**Collaborative Research: An Agent-Based Simulation Environment for Predictive Longitudinal Modeling of High School Math Performance**

Proposal Submitted to the National Science Foundation

By

UC Santa Cruz

In collaboration with

Los Alamos National Laboratory

&

UT Austin

Principal Investigator: Michael Strong, PhD

## Project Summary

The Agent-Based Simulation Environment (ABSE) collaboration is an **exploratory project** designed to create a powerful set of agent-based modeling and visualization tools to examine the precursors for math success at 8th grade and beyond. It addresses Challenges 2 and 4 and is designed to provide information to school personnel that will enable them to take action to improve the provision of math education in their districts. The project will test the hypothesis that a sophisticated and dynamic predictive agent-based model utilizing comprehensive student and school data presented through a visualization tool will effect a better understanding of a) the precursor and concurrent factors that lead to 8th grade Algebra success and failure and b) which factors in 8th grade Algebra curriculum or environment positively or negatively affect consequent terminal math choices in high school. The scalable visualization tool that results from the ABSE collaboration will provide insight into how a wide variety of variables impact student learning in mathematics from 6th grade to 12th grade, leading to improved decision-making by teachers, administrators, and policymakers .

ABSE is a collaborative project comprised of four organizations united in purpose and bringing unique abilities. Experts in agent-based modeling at the Los Alamos National Laboratory will analyze an unusually rich database provided by the San Jose Unified School District in California and conduct a proof-of-concept study that will demonstrate the benefits of applying complex agent-based modeling techniques to educational data in order to predict the factors that are associated with successful trajectories in math learning up to 8th grade and beyond. Educators and math advisors will interpret the LANL output and contribute to the construction and running of simulation models. Advanced visualization techniques will be applied to the model data by experts at the University of Texas at Austin in order to make the findings maximally accessible to educators and policy makers so that they can devise appropriate action.

The **intellectual merit** of ABSE is embodied in the advancement of reliable systemic analytic and interpretation tools for education stakeholders as well as predictive tools to support better decision-making for math education. Current educational analytical efforts often lack either comprehensive, reliable data resources or advanced modeling and visualization technologies. ABSE will employ both, thereby transcending what is currently available to decision makers in the field of education.

The **broader impact** of ABSE's work on this project will be directed first to the school district upon whose data the modeling is based. Insights into the structural and functional correlates of student math performance in the upper grades will provide a basis for constructing and adjusting early interventions, informing their existing early warning system with initial predictions from a more sophisticated modeling technique than has been possible with current resources. Further impact will be felt among representatives from other school districts who participate in planned dissemination conferences, the research community who read project reports and articles or attend academic conferences, and other education decision makers whose goal is to improve the standing of math education in the United States. Not only will this project have the potential for contributing to math learning outcomes, but also, if successful, to other STEM subject areas. The analytic agent-based modeling methods and interpretive visualization tools will inspire further research and development.

# PROJECT DESCRIPTION

## 1. Goals and purpose

This project addresses the following **research question**: What are the school-related predictors of mathematics success of public school students in 8th grade and beyond? In a recent report Hanushek et al. (2010) analyzed international and national data (from the PISA and NAEP databases) to pinpoint how the US Class of 2009 compared with students from other countries in advanced math. No fewer than 30 of the 56 other countries in the comparison had a larger percentage of students who scored at the international equivalent of the advanced level on the NAEP tests. The authors characterize the percentages of US high-achieving math students as "shockingly below those of many of the world's leading industrialized nations."(p.4). Furthermore, the differences are not explained by the numbers of immigrant test takers in the US, and the performance of students with parents having a college degree (and therefore with a presumed advantage) is disappointingly low. These findings are especially troubling since math accomplishment in secondary school is particularly significant for both an individual's and a country's economic well-being. Research such as that by Bishop (1992) and Murnane, Willett, and Levy (1995) suggests that math skills are a better predictor of future earnings and other economic outcomes than other skills learned in high school. Speculating on the sources of the poor math performance of American students, Hanushek and colleagues suggest:

> "Sources of the problem may lie in the lack of initiative among students themselves, anti-educational pressures within the adolescent peer group culture, a lack of parental concern and support, anti-intellectual influences within the entertainment and mass media industries, a substantial minority population, high rates of in-migration, or even broader and deeper societal influences. But even though we suspect that one or more of these factors is at work, some of our findings point specifically to problematic elements within the nation's schools. That even relatively advantaged groups in American society—white students and those with a parent who has a college education—do not generate a high percentage of students who achieve at the advanced level in math suggests, we submit, that schools are failing to teach students effectively." (p. 32).

