Can Data Mining Save America's Schools?

Schools have more data than ever, and there's a major push on to make better

use of it to identify students who need help.

By Marianne Kolbasuk McGee

InformationWeek

Oct. 25, 2008

During the 2007-2008 school year, students in New York City weren't the only

ones getting report cards. So did the city's 1,500 public schools.

New York City's Department of Education, responsible for 1.1 million

children, began issuing annual "progress reports" to each of its schools

last fall, with grades ranging from A to F. Principals of the top 20% of

schools received bonuses from $7,000 to $25,000. Teachers at schools with

high poverty rates qualify for a bonus program. And over time, schools

receiving D's or F's face possible changes in leadership, restructuring, and

even closure.

Yet grading schools is kid's stuff compared with what a growing number of

school districts around the United States think they can do with data mining

and data analysis. Combining standardized test scores, attendance, grades,

and other data sources, districts are trying to spot weaknesses and

strengths of not just schools, but groups of kids and even individual

students. For example, the Plano, Texas, district scanned data across eight

schools and zeroed in on 60 kids who looked at risk of failing a

standardized test, and created plans to help them.

This is just the start. While there's much criticism of the federal No Child

Left Behind legislation--mainly, that it's left teachers teaching to test

requirements, not student needs--it has undeniably created a mountain of

data, all of which can be analyzed. "Without data, we just went on opinion.

There was no data to back up instructional needs of kids," says Cindy

Goldsworthy, assistant superintendent of Derry Township School District in

Hershey, Pa. "It takes what in education was often driven by intuition into

showing quantitative proof."

In New York City, the effort centers on an $80 million Web-based data mining

and business intelligence project called Achievement Reporting and

Innovation System. Beginning this year, all 80,000 of the city's public

school teachers will have access to the ARIS system and get training in the

analysis tools. Parents also will have Web access to data about their

children this year.

The school-by-school grades are based on a complex analysis of an array of

information about each school, including students' year-over-year academic

progress, state test performance, and attendance, as well as surveys of

parents. "Any metric we have in the progress report, you can drill down on,"

says Jim Liebman, who champions ARIS as the chief accountability officer for

the city's schools. So a principal can see the school's grade on ARIS, which

might indicate the school is lagging in math. The principal can drill down

to find the school's math scores are in the bottom third of city schools,

then look further to see the individual students who make up that bottom

group. A step further shows what math skills they're weakest in. Principals

can spot, for example, "these 10 kids" who are having trouble in math and

English and need extra help, Liebman says.

The effort to give teachers these tools began last fall as New York began

rolling out access to ARIS to all principals and small "Inquiry Teams" of

teachers in every school, who are using the tools to analyze the performance

and growth of the most at-risk students.

This year, it's being rolled out to every teacher, and the Inquiry Teams are

asked to hold training sessions at their schools just as all those teachers

get their login information. There's deeper training for teachers who want

it--in two waves, one starting next month and another in the spring, since

new functionality and data will be added later in the year.

The Inquiry Team at I.S. 143 Eleanor Roosevelt School in New York City

Photo by Erica Berger

Those Inquiry Teams have been focusing on identifying troublesome

areas--say, one group of fifth-grade boys in a particular school having

difficulties in math--and diagnosing the problem so the school can consider

how to react. New York also wants teachers to use the system to figure out

what works.

So if a teacher tries a new way of presenting lessons for students

struggling with multiplying fractions, the ARIS system can be used to track

progress from one month to the next based on students' periodic assessment

test results, compared with other student data in preset reports and

customized reports functions.

"These are diagnostic tools for teachers to use every day, not just on the

side," says Liebman.

Later this fall, teachers and principals will be able to share ideas and

information via Web collaboration tools that are part of the ARIS project

rollout. Using a module of the open source content management software

Drupal, teachers will be able to create communities of like-minded

collaborators, using blogs, wikis, and private community spaces. Educators

can add to their profiles to create "instructional identities" to make it

easier for teachers to find others who share interests.

The effort involves up to 100 TB in a data warehouse, with enrollment,

assessment, and biographical data for all 1.1 million students, plus profile

data for every staff member. Today, teachers are tapping mostly preset

reports, which they access through a browser using the same login as the

e-mail system, but by midwinter, the school system expects to have added

business intelligence tools--it has considered software from Cognos, which

is owned by IBM, the project's lead contractor--to allow more complicated

queries.

The New York City teachers' union, the United Federation of Teachers, has

backed the ARIS program, as long as it isn't used to judge teachers and the

school provides teachers with programs that have proven to help with

particular problems. "Teachers want to use it," says Michael Mulgrew, the

union's chief operating officer. "They want to make their instruction

better."

DIG DEEPER

ECONOMIC UNREST

See how other IT leaders are crafting strategy amid an economic downturn and

market uncertainty.

