Evaluation of the CoreModels Project
Final Report

October, 2001

Wendy Friedman, Katie McMillan Culp
Education Development Center/Center for Children & Technology

Introduction

This is a final report on the evaluation of the CoreModels program, a three-year effort to support science teachers in the state of Maryland in developing computer-based modeling curriculum units (using the computer program STELLA) and integrating systems thinking into a range of science courses. This summary report presents data on the program's impact on student learning, and considers these findings in relation to the evaluation's examination of two other themes: the successes and challenges encountered during the program implementation, and the program's impact on participating teachers' beliefs and practices regarding computer-based modeling and systems thinking.

Overview of the program

The CoreModels program was a three-year project, funded by the National Science Foundation. This program was designed to increase the integration of systems thinking and computer-based scientific modeling into science classes across Maryland through the development of curricular materials and teacher training workshops. Although research has strongly suggested that modeling can be a powerful tool for supporting high-level reasoning by high school students in scientific domains, achieving a broad-based adoption of modeling tools and curricula has been difficult and effective implementation in classrooms has proved challenging (Mandinach and Cline, 1994; Roberts & Barclay, 1988; Stratford, 1996). The main challenge for the program, then, was to develop an infrastructure that would facilitate a successful adoption process and support students' and teachers' use of a systems approach to studying scientific phenomena over time. To address these issues, the CoreModels program used several distinct strategies to support their high school science teachers in becoming sustained and thoughtful users of its materials and systems approach, such as creating flexible modeling units tailored to state standards and curriculum goals, and developing close collaborations among teachers of specific scientific domains.

CoreModels grew out of the Maryland Virtual High School project, also funded by the National Science Foundation. This program had two main objectives - to determine if modeling activities might help students in achieving core learning goals, and to determine if a process of peer support might help teachers in implementing these activities in the classroom. Building on what had been learned through the Maryland Virtual High School project, the CoreModels program intentionally emphasized three specific organizational features that it determined to be central components of successful programs with these goals.

- CoreModels invited teachers into using modeling software (STELLA) by providing them with a diversity of curricular materials to facilitate the integration and use of that software within their classrooms (see Year 2 Report for more details).
- CoreModels involved teachers in the development of these classroom materials, and in the training of new participants, using peer-to-peer teams and collaborative groups.
- CoreModels aligned its materials with the Maryland Core Learning Goals, designing them specifically to help teachers meet state goals with their students.

More specifically, CoreModels was designed as a three-tiered organization of teachers, headed by Mary Ellen Verona. The first tier was comprised of three Center Directors, each responsible for one region of Maryland. The second tier was made up of Supporting Teachers, who were expected to help the Center Directors develop materials, as well as to provide one-on-one peer support to Participating Teachers within their regions. The final, largest tier was made up of Participating Teachers. This group was trained to use STELLA and implement CoreModels materials in their classrooms through project workshops. To further facilitate materials adoption, and to help them integrate modeling and new technologies into their teaching as seamlessly as possible, the Participating Teachers were also invited to engage in one-on-one peer support relationships with Supporting Teachers.

The evaluation of CoreModels

Over the course of the CoreModels program, CCT's evaluation has focused on three central themes, which correspond to key components laid out by the program.

- The efficacy of the program's core features as professional development mechanisms for teachers. CoreModels has used mentoring, workshops and online communication as professional development mechanisms for its participating teachers. The evaluation of CoreModels has included an investigation of the effectiveness of these features for meeting program goals related to professional development.
- Changes in teacher practices. CoreModels was designed to introduce teachers to new tools, new curricular materials, strategies for implementing a "systems" perspective in the teaching of science and new ways of working with other educators. The CoreModels evaluation has tracked teacher beliefs and practices related to technology use, materials implementation,

student learning goals and teaching and assessment strategies, through surveys, interviews and observations.

- Impact on student learning. Teachers involved with CoreModels have engaged in a professional development experience with an ultimate goal of engaging students in a "systems approach" to science through the use of modeling to investigate and understand various scientific processes. The CoreModels evaluation has looked at the evolution of teachers' beliefs about what their students can learn through modeling, and the development and use of appropriate assessments.

The research context

An accumulating body of research is demonstrating that computer-based modeling and simulations are potentially powerful tools for improving junior high and high school students' performance in science classes. The results of the most recent NAEP science exam demonstrate significant improvements in science scores for eighth grade students who report using these tools in science classes. For twelfth graders, scores improve significantly when these tools are used as little as once a month or less, and improve again when use goes up to once or twice a month (U.S. Department of Education, 2001).

While these findings strongly suggest the importance of computer-based modeling and simulation in the classroom, they do not help us to understand what it is about using these tools that might be improving student performance on these tests. What can students learn from using these tools? And under what kinds of classroom conditions – what curricular context, what kinds of teaching strategies, activities and work products – are these positive outcomes most likely to occur?