They further speculate that simply allocating more money to public schools is not going to solve the problem.

The Hanushek et al. study may be the most recent but is certainly not the only research report to demonstrate the relatively poor math performance of America's high school students, particularly minorities of African American, Native American, or Hispanic heritage (see NSF report by SRI International, 2008). Hanushek and colleagues posit numerous possible causes for this predicament, and point to the inescapable fact that our schools are falling down on the job of providing adequate education in math. It is clear that no single indicator can provide information about a phenomenon as complex as education (Shavelson et al., 1989) and we should look to a model that examines the relationships among the pertinent characteristics. Over the past few decades, many potentially relevant factors have been studied in an effort to explain math achievement. These include attitude (Ma, 1997), beliefs (Garofolo, 1989; Kloosterman, 1995; Schoenfeld, 1985; Schommer, 1990), gender (Benbow & Stanley, 1980; Fennema & Carpenter, 1981), parent education (Ethington & Wolfe, 1984; Ma, 1997; Tsai & Walberg, 1983), employment (Greenberger & Steinberg, 1986), homework (Keith & Cool, 1992), and school size (Lee & Smith, 1997), as well as numerous others. An efficient model would be one that could examine many or all such factors, looking for significant interactions, and highlighting those that have the most power to predict success.

A very recent development in modeling and simulation of the education system has been the introduction of *agent-based models* (ABMs). Past efforts to model the education system have either used static equation-based models, or system dynamics models that capture the overall flow of students through the system. These types of models are useful for understanding educational systems primarily at a structural level, but are generally unable to capture the link between the classroom experience of students and outcomes at the system level. ABMs are different because they can start by modeling individuals – for example, students and teachers – and their interactions in a given environment, such as a particular classroom, school, or district. The behavior of an individual agent is governed by a set of rules that enables it to make choices or change behavior in response to environmental inputs and interactions with other agents. ABMs can be populated with thousands of these agents. System outcomes are predicted by aggregating the behavior of individual agents, and with proper software architecture are scalable to millions. Unlike other forms of modeling, this enables a connection to be made between the individual pathways students follow through the educational system and the overall performance of the system.

The main **goal** of this proposed project is to initiate the development of such a model for the purposes of enabling a school district to identify, in accordance with NSF's **Challenge Two**, where in their students' educational trajectories appropriate interventions should be made to address early warning signs of potential problems or to nourish early indicators of latent promise. This approach is consistent with the notion of the "complex systems view" of educational policy research suggested by Maroulis et al. (2010). The argument associated with this approach is that:

> "conceptualizing schools and districts as complex adaptive systems, composed of many networked parts that give rise to emergent patterns through their interactions, holds promise for understanding such important problems." (Maroulis et al., p. 38)

The present project will adopt the agent-based modeling approach recommended by these researchers by utilizing the power and experience of scientists at the Los Alamos National Laboratory (LANL) to analyze longitudinal data from an exceptionally rich data warehouse compiled by the San Jose Unified School District (SJUSD). Agent-based, dynamic modeling techniques reveal insights into the behavior of a system over time by separating the specific components and examining them together in an overall model "that is able to address questions pertaining to the paths between equilibrium points, such as whether a transition to choice might make a system worse before it gets better and for how long, and for whom it is worse." (op.cit, p.39). Subsequent simulations whereby individual or groups of components are varied systematically can suggest proposed solutions and interventions, together with their potential costs, that district personnel, policymakers, or other educational leaders can then put into action. The effective presentation of these simulations will be enhanced by the use of sophisticated visualization techniques executed by a collaborative partnership between the Learning Technology Center (LTC), which has a focus on data visualization in support educational research, and the Texas Advanced Computing Center (TACC), home of one of the country's most advanced visualization facilities. Both centers are located at The University of Texas at Austin. In short, this project will bring together a scientific approach to predictive modeling that incorporates its collaborators' capabilities in mathematics, computer science, and domain expertise.

