



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

InformationWeek
Research & Reports

Can Data Mining Save Our Schools?

Schools have more data than ever on students, and there's a nationwide push on to use that to spot problems. It can help—if schools are able to react to what they learn.

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

By Marianne Kolbasuk McGee

IN DEPTH / DATA MINING

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.

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 uses 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

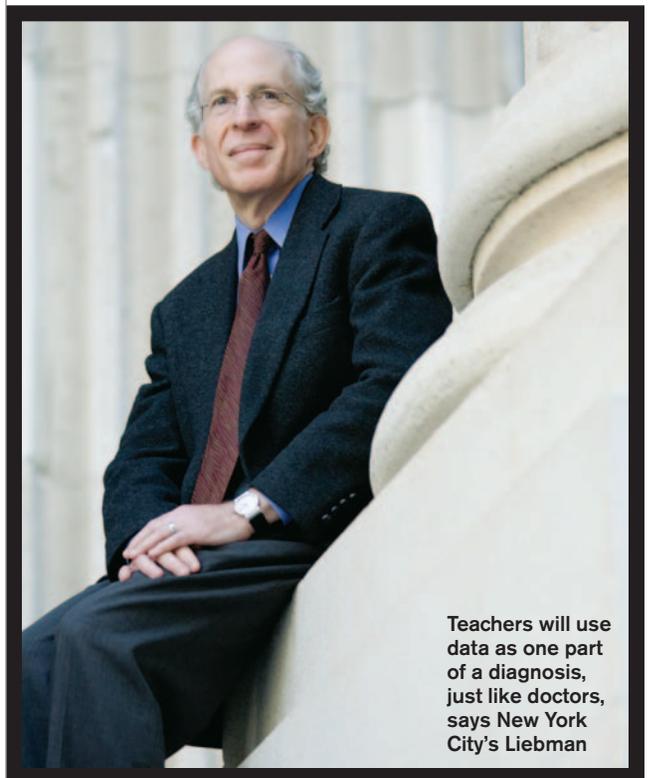
preset reports and customized reports functions.

“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 mid-winter, 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



Teachers will use data as one part of a diagnosis, just like doctors, says New York City’s Liebman

Erica Berger

chief operating officer. "They want to make their instruction better."

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

In The Classroom, Technology Options

IN ITS USE OF CLASSROOM technologies, Arcadia Unified School District in California's San Gabriel Valley could be considered a role model for other school districts—if they all happen to have the means that Arcadia has. Because when it comes to giving teachers and students technologies, this well-off community 20 miles north of Los Angeles just about has it all.

The debate over the best ways to bring technology into the classroom is by no means settled. Attempts championed by politicians a decade ago to have "a computer in every classroom" often meant an idle computer in every classroom, if the PCs came with no teacher training or inadequate tech support. While most schools can't afford every tool Arcadia has, a look at its experience shows the widening possibilities for technology in the classroom.

In 1996, the district installed fiber optics in all classrooms, providing Internet

access and a broadband pipe for other media content. Two years ago, voters approved a \$218 million bond issue to renovate buildings and build new ones, with the biggest single chunk of money designated for technology upgrades.

With that, Arcadia is building "classrooms of the future," says Rob Leri, the district assistant superintendent who was recently promoted from Arcadia's director of technology and information services, a job that reports directly to the superintendent. All the schools will have wireless Internet so students can use their own devices, from laptops to smartphones with Wi-Fi. Students without computers can get access to school computers. Arcadia experimented with providing laptops to a group of students, but that proved inefficient, given how many students already had PCs at home.

Coyla Grumm, an Arcadia middle school classroom teacher for 33 years, is working part time providing tech train-

ing to other teachers. Before retiring two years ago, Grumm was involved in a pilot program that wired 13 classrooms. Now, 400 classrooms in 10 schools are being equipped with whiteboards, laptops, and document cameras, which take pictures of images in a textbook, display them on the whiteboard for discussion, and publish them on the class' Web site. That lets students access from home something they might have missed, and the images can be archived to share with other teachers.

