# A Computational Social Model of the Education System

**Stephan Eidenbenz,**
**Patrick Kelly, CCS-3;**
**Benjamin Sims,**
**Joanne Wendelberger, CCS-6**

The question of how to improve the US education system, particularly in science, technology, engineering, and mathematics (STEM) fields, is an important policy issue. Yet there is little consensus on what policy interventions actually work: Smaller classes? Placing top teachers in failing schools? Tracking students by ability level? Education researchers have examined many interventions through rigorous statistical analysis and modeling [1,2]. However, these efforts are often complicated by the difficulty of isolating the effects of a particular intervention from a wide array of other variables—a common problem with assessing "social experiments" in real-world systems. Computational social models represent an alternative way of carrying out such social experiments. Although they cannot capture all the dynamics of a complex social system, such models can incorporate the most important factors. These factors can then be manipulated and monitored without the difficulties associated with doing so in real-world settings. These virtual experiments, so long as their limitations are understood, can be a valuable tool for assessing the potential impacts of policy changes.

This year, LANL completed the initial phase of an exciting collaborative research effort to construct a computational social model of the US education system. Funded by the National Science Foundation and led by the University of California, Santa Cruz, this project also involves education experts from the University of Texas at Austin, the National Laboratory for Education Transformation, the San Jose Unified School District, and the education evaluation firm Gargani and Company. The goal was to have these domain experts interact with LANL statisticians and computational modelers throughout the modeling process to ensure that the final result is not only computationally rigorous, but also relevant to the knowledge and concerns of the education community. This modeling effort builds on LANL's past success in modeling human behavior in socio-technical systems such as transportation and communication infrastructure [3].

The project takes a three-pronged approach involving conceptual model development, computational model development, and statistical analysis [4].

Conceptual model development is a process in which subject matter experts work with model developers to lay out the key variables that are relevant to a computational model and some of the relationships between the variables. In the current project, this initially took the form of informal conversations among the collaborators, but eventually involved a formalized knowledge elicitation process. A wide range of possible variables was considered, but in the end was cut down to a basic set of entities and properties shown in Fig. 1.

Computational model development is the process of implementing the conceptual model in an agent-based modeling framework. The progress of an individual student through the model in a particular subject area (say, math) is shown in Fig. 2. In order to run the model, a school system is specified consisting of schools, students, teachers, courses, classrooms, and individual classes. The model then assigns teachers and students to individual classes and progresses from one year to the next. Say student X has a medium ability to learn math. Each year, they are assigned a math test score based on their ability, their teacher's ability, the average ability of the class, and their math score from the previous year (if any). As a result, Student X struggles in Grade 1, but the influence of teachers and peers enables him/her to improve to a high math score by Grade 4.

Statistical data analysis is used to characterize San Jose Unified School District data in order to generate realistic inputs to the agent-based model. For this initial effort, we identified a set of students who could be tracked continuously from Grade 2 to Grade 8 and used Grade 2 test scores to approximate their initial learning ability. We also generated teacher effectiveness scores based on the average improvement in the scores of their students over their scores in the previous year. Based on these inputs, the model was then run for grades 3–8. The model was run numerous times with different weights assigned to the model variables. For each model run, the distribution of predicted student math scores in Grade 8 was compared to the actual distribution of Grade 8 scores in order to identify which set of weights generated the most realistic results. While still in its early stages, this modeling collaboration has already generated significant interest in the education community, and is

**For more information contact Joanne Wendelberger at joanne@lanl.gov.**

**Funding Acknowledgments**
National Science Foundation

likely to continue. Future efforts will focus on expanding the number of variables in the model, refining the relationships between variables, and using statistical methods to generate realistic synthetic populations that can be tailored to address specific research or policy questions [5].

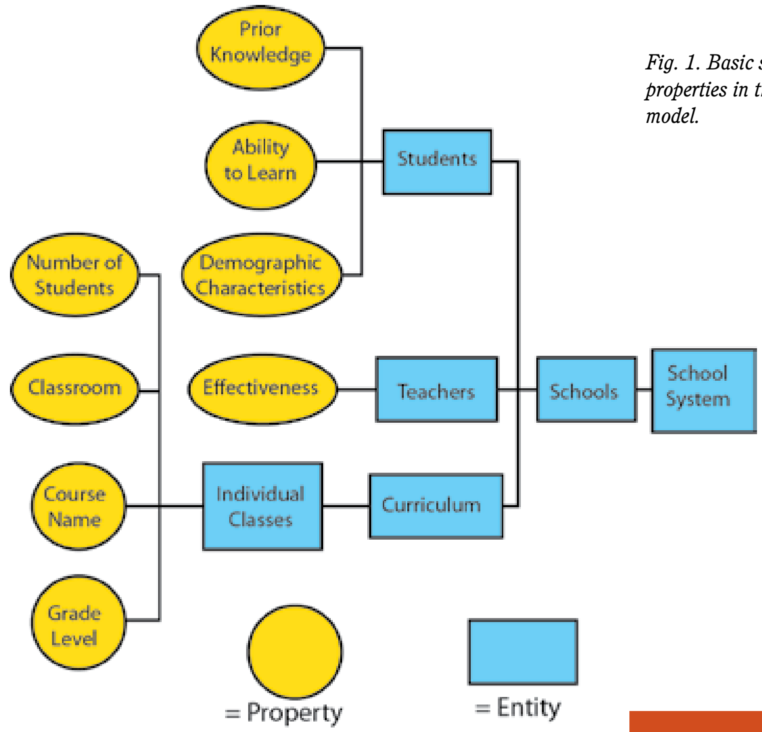


*Fig. 1. Basic set of entities and properties in the conceptual model.*

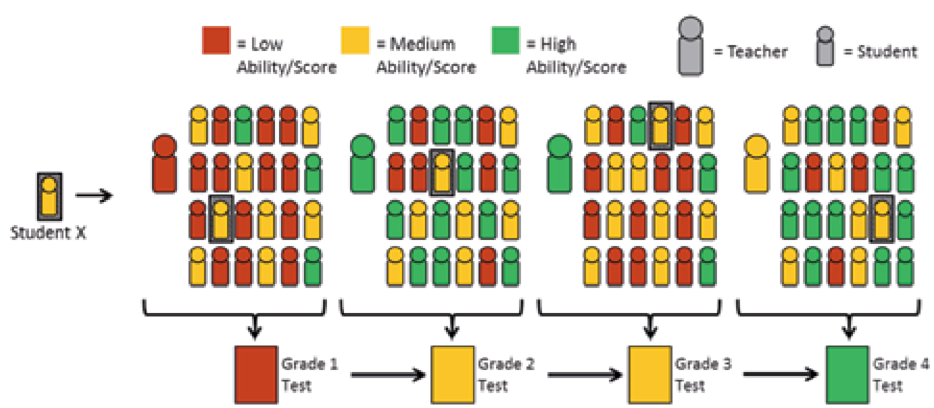


*Fig. 2. Student X's progress through the computational model.*

[1] Oppendakker, M. et al., *Sch Effect Sch Improv* **13**, 399 (2002).

[2] Young, D.J. et al., *J Educ Res* **89**, 272 (1996).

[3] Galli E. et al., "ActivitySim: Large-Scale Agent-Based Activity Generation for Infrastructure Simulation," *Proceedings 2009 Spring Simulation Multiconference*, (2009).

[4] Sargent, R.G., *Proceedings 2005 Winter Simulation Conference*, **130** (2005).

[5] Beckman, R.J. et al., *Transport Res Pol Pract* **30,** 415 (1996).