The dynamic modeling approach has been used by others to address STEM issues. For example, the Raytheon Company developed a STEM model designed to help increase the number of students who pursue majors in the STEM fields at college (BHEF, 2010). Likewise, access to ever-increasing capacity for storage and computing power, improved database technology and data mining techniques, together with an urgent need for speedy solutions to complex problems have opened up advanced modeling and data mining to the field of education. As defined in a chapter on data mining by Luan (2002) referring to the Gartner Group (2000), data mining is: "the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques." (Luan, p.19). Luan used data mining models to analyze data to monitor and predict community college students' transfer to four-year institutions. As these two examples demonstrate, educators who have discovered data mining and dynamic modeling have applied it mainly to higher education settings. We propose to use agent-based modeling to improve math education at the primary and secondary levels.

Independent of and prior to this proposal, LANL has focused on three main areas for its initial efforts to model education, beginning first in New Mexico. First, they have worked to develop an understanding of the important issues facing the educational system, based in interactions with education experts and officials in New Mexico, and through review of the research and policy literature in the education field. Second, they have worked to obtain and interpret publicly-available education data for New Mexico. They have analyzed these data to develop statistical models of the student population by school district across the state, and the relationship between student demographic characteristics and test scores. These statistical models provide input data for their third and most important effort, which is to develop an agent-based model for predicting the influence of key student, teacher, and classroom variables on education outcomes. This prior experience will be critical in accelerating the work on the project proposed here.

Combining visualization with data mining and agent-based modeling will enable the compiling and presentation of complicated information in ways that non-experts can understand, making it more likely that school district personnel will be motivated to take action. There also exists a precedent for modeling and visualization of large-scale, multivariate educational data using tools familiar to complex systems and nonlinear dynamics. Techniques typically used to model convective and diffusive flows have been effectively applied to trajectories of student high stakes test scores (Marder & Bansal, 2009; Shah & Burke, 1999). In addition to the numerical modeling involved, the visualization techniques from that particular discipline effectively convey student pathways, including diffusion, disappearance, and retention, as a function of economic class and grade level. Using the agent-based models developed as a result of the work proposed here, the collaborators at LTC and TACC will use their visualization capabilities to develop the most effective ways to represent the patterns, trends, and relationships among the wide variety of variables included in the models to facilitate a better understanding of the precursor and concurrent factors leading to 8$^{th}$ grade Algebra success and subsequent math choices.

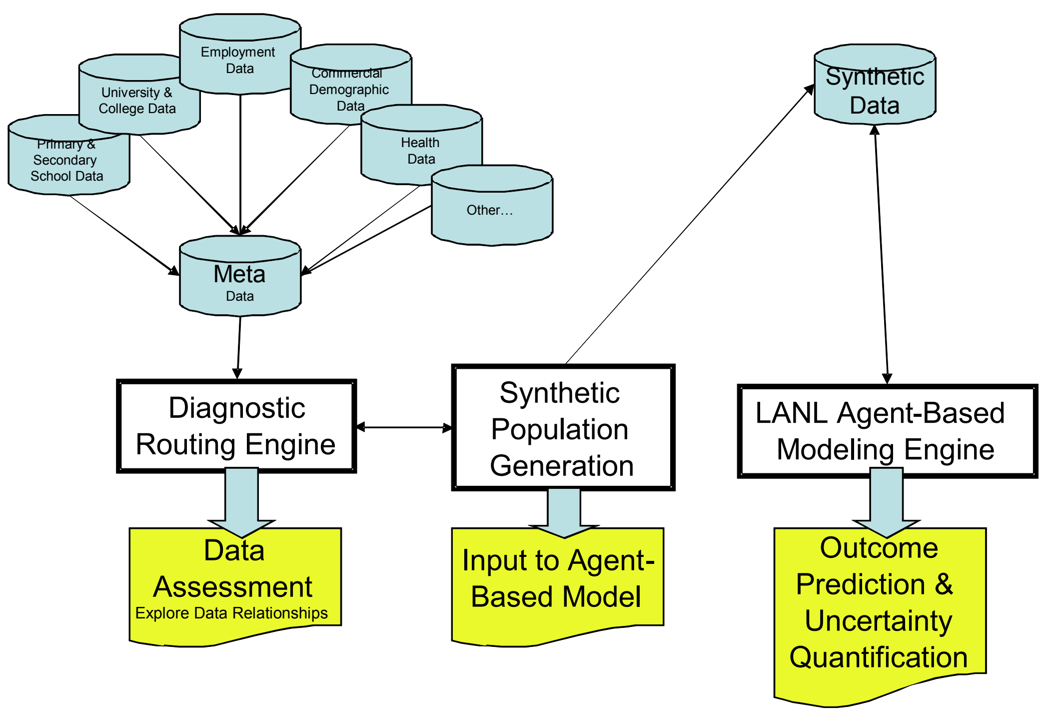
2. **Results from prior NSF support – N/A**
3. **Research and Development Design**