Download this

InformationWeek Report

>> See all our Reports <<

For those who see education's rush to data analysis as a bad thing, as just

a more individualized way to "teach to the test," Liebman has little

patience. "This process is no more 'teaching to the test' than a doctor

diagnosing and then treating a patient for a bacterial infection of the

kidney is 'treating to the test,'" says Liebman, who's also a law professor

at Columbia University. Teachers will consider the data along with

everything else they observe and change their "treatment" if the student

continues to struggle. "This is what professionals do," he says.

U.S. schools need change. The Program for International Student Assessment,

which gives math and science exams every three years to about 400,000

teenagers in 30 countries, found U.S. students ranked 24th in math and 16th

in science in 2006. U.S. graduation rates, long thought to be about 85%,

could be as low as 70% finishing in four years, concludes a shocking report

issued in April by the nonprofit Editorial Projects in Education, with

backing from America's Promise Alliance and the Bill & Melinda Gates

Foundation. In the largest cities, it's only 50%, and in some of them it's

35% or lower.

PLANO'S PLAN

Using BI tools only to produce more elegant reports on No Child Left Behind

mandates amounts to a wasteful "autopsy report," says Jim Hirsch, associate

superintendent of academic and technology services for the Plano Independent

School District. Instead, the Texas district with 68 schools and 54,500

students implemented a SAS Institute analytics system so it could draw in

other measurements beyond the annual state standardized tests, including

data from the schools' periodic student assessment exams, and try to predict

what problems might lie ahead.

The district has been using SAS Enterprise Intelligence Platform BI tools

for four years, but this is just its second year using them to not just look

back at student performance, but also to "give insight into elevating

performance," says Hirsch, who began a 34-year career in education as a math

and programming teacher and has been in administration for 22 years.

With help from SAS, Plano created a data mart that brings in several sources

of information, including Texas' annual state standardized test results and

the Measures of Academic Progress testing results that are given to students

in grades kindergarten through 10 multiple times a year.

Plano uses the SAS tools to analyze a variety of student data, looking at

performance of entire schools, grade levels, groups of students (including

subgroups, like those who speak English as a second language), and even

individual students. Plano uses the software to create trajectory graphs

showing how children are expected to perform several years ahead, taking

into account their current strengths and weaknesses.

In one significant study in eight of its schools, it used that trajectory

analysis to identify 60 students at risk of failing state standardized

tests, and teachers developed plans to address their needs. Only 10 ended up

doing poorly. "It was a huge success story," says Hirsch.

The technical challenges are less infrastructure-related and more about

building an effective predictive model, Hirsch says, something that took

Plano about two years to develop. "The highest priority is understanding

what questions you want answered, what data is necessary to answer those

questions, and to take advantage of the analytics," he says.

Parents access all student information via Plano's parent portal, where each

family has individual, Web-based accounts that recognize them to provide

appropriate access to children's records, says Hirsch. Parents can't run

queries--"there's no legitimate way we could educate over 37,000 families in

the proper way to combine variables and interpret the results," says

Hirsch--but the reports include data visualization features such as learning

growth charts.

The SAS system cost Plano $300,000, including license fees, hardware, and

services, says Hirsch. Since it was deployed, Plano has been applying the

software to analyze other problems, such as tracking credentials of

teachers.

SAS also offers a hosted service for education analytics, but Hirsch sees an

in-house data mart letting the district ask questions on the fly, build and

test new analytical models more quickly, and integrate add-ons such as

Futrix, multidimensional cube analysis software.

What questions might Plano explore? It's researching factors tied to

training and credentials, such as what effect a science teacher's specialty

in biology has on student performance. "There are many causal links to

student achievement that need to be investigated now that we can correlate

more data variables," says Hirsch.

(click image for larger view)

ANALYTICS AS A SERVICE

However, many school districts have trouble retaining a deep enough IT bench

to keep networks running, let alone run a sophisticated data warehouse and

analytics operation. SAS is targeting those districts with online analytical

services that, for about $2.50 a student, will use student data to create

graphical trajectories forecasting how students will perform in the future.

Districts send SAS the data, SAS does the analysis in its data center, and

districts access results online.

The analysis is based on the Tennessee Value Added Assessment System, or

TVAAS, algorithms and methodologies developed by William Sanders and

colleagues at the University of Tennessee, where Sanders was a professor for

34 years. Sanders now works for SAS, marketing a variation of the

methodology dubbed EVAAS, while also a research fellow at the University of

North Carolina. "If a school district has 10,000 kids, you wouldn't be able

to pay a secretary $25,000 a year to do this complex analysis," Sanders

says. TVAAS uses math, language, reading, demographic data--"anything we can

get our hands on," Sanders says--from school districts and states as the

basis for analysis.