Work done by David Hestenes and his colleagues at Arizona State University (see http://modeling.la.asu.edu/modeling-HS.html for information on the Physics Modeling Project) provides one set of answers to these questions (Wells and Hestenes, 1995). This project focuses on refining our understanding of how modeling can improve student understanding of specific scientific concepts, and has generated important evidence about the value of modeling in science education. Hestenes and his colleagues have demonstrated that students whose teachers closely follow a university-designed modeling curriculum and pedagogical approach produce significant gains on the Force Concept Index as compared to students in other classes. Additionally, Hestenes and his colleagues demonstrate that student achievement on this measure drops in correspondence with how closely individual teachers followed the program's recommended course of implementation, which includes specific instructions on what content should be covered, how to structure daily activities, how to lead discussions, and other pedagogical and classroom management issues.

Hestenes' work is important because it provides an answer to one of the two questions raised above – what is it, exactly, that students can learn from modeling activities? Their research makes a strong case for the potential power of modeling in the science classroom. It demonstrates that a properly designed and implemented modeling course can enhance students' understanding of core physical science concepts. The training model used in Hestenes' program, however, is difficult and costly to implement on a large scale and does not attempt to demonstrate how it could be scaled across diverse classroom contexts.

The CoreModels program was created to address this challenge – how can the promise of computer-based modeling to be spread broadly across a range of scientific domains, curricular areas, and school and classroom contexts? This evaluation, in turn, explored the successes and challenges the CoreModels faced in implementing this program; the evidence of student learning related to the modeling curricula developed and delivered in this program; and begins to address the second question raised above -- under what kinds of classroom conditions are positive outcomes most likely to occur?

Rather than delivering a set curricula and set of practices, as Hestenes and his colleagues did, the CoreModels program sought to develop a professional development model based on constructivist principles. The program focused from the beginning on inviting teachers into an engaged learning experience in which they would drive their own explorations into the cognitive purposes of modeling, the relationship between modeling and content knowledge, and the role of classroom discourse in generating and supporting engaged student learning. The CoreModels project has asked, how can teachers without extensive access to specialists work with one another to bring a promising and innovative practice like modeling into their science classrooms? In addition to addressing the major challenge facing advocates of the value of using computer-based modeling in science classrooms, this approach addresses two of the key challenges to improving science learning in K-12 contexts:

- The challenge of *scale*: how can promising practices, tools and curricula move beyond testbed programs and be disseminated to a broad population of teachers, while retaining the structural and conceptual qualities that have made them successful in more controlled settings?

- The challenge of *locally-driven quality professional development*: How can individual schools and districts establish programs of professional development that are cost-effective and not dependent on outside experts, but that retain the characteristics of high-quality professional development? Sustained engagement with a peer group around a set of problems, opportunities to develop curricula that fit local needs and priorities, adequate opportunities for experimentation, and substantial guidance from peer mentors are all qualities identified in other research as being key to causing long-term change in teacher practices. However, these qualities are difficult to support in cost-effective and locally sustainable ways.

Because the CoreModels project focused on this set of challenges, examining student learning in the classrooms of CoreModels teachers sets a very high standard for judging the value of modeling in the science classroom. This data tells us not only about what is *possible* to accomplish with students using modeling software and curricula (as does the work of Hestenes, et al), but what was achieved in the classrooms of teachers who designed, built, and sustained the conceptual framework of the program, the curriculum, and their own network of professional development and shared professional expertise. CoreModels teachers have a level of ownership over their modeling curriculum and classroom activities that is central to their teaching of these concepts. Studying a program with this quality of local, teacher-based ownership of this innovation is crucial to establishing the real value of computer-based modeling, because scalability and sustainability of such an innovation depends on establishing and sustaining these conditions at the local level.

Evidence of Program Impact on Student Learning

This student learning study uses complex assessments to capture student achievement related to a highly diversified, teacher-driven program. Rather than being designed to maximize evidence of student learning (which would have required focusing on a narrower range of curricular topics, a more uniform implementation model for teachers, and a more structured form of assessment), this study asked what evidence of student learning would be produced from a program that had intentionally emphasized a highly naturalistic implementation model, one designed to privilege communal knowledge-building by teachers over time, and the application of modeling to a broad range of curricular topics.

Teachers involved in CoreModels worked together over the course of the program to develop and implement instructional activities that integrated modeling approaches into science curricula. Each activity was designed to meet the Maryland High School Science Core Learning Goals as well as the AAAS Project 2061 Benchmarks. At the same time that MVHS teachers were implementing these modeling activities, the Maryland State Board of Education was field-testing the Maryland High School Assessment (HSA) tests, the final piece of the state's systemic reform plan. The HSA includes both selected response items (i.e. multiple-choice) and constructed response items which require the analysis, synthesis, and written expression of ideas.