Our hypothesis is that agent-based modeling of rich school district data can be used to improve math education. In this proposed project, LANL, guided by education professionals on the project team, will explore innovative use of previously developed sophisticated agent-based modeling techniques in order to predict U.S. education outcomes. Using data from SJUSD, LANL will deliver within one year from project start, the completion of a proof-of-concept study that demonstrates the efficacy of using agent-based modeling to 1) predict student math outcomes at 8th grade based upon precursor and concurrent factors and 2) to predict high school math success, stasis, or failure based upon 8th grade math experiences and other concurrent factors. LTC and TACC will display the model's output using sophisticated visualization technologies.

The proof-of-concept study will make use of existing modeling and analytic experience that LANL staff already has in-house. Staff background and LANL tools are available from having performed various infrastructure modeling projects on behalf of national security. The team's recent experience gained from analyzing and modeling the education data of New Mexico will be leveraged, too. Most critically, LANL's uniquely powerful experience in quantifying the uncertainty of simulations will be employed to demonstrate the trusted capability for predicting educational outcomes.

One way of picturing the architecture of this approach is shown in Figure 1. Three base

**Figure 1: Architecture Summary**




actions will be undertaken in the execution of the model with emphasis on the initial data assessment, incorporation of distributional information, and quantification of uncertainty. Preliminarily, a diagnostic engine will be developed. This engine will assess the data, exploring relationships that will inform the agent-based model. Following this, a population is constructed with agents (various factors that are present in education, as explained further below), and finally, the agent-based model engine will be developed which will provide predictive trajectories on student, teacher, or district performance. The model will allow end users to manipulate, or turn on or off, factors and families of factors to review trajectories under varying conditions, resulting in a proof-of-concept study applying these approaches to school district data for the purposes of predicting 8th grade math outcomes.

Two LANL technical teams – Statistical Sciences and Socio-Technical Modeling – will work under the leadership and professional guidance of education professionals from the University of California Santa Cruz (UCSC) and the SJUSD, together with input from the Advisory Board. The LANL teams' roles are to 1) demonstrate unique analytic capability that goes beyond existing practice; 2) to use existing experience with national security and education data to develop a functional prototype of an infrastructure simulation for math education.

LANL analytic and modeling development will be steered by directed research questions articulated by the education profession team members. The initial set of guide questions are: *What are the school-related predictors of mathematics success of public school students in 8th grade and beyond?*

*a) as measured by mathematics performance (test scores) broken down by different mathematical skills*

*b) as measured by enrollment in algebra class (8th grade and high school)*

*c) as measured by algebra and mathematics grades in 8th grade and high school*

## The Core Modeling Proposition

LANL proposes to explore the use of an agent-based modeling method of modeling socio-technical infrastructure to anticipate systemic responses to control inputs, and also responses to shock. One might characterize curriculum selections and class size policies as examples of control inputs, and performance pay changes and budget cuts as shocks. The "agents" themselves will be populations of teachers, students, parents, etc., based upon actual demographics and performance data. Agents may also be inanimate objects, such as textbook contents, curriculum content, tests, facilities, etc. Populations may be constructed from actual records and census based individuals, or may be synthetically generated to reflect general statistics. Synthetic generation is used to explore alternative scenarios as well as to bridge gaps in existing data. LANL has automated tools to assist in the controlled and rapid generation of synthetic populations. The data available from SJUSD promise to provide LANL the best possible basis for conducting this work

## The Plan of Work

**The Data:** LANL will receive a copy of the extensive database of school, student, and teacher information going back 14 years from the SJUSD. This comprehensive and robust data warehouse comprises 33GB of education data, with 61 objects and 2,707 attributes. Currently, the warehouse includes information on the following variables:

- Student variables: unique student identifiers; comprehensive student information (demographics, grades, credits, GPA, attendance, coursework, program/intervention participation, etc.); a wide array of summative and formative assessments; office disciplinary referrals/behaviors; health and physical fitness.
- Teacher variables: certification status, ethnicity, years of service, professional development courses taken, and endorsements.

- School Climate variables: SJUSD's Annual Student Survey (approximately 20,000 student surveys, grades 3-12 and 1,700 teacher surveys) captures students' and teachers' perceptions of their learning environments.

