The Computational Literacy Final Report: Major Research and Education Findings

Overview

Computational Literacy: A Study of the Efficacy of Computational Science in High School Biology and Earth and Space Science Classrooms was a three year research and development project funded by an NSF / IERI grant. A multidisciplinary research and development team assembled by the Education Development Center's Center for Children and Technology (EDC/CCT) in collaboration with the Krell Institute, Maryland Virtual High School, and the University of Northern Iowa worked at the intersections of education and computational science to address the challenge of building students' awareness of and facility with computational models, particularly simulations; developing both teachers' and students' computational literacy.

The evolving notion of computational literacy is defined by the project as:

An individual's capacity to understand the relationship between domain knowledge and the mathematical and visualization/modeling processes that are the building blocks of computational science

Over the course of the three-year project, the interdisciplinary project team worked to satisfy the requirements of two interdependent components of the project: research and development. Project staff have designed, developed, and:

- Tested four topic specific computational simulations and the supporting materials that combine to create topic modules
- Tested the Computational Laboratory, the project's website that houses the topic modules
- Conducted formative research and field-testing of the student outcome measures, the topic modules, and the Computational Laboratory
- Distributed project related student outcome measures, a web-based teacher survey, research study protocols for participating teachers, and a number of materials essential for the implementation of the research study
- Implemented an experimental research study focused on four computational simulations developed and enhanced by the project in years one and two. Two of the four simulations were associated with biology: *Population Dynamics* (PD) and *Spread of Disease* (SoD) and two were associated with Earth / Space Science: the *Carbon Cycle* (CC) and the *Rock Cycle* (RC).

The research component of the Computational Literacy project culminated in the 2006 – 2007 Computational Literacy Experimental Research Study conducted in three states: Maryland, Iowa, and Tennessee, with a total of 44 participating teachers, 1,542 participating students, and affecting an additional 1,639, all from 21 schools. The analysis of data collected as a result of the study is described in this section of the project's final report including:

- Findings about the alignment of research study science topics and the high school science courses that accommodated those topics in participating schools and classrooms
- Teacher participants, participation, and impact on the research study
- Student participants, analysis of and findings from study outcome measures
- Technical environments that either supported or were found to hinder the Computational Laboratory, the project website, and
- Limitations of the experimental research study

In addition, this report section documents:

- Issues related to teacher recruitment, retention, and participation
- Lessons learned about the design and implementation of an experimental research study dependent on the recruitment, retention, and participation of science teachers of diverse high school students in a variety of educational settings, spanning different states.

~ NOTE: Findings in this section are indicated by italics preceded by a tilde (~) similar to this note.

Computational Literacy Research Study Design

The experimental research study examined two research groups. Those teachers and students participating as part of the "treatment" group used the project's computational simulations to test the:

Computational Literacy project hypothesis that:

Students who use our topic simulations will show greater understanding of the science content as evidenced by their pre- and post-test scores, and better critical thinking and problem solving skills as evidenced by their far transfer test scores (our Logic Test).

Those teachers participating as part of the "control" group taught as they normally do, covering the same topics as those participating as part of the treatment group. Their students engaged with the relevant science topic/s (PD, CC, SoD, RC) without benefit of the project's computational simulations.

The primary goal of the Computational Literacy Experimental Research Study was to test the impact of computational models on science learning with students from diverse academic, ethnic and socio-economic backgrounds as measured by three different outcome measures:

- PRE- AND POST- TEST -– measured the extent to which there were pre-existing differences; baseline knowledge and science learning (treatment and control)

- PERFORMANCE ASSESSMENT -- our Driving Questions aimed to gauge students' conceptual understanding of the computational model and their ability to recognize real-world connections, as well as the affordances and limitations of the model (treatment only)

- A MEASURE OF FAR TRANSFER -- our Logic Test measures students' ability to solve novel problems that are different from the problems they have been solving in their science classes (treatment and control)

The sample of treatment schools and control schools was randomly selected by state using the SPSS sample command ("Select cases→Random sample of cases→Exactly _ cases from the first _ cases") that is functionally equivalent to using a table of random numbers to select schools. Specifically, schools were sorted by the total number of units (RC, PD, SoD, CC), projected to be taught in order to balance the breadth of content coverage. Selected schools were assigned to the treatment group and non-selected schools were assigned to the control group. This process was repeated for each state individually. Those teachers and students participating as part of the "treatment" group used the project's computational simulations to test the research study's hypothesis. (See above.)

As part of the research study, students were administered a pretest in each topic encountered in their science class that their teacher had decided to implement with her students. Ideally, topic pretests were administered in the first week or two of the school year or semester. A posttest was administered after all topic lessons had been completed and/or after students had engaged with a topic simulation and participated in a follow-up / debriefing.

The measure of far transfer was administered to students at the end of the semester or school year when fidelity of implementation was adhered to.

<u>Methods: Data Collection / Data Sources</u>

Data for the experimental study was collected from a number of sources. Publicly available School Report Cards were collected for each participating treatment and control school when the information was available. Teachers were asked to complete an online survey and to fill out and submit Topic Teacher Reports (fidelity of implementation documents) for each topic addressed with students – treatment and control. Class rosters, student permission slips, pre-, post-, and Logic Tests combined to create a profile of participating treatment and control students. Driving Questions offered distinct information about treatment students as they engaged with the simulations. The existence of a set of Driving Questions also acted as confirmation that students had actually engaged with a topic simulation.

Computational Literacy Experimental Research Study Participation

The Computational Literacy Experimental Research Study conducted during the 2006 – 2007 school year was implemented in three states: Maryland, Iowa, and Tennessee. These three states were selected because of their previous work with high school students in modeling and visualization in science; Maryland because of its involvement in the NSF funded *CoreModels*[1] program which was built on what had been learned through their *Maryland Virtual High School* project[2]; Iowa and Tennessee because of their previous involvement with *Adventures in Supercomputing* (AiS)[3]. In addition, members of the Computational Literacy project team had identified, acquired, and maintained advocates, allies, and supporters of high school computational science, in the three selected states.

Prior to recruiting science teachers for the study, we projected that when the research study was completed there needed to be 36 treatment teachers and 36 control teachers; 18 biology treatment / 18 control; 18 earth/space treatment / 18 control. In August 2006, the beginning of the school year and the beginning of the research study there were potentially 82 teachers 41 treatment and 41 control from across the three participating states. The table below summarizes both projected and actual participant numbers. By the end of May 2007, the end of the school year and the end of the research study, there were a total of 44 science teachers who had participated in the research study. Of the 44 science teachers, 30 had participated from Maryland, 14 treatment and 16 control. In Iowa, a total of 9 science teachers remained -- 3 treatment and 6 control and Tennessee's total number of participating science teachers was 5 -- 2 treatment and 3 control. Table 1 compares the projects projected numbers to the actual number of participants in the Computational Literacy Research Study.

Table 1: Projected Participation vs. Actual Participation

	August 2006 – Projections (T + C = Total)		
	Schools	Teachers	Students
Maryland	5 + 5 = 10	28 + 28 = 56	2500 + 2000 = 4500
Iowa	4 + 4 = 8	6 + 7 = 13	400 + 600 = 1000
Tennessee	5 + 5 = 10	7 + 6 = 13	600 + 500 = 1100
	14 + 14 = 28	41 + 41 = 82	3500 + 3100 = 6600
	May 2007 – Actual (T + C =Total)		
	Schools	Teachers	Students
Maryland	5 + 5 = 10	14 + 16 = 30	417 + 645 = 1062
Iowa	2 + 4 = 6	3 + 6 = 9	175 + 198 = 373
Tennessee	2 + 3 = 5	2 + 3 = 5	43 + 64 = 107
	9 + 12 = 21	19 + 25 = 44	635 + 907 = 1542

Issues Related to Teacher Recruitment, Retention, and Participation

A number of issues impacted the project's ability to both recruit teachers for the experimental research study and then retain them once they had decided to participate. Issues surrounding

---

[1] *CoreModels* was funded by the National Science Foundation (NSF) from 1997-2000

[2] Maryland Virtual High School project also funded by the NSF (RED-9355806)

[3] Adventures in Supercomputing (AiS) was funded by the Department of Energy from (1991-1999)

teacher recruitment for participation in the study included: 1) the failure of science coordinators and/or school administrators to involve potential teacher participants in the original decision to participate, 2) teacher's anticipation of high stakes testing, the timing of the experimental research study, and personal considerations, 3) dissatisfaction with the assigned study group and 5) issues surrounding the technology to be used.

~ The failure of some science coordinators and/or school administrators to involve potential teacher participants in the original decision to participate in the Computational Literacy Study resulted in a reduced number of teacher recruits and/or fewer teacher participants at the end of the study.

1) Of the thirty-two science teachers in Maryland who signed up and became part of the treatment group fourteen dropped out leaving 18 treatment teachers. Seven of the 14 were from one school where the department head had made the decision to join the study without any input from her department. Consequently, she was the only one to follow through and actually use the simulations.