The program directors and the evaluation team decided together to construct student assessment measures that would be short and open-ended and that focused on students' ability to:

- meaningfully interpret the graphical representation of data
- understand the relationship of a model to the real world behavior it represents

There were several reasons for doing this, including:

- Anecdotally, teachers reported that they especially saw improvement in their students' level of sophistication in these two areas specifically;
- The emphasis in the evaluation on the role modeling played on shaping students' conceptual understanding, as opposed to their technical mastery of either the software or of particular methods of data manipulation. Open-ended quesitons were our best opportunity to collect evidence that would allow us to assess this aspect of student learning.
- Many MVHS teachers were concerned that, because their students were not accustomed to expressing themselves in writing in science classes, they would have difficulty with the constructed response items on the HSA. By using at least one constructed response question, it was felt, the CoreModels assessments could serve as test-taking practice for these teachers and their their students. By meeting this peripheral need among the MVHS teachers, the task of using the CoreModels assessments in their classrooms would be made more relevant to their teaching.

<u>Methods</u>

Measuring student outcomes for the CoreModels project was a core goal of the evaluation, and by Year Three of the project we had strong evidence, both from the reports of participating teachers and from the existing research base, about what aspects of student learning were most important to assess. The two primary issues we chose to investigate were:

- Students' ability to interpret visual representations of data;

- Students' awareness of and ability to explain the heuristics of a model. That is, how a model does and does not encompass the complexities of the system it represents.

These two issues are key standards in the AAAS 2061 Science Standards (CITE) for modeling and visualization. Interpretation of graphs was also an important issue to many teachers involved in the program. Student awareness of the heuristic value of a model was not a major issue for some teachers in the program, although it was most likely to be of interest to those teachers who had been involved in the program the longest (see the discussion of changes in teacher practice, below).

To many teachers, the assessment question of greatest interest was whether use of the STELLA exercises increased student understanding of the particular content being addressed by the exercise. However, because the goal of this evaluation was to examine whether modeling was making a unique contribution to student learning *across* content areas, we did not focus on this question. Rather, this study was designed to examine whether modeling activities implemented in connection with a wide range of scientific content would produce discernible improvements in students' understanding of scientific concepts that cut across the specific content areas.

Finally, since teachers had reported, anecdotally, seeing improvement in their students' ability to meet these learning objectives as more MVHS activities were used in classroom instruction, we decided to measure the effect of multiple uses of MVHS activities on student learning.

<u>Subjects</u>. 449 students participated in the assessments that are reported on here (358 biology students and 91 physics students). They represent biology and physics classes taught by 11 biology teachers and 4 physics teachers in 11 schools (6 rural, 4 suburban and 1 urban). The academic levels of the courses range from Basic Skills to Advanced Placement, and the courses were taught by teachers with between one and four years of experience using STELLA in the classroom.

The 15 teachers who participated in this study were recruited in the fall of 1999 from within the CoreModels program. All participants were required to meet the following conditions:

- Cover three MVHS activities during the second semester of the 1999-00 school year;
- Administer a CoreModels assessment after each of the three required activities;.
- Sumbit the original assessments to the program sstaff, and return a copy to to the students;
- Score the assessment copies according to the Maryland High School Science rubric;

- After scoring, return the assessments to the students and discuss the results prior to administering the next assessment.

Instruments. Two open-ended assessment questions were designed for each activity in biology and physics. The first question presented the student with a graph related to the topic recently studied and asked the student to explain its meaning. The second question asked the student to evaluate the ability of the model to represent real world behavior. See Appendix A for copies of sample assessments. Both questions were scored using the 5-point Maryland High School Science Rubric used on the Maryland High School Assessment exams.

Procedures. For biology classes, the time that passed between consecutive assessments ranged from same day to 12 weeks. The assessments covered 8 different topics, of varying levels of content difficulty. Teachers did not teach the topics in the same order. Since the sequence of quiz topic administration was counterbalanced across all classes, conceptual difficulty was not a confounding factor in the analysis done to assess the effect of exposure to modeling activities on students' performance. For the physics classes, the time that passed between assessments ranged from one week to 13 weeks. The assessments covered 6 different topics, of varying levels of content difficulty. Due to the sequential nature of the physics curriculum, everyone covered the same topic for Quiz 1 (Simple Kinematics). However, the four teachers then varied in which of the other models and associated quizzes they chose to use.

Three of the four physics teachers had classes in schools using a block schedule in which each class meets for a 90 minute period each day, covering a year's worth of physics in one semester. Since the assessments were being administered during the second semester of the year and the first topic of the physics course is simple kinematics, the MVHS activities fit seamlessly into the course content. The fourth teacher's class met for a 45-minute period each day and began in the fall, rather than the spring semester. Due to the fall starting time for this class, the MVHS activities were a review of topics already covered. Because of these differences, data from this teacher have been analyzed separately.