As a collaborator on this project, SJUSD has agreed to transfer files to LANL in an acceptable, previously approved format.

**Statistical Sciences Tasks:** The Statistical Sciences team will employ its expertise in the statistical assessment of data and uncertainty quantification for models to provide a rigorous analysis of sources of variability and resulting uncertainty in model estimates and predictions. This effort will be carried out through three primary tasks: (i) data assessment, (ii) incorporation of distributional information into the modeling process, and (iii) quantification of uncertainty.

***Task (i) Data Assessment:*** A preliminary assessment of available data will be conducted to obtain distributional information about the different variables and possible relationships between variables. This exploratory phase will employ both analytical and visual approaches to extract distributional information and identifiable structure within the data. The goals of this initial phase will be to develop an understanding of the variables in the database and to develop distributional information required to generate synthetic data for the simulation model.

***Task (ii) Incorporation of Distributional Information:*** The Statistical Sciences Team will work closely with the Socio-Technical Modeling Team to develop information about populations required for model development. Previous statistical research conducted at LANL in support of infrastructure modeling involved the development of methods for synthetic population generation. These methods use information about the distributions of several variables to generate multivariate samples from a population that is consistent with the available marginal data for one or more variables. Samples generated in this manner provide synthetic populations that can be used in the agent-based model. These samples preserve the distributional structure present in the original data, address confidentiality concerns that would arise if actual data was used directly in the model, and provide opportunities to examine a variety of what-if scenarios for proposed alternatives for improvements.

***Task (iii) Quantification of Uncertainty:*** The modeling process will make use of actual data from the San Jose database. As in any analysis involving data, this will inject stochastic behavior into the modeling process that needs to be appropriately modeled and quantified. The Statistical Sciences Team will develop approaches for understanding and quantifying sources of variability as well as the resulting uncertainty in model estimates and predictions. This effort will build on previous expertise in identifying and estimating sources of variability, understanding the propagation of uncertainty within a modeling framework, and quantification of uncertainty in results obtained using a specified model. Development of uncertainty estimates will require selection of various model input settings, which will then be used to generate model outputs corresponding to the range of specified input conditions. Variability in the outputs will be analyzed and attributed to various sources of uncertainty present in the modeling process. This inventory of variability will aid decision-makers in understanding simulation results and in choosing between alternative strategies for improvements.

**Socio-Technical Modeling Tasks:** The Socio-Technical Modeling team will leverage its long-standing experience in designing light-weight agent-based simulations in the socio-technical domain to design and implement a prototype model, dubbed the Agent Based Simulation Environment model (ABSE), with a focus on $8^{th}$ grade math as a test study. The tasks are: (i) Model Design, (ii) Model Implementation, (iii) Data Population, (iv) Parameter Calibration, and (v) Scenario Execution.

**Task (i) Model Design:** LANL researchers will use their modeling paradigm SimCore that requires the definition of agents (e.g., teachers), processes (e.g., teaching), and information bits called infos (e.g., math) that get exchanged. Processes, or services, run on agents, or entities; a service sends infos to other services on different entities which then process these infos and create more infos.

In this instance, researchers will define different agent types. The main types would be students and teachers, and additional types might be principal, parent, school board member, classroom, or even textbook. Each agent type has a set of attributes, such as ability, effort, grade, knowledge level, etc. Defining which agent types are needed and what their exact attributes should be will be a key part of the model design task.

Different agent types execute different processes. The main processes are: *Teach* for teachers, *Learn* for students. These two processes are coupled through infos, where each info might represent a lesson of, say, Algebra 1. The result of executing the Learn process is that the knowledge level of a student increases.

The exact nature of this increase in knowledge is a function of a variety of parameters and its precise nature will need to be calibrated in the Parameter Calibration steps. In previous work, we have made this knowledge increase a function of the student's previous knowledge, class size, student effort, teacher ability, student ability, teacher effort. The synthetic population generation methodology incorporates empirical information about the multivariate distributions of variables in order to generate synthetic data that has correlation structure consistent with the available data. We will experiment with more complex definitions even of student knowledge by looking at knowledge as a set or vector of data items that a student knows. Such highly discretized curricula are readily available in particular in mathematics, where one such knowledge item could be "rounding of three digit numbers". The advantage of such an approach is that we can make knowledge acquisition depend on the exact knowledge item that the student already has; we can even make it depend on the (simulated) time that has elapsed since the student last learned the most relevant knowledge items that are required to understand a new knowledge item. We can thus build and integrate a dependency graph among these knowledge items. Introducing such detailed knowledge acquisition may be suitable to answer some questions, whereas a simpler representation (such as just a simple number or levels similar to grade levels) may in some cases be more suitable, in particular for a less-technically astute audience.