~ Teachers' anticipation of high stakes testing, the timing of the experimental research study, and personal considerations were additional issues contributing to lower teacher recruitment and retention numbers.

2) In Maryland, three of the teachers who dropped out were new to their school, had not attended the information meeting in the spring, and felt overwhelmed by their required teaching duties. Two other teachers had attended the spring information meeting and were not new to their school, but still failed to follow through with the study materials. Other teachers were reluctant to invest the time, couldn't spare the extra lab time required by the simulations or ran out of time in their courses. They may simply have forgotten or deliberately left them out. Even though a time commitment of 15 minutes per pre- and post-test administration seems manageable, some teachers did not feel it was worthwhile to include these items in their teaching schedule. Personal life issues (e.g., pregnancy, graduate school enrollment) also impacted teacher recruitment and retentions. One recruit declined because the substitute teacher who would replace her did not have the expertise to complete the study. For others, high-stakes testing presented barriers to participation.

~ The teacher dropout rate was further impacted by potential teacher participants' disappointment with their assigned study group. Some teachers interested in using the simulations with their students as part of the treatment group but were assigned to the control group opted out before the study got started.

3) This is evident in Maryland where of the thirty-two teachers in control schools, sixteen dropped out; but nine of those were from one school that lost interest once they were selected as a control school instead of a treatment school. In spite of the fact that they recognized that control data was essential to the study.

~ Issues surrounding the technology used in the research study acted as a deterrent for some potential participant teachers

4) Initial performance issues with the Computational Laboratory's online data collection was discouraging to some teachers who dropped out after doing only one simulation. Others didn't even try once they had heard bad reports from others. Additional information on issues related to the technical environments is available later in this section of the report.

Science Courses, Computational Literacy Research Study Topics, and the Alignment of the Two

The research study focused on four science topics. Initial research into the alignment of topics and science courses indicated that two of the projects four topics *Population Dynamics* (PD) and *Spread of Disease* (SoD) were commonly taught in biology classes; and two, the *Carbon Cycle* (CC) and the *Rock Cycle* (RC) were commonly taught in Earth / Space Science classes. For participants in the treatment research group, each topic was focused on a computational simulation. Those teachers participating as part of the control research study group taught as they normally do, covering the same topics as those participating as part of the treatment group. Their students engaged with the relevant science topic/s (PD, CC. SoD, RC) without benefit of the project's computational simulations. Table 2 below illustrates the number of project topics covered by science teachers in each participating state by research study group. Note that in Iowa, there was no participation, treatment or control in the *Carbon Cycle* topic.

Table 2: Research Study Topics by State and Research Study Group – Treatment (T) / Control

	Topic	Population Dynamics			Carbon Cycle			Spread of Disease			Rock Cycle				
	Group	T	C	Total PD	T	C	Total CC	T	C	Total SoD	T	C	Total RC		Total State
State	Maryland	9	8	17	7	10	17	4	4	8	3	1	4		46
	Iowa	0	2	2	0	0	0	3	4	7	0	3	3		12
	Tennessee	2	3	5	0	1	1	2	1	3	0	1	1		10
Total topics		11	13	24	7	11	18	9	9	18	3	5	8		68

The overwhelming majority of courses accommodating the Computational Literacy science topics were biology courses. While 77% (n=34) of science teachers participating in the Computational Literacy Experimental Research Study were identified with biology courses, a scant 6% of courses were identified as earth science or earth space systems, the courses originally attributed to the Carbon and Rock Cycle topics. While earth science / earth space systems courses were represented in significantly small numbers (only 3 courses overall), almost 30% of participating teachers were identified with environmental science courses, an unanticipated addition to the roster of courses accommodating the Computational Literacy science topics. A small number of general science-type courses rounded out the course roster at the same scant 6% as earth science / earth space systems courses.

Table 3 below details the science courses across the three states that accommodated the Computational Literacy science topics in the 2006 – 2007 research study and the number of teachers associated with each course.

Table 3: Science Courses Accommodating the Computational Literacy Topics

Class	Course	# courses	Class	Course	# courses
Biology	Anatomy	2	Environmental Science	Environmental Science (Regular)	7
	Anatomy & Physiology	1		Environmental Science (Inclusion)	1
	Biology (Regular)	18		Environmental Science (Honors)	1
	Biology (Honors)	7		Environmental Science (Advanced Placement)	3
	Biology (Advanced)	2		Total Environmental Science Courses	12
	Biology (International Baccalaureate)	1			
	Biology (IGCSE Cambridge)	1	Earth Science	Earth Science (Regular)	1
	Biology (Signature)	1	Earth Space Systems	Earth Space Systems (Regular)	2
	Biology (Advanced Placement)	1		Total Earth / Earth Space Courses	3
	Biotechnology	1			
	Total Biology Courses	35			
			General Science	Applied Science	1
unknown	unknown	1		Interactive Science	1
	Total unknown	1		English Language Learners Science	1
				Total General Science courses	3
Total Science Courses Accommodating Computational Literacy Topics = 54					

~ The association of research study topics with specified high school science courses did not match up as expected.

As mentioned earlier, when the Computational Literacy project began, the topics selected were usually covered in the areas that the project associated them with: *Population Dynamics* (PD) and *Spread of Disease* (SoD) were covered in biology classes and the *Carbon* (CC) and *Rock* (RC) *Cycles* were covered in Earth / Space Systems classes. As recruitment for the study progressed, it was discovered that the alignment of topics with school subjects was not as expected. For example, in Maryland, it was reported that the Spread of Disease (SoD) was no longer part of the regular biology curriculum. It was reported that SoD had become part of anatomy and physiology. In response to this information, the project broadened the scope of biology classes to include anatomy and physiology, while taking into consideration that generally biology students are 9th and 10th graders while anatomy students are 10th and 11th graders. Both biology students and anatomy students can be found in regular and honors classes but anatomy students already had a biology course. Any concerns or considerations turned out to be inconsequential because as illustrated in Table 3 above, the participation by biology teachers was robust, and as evidenced in Table 6 on a following page, science teachers across the biology spectrum accommodated those simulations and topics (PD, SoD) originally attributed to biology.

~ Earth Science and Earth Space Systems courses were represented in more limited numbers than originally anticipated. The limited availability of Earth Science and Earth Space Systems courses adversely impacted the implementation of the Rock Cycle topic / simulation.

While there were not enough earth / earth space systems courses to accommodate the topics previously associated with them (CC, RC), the addition of environmental science courses helped provide additional classes where the Computational Literacy topics could be implemented. The 30% of courses represented by the environmental sciences courses addressed topics attributed to both biology and earth science / earth space systems – but the Rock Cycle topic / simulation suffered and a hole was created that was never filled.

~ The addition of environmental science courses to the research study's roster of science courses accommodating Computational Literacy topics though unanticipated during development and early testing helped to partially fill the void created by limited earth science / earth space systems courses.

Analysis of the courses accommodating the Computational Literacy study revealed that biology courses comprised 77% of the courses taught during the study. With more courses taught (54) than participating teachers (44), the numbers indicated that some participating science teachers taught more than one course during the study. Even though the overwhelming majority (75%) of teacher study participants submitted materials for students from a single science course, about 18% of teacher participants reported teaching two courses. Only 2 of the 44 teachers (about 4%) reported being responsible for teaching three science courses. Table 4 below details the number of teachers in the study, by state and research study group, who were responsible for teaching one, two, and three science courses during the implementation of the Computational Literacy Experimental research study.

Table 4: Number of Courses Taught by Participating Teachers

# courses taught	unknown		1 science course		2 science courses		3 science courses		Total science teachers	
State	T	C	T	C	T	C	T	C	Treatment	Control
Maryland	1	0	8	14	5	0	1	1	15	15
Iowa	0	0	3	5	0	1	0	0	4	5
Tennessee	0	0	1	2	1	1	0	0	1	4
Total States	1	0	12	21	6	2	1	1	20 (45.5%)	24 (54.5%)
Total courses	Total unk = 1 (2%)		Total 1 taught = 33 (75%)		Total 2 taught = 8 (18%)		Total 3 taught = 2 (4%)		Total Science Teachers = 44 100%	
Total science courses taught by participating teachers = 54 100%										

Teachers participating in the research study used all four of the simulations or covered the comparable topics in their classrooms with their students. *Population Dynamics*, our prototype topic proved to be one of the most popular with participating teachers in Maryland and the most popular with teachers in Tennessee. *Carbon Cycle* was equally popular in Maryland, non-existent in Iowa, and almost non-existent in Tennessee. *Spread of Disease* was well represented in all three states but relative to their participation with other topics, proved to be the most popular in Iowa. Table 5 below details the number of study topics by state, school and research study group.