In the summer of 2000, these teachers met together to do a formal scoring of the assessments. Each question was subjected to blind scoring by two teachers. A third teacher was used to resolve discrepancies.

Methods of data analysis. Because there was some disparity across classrooms in the number of units and assessments implemented and the timing of distribution, sub-groups were formed for the purpose of further analysis. For Biology, three final groups were compared, which represented: 1) the entire group of students (n=358); 2) students who took completed all three units and assessments (n=160); 3) students who finished all three units and assessments before May 1 (n=30). For Physics, two groups of scores were analyzed – one including students from four classes, taught by three different teachers in three schools, and one including students from the class that was reviewing the content covered in the assessments. The Repeated Measures ANOVA Pillai's Trace test was used to measure the effect of exposure to modeling activities on student scores on the assessments.

Results - Biology

Although the biology teachers tried to meet the goal of three assessments, 4 of the 11 teachers only gave 2 assessments. Figures 1 and 2 present mean scores for the full cohort of students taking at least two biology quizzes and for the subgroup that took three assessments. Tables 1 and 2 list the means that are described in the following analysis.

Figure 1: Mean Scores Graph Interpretation Question Biology

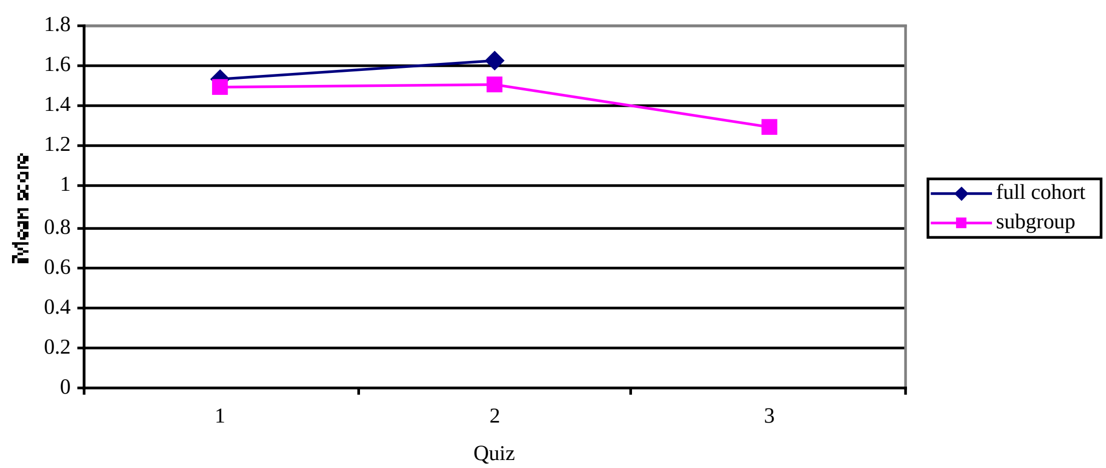


Figure 2: Mean Scores Modeling Heuristic Question Biology

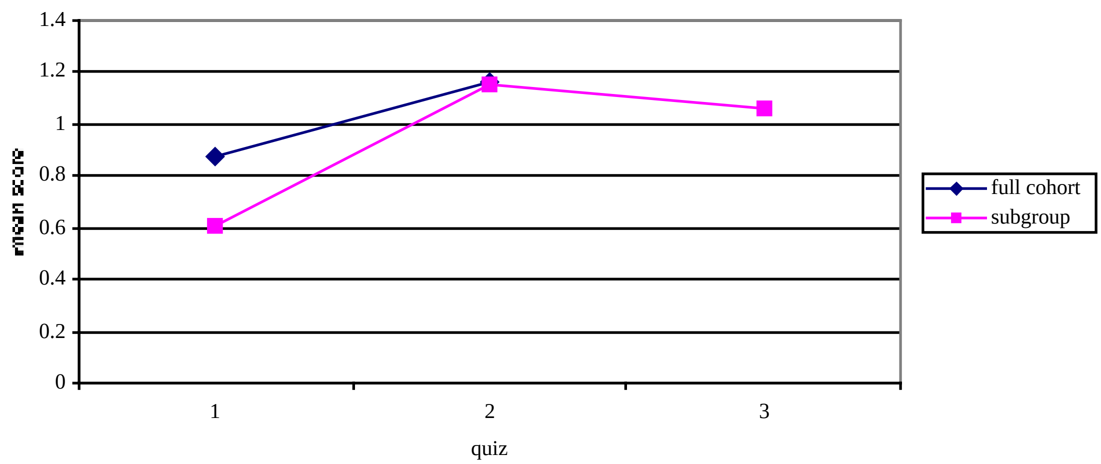


Table 1: Student Mean Scores, Graph Interpretation Question, Biology Assessments

	Full Cohort (2 or 3 assessments)	Subgroup (3 assessments)
	N=358	N=160
QUIZ 1	1.54	1.49
QUIZ 2	1.63	1.51
QUIZ 3	not applicable	1.30