The *Teach* and *Learn* processes are the core of our agent-based model, however, other processes should be added as well, in particular those that can be changed by policy makers. Such processes include recruiting, course selection (which is ultimately crucial to study STEM), extra-curricular activities, social interaction, curriculum changes, etc. We will also model the unfortunate but very real process *Forget* (or knowledge loss).

**Task (ii) Model Implementation:** The scalable discrete event simulation engine SimCore will be used to implement the ABSE model. SimCore applications of the past include a large number of the National Infrastructure Simulation and Analysis Center (NISAC) infrastructure sector models, which are unique to LANL in their complexity and their scaling properties. SimCore uses a heavily object-oriented and templated C++ to define base classes that are then overwritten by application classes. The complexity of implementation in SimCore is reduced if the model design is constructed within SimCore modeling paradigms.

**Task (iii) Data Population and Task (iv) Parameter Calibration:** In close coordination with the statistical analysis effort, the modeling team will analyze the SJUSD data sets, using disaggregation techniques, wherein the datasets are separated into component parts, to populate the ABSE agents and set initial values for parameters. These parameters will need to be calibrated against the SJUSD data. The key value that an agent-based model provides is its ability to simulate the effects of various scenarios and choose from a variety of options. In order to achieve credibility the model needs to predict observed data reasonably well. It is necessary to distinguish clearly between data that constitute input to ABSE and emergent data that ABSE produces, but for which we also have real-life data to compare and calibrate against. All ABSE parameters need calibration, but the knowledge acquisition function in particular will be the main focus for calibration.

**Task (v) Scenario execution:** ABSE will then study the model output for guide questions posed above. In particular, the team will set all parameters to match SJUSD data as much as possible and identify the root causes for math success or failure as well as mitigating actions through a large set of simulation runs, where the modeling team will vary parameters in a design experiment in coordination with the statistical analysis team to allow the model to come to statistically valid conclusions with respect to causation.

**Core Visualization:** A profound challenge that accompanies modeling and simulation of complex systems is how to present models and results in ways that are comprehensible and informative to the non-specialist audience. In the case of the ABSE, visualization challenges will largely revolve around representing the complex interplay of factors at work in the math pipeline as well as the predictive LANL models that describe these interactions. Specifically, the team at The University of Texas will focus on how these visualizations can best inform decisions relating to instruction, curriculum, assessment, and policy.

With access to some of the most powerful visualization resources in the nation, the LTC/TACC team will explore the most effective means of presenting this information (varying dimensionalities, static vs. dynamic representations, etc.) so that it facilitates understanding of these interactions and optimizes their representation in ways that best serve various educational stakeholder groups (teachers, administrators, policymakers, students, parents, and professional organizations). The LTC/TACC team is uniquely positioned to accomplish this task, as TACC houses one of the most sophisticated visualization centers in the world, including the world's highest resolution tiled display. Additionally, the soon-to-be-completed Learning Technology Center Visualization and Discovery Environment (LTC-VDE) will provide advanced visualization capabilities specifically designed to support educational research. Much of the visualization work will be done in the LTC-VDE supported by the vast TACC visualization expertise and resources. The LTC will have direct access to TACC via the Longhorn Visualization Portal, an internet gateway to the Longhorn visualization cluster and its easy-to-use interface for scientific visualization.

## 4. Evaluation

The project will be evaluated by Gargani + Company (GCO), an independent evaluation consulting firm with twenty years of experience evaluating educational programs and technology. The formative evaluation will be organized around three questions:

1. <u>Progress</u>: Is the project progressing as planned?

2. Use: How do school administrators, teachers, and other stakeholders use the information produced by the project?
3. Perceived Efficacy: Do school administrators, teachers, and other stakeholders believe that the information produced by the DMME project will help schools be more effective?
4. Scalability: How can this project be scaled up and/or replicated?