Table 5: Topics Implemented by State, School, and Research Group

Topic		Population Dynamics			Carbon Cycle			Spread of Disease			Rock Cycle			
Group		T	C	PD sum	T	C	CC sum	T	C	SoD sum	T	C	RC sum	School sum
State	Sch													
MARYLAND	001	0	0	0	0	0	0	1	0	1	1	0	1	2
	002	2	0	2	3	0	3	1	0	1	1	0	1	7
	003	5	0	5	2	0	2	0	0	0	1	0	1	8
	004	0	1	1	0	1	1	0	1	1	0	0	0	3
	005	1	0	1	1	0	1	1	0	1	0	0	0	3
	006	0	1	1	0	2	2	0	1	1	0	0	0	4
	007	0	1	1	0	1	1	0	1	1	0	1	1	4
	008	0	0	0	0	5	5	0	0	0	0	0	0	5
	009	0	5	5	0	1	1	0	1	1	0	0	0	7
	010	1	0	1	1	0	1	1	0	1	0	0	0	3
Maryland sum		9	8	17	7	10	17	4	4	8	3	1	4	46
IOWA	021	0	0	0	0	0	0	0	1	1	0	1	1	2
	022	0	0	0	0	0	0	1	0	1	0	0	0	1
	024	0	0	0	0	0	0	0	2	2	0	1	1	3
	025	0	1	1	0	0	0	0	0	0	0	1	1	2
	026	0	1	1	0	0	0	0	1	1	0	0	0	2
	028	0	0	0	0	0	0	2	0	2	0	0	0	2
I owa sum		0	2	2	0	0	0	3	4	7	0	3	3	12
TENNE SSEE	013	1	0	1	0	0	0	1	0	1	0	0	0	2
	014	0	1	1	0	0	0	0	0	0	0	0	0	1
	017	0	1	1	0	1	1	0	0	0	0	0	0	2
	018	0	1	1	0	0	0	0	1	1	0	1	1	3
	019	1	0	1	0	0	0	1	0	1	0	0	0	2
Tennessee sum		2	3	5	0	1	1	2	1	3	0	1	1	10
Total		11	13	24	7	11	18	9	9	18	3	5	8	68

Table 6: Simulations / Topics by Science Course and Research Study Group

TREATMENT								CONTROL						
SIMULATIONS						T	Science Courses	C			TOPICS			
PD	CC	SoD	RC								RC	SoD	CC	PD
0	0	0	0	0		0	Interactive Science	1		1	1	0	0	0
0	0	0	0	0		0	English Language Learners Science	1		1	0	0	0	1
0	0	0	1	1		1	Applied Science	0		0	0	0	0	0
0	0	0	1	1		1	Total General Science	2		2	1	0	0	1
0	0	2	0	2		2	Anatomy / Anatomy & Physiology	1		1	0	1	0	0
3	0	6	0	9		8	Biology (regular)	10		10	0	2	4	4
3	2	1	0	6		3	Biology (honors, IB, advanced, signature)	9		15	1	3	5	6
1	1	1	0	3		1	Biology (AP)	0		0	0	0	0	0
0	0	0	0	0		0	Biotechnology	1		1	0	1	0	0
7	3	10	0	20		14	Total Biology	21		27	1	7	9	10
0	2	0	1	3		2	Earth / Earth Space Systems	1		1	1	0	0	0
0	2	0	1	3						1	1	0	0	0
3	3	0	0	6		3	Environmental Science (regular)	4		5	1	1	2	1
1	1	0	1	3		1	Environmental Science (honors)	0		0	0	0	0	0
2	2	0	0	4		2	Environmental Science (AP)	1		4	1	1	1	1
1	0	0	0	1		1	Environmental Science (Inclusion)	0		0	0	0	0	0
7	6	0	1	14		7	Total Environmental Science	5		9	2	2	3	2
0	0	0	0	0		0	Unknown	1		1	0	0	0	1
14	11	10	3	38						39	5	9	12	13
						24	Total Science Courses	30						
				38						39				
Total Topics Taught = 77 = Total Classes Taught														

During the Computational Literacy Experimental Research Study, 44 science teachers from 21 schools across three states, taught a total of 68 research study topics / simulations (see Table 5) through 54 individual science courses (see Table 3) to a total of 77 science classes (see Table 6).

Teacher Participation in the Computational Literacy Experimental Research Study

Teachers were central to the research study, often is ways unanticipated. They taught the science topics that the simulations related to. They decided to participate or not to participate in the research study. Their participation determined the study's student numbers, topic implementation and distribution, and teachers were responsible for the lion's share of data collection for the study. The degree to which they adhered to the research protocols had the potential to impact student engagement with the topics, students' willingness to participate, bring in signed consent slips, and ultimately student outcome scores. In many ways teachers are the roots of the Computational Literacy Experimental Research Study. Students are the product of those roots.

<u>Teacher Survey Results</u>

To investigate the pedagogical knowledge, backgrounds, and beliefs of teacher participants in the Computational Literacy Experimental Research Study, as well as to look at their habits and uses for technology in general, a teacher survey was created. Specifically the project was interested in teasing out teachers' perspectives on the use of simulations and their philosophies on the type of student simulations can reach. Additional interests included whether teachers' primary use of technology was to reinforce basic skills, for exploration, or for other uses.

Survey items were created in-house, or culled (and modified) from a number of sources (see activities section of this report for a list of survey sources). An online version of the final survey was created using Perseus Development Corporation's SurveySolutions 6 program, and hosted on the Educational Development Center's server.

A Glimpse Into All Survey Takers

Forty-nine science teachers from the 3 participating states completed the Computational Literacy web-based teacher survey. By the end of October 2006, all surveys were completed except one from a teacher who completed the survey in January of 2007. Of the 49 science teachers who completed the survey about 63% were female, 35% were male. One participant did not include any information on gender. Of all the teachers participating in the survey the overwhelming majority of them reported having home access to a computer and almost all reported having access to a computer in their classroom. Of the 49 science teachers who completed the survey, 13 did not participate in the research study in any way, an additional 4 were not included in the final analysis because it was believed that their students had not submitted the necessary permissions to be included in the study and hence were believed to be ineligible for the study. The survey responses of 32 teacher research study participants were analyzed. They represented about 73% of those teachers who had participated in the research study.

The gender breakdown and access to technology of all teachers completing surveys are mirrored in the subset of teacher research study participants whose surveys were analyzed and used in this section to describe Computational Literacy teacher research study participants.

A Glimpse into Teacher Research Study Participants

<u>Teacher Participation</u>

There were more research study teachers in the control group than in the treatment group. Of the 44 science teachers participating in the research study about 43% belonged to the treatment group and 57% to the control group. Teacher research participants were overwhelmingly female with female participation out numbering male participation by more than 2:1. Of the 44 science teachers participating in the research study about 68% were female and 32% were male.

The results for the set of teacher participant surveys analyzed, is very similar in both gender and group as the complete collection of all teacher survey takers. The set of surveys analyzed represents about 73 % (n=32) of those participating in the study. Results from that analysis indicate that of the set of teachers whose surveys were analyzed, the majority (56%) were control group participants; 44% were from the treatment group. The overwhelming majority of participants in the set were female with females out representing males 3:1. When gender was analyzed, by research group the results were similar. Within the control group, female participation was at 72% and male at 28%. Within the treatment group female participation was at 79% and male at 21%.

<u>Teachers' Pedagogical Knowledge, Background and Beliefs</u>

~ Control group teachers tended to have more years teaching experience than their treatment group counterparts and they reported being more confident with their knowledge of the science topics that they taught.

The average number of years teaching by control teachers was more than that of treatment teachers by almost 4 years. Control teachers had also been at their schools longer by almost 3 years and reported being more confident with their knowledge of the science topics they taught.

Fifty-six per cent (56%) of the control group teachers registered at the high end of confidence with their knowledge of the science topics they taught compared to about 44% of treatment teachers at the same level of confidence. This was only slightly offset by larger classes than treatment teachers.

~ Treatment teachers reported being more confident than control teachers in a few areas including: trying new techniques for teaching science in their classrooms, implementing inquiry learning and implementing innovative technology-rich learning.

Almost all participating treatment teachers were at the upper end of the confidence scale while control group teachers leveled at mid range. Fifty-seven percent of treatment teachers were very confident in their ability to implement inquiry learning. Only 33% of control group teachers reported feeling as confident. Thirty-six percent of treatment teachers were very confident in their ability to implement technology-rich learning while only 22% of control group participants reported feeling as confident.

~When asked to report on their beliefs and goals for their classroom, participating teachers' responses at times seemed to be at odds.

Both groups thought that it was important to identify students' strengths and encourage participation by all students. About 30% of control teachers reported that preparing students for standardized tests was an important goal in their classrooms. About 23% of treatment teachers agreed. Treatment and Control groups felt similarly about the importance of students mastering a uniform body of core content with about 35% of control group teachers ranking this as very important compared to the treatment group's 30%. A greater percentage of control teachers (control = 64%; treatment =36%) believed that technology was essential to provide students with higher levels of interactivity. Yet twice as many treatment teachers (66.7%) as control teachers (33.3%) believed that too much technology could interfere with real learning.

