teachers felt rushed and needed more time, according to the interviews. During the majority of reflection activity time, teachers were observed directly assisting students. Also as expected, observations showed that teachers were more actively engaged and involved with the activities than with the assessments.

Student Interaction and Engagement

Assessments

According to the observations, in all classes save one, students were actively engaged in the assessments more than 75% of the time. Interviews provided evidence that teachers found the simulated computer assessment more engaging for students than traditional assessments.

Activities

During reflection activities, we witnessed students co-constructing meaning for about one third of the class period. We also observed passive students with few responses, especially if one or a few students dominated the activity.

Accommodations

Out of 1,497 students observed in all schools, there were only 8 ELL students and 45 students with disabilities. Although teachers could have made SimScientists' accommodation features available to these students, they did not necessarily do so. Only two students used text-to-speech, and no student used screen magnification. In one school, there were 4 ELL students in one class and an ELL teacher allocated to just those students. (See the appendix for additional accommodations information.)

Recommendations

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

Embedded and Benchmark Assessments

1. Teachers were very positive about their SimScientists experience and indicated that, if available, they would use it with more instructional units. However, they noted the challenge of securing regular access to a sufficient number of computers for full classroom implementation. While pairing students on computers is one approach to improving access, our limited observation of students working in pairs revealed other logistical issues. Teachers seemed resistant to the strategy, particularly if they were able to access computer labs with a sufficient number of computers for each student. Schools should consider how access may be increased so that each student has his/her own computer during the simulation.

4. SimScientists developers aligned their materials with state and national science standards. However, the close alignment between SimScientists and their state standards was not always apparent to teachers. Moreover, there were instances of inconsistency in terminology between specific state standards and concerns that SimScientists did not provide adequate practice in core, topic-related vocabulary likely to appear on state tests. Program developers should consider how the alignment between SimScientists and specific state standards can be more specifically communicated to teachers. Consider also whether there might be minor, state-specific adaptations to deal with variations in terminology and whether it would be valuable to incorporate vocabulary work as a pre-assessment activity.

5. While teachers and students gained experience with SimScientists as they used it, they tended to have difficulty with the first assessment. Program developers should consider whether it would be valuable to add a tutorial or practice opportunity prior to the first assessment.

6. Teachers were positive about the potential value of the accommodations included in SimScientists, yet they rarely assigned or used them. There was confusion about who could use the accommodations, and there were differences in pedagogical beliefs about the value of text-to-speech accommodations for ELLs or struggling readers. Program developers and schools should consider how to communicate more clearly to teachers the value of the accommodations and the decision rules for assigning them. In addition, a few teachers suggested that SimScientists add automated glossaries to the program, hot buttons that would enable EL students to access the meaning of difficult vocabulary words.

7. SimScientists included interesting reporting features. Using color-coded displays, students could see at a glance how they did on the embedded assessments and where they needed to improve. However, the reports were little utilized by students. Developers should clarify the purpose of the reports and consider providing directions to teachers for better engaging students with them.

8. While SimScientists suggests that students work in pairs for the assessment, just one teacher used this organization format and was confused about the meaning of results. For example, did the results apply equally to both students? More explicit guidance would be helpful about using results in this context.

Reflection Activities

1. Teachers were positive about the reflection activities but were challenged to complete all components of the activity. Program developers should either recommend that the activity be conducted over two class periods or consider how the activities might be simplified for completion in a single period.

2. Teachers appreciated the opportunity for differentiated instruction, but frequently the number and distribution of their students did not match the activity plans (e.g., too few B students or insufficient students for the nine-part story). Consider how to make the reflection activities more flexible.

3. While teachers became more comfortable after using the activities with several classes, more training in how to implement the reflection activities would have been beneficial. Developers should consider how training and support for the reflection activities could be strengthened (e.g., through additional demonstrations, simulated practice, or video training). Frequently Asked Question files could be used to help teachers adapt the materials to their classroom contexts or deal with common problems.

4. In addition to giving explicit directions about report usage, specific information needs to be given about accommodations. For example, one ELL, rather than use text-to-speech, had another student read to him. Policies need to be outlined for how students with accommodations might work together if this is to be utilized in the future.