**TABLE 1: Relationship of evaluation questions, data sources, and who or what the data describe.**

| | Milestone Meetings | Surveys | Interviews |
|---|---|---|---|
| Progress | **ABSE** | **ABSE** | **ABSE** |
| Use | | School administrators, teachers, and other stakeholders | School administrators, teachers, and other stakeholders |
| Perceived Efficacy | | School administrators, teachers, and other stakeholders | School administrators, teachers, and other stakeholders |
| Scalability | | | |

Consistent with the purposes of the project, GCO will work collaboratively with LANL and other collaborators to answer these questions and provide rapid feedback that will support the model development process. The data for the evaluation will come from three sources—surveys, interviews, and regular milestone meetings. Table 1 relates these data sources to the three evaluation questions and who or what the data describe. The surveys and interviews of school administrators, teachers, and other stakeholders will be developed alongside the model in order to ensure the model's potential uses are fully represented. The milestone meetings will follow a format used successfully by GCO in other projects to monitor, document, and rate the progress of work.

The evaluation will be conducted over a 13-month period. Table 2 outlines GCO's schedule of work, which will involve monthly milestone meetings and repeated surveys and interviews of school administrators, teachers, and other stakeholders.

**Table 2: Schedule for efficacy study.**

| | S | O | N | D | J | F | M | A | M | J | J | A | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Milestone meetings | X | X | X | X | X | X | X | X | X | X | X | X | |
| Survey & Interview Development | X | X | X | X | X | | | | | | | | |
| Survey & Interview Pilot Testing | | | X | X | | | | | | | | | |
| Survey & Interview Implementation | | | | X | | X | | | X | | X | X | |
| Formative feedback | X | X | X | X | X | X | X | X | X | X | X | X | |
| Data Analysis and Final Reporting | | | | | | | | | | | X | X | X |

The information gathered from these sources will be used to provide timely formative guidance to project collaborators, and at the end of the year it will be documented in a final report.

## 5. Dissemination

The ABSE project has three methods of dissemination. These comprise distinct strategies for staff and administrators conducted by the Krause Center for Innovation in Education (KCI), strategies for policymakers, higher education and researchers conducted by the partners (LTC, TACC, LANL, SJUSD), and strategies for the broader education community conducted by the National Laboratory for Education Transformation (NLET) a new non-profit organization.

The primary teacher, staff and administrator sessions will be conducted by the Krause Center, a state-of-the-art teacher development facility based on the campus of Foothill Community College in Los Altos Hills, CA. Since 1999, more than 12,000 teachers have participated in KCI's programs, courses and workshops. KCI will hold three sessions spaced between the mid-point of the project and its conclusion. The first session will be to explore with a small group of teachers, staff and administrators their awareness, attitudes and experience about considering math learning trajectories. The second session several months later will be a small planning session to examine preliminary results from the ABSE and to interpret those for the final and larger session with educators. In this second session, discussion and planning will center on understanding the work of LANL and the LTC and TACC, and interpreting this into language where the ABSE methodology and findings can easily be communicated to a larger gathering of educators in accordance with the end goals of wider dissemination and applying what has been learned in school systems. The final session near the end of the project, will bring together approximately 30 select educators to discuss the findings, their application and their wider dissemination. At that session there will be a solicitation for volunteers to help carry on the work discovered in this project.

Dissemination to policymakers, researchers and higher education will be conducted by the core partners in the ABSE; LANL, LTC/TACC, and SJUSD. The LTC works closely with the Math and Science teacher certification programs, as well as with the Principalship and Superintendency certification programs, at The University of Texas at Austin. The LTC also initiated and hosted the Invitational Summit for Redefining Teacher Education for Digital Age Learners (http://redefineteachered.org), bringing together 100 educational leaders to generate a consensus vision and set of policy recommendations for redefining teacher education to support 21st Century teachers and learners. The LTC will work these and other stakeholder groups that are part of its ongoing outreach efforts to disseminate the findings and products of this work.

Each facility will arrange dissemination activities as part of their core outreach efforts to discuss the construction of the ABSE, to announce its findings, and to solicit further dissemination and participation in subsequent work. Each of the partners has substantial education practice and policy networks that can be involved in this activity.

Dissemination to the broader education community and to the general public will be conducted by NLET. NLET will design and produce an initial DMME website, several short web-based videos, and create a methodology for disseminating the project findings. NLET and LTC/TACC will be responsible for sharing research and development designs, findings, and reporting this information to the DR K-12 Resource Network and to the online data system. NLET board members and advisors will assist in dissemination and further planning for ABSE.