Habits and Uses of Technology

~ Treatment group teachers tended to be only slightly more experienced with computers and more reported greater access to computers than their control counterparts.

Although all participating teachers, for the most part, reported having experience using computers, and having access to computers both at home and in their classrooms, treatment group teachers tended to be slightly more experienced and more had access than their control counterparts. All treatment teachers reported having ten or more years experience using computers while only 78% of control group teachers made the same claim. All of treatment group teachers reported having both computer access at home and in their classrooms. All control group teachers reported having access in their classrooms, while 94% percent of them reported having home access.

~ The two research study groups were comparable in their confidence in the quality of their schools installed technology but differed in their confidence to get access to technology when needed.

Both the treatment and the control group expressed only moderate confidence in the quality of their school's installed technology. But control teachers tended to be more confident in their ability to gain access to that technology than treatment teachers.

~ The two research study groups were comparable in their reported use of a computer with a projection system to deliver instruction to as class. Just under 30% of teachers from both research study groups reported daily use and about 35% of both groups reported using the set-up about once a week.

Perspectives on the Use of Simulations

~ Both treatment and control teachers were comparable in their belief that computer-based simulations can help all students.

~ How often teachers reported using classroom resources like graphic representations, physical models, and simulations varied by research study group. Whether they used these resources and the significance of them seemed to be in transition.

The frequency with which teachers use certain resources in their classrooms was identified. Forty-eight percent of control group participants use graphic representations daily. Only 29% of treatment participants report the same level of use. Seventeen percent of control teachers report using physical models daily but 39% report using them once a week. A mere 6% of treatment teachers report using physical models daily but 39% of them use physical models once a week.

When asked about the frequency of use of computer-based simulations once a week, use of this resource by both treatment and control was comparable but all other frequency options varied by group. Thirty-five percent of control group teachers reported using visualization less than weekly; the percentage for treatment teachers reporting the same frequency was 47%. Thirteen percent of control participants and 6% of treatment participants reported never using computer-based visualizations.

Treatment and control participant responses were comparable when asked if they perceive simulations as very useful motivational tools and whether they use computer based models as unit cores and develop the rest of their lesson around them. Only one control teacher reported using simulations about once week. Control teachers use computer based models when they have available time within units and the models seem to fit in with what they're teaching: once a week -- control is at 13%; treatment at 6%; less than weekly control is at 44%; treatment is at 53%. Both groups are comparable in the percentage of teachers who have never used a computer-based model when time and fit align (C= 22%; T = 24%).

Fidelity of Implementation: Topic Teacher Reports

Each participating science teacher was asked to complete a Topic Teacher Report (the project's fidelity of implementation document) for each classroom for each study topic. The Topic Teacher Reports were the only indicator researchers had that the study had been implemented as designed. Blank copies of the instrument were included in the implementation package mailed to each teacher prior to implementation of the study that included all the materials that teachers needed to implement the study in their classrooms. Of the 44 participating science teachers, 36 returned Topic Teacher Reports, an 81% response rate. Table 7 below details which topics the project received reports for and the number received by state, topic, and research study group.

Table 7: Topic Teacher Reports Submitted by State, Topic, and Research Study Group

	Treatment					Control				
	PD	SoD	RC	CC	Total	PD	SoD	RC	CC	Total
Maryland	6	4	3	6	19	8	4	1	8	21
Iowa	0	3	0	0	3	2	3	3	0	8
Tennessee	2	2	0	0	4	1	0	0	1	2
Total	8	9	3	6	26	11	7	4	9	31
57 Topic Teacher Reports received in total										

Quality of Submitted Topic Teacher Reports

~ There was wide variance in the quality of the Topic Teacher Reports submitted and in the information the documents communicated to research staff. The majority of them did not supply the information expected or anticipated nor did they meet requirements for fidelity.

Some of the more common issues with the submitted reports included:

- Missing dates and times. The amount of time students were exposed to a topic can significantly impact their understanding of the topic, which in turn can directly impact their performance on the administered tests. In addition to missing dates and times for topic related lessons, a number of reports failed to identify the amount of time it took students to complete the tests.

- Inadequate descriptions of topic related lessons. Many teachers were too concise in their descriptions of the topic related lessons their students were exposed to. Some teachers only included a title for a lesson instead of the description. Others offered a brief, cryptic line that was almost meaningless to study researchers.

- Missing post-simulation follow-up/debriefing for treatment group students. A class debriefing after students' exposure to the simulation/s was required in order to wrap-up the topic and to gauge students' understanding. Seven of the 17 treatment teachers who submitted Topic Teacher Reports did not indicate whether or not they had included a follow-up/debriefing.

- Minimum required 90 minutes for student exploration of simulation not met by some treatment participants. In order for the simulations to have even a small chance of impacting students' knowledge and understanding of the project related topics, a minimum of 90 minutes exposure to the topic simulations was necessary. Four of the 17 treatment teachers did not include the duration of time that their students interacted with the simulation and 3 entered times less than what was suggested. Others offered explanations for their failure to meet this study requirement. A few teachers encountered technical obstacles that included difficulties with the Computational Laboratory, the project's website, lack of adequate sInternet access, and difficulty gaining access to their schools' computer labs. Two Maryland teachers reported that due to slow response time the simulation had to be presented to the whole class using an LCD projector. Technical issues surrounding the Lab and access to the Lab are discussed in "Findings Related to the Technical Environments and the Computational Laboratory" below in this section of the report.

- Multiple project related topics crammed into a single Topic Teacher Report instrument. Several teachers addressed more than one topic in their classrooms, and a few of them did not create a report for each topic addressed, instead they crammed all of them into a single document making it extremely difficult to understand what lessons applied to which topics and the amount of time devoted to each. These reports were especially confusing.

- Identical reports received from multiple participants. A number of identical reports from teachers at the same school caught our attention and prompted us to speculate. These reports were for the same topic (PD), contained the same dates, times, and session descriptions. Perhaps the teachers decide the sequence for their classes as a unit, so the details of anyone of their implementations would reflect the implementation of those remaining. Or, perhaps they just copied from one another or one teacher filled out all the reports. These were all control teachers.

Fidelity was met when minimally all students took a pre- and post- test in each topic that their teacher signed on for and a Logic Test for each semester's participation. Additionally each student in the treatment group was exposed to an introductory lesson before using the Driving Questions to guide their explorations with the simulations and a follow-up / debriefing lesson after interaction with the topic simulation. The table below details the fidelity of implementation protocols for the Computational Literacy Research Study.

Table 8: Fidelity of Implementation Protocols

Treatment	Control
• The pretest was administered before any topic lessons were taught, optimally during the first or second week of the school year or semester.	• SAME AS TREATMENT
• Teacher indicated that at least an introductory lesson took place before students worked with the simulation	• Teacher indicates the sequence of topic specific lessons taught in her science class
• Simulation indicated for the duration of at least 90 minutes (Driving Questions in support)	N/A
• Teacher indicates that there was at least one follow-up / debriefing lesson after students interacted with the topic simulation	N/A
• Posttest given after all topic specific lessons	• SAME AS TREATMENT
• Logic Test administered after everything	• SAME AS TREATMENT

Compliance criteria were determined by the project, after the completion of the study, once it was discovered that teacher fidelity was very low. At a minimum, all students were expected to take a pre- and post- test in each topic that their teacher signed on for and a Logic Test for each semester's participation. Additionally each student in the treatment group would use the Driving Questions to guide their explorations with the simulations. Teachers who met these minimum criteria were identified as in compliance with the study. Table 9 below details the implementation compliance criteria established.

Table 9: Implementation Compliance

Treatment	Control
• The pretest was administered before any topic lessons were taught	• SAME AS TREATMENT
• Simulation indicated (Driving Questions in support)	• Teacher indicates the sequence of topic specific lessons taught in her science class
• Posttest given after all topic specific lessons	• SAME AS TREATMENT
• Logic Test administered after everything	• SAME AS TREATMENT

Of the 44 teachers participating in the research study, 17 qualified as having implemented the research study with fidelity -- a mere 39%. Of those who qualified for fidelity, only 4 were part of the treatment group. The remaining 13 were part of the control group. In an effort to improve the contextual information available for additional student scores, project researchers invented a compliance option (described above). Three teachers from treatment and 3 from control fulfilled the compliance criteria. But in the end there still were not enough students for analysis of either fidelity or compliance. Table 10 below illustrates the numbers and percentages of teachers by group who implemented the research study with fidelity or compliance.