Appendix:

Embedded and Benchmark Assessments

Table A1

Summary of Assessments and Classroom Composition

TID	Number of periods	Length of period (min)	Average time taken for the assessment (min)	Grade	Average number of students	Average number of female students	Average number of male students	Topic
Embedded assessments								
T200	5	52.80	44.00	6	32.40	14.40	18.00	Eco
T201	5	52.80	36.00	6	31.60	16.00	15.60	Eco
T202	2	45.00	40.00	7	28.00	19.50	8.50	Eco
T300	5	49.00	45.00	7	20.80	9.20	11.60	FM
T301	5	49.00	41.00	7	20.20	10.60	9.60	FM
T101	7	48.00	42.14	8	29.86	14.57	15.29	Eco
T102	3	46.00	35.00	8	30.67	12.00	17.50	Eco
Total/Avg	32	49.44	40.94	7	27.56	13.42	13.97	
Benchmark assessments								
T200	5	52.80	43.60	6	31.80	14.20	17.60	Eco
T201	6	53.17	41.67	6	29.33	13.50	14.50	Eco
T202	1	45.00	30.00	7	34.00	17.00	17.00	Eco
T300	5	37.80*	33.80	7	20.25	9.50	10.75	FM
T301	5	34.00*	30.80	7	20.80	10.80	9.80	FM
T100	2	72.00	45.00	8	30.50	19.50	11.00	Eco
Total/Avg	24	47.13	37.96	7	26.74	13.00	13.19	
All assessments								
T200	10	52.80	43.80	6	32.10	14.30	17.80	Eco
T201	11	53.00	39.09	6	30.36	14.89	15.11	Eco
T202	3	45.00	36.67	7	30.00	18.67	11.33	Eco
T300	10	43.40	39.40	7	20.56	9.33	11.22	FM
T301	10	41.50	35.90	7	20.50	10.70	9.70	FM
T100	2	72.00	45.00	8	30.50	19.50	11.00	Eco
T101	7	48.00	42.14	8	29.86	14.57	15.29	Eco
T102	3	46.00	35.00	8	30.67	12.00	17.50	Eco
Total/Avg	56	48.45	39.66	7	27.22	13.25	13.65	

Note. *Delayed opening.

Table A2

Classroom Organization

All assessments									
	5 min	10 min	15 min	20 min	25 min	30 min	35 min	40 min	45 min
Individual students working alone	44	47.5	53	53	53	51	37	26	19
Pairs of students	2	2	2	2	2	2	2	1	0
Others	10	6.5	1	1	1	1	1	1	1
Missing*	0	0	0	0	0	2	16	28	36
Total	56	56	56	56	56	56	56	56	56
Embedded assessment									
	5 min	10 min	15 min	20 min	25 min	30 min	35 min	40 min	45 min
Individual students working alone	26	28	30	30	30	29	19	14	10
Pairs of students	1	1	1	1	1	1	2	1	0
Others	5	3	1	1	1	1	1	1	1
Missing*	0	0	0	0	0	1	10	16	21
Total	32	32	32	32	32	32	32	32	32
Benchmark assessments									
	5 min	10 min	15 min	20 min	25 min	30 min	35 min	40 min	45 min
Individual students working alone	18	19.5	23	23	23	22	18	12	9
Pairs of students	1	1	1	1	1	1	0	0	0
Others	5	3.5	0	1	0	0	0	0	0
Missing*	0	0	0	0	0	1	6	12	15
Total	24	24	24	25	24	24	24	24	24

*Missing at the end of observation usually means class ends early or all the students finish the assessment early.

Table A3

Summary of Assessments

TID	Embedded assessment	Benchmark assessment	Total
T200	5	5	10
T201	5	6	11
T202	2	1	3
T300	5	5	10
T301	5	5	10
T100	0	2	2
T101	7	0	7
T102	3	0	3
Total	32	24	56

Class Organization Figures

This serves as an example to show that there is no significant difference between Embedded and Benchmark assessment to report them separately.