Table 2: Student Mean Scores, Modeling Heuristic Question, Biology Assessments

	Full Cohort (2 or 3 assessments)	Subgroup (3 assessments)
	N=358	N=160
QUIZ 1	0.87	0.61
QUIZ 2	1.16	1.15
QUIZ 3	not applicable	1.06

For the subgroup that completed three assessments, mean scores on the graph interpretation question dropped significantly ($p<.01$) between time 1 and time 3, while mean scores on the modeling heuristic question rose significantly $<.01$). When the full cohort is considered, the increase in scores (from time 1 to time 2) on the modeling heuristic question holds, and scores on the graph interpretation question increase at a rate approaching significance ($p=.055$). The drop in scores for the subgroup on the graph interpretation question may be attributable to the fact that most third quizzes were given near the end of the school year, when student motivation was low.

Results - Physics

Figures 3 presents mean scores for the full cohort of students taking three physics quizzes. This data was gathered on 91 students, representing 4 classes taught by 3 teachers at 3 different schools.

For the physics cohort (n=91), mean scores for the graph interpretation question decreased significantly between time 1 and time 3 ($p<.05$), while mean scores for the modeling heuristics question increased significantly ($p=.005$).

We had hypothesized that student achievement would increase on both quiz questions as exposure to modeling activities increased. However, in both biology and physics classes, scores on graph interpretation questions tended to fall over time, while scores on questions about modeling heuristics tended to rise. This suggests that students' ability to interpret graphs was not improving over time (although they were being introduced to increasingly abstract physical and biological concepts), making graph interpretation progressively more difficult for them. In contrast, their mastery of overarching modeling concepts may have been growing over time, so their ability to answer questions about modeling principles in the context of various specific content did improve across multiple assessments.

We conducted an addition analysis to explore the relative impact on the introduction of new conceptual material to student scores on the graph interpretation question and the model heuristics question. We hypothesized that scores on the graph interpretation question would be more closely correlated to movement through more complex conceptual material from unit to unit, while scores on the model heuristics question would grow steadily regardless of the content covered in the units. Scores from a class of students (n=24) who covered, and were assessed on, eight units, provide an opportunity to unpack this relationship. This teacher saw student performance increase significantly on both the graph interpretation question and the model heuristic question over the first four quizzes, which all covered units addressing aspects of kinematics. However, when the concept of force was introduced on the fifth quiz, scores dropped dramatically on the graph interpretation question, but less so on the model heuristics question. Then, when the sixth unit reinforced the new material covered in unit five (this was an activity about elevator movement and again addressed concepts of force), means on both the graph interpretation and the model heuristics question increased significantly.

Analysis of scores from this one class suggest that the introduction of a new concept has a major impact on students' comprehension of graphical representations of the relevant material, while their mastery of broader modeling concepts may be growing steadily. This finding, while far from exhaustive, suggests that the kinds of modeling curricula used in this program may not be having a major impact on the depth of students' content knowledge, but may be quite effective at exposing students to modeling concepts and principles.

Table 3: Mean Scores, Physics

	Graph Interpretation Question	Modeling Heuristic Question
Quiz 1	1.73	1.13
Quiz 2	1.4	1.2
Quiz 3	1.2	1.5

Figure 3: Mean Scores for Graph Interpretation and Modeling Heuristic Questions, Physics