Table 10: Fidelity and Compliance by Research Study Group and Percentage

		N	% Total	% Fidelity	% Compliance
Teacher Participants	Treatment	19	43		
	Control	25	57		
	Total	44	100		
Fidelity	Treatment	4	9	23.5	
	Control	13	29.5	76.5	
	Total	17	38.6	100	
Compliance	Treatment	3	7		50
	Control	3	7		50
	Total	6	14		100

Student Participation in the Computational Literacy Experimental Research Study

The Computational Literacy Experimental Research Study was conducted during the 2006 - 2007 school year in three states: Maryland, Iowa, and Tennessee. The program was implemented in a total of 21 schools with 44 teachers and 1542 students. An additional 1,639 students experienced part of the intervention but were NOT included in the study itself.[4]

Table 11: Participating Teachers, Student Participants, and Affected students by State, School, and Research Group

		Study Teachers			Study Students			Affected Students		
State	School	T	C	Teach Sum	T_1	C_1	Stu sum	T_2	C_2	Stu sum
Maryland	001	1	0	1	32	0	32	67	0	67
	002	3	0	3	165	0	165	207	0	207
	003	5	0	5	102	0	103	286	0	286
	004	0	1	1	0	51	51	0	20	20
	005	2	0	2	27	0	27	78	0	78
	006	0	2	2	0	45	45	0	41	41
	007	0	1	1	0	54	54	0	4	4
	008	0	5	5	0	162	162	0	261	261
	009	0	7	7	0	333	333	0	190	190
	010	3	0	3	90	0	90	182	0	182
M sum	10	14	16	30	416*	645	1062	820	516	1336
Iowa	021	0	1	1	0	19	19	0	17	17
	022	1	0	1	103	0	103	37	0	37
	024	0	3	3	0	80	80	0	84	84
	025	0	1	1	0	72	72	0	50	50
	026	0	1	1	0	29	29	0	7	7
	028	2	0	2	72	0	72	13	0	13
I sum	6	3	6	9	175	200	375	50	158	208
Tennessee	013	1	0	1	27	0	27	0	0	0
	014	0	1	1	0	0	0	0	63	63
	017	0	1	1	0	42	42	0	7	7
	018	0	1	1	0	22	22	0	10	10
	019	1	0	1	16	0	16	14	0	14
T sum	5	2	3	5	43	64	107	14	80	94
Total	**21**	**19**	**25**	**44**	**634**	**907**	**1542**	**884**	**754**	**1639**

Of the 3181 students in the Computational Literacy Research Study database, a total of 1,542 students had submitted signed permission slips to participate in the study, 1639 did not. These 1542 students comprised the set of student research study participants and from whom data was

[4] All student participants in the research study have, on file, a permission slip signed by a parent or legal guardian.

collected and analyzed. Control group participants comprised the larger of the two groups. Control group participants were 59% of participating students and treatment group participants comprised 41%. Maryland had the majority of participants, with 70% of participating students from Maryland, 24% from Iowa and 6% from Tennessee. Overall there were more females than males and tenth graders were also the greatest percentage of students participating in the study by grade. Tenth graders were the largest percentage of participating students in Maryland and Iowa. In Tennessee there were more 9th graders. There was also more female than male participation in Maryland and Tennessee, but in Iowa there were more males than females. In each of the three states there were more control students than treatment.

Analysis of Data Collected from Students Participating in the Computational Literacy Experimental Research Study

The unit of assignment was at the school level, in order to prevent contamination between treatment and control classrooms. Blocking within state, schools were randomly selected (from a pool of willing schools) from within each state and assigned to treatment and control groups. Randomization was accomplished using the SPSS program's "random select" feature.

As part of the research study, students were administered a pretest in each topic encountered in their science class that their teacher had decided to implement with her students. Ideally, topic pretests were administered in the first week or two of the school year or semester. A posttest was administered after all topic lessons had been completed and/or after students had engaged with a topic simulation and participated in a follow-up / debriefing. Pre-test and post- test data was analyzed using a Repeated Measures ANOVA and a Repeated Measures Hierarchical Linear Modeling technique, taking into account the two time points for each participant and the nested structure of the classroom intervention.

RESULTS: Repeated Measures ANOVA

Population Dynamics

The mean for the control group was higher on the pre-test and post-test for *Population Dynamics* than the mean for the treatment group. This pattern of higher scores for the control group may indicate that the two groups were not equivalent in knowledge prior to the intervention (pre-test scores were not used as a criteria during the randomization). Including all three states in the analysis, there was a significant difference in favor of the control group, $F(1, 582) = 35.261$, $p < .001$.

Table 12: Population Dynamics Pre-/Post- Test Results

		N	Mean	Std Deviation
PD_Pre	Treatment	165	5.34	1.952
	Control	417	5.98	1.797
	Total	582	5.80	1.863
PD_Post	Treatment	165	6.04	2.155
	Control	417	7.15	1.793
	Total	582	6.83	1.966

Carbon Cycle

The mean for the control group was higher on the pre-test and post-test for *Carbon Cycle* than the mean for the treatment group; however, the treatment group did show modest gains from pre-test to post-test. This pattern of higher scores for the control group may indicate that the two groups were not equivalent in knowledge prior to the intervention (pre-test scores were not used as a criteria during the randomization). Including all three states in the analysis, there was a significant difference in favor of the control group, $F(1, 456) = 11.861$, $p < .001$.

Table 13: Carbon Cycle Pre-/Post- Test Results

		N	Mean	Std Deviation
CC_Pre	Treatment	170	2.25	1.152
	Control	286	2.65	1.062
	Total	456	2.50	1.111
CC_Post	Treatment	170	2.70	1.273
	Control	286	2.94	1.149
	Total	456	2.85	1.201

Spread of Disease

The mean for the control group was slightly higher on the pre-test for *Spread of Disease*, but this difference was not meaningful. The mean for the treatment group was slightly higher for the post-test, but this difference was not significant. Including all three states in the analysis, there was not a significant difference between groups, $F(1, 452) = .06$, $p > .05$.

Table 14: Spread of Disease Pre-/Post- Test Results

		N	Mean	Std Deviation
SoD_Pre	Treatment	245	2.86	1.120
	Control	207	2.88	1.250
	Total	452	2.87	1.180
SoD _Post	Treatment	245	3.02	1.240
	Control	207	2.99	1.155
	Total	452	3.00	1.201

Because of the limited participation in the *Rock Cycle* topic / simulation there were no results for the *Rock Cycle*.

<u>RESULTS: Hierarchical Linear Modeling</u>

Maryland student test scores were analyzed using hierarchical linear modeling. Maryland had 70% of all the students in the study and had data available through public sources to use as covariates in the model, so only students from this state were chosen for this analysis. School level data not available through the research study was obtained primarily from school report cards. The analysis used a three level model with student study participants, nesting pre- and post- test scores (level 1) within students (Level 2), and finally within schools (level 3).

Table 15: Topics Implemented in Maryland by School and Research Group

Topic		Population Dynamics			Carbon Cycle			Spread of Disease			Rock Cycle			
Group		T	C	PD sum	T	C	CC sum	T	C	SoD sum	T	C	RC sum	School sum
State	Sch													
MARYLAND	001	0	0	0	0	0	0	1	0	1	1	0	1	2
	002	2	0	2	3	0	3	1	0	1	1	0	1	7
	003	5	0	5	2	0	2	0	0	0	1	0	1	8
	004	0	1	1	0	1	1	0	1	1	0	0	0	3
	005	1	0	1	1	0	1	1	0	1	0	0	0	3
	006	0	1	1	0	2	2	0	1	1	0	0	0	4
	007	0	1	1	0	1	1	0	1	1	0	1	1	4
	008	0	0	0	0	5	5	0	0	0	0	0	0	5
	009	0	5	5	0	1	1	0	1	1	0	0	0	7
	010	1	0	1	1	0	1	1	0	1	0	0	0	3
Maryland sum		9	8	17	7	10	17	4	4	8	3	1	4	46

Population Dynamics

One of two popular topics in Maryland, the implementation of *Population Dynamics* in the state represented 37% of the total number of topics implemented in Maryland. Of the 10 participating schools 8 (4 treatment, 4 control) implemented the topic with their students. In each of two schools (1 treatment, 1 control), 5 teachers (a total of 10 in both schools) represented almost 57% of the state's participating teachers who implemented the topic in their classrooms.

For *Population Dynamics* (PD), school level means for socioeconomic status, percentage of English Language Learners (ELL) in the school, and percentage of Special Education students were not significant predictors of outcome measures. Class type, meaning honors vs. non-honors classes, was also not a significant predictor. The final model included students' gender (L2) and research study group (L3), yielding an intercept of 6.50 for male students in the control group. Male students in the treatment group scored lower, with a slope of -1.22, and female students scored even lower, with an additional decrease in slope by -.41. This model's fit was significantly better than a three level model that only included group as a covariate (p=.017). The explanatory value of gender was only indicated within the topic of *Population Dynamics*. It should be noted that *Population Dynamics* had the widest range of scores, from 0-9.

The initial interclass correlation (ICC) for the null model was 38% and the final ICC was 37%, indicating a high degree of within group nesting; however, there were no other student level variables available to further explain similarities between students.

Table 16: Population Dynamics Maryland Mean Pre-/Post- Test Scores by Gender and Group

Group	Gender	Mean Score
Control	Male	6.497321
	Female	6.088872
Treatment	Male	5.273504
	Female	4.865055

Spread of Disease

For *Spread of Disease* (SoD), while eight of the 10 schools participating from Maryland actually implemented the SoD topic with their students, in each participating school only one course was indicated as having implemented the topic with students. Four of the schools had insufficient data and further analysis on this data using a three level hierarchical linear model was not possible.