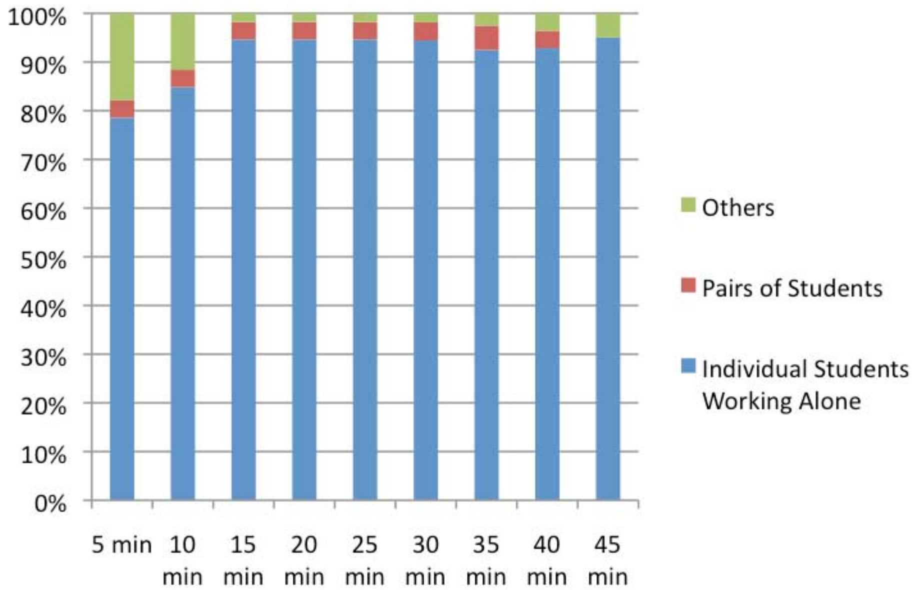

Figure A1. All assessments.

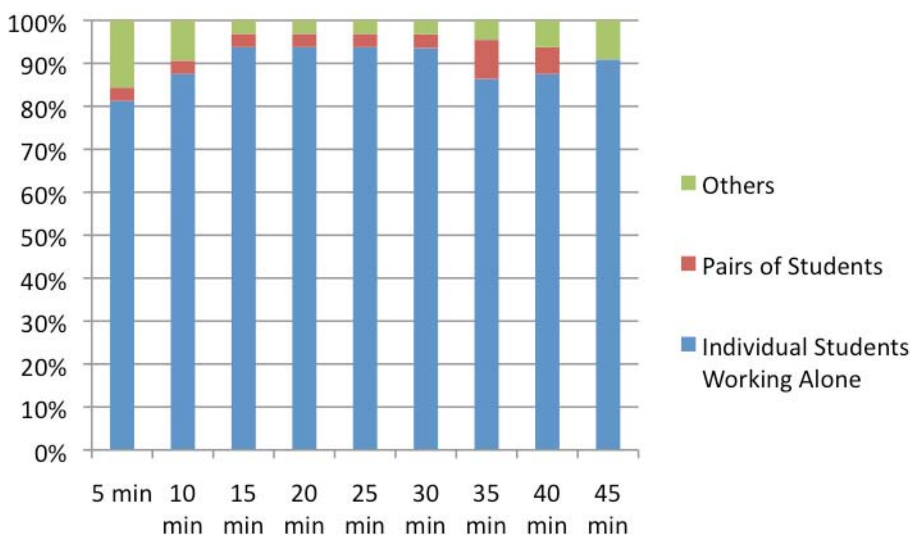

Figure A2. Embedded assessments.

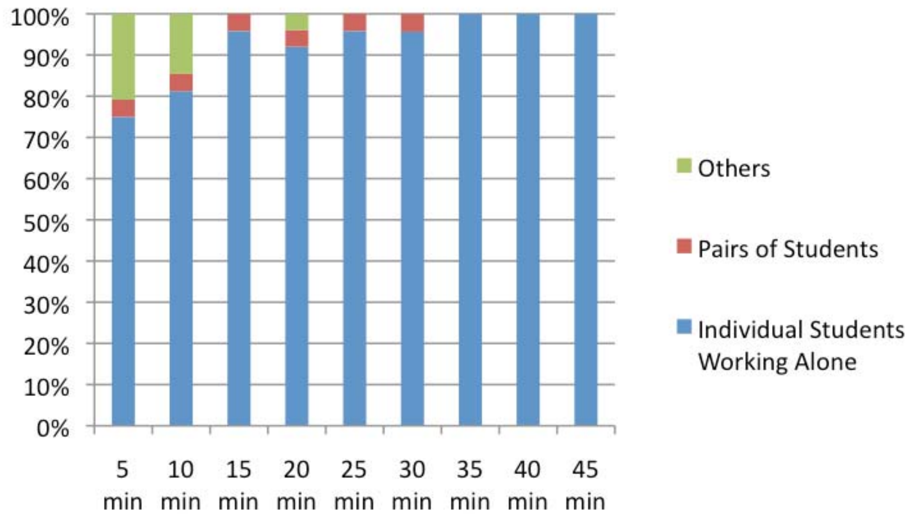

Figure A3. Benchmark assessments.