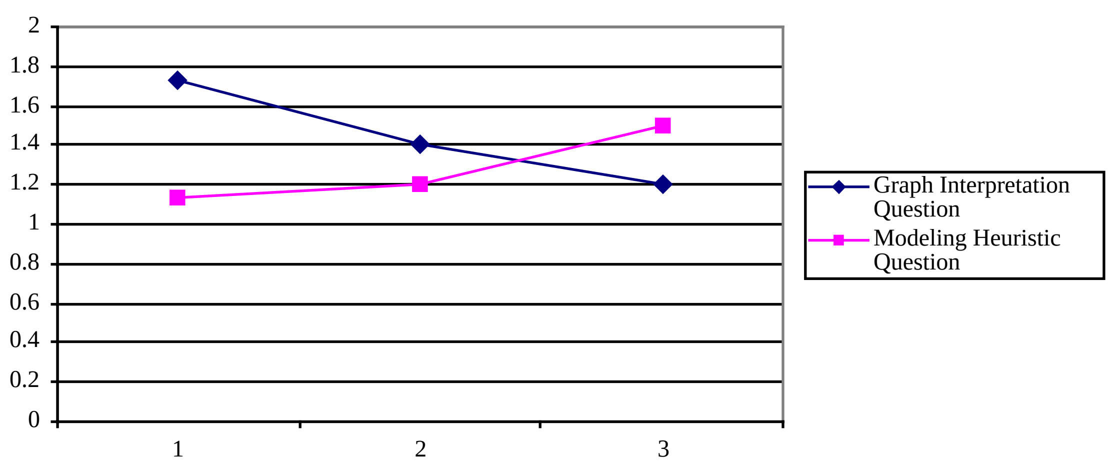
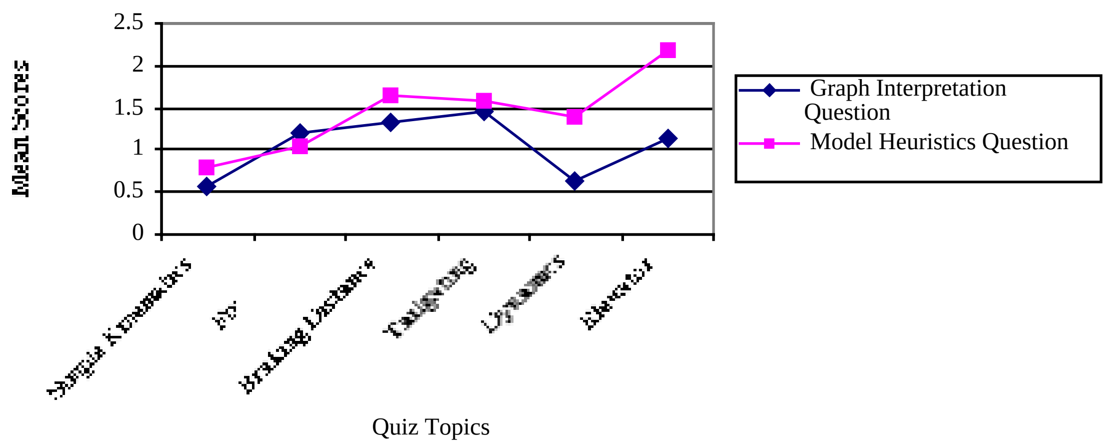


Table 4: Mean Student Scores for Individual Physics Class With Extra Quizzes

Quiz Topics	Graph Interpretation Question	Model Heuristics Question
Quiz 1: Simple Kinematics	0.58	0.79
Quiz 2: Free Fall	1.21	1.04
Quiz 3: Braking Distance	1.33	1.63
Quiz 4: Tailgating	1.46	1.58
Quiz 5: Dynamics	0.63	1.38
Quiz 6: Elevator	1.13	2.17

Figure 4: Mean Scores, Graph Interpretation and Modeling Heuristic Question, Across Multiple Physics Quizzes



Discussion

The student learning data presented above suggests that in the context of a peer-driven professional development program like CoreModels, teachers are able to use computer-based modeling in a range of curricular contexts to improve student understanding of some of the core scientific concepts underlying modeling as a scientific practice. The program is less successful in supporting teachers in improving their students' abilities to interpret visual representations of data. One conclusion that can be inferred from these findings is that central modeling concepts, such as the heuristic relationship of models to the physical world, seem to be relatively transferable concepts that can be elaborated across curricular content areas, while interpretation of visual representations of data remains, at least in this context, a more content-dependent skill that is not easily transferred from one content area to another. These findings demonstrate that CoreModels was successful in building teachers' understanding of, and ability to teach about, modeling not only as a way to explore specific content areas but as a particular conceptual approach to the task of scientific inquiry.

This program did not use a lockstep curriculum, and provided teachers with an enormous amount of flexibility to determine when and how modeling might fit into their teaching. But rather than resulting in a "lowest common denominator" set of practices among participants, this approach seems to have supported teachers in engaging with the conceptual underpinnings of computer-based modeling, and in turn to engage their students with these concepts as well. What features of CoreModels allowed this to happen?

Both our qualitative data and findings from our surveys suggest that many teachers in this program began with a highly procedural or technical perspective on modeling, and were interested in it primarily as a vehicle for reinforcing content knowledge they wanted to provide to their students. However, over time these teachers were able to appropriate another set of learning goals for their students as they participated in modeling activities. In the rest of this section highlight some of the key findings from three years of data collection that illustrate this process of gradual appropriation, and the features of the program that worked most effective to support this learning process for participating teachers.

Workshops. Program workshops were consistently received with enthusiasm by program participants. Participating teachers had a high regard for the teachers leading the program, as well as for the Supporting Teachers who emerged as leaders for each curricular area (these teachers were ones who took prominent roles in curriculum development and in workshop discussions). Their enthusiasm for these workshops was consistently reflected in interviews and survey data.