Carbon Cycle

One of two popular topics in Maryland, the implementation of the *Carbon Cycle* in the state, represented 37% of the total number of topics implemented in Maryland. Of the 10 participating schools 9 (4 treatment, 5 control) implemented the topic with their students.

For *Carbon Cycle*, the null model indicated an ICC= 30%, supporting the use of a nested model. Even in the final model the ICC=26.46% meaning that there is still variability left to be explained, which may be due to the lack of student level demographic information. Several covariates were not significant predictors, including group, socio-economic status (ses), special education, and gender; however group was maintained in the final model for theoretical reasons. The percentage of school ELL was significant in preliminary analyses, but was dropped from the final model because it did not significantly improve the model estimation or explain variability, an outcome that was similar in the other topics tested.

In the final model, group (treatment or control) was entered as a school covariate (Level 3) and class type (honors or non-honors) was entered as a student level covariate, which was statistically significant (p=.008). Students in the control group taking non-honors courses scored the lowest (2.050033), followed by students in the treatment group taking non-honors courses (2.224419).

Students in the control group taking honors courses scored the third highest (2.759503) and students in the treatment group taking honors courses scored highest (2.933889). Honors students outperformed their non-honors counterparts regardless of whether they were in the treatment or control group; however, those in the treatment group performed best overall after the intervention. This model's fit was significantly better than a three level null model (p=.000).

Table 17: Carbon Cycle Maryland Mean Pre-/Post- Test Scores by Class Type and Group

Group	Class Type	Mean Score
Control	Non-honors	2.050033
	Honors	2.759503
Treatment	Non-honors	2.224419
	Honors	2.933889

Table 18: Level-1 Descriptive Statistics

VARIABLE	N	MEAN	SD	MINIMUM	MAXIMUM
INDEX1	1486	1.50	0.50	1.00	2.00
FL_PD	743	7.20	2.22	0.00	10.00
FL_DS	103	3.7	1.31	0.00	6.00
FL_CC	657	4.13	1.49	0.00	7.00
RE_PD	743	5.91	2.06	0.00	9.00
RE_DS	103	2.80	1.20	0.00	5.00
RE_CC	657	2.47	1.19	0.00	5.00

Table 19: Level-2 Descriptive Statistics

VARIABLE	N	MEAN	SD	MINIMUM	MAXIMUM
FEMALE	743	0.56	0.50	0.00	1.00
GROUP	743	0.38	0.49	0.00	1.00
SES_CTR	743	1.93	4.64	-3.49	10.61
SPED_CTR	743	0.16	2.02	-3.90	2.70
ELL_CTR	743	0.92	2.69	-5.83	5.17
SELECT	743	1.00	0.00	1.00	1.00
CLASS	743	0.36	0.48	0.00	1.00

Table 20: Level-3 Descriptive Statistics

VARIABLE	N	MEAN	SD	MINIMUM	MAXIMUM
STUD_ID	7	10632290.14	314284.76	10203005.00	11037003.00
GROUP	7	0.57	0.53	0.00	1.00
SES_CTR	7	2.78	5.28	-3.49	10.61
SPED_CTR	7	0.40	2.43	-3.90	2.70
ELL_CTR	7	0.83	3.80	-5.83	5.17

Measure of Far Transfer: Student Logic Model Test Results

The far transfer test was designed to assess students' mastery over two main skill sets—systems thinking, (e.g., understanding the relationship between variables) and data interpretation, including the ability to read graphs and tabular data, and to make recommendations based on such data. In contrast to the near transfer test (see pre/post section elsewhere in this report), in which students were asked to solve problems similar to those encountered in the *Driving Questions* but involving characters and situations slightly different from those found in the simulation (e.g., wolves and deer instead of foxes and rabbits in the case of the predator prey simulation), the logic tasks test students' ability to apply these skills to novel situations.

The eleven questions contained in the Logic Test booklet are all multiple choice. In the *Clothes in the Closet* task, students are asked to apply their background knowledge to a problem to determine direct and inverse relationships and then to determine the corresponding graph that would represent the particular situation. In the *College Enrollment* task, students are again given data in table and graph form, but the data are counterintuitive—enrollment goes up when applicants and acceptances go down and enrollment goes up even when tuition increases. This counterintuitive relationship is included to test whether or not students are relying on data to answer the questions, as opposed to using common sense. As mentioned earlier, in the Downloading Music problem, students are given data in table and graph formats and are expected to draw conclusions from them.

Of the 1542 students who participated in the Computational Literacy Experimental Research Study, 952 had Logic Tests results entered into a data file. A full 590 students did not have Logic Test scores available for analysis. The 952 students with entered scores represented 61% of the total number of participating students.

The 14.84 mean score for treatment students was lower than the 16.86 mean for control students. There was a significant difference in favor of the control students. It is important to note that there were almost twice as many control students as there were treatment students. The asymmetrical non-alignment of research study participants is the consequence of many more control participants than treatment, and is a recurring theme throughout the research study.

		N	Mean	Std Deviation
	Treatment	341	14.84	6.271
	Control	611	16.86	6.810
		952	16.14	6.690

Findings related to the Technical Environments and the Computational Laboratory

The Computational Laboratory houses the four computational simulations that were the focus of the project's experimental research study. The simulations along with their supporting materials combine to create a topic module. Two of the four topic modules were associated with biology: *Population Dynamics* (PD) and *Spread of Disease* (SoD) and two with Earth / Space science: the *Carbon Cycle* (CC) and the *Rock Cycle* (RC).

Because the research study involved a variety of school districts with limits on technology staff and security concerns, a web-based interaction approach was decided on as the most common and universal mechanism for content delivery. The computational simulations housed in the Computational Laboratory were developed to be delivered by a set of Java applets that provide a universal platform and were tested on a variety of platforms, operating systems, and web browsers including: Mac OSX, Windows XP/Vista, and GNU/Linux operating systems using Safari, Explorer, Firefox, and Konqueror. The Java platform relies on plug-ins that can be easily installed and run on nearly any modern web browser.

The decision to use this universal mechanism for content delivery provided a method for 1) monitoring and reporting statistics on a per-student basis, real-time data collection via centralized servers allowing for full collection and analysis of a breadth of data and 2) delivery of the computational simulations to the states, schools, classrooms, teachers, and students who participated in the experimental research study. Issues relating to the technical environments required considerable attention during the research study year. Many of the logistical issues, concerns, and limitations (e.g., different technical set-ups in districts and schools) that were identified required adjustments to the online / technical environments during the actual experimental research in classrooms in year three. A number of those issues, concerns and limitations are highlighted below.

~ Even though the Java applets provide a universal platform for content delivery there can be issues associated with their required use.

Non-homogenous classroom technical environments were largely anticipated, but one main concern introduced by using this universal mechanism for content delivery was the ability to work with disparate versions of the Java runtime environment found amongst the participating institutions. Limiting factors arose as a result of the project's technical environment and our design for collecting student answers online and tracking student activities over the course of the study.

Limitations of Architectural Components. In order to accomplish collecting student responses online and tracking student activities during their use of the simulations, we put in place an architecture that relied on "Web Services": a method that allows communication of a packet consisting of student activities and answers to Driving Questions -- back to web servers located at the Krell Institute and the University of Northern Iowa. In order to enable this type of communication, we had to require a certain minimum version of the Java runtime: in this case, Java 1.4.2. This occasionally caused problems, making the simulations unavailable to the few student computers we ran into that were incapable of running this later version of Java. (Java 1.4.2 was released on June 26, 2003)

Limitations of Operating System Versions. A primary case of a limitation imposed by our requirement for Java 1.4.2 was a school in Iowa that had a laboratory of computers that were all running an older version of Mac OS X. (OS X 1.2) Apple had released a supported version of Java only up to Java 1.4.1 for Mac OS X 10.2, which contained a bug fixed by Java 1.4.2 that would allow the graphics and web services infrastructure to work. Unfortunately, the school had no budget for an operating system upgrade (and was not planning one in the near future), so we had to transition the school from being a 'treatment' school to a 'control' school. This was especially difficult given the already lower number of treatment schools.

~ Limiting broader network access to a single point of access, a network option used by some participating districts, acted like a 'choke point' slowing the network connection for whole classrooms of students trying to download the computational simulations (Java applets).

Limitations Due To Network Architecture. Another limitation the team encountered was various network architectures deployed by schools that were different than our test runs of the simulations, or limitations that we encountered when we ran a classroom of students during the school day when network traffic was at a peak. The primary example of this was a county in Maryland that limited broader network access to a single access point: this allowed the school to make networking to the broader Internet easier and provided the capability of central administration. We encountered this single point of access as a 'choke point' that slowed the network connection for a classroom of students to download the equivalent of a large file (the Java applet) on several individual computers.