The state of Pennsylvania began a pilot with 100 districts in 2002, and now

all 501 school districts can access EVAAS data analysis on student state

assessment reports, says Kristen Lewald, director of the Pennsylvania Value

Added Assessment System statewide project.

"The services provide a red flag about what kids run the risk of failing in

high school" or dropping out based on trajectories from the assessment

analysis, she says. The analysis helps project the performance growth of

average and high-achieving kids if they're not provided with appropriately

rigorous schoolwork. Each superintendent decides who can access the data,

and some districts provide the trajectories to teachers and parents to get

them engaged in the kids' performance, Lewald says.

Sanders and his methodology naturally have detractors, who question

statistical concepts used in TVAAS. Some academic research argues its

growth-measure methodologies penalize schools with higher-performing

students because there's less room for top students to show year-to-year

growth, and just one or two wrong answers by a top student can overly skew a

growth path downward. SAS customer Hirsch says Plano doesn't use the TVAAS

models in its on-site analysis using SAS tools, saying it doesn't give an

accurate enough picture for individual students in comparing their "starting

point versus their individual growth" during a time period. Plano does,

however, use student growth models based on another methodology. Count on

more such models to proliferate, perhaps making it all the more difficult

for districts to compare and benchmark performance.

Teachers will use data as one part of a diagnosis, just like doctors, says

New York City's Liebman

Photo by Erica Berger

Nonetheless, the foundation of EVAAS concepts and methodologies are

attracting districts and state departments of education in North Carolina,

Ohio, Pennsylvania, Tennessee, and other states.

In Tennessee, teachers get professional development credit for taking

sessions that teach how to make use of EVAAS data. Schools use the EVAAS

analysis for diagnosing and addressing student performance--and, to a

limited degree, evaluating teachers.

That's controversial because it's one of the few states where school

districts use EVAAS-type analysis in any way as part of teachers' evaluation

process, says Keith Brewer, executive director of the Tennessee Organization

of School Superintendents. Student performance data analysis can't be used

to terminate teachers, but it is used to identify professional development

needs.

"I can show a teacher that their teaching methodology is good for average

students but is missing the below-average or above-average students," Brewer

says. For instance, if data indicates that certain groups of students of a

teacher over three consecutive years have not made significant progress in

certain skills, a district can offer professional development to help that

teacher put together different kinds of lesson plans or modes of delivery

for lessons.

CHALLENGES AHEAD

Tennessee's mild example just hints at the controversy ahead as data becomes

a bigger part of education. What happens when the data continually shows a

teacher underperforming, regardless of the "professional development" that's

been offered? Will pricey BI systems compete with more direct student needs,

from in-classroom technologies and e-learning to smaller class sizes and

better-paid teachers? Will a new digital divide emerge, where those students

who've built a good data history have a better safety net than those who've

shuttled among schools? Could parents become as obsessed about trajectory

graphs as they are about SAT scores, spawning a prep-course industry for

middle school? And as educators start identifying the learning needs of

individual students, how do teachers prioritize to address dozens of kids'

unique needs?

Steering through such questions will be up to a new generation of data-savvy

administrators and educators, as well as parents and policy makers.

Industries such as health care are facing similar slow-motion revolutions,

trying to figure out the best way to make digital health data part of

treatment.

The pressure is clearly growing on school districts to do more with the data

they're collecting. "We are in the early stages of evaluating data mining

and reporting tools," says Lenny Schad, CIO of the 55,000-student Katy

Independent School District, outside Houston. Schad, who joined the district

in 2003 after a 15-year IT career in the oil and gas business, including a

CIO stint, says the district's goal is to generate dashboards specific for

each level in the organization, with drill-down capability to provide

further analysis.

All these tools, however, will only work if teachers believe in the data

analysis and can easily translate it to what students need to learn.

That leads to a problem the business sector knows well: how to get analytics

tools spread broadly to give decision makers the insights they need.

Business intelligence vendors including IBM Cognos, SAP Business Objects,

SAS, and SPSS all provide BI or analytics products aimed at the K-12

education segment. But despite a lot of talk about "BI to the masses" in

business, the typical private-sector company might provide "150 power users"

with the sort of high-level business insight you get from business

intelligence, says Gartner analyst Bill Rust. In a large school district,

"you might have 7,000 power users." In New York City, they're hoping for

80,000 power users.

In Hershey, Pa., the effort starts with staff meetings devoted to discussing

data and its uses, and for specifically carving out time for teachers to use

the data analysis tools. Assistant superintendent Goldsworthy has spent 35

years in education and has seen big strides only in the last five years in

using data well. "Data without analysis doesn't teach you anything, and

analysis without action doesn't change anything," she says.

Goldsworthy describes public education as "in its infancy in using data."

That's true. But all indications are it's going to need to grow up fast.