## 6. Expertise

Four interdisciplinary teams will collaborate to conduct the research project described here. The Principal Investigator, Michael Strong, is a senior educational researcher at UCSC, with expertise and long background in the study of teachers and teaching. The central component of the work on agent-based modeling will be conducted by a team of modeling experts from LANL, led by the Co-Principal Investigator, Joanne Wendelberger, an expert in statistical sciences. The essential collaboration of the SJUSD is led by Marcy Lauck, Manager of Continuous Improvement Programs for the District. The visualization component of the project is the contribution of members of the LTC at UT Austin, headed by Paul Resta, Ruth Knight Milliken Centennial Professor. Dissemination activities will be coordinated by Gordon Freedman and NLET.

LANL has been selected as a collaborator on this project because of their innovative development and application of statistical methodology to problems of national importance and unparalleled capability in socio-technical modeling. The nationally recognized Statistical Sciences Group has received multiple awards for its collaborations with academia and industry. In addition to its efforts in modeling New Mexico education data, LANL has worked with cities such as Portland, Houston, and Jacksonville to model potential city planning impacts. LANL saved billions of dollars for the city of Houston through modeling a potential new freeway project. Finally, LANL's socio-technical modeling teams have worked extensively with the Department of Defense and other agencies to help facilitate understanding of potential foreign military, economic, cultural, and psychological reactions to U.S. defense actions. LANL has demonstrated a commitment to support American STEM education by sponsoring numerous outreach programs in education such as the *Supercomputing Challenge, the Northern New Mexico Math and Science Academy, the Bradbury Science Museum*, *collaboration* with Lincoln Interactive to develop science and engineering curriculum and support materials, and in the recruitment of students to work with permanent staff on scientific projects in summer sessions or full-time paid employment.

**References**


BHEF (Business-Higher Education Forum) (2010). *Increasing the number of STEM graduates: Insights from the U.S. STEM education & modeling project*. Washington, DC: BHEF.

Hanushek, E.A., Peterson, P.E., & Woessmann, L. (2010). *U. S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?* PEPG Report No.: 10-19. Harvard University and Education Next.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 12*, 1262–1264.

Bishop, J.H. (1992). The impact of academic competencies of wages, unemployment, and job performance. *Carnegie-Rochester Conference Series on Public Policy, 37*(December): 127-194.

Ethington, C. A., & Wolfe, L. M. (1984). Sex differences in a causal model of mathematics achievement. *Journal for Research in Mathematics Education, 15,* 361–377.

Fennema, E., & Carpenter, T. P. (1981). Sex related differences in mathematics: Results from a national assessment. *Mathematics Teacher, 74,* 554–559.

Garofalo, J. (1989). Beliefs and their influence on mathematical performance. *Mathematics Teacher, 82,* 502–505.

Gartner Group. (2000). *The GartnerGroup CRM Glossary*. [http://www.gartnerweb.com /public/static/hotc/hc00086148.html].

Greenberger, E., & Steinberg, L. (1986) *When teenagers work: The psychological and social costs of adolescent employment*. New York: Basic Books.
Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly, 7*(3), 207–226.
Kloosterman, P. (1995). Students' beliefs about knowing and learning mathematics: Implications for motivation. In M. Carr (Ed.), *Motivation in mathematics* (pp. 131–156). Cresskill, NJ: Hampton Press.
Lee, V. E., & Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis, 19*(3), 205–227.
Luan, J. (2002). Data mining and its applications in higher education. *New Directions for Institutional Research, 113*(Spring), 17–36.
Ma, X. (1997). Reciprocal relationships between attitude toward mathematics and achievement in mathematics. *The Journal of Educational Research, 90,* 221–229.
Marder, M., & Bansal, D. (2009). Flow and diffusion of high-stakes test scores. *Proceedings of the National Academy of Sciences of the United States of America, 106(41),* 17267-17270.
Maroulis, S., Guimerà , R., Petry, H., Stringer, M.J., Gomez, L.M., Amaral, L.A,N., & Wilensky, U. (2010). Complex systems view of educational policy research. *Science, 330*, October, 38-39.
Murnane, R. J., Willett, J.B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251-266.
Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.
Schommer, M. (1990). The effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82,* 498–504.
Shah, C., & Burke, G. (1999). An undergraduate student flow model: Australian higher education. *Higher Education, 37,* 359-375.
Shavelson, R., McDonnell, L., & Oakes, J. (1989). *Indicators for monitoring mathematics and science education.* Santa Monica, CA: RAND.
Tsai, S. L., & Walberg, H. J. (1983). Mathematics achievement and attitude productivity in junior high school. *The Journal of Educational Research, 76,* 267–272.