A solution to this problem was to deploy a 'stripped down' version of the simulations that did not include the Java libraries for networking or access to student systems - these simulations were then limited in functionality. The students using these simulations were unable to submit answers through our tracking system, so teachers distributed to students the hard copy set of Driving Questions included as part of the materials mailed to teacher implementing the research study in their classes. Students handed in the hard copies at the completion of the simulation.

~ Future projects will have to take into account the growing issue of Internet monitoring and filtering and the strategies that school districts have been gradually deploying to protect students and resources witnessed over the three years of the Computational Literacy project.

During the course of the study, we bumped up against school districts deploying filtering equipment that limits the type of access that students or applications have to the broader Internet. This filtering and monitoring is meant to prevent malicious users or applications from collecting information about students or the resources they are using. (This is particularly important in the age of 'spyware' that may gather information or potentially use the computer resources for malicious purposes.)

The limitation we encountered was in the use of web services (using XML packets to communicate from a deployed applet to a central server) to enable a communications architecture. The documents that were being exchanged between the applet and the central Krell and University of Northern Iowa servers for the purposes of tracking were similar in structure to documents and information that could be potentially exchanged by spyware, adware or other malicious applications.

~ The amount of bandwidth required to serve a classroom of student computers was a real impediment to data-collection, sometimes impacting the amount of available in-class time for students to experience the topic simulations.

The medium used for the simulation environments and online data collection consists of Java-based applets and data communications relating to the students' interactions with the computational models. The Java applets and data reporting are accomplished using persistent network connections to centralized servers located at the Krell Institute and the University of Northern Iowa. Both of these locations are in central Iowa. For a typical-use scenario, this approach is appropriate and the simulations are very responsive. However, when classrooms of students simultaneously access the centralized servers to serve up the modest-sized Java applets, bandwidth at the local school systems can become saturated very quickly. When this occurs, a considerable delay in the time required to initialize the Java applet on the student's computer is incurred. These in-class delays are significant enough to impact the amount of available classroom time to the point where students are unable to complete the session.

~ Future classroom applications will have to be designed with bandwidth limitations in mind, even for schools that apparently have a high-bandwidth connection to the broader Internet because the students using the simulations often have to share resources with other students that could be using bandwidth intensive applications, such as streaming video, etc.

Potential Solutions. The main bottleneck discovered early in the implementation phase of the research study occurred when a classroom of students were concurrently initial downloading the Java applet codes that make up the topic simulations. In order to circumvent this constraint caused by a combination of Java bytecode size and a limited bandwidth, development turned toward solutions that would still be acceptable within the administrative constraints of the systems (no administrative rights required, for example) and that would alleviate the chokepoint behavior.

~ The bottleneck caused by network bandwidth was addressed in the project's last year. The potential solution to the problem involves running Apache, Tomcat and MySQL on the local system using a local USB drive all within non-administrative privileges.

The approach taken to alleviate this problem has been to serve up the Java applet codes that make up the models using removable media. Instead of serving up the Java applet codes from the remote servers, student computers can load the Java applet code from a local USB drive or CD drive. This approach would eliminate the network bottleneck, but still presented some technical challenges.

The crux of the technical issue with serving up the Java applet code from a local drive instead of over the network involves the core use of the Tomcat environment for the web content delivery and an internal MySQL database to manage dynamic content for tracking based upon school districts and instructors. Tomcat is a Java content delivery environment used on the central servers for managing the Java aspect of the web content (http://tomcat.apache.org). In order to maintain usability and tracking of the students' interactions with the models when leveraging the local drive to provide the Java applet codes, the Tomcat functionality needs to be provided on the local system as well.

Further, school, student, and tracking information that is communicated over the common network connection back to the central servers use MySQL database connections internally to maintain state and facilitate tracking. The centralized servers communicate with the MySQL database privately. However, with the Tomcat functionality running off of a local USB drive on the student's computer, the MySQL functionality was also required locally.

In summary, the bottleneck caused by network bandwidth was addressed in the project's last year. The solution to the problem involves running Apache, Tomcat and MySQL on the local system using a local USB drive all within non-administrative privileges. USB drives are provided to the local teacher and startup of these services requires running a simple automated startup script. The startup script that configures the Tomcat temporary file locations and starts up the services is executed when the drive is inserted into the computer's USB drive if the system supports an AutoRun feature. If the AutoRun feature is disabled, the script is run by double-clicking on the autorun icon. The USB drives are collected after the classroom session, and the MySQL database information is read off of the USB drives and merged into the common database.

Limitations of the Study

The Computational Laboratory containing the four topic modules *Population Dynamics, Carbon Cycle, Spread of Disease*, and *Rock Cycle*, each centered on a computational simulation, provided a safe place for students to scientifically test out alternatives, to make predictions and to gain experience with "cause and effect" and "if-then" reasoning without learning background knowledge specific to a software package. Working through the scenarios in each module gave students practice in mathematical problem-solving strategies such as trial and error, searching for patterns and analyzing data in tables and/or graphs. After working through module content, teachers were asked to give students the opportunity to explain their strategies and outcomes. Through their interactions with the simulations, students could build their skills in graph interpretation, data analysis and mathematical modeling as well as critical thinking skills, the ability to use science content knowledge and the ability to interpret and evaluate scientific information. Yet, when the study was completed and all the data collected and analyzed, the Computational Literacy Experimental Research Study was unable to lend any validity to our hypothesis that:

> *Students who use our topic simulation will show greater understanding of the science content as evidenced by their pre- and post- test scores, and better critical thinking and problem solving skills as evidenced by their far transfer test scores (our Logic test).*

The study, conducted with 44 teachers, 19 treatment and 25 control; 1542 students, 635 treatment and 907 control was affected by a number of limitations, one -- the asymmetry between treatment and control groups. The asymmetrical non-alignment of research study participants was the consequence of many more control participants than treatment, and is a recurring theme throughout the research study.

<u>Teacher Participation</u>

Ultimately, our research study, while specifically targeting high school science students in the three states, Maryland, Iowa, and Tennessee, is primarily a story of the teachers who taught those students. The study's impact was greatly affected by those teachers who choose to participate, as well as by those who opted out along the way.

When the teacher surveys were analyzed to get a sense of the study's participating teachers, it was learned that the group was overwhelmingly female; control teachers were more experienced than their treatment counterparts by almost 4 years. They had been at their current schools longer by almost three years and they reported being more confident in their knowledge of the science they taught. Both treatment and control teachers were almost comparable in technology experience and access to technology with treatment teachers showing a slight advantage. They were also comparable in their belief that all students could benefit from the use of simulations.

In addition to the asymmetry between treatment and control participants, another limitation of the study was the poor fidelity of implementation by participating teachers in general, but to a greater degree by treatment teachers. Of the 44 participating science teachers, 36 returned fidelity of implementation documents (Topic Teacher Reports), equivalent to an 81% response rate. Control teachers in general documented more topic related lessons, their descriptions were richer and contained more usable information. Treatment teachers tended to report using the simulations alone with few or no supporting lessons. Seventeen of the teachers who returned Topic Teacher Reports qualified as having implemented the study with fidelity -- a mere 39%. But treatment teachers were only 4 of the 17, representing less than one quarter of the fidelity group but only 9% of the total group of participating teachers and 21% of treatment teachers.

In sum, more experienced and confident in their science knowledge, control teachers reported greater fidelity of implementation of their topics that were addressed to almost twice as many students than their treatment counterparts.

Student Participation

By the time the student data had been cleaned and reconciled there were 3181 students entered into the database. Of those students 1639 (52%) did NOT have a valid permission slip, signed by a parent or guardian, allowing them to participate in the study. The 1542 (48%) students who did have permission slips were our student study participants. This highlights another limitation of the study -- our limited ability to get high school students to return permission slips signed by a parent or guardian.

The student data was analyzed using Repeated Measures ANOVA. The data indicated that treatment students using *Population Dynamics, Carbon Cycle,* and *Spread of Disease* did not perform better than their control counterparts. We believe that the pattern of higher scores for the control group might indicate that the two groups were not equivalent in knowledge prior to the study. Because of limited participation in the *Rock Cycle* topic and simulation there were no results available for that topic.

The student results for Maryland were also analyzed using Hierarchical Linear Modeling. Maryland had 70% of all the students in the study and had school report cards that supplied covariates that could be used in the analysis. There was insufficient data for Spread of Disease, but for Population Dynamics, gender was a significant explanatory value for the topic. All treatment students performed below control students but female treatment students scored the lowest of all students participating from Maryland using Population Dynamics.

For the Carbon Cycle, class type (honors or non-honors), used, as a student level covariate was statistically significant. Students in the control group taking non-honors courses scored the lowest, followed by students in the treatment group taking non-honors courses. Students in the control group taking honors courses scored the third highest and students in the treatment group taking honors courses scored highest. Honors students outperformed their non-honors counterparts regardless of whether they were in the treatment or control group however the treatment group performed best overall when using this topic.