Interviews and observations allowed us to observe the evolution of program workshops over the three years of the project. In the early phases of the program, teachers were focused on discussions of the technical and logistical hurdles to using computer-based modeling activities. Limited computer access, lack of familiarity with STELLA, and concern about classroom management during independent work periods were dominant topics of conversation during these early periods. Over time, these concerns evolved to a second stage, in which logistical and technical hurdles were still prominent, but they had evolved to reflect the challenges teachers faced once they had gained some experience with modeling curricula and wanted to improve how the units unfolded in their classrooms. During this period, discussions often focused on topics like the layout of computer labs or the placement of computers in classrooms, as teachers began to want to move students back and forth between independent or group work on computers and class discussion. Questions about technical aspects of STELLA grew more advanced and concerned with tailoring use of STELLA to specific learning goals. However, during this period the group maintained, overall, a focus on modeling as an extension activity that could potentially enhance their students' mastery of specific content knowledge.

In a third and final phase, group discussions in workshop settings began to move away from technical and logistical discussion and toward the discussion of teaching practices, student learning goals and resources to support writing original curriculum. CoreModels participants came to a general consensus that having students engage in the model-building process is the preferable classroom experience, as opposed to exploring prefabricated models, an activity that many teachers had found preferable for their purposes early in the program. Teachers now viewed work with prefabricated models as being more appropriate to sessions early in the year when students are first using STELLA. In this phase, assessment arose naturally as a topic for discussion, as teachers began to incorporate modeling concepts into their learning goals for their students (in addition to content knowledge goals) and began to debate how best to evaluate student understanding of those concepts.

Program participants clearly brought their workshop experiences back to their classrooms. Survey data demonstrates that, over time, program participants increased their skill and confidence in using technology with their students and were more frequently using technology to support their teaching. Supporting Teachers also reported, in surveys and interviews, a growing reflectiveness about student learning goals for instruction and appropriate forms of student assessment. Participant teachers reported using the new curricula developed in the program (not only units that they developed themselves), teaching new concepts, expanding their use of technology and, in some cases, covering new content or covered existing curricular areas in more depth.

Peer support. CoreModels began with a carefully designed program of peer support, which involved pairing Supporting Teachers with Participant Teachers and encouraging regular classroom visits and discussions between each Supporting/Participating Teacher pair. However, these relationships rarely grew into the productive collaborative partnerships originally intended by the program. Because pairs were originally matched across schools, logistical challenges (travel time, being willing or able to miss one's own class time to visit another teacher's class) played a major role in keeping these relationships from developing. Further, teachers were generally unwilling to comment on or critique one another's practices, which they understood to be a primary purpose of observing one another's classrooms. This was reflected in "Supporting Teachers" rejection of the term "mentors."

Other professional development programs have had success with collaborative partnership models like this one. It is impossible to determine exactly what combination of local circumstances, program emphasis, and support structures might have been responsible for the relative lack of prominence of this feature of the CoreModels program over time. However, what we did find was that teachers gradually shifted to intra-school, more informal forms of peer support, and the program followed the teachers' lead and instituted cross-discipline, within-school, team-oriented peer support structures during Year 3 of the program. This model seemed to function more productively for teachers.

A parallel development in the program was the growth, over time, of some of the Supporting Teachers as program leaders. A subset of the Supporting Teachers took on increasingly prominent leadership roles within the project and within their schools, and became deeply invested in acting as mentors and collaborators around systems thinking curriculum issues with other teachers. The professional growth of this subset of program participants resulted in an expanded core group of teachers who were effectively leading the program and providing guidance to the larger cohort of teachers, strengthening an already strong group of teacher leaders and contributing to the persistent, gradual progress of the level and content of teachers' discussion of modeling over the life of the program.

Curriculum development. Another key feature of this program was its emphasis on teachers' development of modeling units that would fit into existing Maryland science curriculum. The goal of this dimension of the project was to avoid making modeling an "extra" curricular area that teachers would not find time to use, and instead to integrate modeling into existing curriculum structures as a mode of inquiry, rather than an independent course of study.

Curriculum development was a radically more time-consuming process than the program leaders had originally envisioned it to be. Developing, testing, refining and disseminating units became an all-consuming task for a small subset of the teachers, while many other teachers dipped in and out of the process, devoting considerable energy to the process when a unit was under development that fit particularly well into their curriculum.