Accommodations

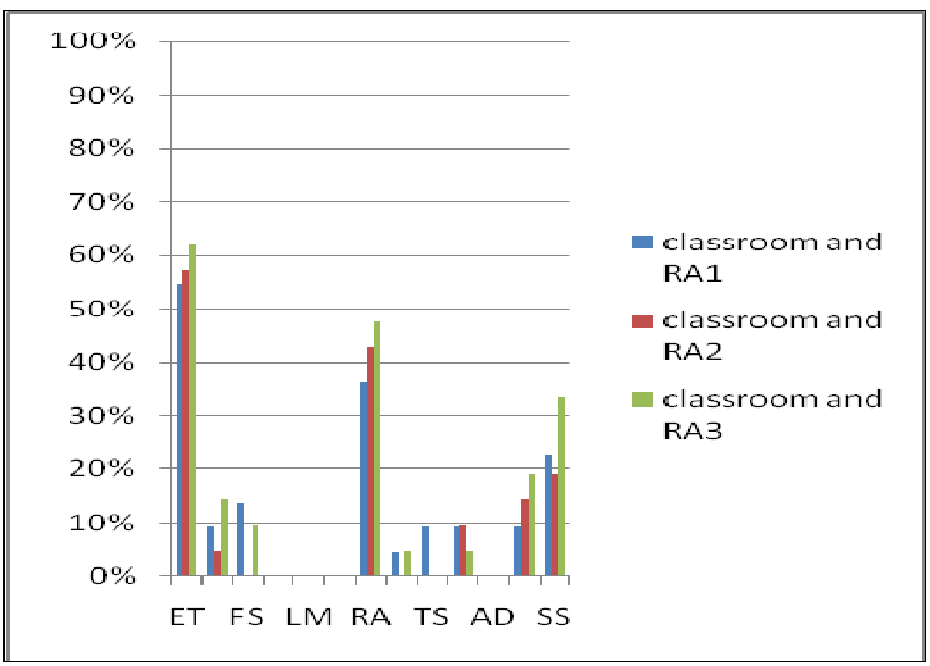

Figure A4: Accommodations during science classroom (including reflecitve acitvity) for force and motion.

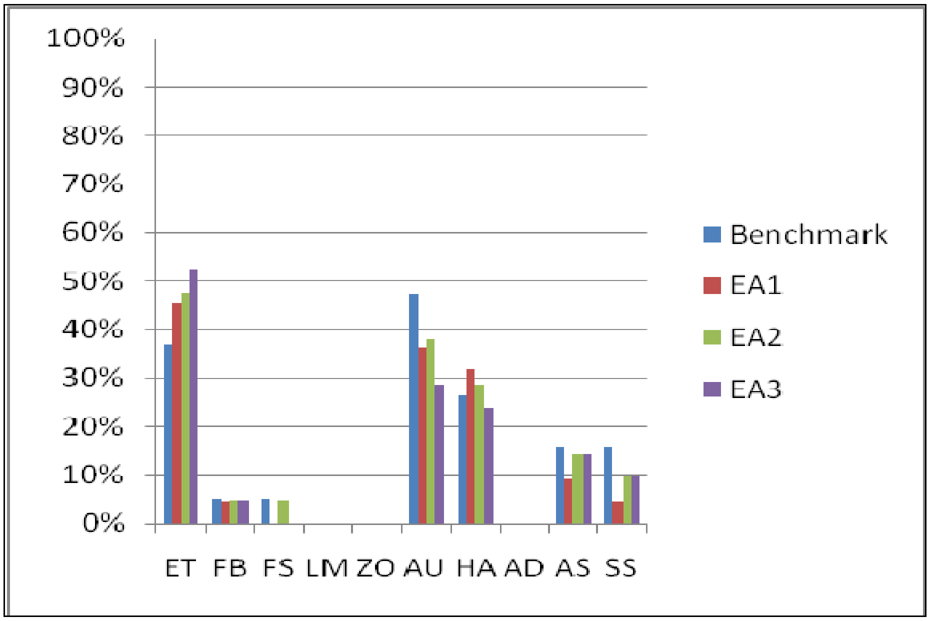

Figure A5. Accommodations during assessment for force and motion.