Of the 1542 students participating in the research study, 952 had Logic Tests (the measure of far transfer) results, representing 61% of participating students. Thirty-nine per cent of participating students did not have results. Of the 952 tests, 611 were control and 341 were treatment. The results of students' performance on the Logic Test indicated that control students scored higher on the measure but there were also twice as many of them. Once again the recurring research study theme of the asymmetrical non-alignment of research study participants.

Theoretically, the randomization of participants into control and treatment groups encourages comparability between groups. But it is likely that the two groups of teachers participating in the study were not comparable. When their students were examined almost across the board (with the exception of honors students in Maryland, using the *Carbon Cycle*) control students outperformed treatment students. It is likely that in this research study, the two student participant groups were not comparable in knowledge prior to the study.

The Technical Environment

Probably one of the biggest technical limitations of the study was student’s inability to submit answers to the Driving Questions via the project's online tracking system. Almost across the board teachers submitted student performance tasks on the hard copies supplied as back up when materials were mailed to them at the onset of the implementation of the research study. Other technical limitations included the amount of bandwidth required to serve a classroom of student computers, limited network architecture in some schools and monitoring and filtering constraints employed by schools and districts. Often these restrictions increased the amount of time it took the simulations to mount on an individual computer.

Lessons Learned…

(…about the design and implementation of an experimental research study dependent on the recruitment, retention, and participation of science teachers of diverse high school students in a variety of educational settings, spanning different states)

~ The development of additional strategies and tools to support evolving computational literacy was addressed during the formative and field testing phases of the project, but could not be used as part of the research study. In the future, small studies that focus on the developed strategies and tools would allow researchers to explore their effects and effectiveness in supporting the simulations that are the center of the topic modules.

The Computational Literacy Experimental Research Study was able to identify clear-cut findings with respect to the project's hypothesis that:

> *Students who use our topic simulation will show greater understanding of the science content as evidenced by their pre- and post- test scores, and better critical thinking and problem solving skills as evidenced by their far transfer test scores (our Logic test).*

We were much less successful in identifying findings with respect to meeting the challenge of building teachers' and students' evolving notion of computational literacy defined by the project as:

> *An individual's capacity to understand the relationship between domain knowledge and the mathematical and visualization/modeling processes that are the building blocks of computational science*

Students' ability to describe the affordances and limitations of the models with respect to real world situations and their conceptual understanding of the dynamic and interactive nature of the phenomenon or model were probably not sufficiently addressed either in terms of the performance assessment or the supporting materials. In the science classrooms that participated, learning the requisite domain knowledge was clearly the priority. Discovering underlying computational literacy inherent in the use of the simulations was secondary at best but in many cases non-existent for a variety of reasons (e.g., lack of knowledge, lack of time, insufficient awareness). Devising strategies and tools to support evolving computational literacy was one of the tasks taken on by the project. Yet, in the end, there was no place for the additional resources in the experimental research study. In the future, it would be beneficial to conduct a series of small studies that investigate the effects of the developed strategies and tools in supporting the topic simulations.

~ Though not acknowledged when designing the research study, as implementation of the study progressed project team members recognized that teacher requirements for participation (particularly as part of the treatment group) could be perceived as burdensome.

Research study data was collected from a number of sources. All of the data sources except one – school report cards, teachers participating in the study were asked to complete and/or keep track of -- making teachers responsible for almost all of the research study's data collection. Teachers were asked to:

- *Complete an online teacher survey* whose goal was to investigate their pedagogical knowledge, backgrounds, and beliefs as well as to determine their habits and uses for technology in general.
- *Fill out and submit*: Topic Teacher Reports (the study's fidelity of implementation document); teacher permission slip signed by each potential research study teacher participant; class rosters listing all students in each class section taught by a participating teacher were

requested once teachers had agreed to participate in the study; a payment voucher at the close of teacher's participation in the study for remuneration for participating in the research study.
• *Distribute and collect student permission slips* that had to be filled out and signed by each participating student's parent/guardian
• *Administer and submit:* pre-/post, and Logic Tests; Driving Questions offered distinct information about treatment students.

~ One of the projects biggest challenges was getting high school student participants to return permission slips signed by a parent or guardian. And even those permission slips that were returned would have been useful in a number of ways if a standard requirement for student demographic information had been established.

Permission slips are an essential part of the data collection effort, but there was information that needed to be included on the forms that would have improved our data collection and reconciliation efforts. Permission slips should have included the student's teacher's name, the student's gender, age, and grade.

~ The absence of a set of standard requirements when requesting demographic information on students (e.g., gender, age, grade) across instruments and data sources (e.g., via student permission slips, class rosters, etc.) negatively impacted our data analysis efforts.

As the study data was being collected, the project identified a number of opportunities for incomplete and inconsistent data collection. For example, the project requested class rosters and usually got them but the information they contained was never consistent. Sometimes they included age, grade and gender but just as often they did not. Because teachers, schools and districts handle things in their own way and we were not always specific in our requests, our data collection was inconsistent and affected our ability to analyze research study outcome data.

~ A lack of on the ground local support adversely affected teacher retention, participation, and the quality and quantity of data returned by teacher research study participants.

With so many demands on a teacher's time, an email from someone never met is just not the most effective means of supporting a large-scale research study. Had there been representation from the research study in each participating school encouraging teacher participation, available to trouble shoot technical issues related to the simulation and the computational laboratory, reminding and nudging teachers to get things done in a timely fashion, data returns might have been better.

~ Alternative options over traditional face-to-face orientation need to be developed for teacher training/ preparation. One of the objectives of the project became, by default, to see if teachers could pick up the topic materials and make the simulations valuable in their classrooms without a lot of outside support.

1) Interested in making teachers feel prepared and supported as they implemented the research study in their schools, the project initially planned for an orientation to the project for teachers in each participating state. We anticipated assembling the necessary resources, identifying a centrally located site and offering a half-day to a whole day's training that would allow teachers to get familiar with the simulations and the assessment instruments, to ask any questions that may come up and to have experienced using the simulation/s before actually using it / them with their students. Among the three states, school start dates vary. Tennessee schools begin the earliest, then Maryland and Iowa schools. Research study materials needed to be available to teachers as soon as school started or at least within the first couple of weeks of the school year. As the implementation dates drew closer and the reality of funding state wide orientations became apparent, the project decided that the teacher training could reasonably consist of each teacher agreeing to go through each simulation while completing the online version of the driving questions herself before using it with her students. In theory we would be able to track / acknowledge their initial participation through the online data gathering mechanism put in place by the project. In

reality, this became a suggestion that we could not enforce and one that few if any participating teachers actually followed. A formal training / orientation did not take place. In thinking ahead, a plan to develop a web-based orientation to project resources would benefit those teachers considering using the simulations in their science classes.

Because documentation of implementation of the topics and simulation use in classes was so crucial to the study, in the future it will be important to provide participating teachers with alternatives/options that help them overcome their reluctance to provide fidelity of implementation documentation.

To add some context to student outcome scores, measures of fidelity were built into the study's data collection strategy. Fidelity implementation documents, the study's Topic Teacher Reports, had the potential to offer a glimpse into the classrooms that participated in the research study. In similar circumstances, a study would probably benefit from making the fidelity of implementation documents available through a variety of media options including the web, email as well as supplying participants with a digital version of the hard copy version they received in their original implementation packages. To improve the quality of the documents, more explicit instructions possibly would encourage participants to be more specific in supplying the information needed to provide some classroom context, and encouraging participating teachers to describe the obstacles they encountered throughout the study could significantly contribute to the project's understanding of context.

At the onset of the project, the team hoped to establish some fidelity validation through classroom visits but teachers were reluctant to invite us in for observations -- maybe because they were new to the material and/or because they didn't know us. The original research plan was to augment the self-report document with a sample of direct classroom observations in 4 treatment classrooms in each of the three participating states and with teacher interviews. This did not happen because teachers were reluctant to invite us to observe, had difficulty pinpointing the days that they would be implementing the simulations in their classrooms, or just did not respond to our requests. The addition of this type of information would have added to what we learned from the research study since so many questions evolved as a result of working with the data

~ The limited number of questions on the pre- / post- test measures jeopardized the tests' reliability and restricted the project's data analysis efforts, leaving little "wiggle" room for alternative strategies.

~ Failure of the online data collection strategy meant that student participants had to complete hard copies of the Driving Questions instead of being able to work completely on the computer. This also meant that instead of the questions being available digitally, evaluation of the documents had to be conducted by hand on students' hard copy replacements.

~ An original project goal was to work with teachers who had had limited exposure to computational science, yet our initial development work was only informed by master teachers who were very experienced in computational science.

The project team conducted focus group discussions with Maryland Virtual High School master teachers regarding which of the topics identified were challenging for their students to understand and for which a simulation would be a valuable teaching tool. As we developed and tested the topic modules and assessment instruments, it probably would have been beneficial to the project and the research study to have cast a wider net to identify, include, and talk to teachers who more closely resembled our target audience.