Despite the difficulty of this aspect of the program, curriculum development actually played a crucial role in driving the evolution of teachers' thinking about modeling and systems thinking over time. Discussions during reviews of new curricular units, or during teachers' reports on their trials of new units with their students, were often occasions for particularly productive discussions of the function and purpose of modeling and the learning goals teachers were associating with the units. During the latest stages of the program, these discussions about learning goals led to discussions of the appropriate assessment practices to associate with these curricular units, and these conversations were the crucial turning point in the group's overall shift toward understanding modeling as a way for their students to pursue open-ended inquiries rather than as a tool for reinforcement of content knowledge. In sum, these teachers used the curricular writing and revision process to develop and clarify their own goals and expectations for modeling and systems thinking within their curricula.

Conclusions

Student learning

This evaluation demonstrates that computer-based modeling, embedded within traditional science curricular sequences, can support students in developing an understanding of the core concepts underlying modeling and systems thinking. Specifically, this study investigated students' ability to articulate and explain the heuristic nature of a model in relation to the physical system it represents (a features of systems thinking stressed in the national science standards), and found that student performance improved in this area as their exposure to modeling and systems thinking increased. Further, this finding held across multiple scientific disciplines (biology and physics) and across a range of specific content areas.

This evaluation also demonstrated that modeling and systems thinking, even when they consistently engage students with graphical representations of data, do not necessarily lead to students' improved performance in graph interpretation. In the context of this study, students' graph interpretation skills appeared to remain closely tied to their mastery of the content of the task, and was not affected by growth in students' mastery of modeling and systems thinking concepts.

Teacher practices

As is indicated by the areas of positive change in the student learning study, these teachers did develop new beliefs about topics including assessment, learning goals for science classes, and classroom management during technology-rich activities through their participation in this program. Those changing beliefs were reflected, in turn, in changes in their classroom practices. Teachers participating in this program increased their use of technology with students, became more comfortable with student-centered use of the technology and with students engaging in sustained project work. They became more reflective about their goals for students' learning in their science classes, and in many cases decreased their emphasis on mastery of content knowledge in favor of an increased focused on inquiry skills (such as systems thinking).

This program also engaged participating teachers in a range of activities that ensures that their professional growth had an impact beyond their own classrooms. Teachers involved in this program developed curricula that were shared within the project cohort and within the science departments of the teachers' respective schools. They shared modeling approaches and STELLA techniques with other science teachers in their schools. Many presented on modeling and systems thinking at local and regional conferences and professional development workshops. As a cohort, CoreModels teachers increased their leadership activities at the school, district and state level across the life of this project.

The professional development model

The CoreModels program began with a strong commitment to a peer-mentoring model of professional development. Over time, this model was largely abandoned in favor of two overlapping systems – an informal, statewide network of teachers deeply engaged in systems thinking and modeling curriculum development, and a series of school-centered cohorts of teachers supporting one another as they experimented with those curricula in their classrooms.

These two systems interacted primarily during face-to-face workshop sessions, although some individuals operated in both systems, and some online conversation and some school visits also occurred to supplement the interactions at the workshops. During workshops and these other interactions, teachers' discussions of their on-the-ground experiences with the curricula intersected with the developing thinking of the core curriculum team. These conversations were the key moments when the "big ideas" of systems thinking and modeling (especially about learning goals and optimal curricular and pedagogical approaches) combined with teachers' "real life" concerns and enthusiasms. The culture of this program framed both the "big ideas" and the "real life concerns" as being valid and important issues. Consequently, these conversations allowed teachers to explore, over time, how they might change their beliefs and practices in order to work more effectively with their students with these tools and curriculum.

In conclusion, the CoreModels program demonstrates that a professional development program can support teachers in using modeling with their students in ways that will have a positive impact on their understanding of modeling concepts that are important and difficult to master areas of the national science standards. Key factors that made the success of this program possible included the close-knit, peer-to-peer structure of the CoreModels community; the presence of a core group of relatively senior and especially dedicated teachers who were able to act as a vanguard in exploring the relative value of student-driven model construction and open-ended inquiry into systems; its sustained, long-term commitment to curriculum development, testing and revision; and the program's open structure, which allowed teachers to adapt and adopt modeling curricula in ways that were realistic for their own particular classrooms.

References

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. Project 2061. New York: Oxford University Press.

Mandinach, E., & Cline, H. (1994). Classroom dynamics: Implementing a technology-based learning environment. New Jersey: Erlbaum Associates.

Roberts, N. & Barclay, T. (1988). Teaching model building to high school students: Theory and reality. *Journal of Computers in Mathematics and Science Teaching*, 8(4), 13-24.

Stratford, S. J. (1996b). Investigating processes and products of secondary science students using dynamic modeling software. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

U.S. Department of Education, National Center for Education Statistics (2001). Report on the 2000 National Assessment of Educational Progress: Science Exam.

Wells, M., and Hestenes, D. (1995). A modeling method for high school physics instruction. The American Journal of Physics, July, 1995, 63(7), 606-619.