Among the instructional tools used at Arcadia is reading assessment software and also MyAccess, an online service that provides students with feedback on the grammar and structure of their writing. Tested in Arcadia's middle schools for two years, it's now being rolled out to fourth- and fifth-graders. Arcadia also is using reading assessment tools and computerized testing programs that let students move on to more difficult levels of material only after they've succeeded

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 ana-

Expand

at the previous level.

Arcadia also subscribes to Discovery Learning, which provides streaming video content that teachers can use in their classrooms or that students can add as clips to their own projects.

HOMEWORK 2.0

Collaborative Web 2.0 technologies also are playing a role, including a gated student blog site designed to give Arcadia students a vehicle to publish what they're learning.

The site, which has been up three years, recently provided a place for exchanges between a fourth-grade Arcadia class and Alexis O'Neill, an author in the midst of writing a children's book. The blog discussion elicited ideas from the kids about how the author could make Victor, one of the main characters, "cooler," says Grumm. Also, after a recent career day, kids blogged for two days about what they learned.

Arcadia also provides wikis for stu-

dents and teachers. One Arcadia teacher recently had students create a wiki about the California gold rush, in which each student contributed facts about the topic.

Aside from computer technology, each Arcadia classroom also has microphones and four speakers in each ceiling. "Everyone hears the same thing," Grumm says. "Kids in the back aren't falling asleep." Each school also has a "presentation cart" that includes a Webcam that can let teachers stream class activities for parents to watch from their home or office computers.

On the administrative side, all of Arcadia's student information is electronic. Teachers take attendance electronically, clicking on each absent child's photo in an electronic seating chart, with the information automatically reported to the school office. Grades also are recorded electronically.

For student data, Arcadia does data

mining at the district level, using SPSS, DataDirector, or simply Crystal Reports connected to its SQL database to run queries and generate reports. Teachers can search data on their students' performance using a spreadsheet interface.

Perhaps most impressive of all, Arcadia has a staff of 10 to 12 who support technology in the schools. Most school districts don't have that kind of staff or capital budget for technology.

Tessa Jolls, president of the Center for Media Literacy, an organization that provides resources and consulting to schools, says staffing's often the biggest obstacle to deploying technology well. "Schools put funding for buying the equipment, not for support," she says. A teacher left stuck with an IT problem in front of 25 squirming students, with no one to call for help, is going to think twice about using that tool again.

—MARIANNE KOLBASUK MCGEE

IN DEPTH / DATA MINING

lytical 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.

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

Impact Assessment: Analytics In Education

● **Benefit**

● **Risk**

IT organization

These analytics efforts are tied directly to student performance, so they're highly relevant. As districts across the United States launch new data analysis efforts, it could mean new career opportunities in the education sector.



Many districts don't have a deep pool of IT talent, especially in high-end tools such as analytics, and IT pay in education generally isn't as high as private-sector IT jobs, so recruiting will be a challenge.

Business organization

Could keep fewer students from falling through the cracks. By flagging weaknesses and strengths of students, it's more likely educators can predict problems. Also, ongoing assessment analysis can show what intervention efforts really work.



Analysis is only as good as the data is accurate. Also, districts should prepare for finger pointing if ongoing analysis shows schools aren't making progress despite their arsenal of data tools. There's also a big training need.

Business competitiveness

Used effectively, these tools can help improve education for students at every level, something critical to schools' mission and the country's global competitiveness.



Just having the data and analysis isn't enough; schools need to take the right corrective action. Also, access to these tools and solutions needs to be spread across all schools, not just among the wealthiest.



Bottom Line

Data analysis can help teachers and administrators anticipate and correct problems students face and gives schools tools to measure the success of their educational efforts. In the bigger picture, anything that can help improve the U.S. educational system helps the country's competitiveness globally.

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.

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.

Write to Marianne Kolbasuk McGee at mmcgee@techweb.com.

DIG DEEPER

ECONOMIC UNREST See how other IT leaders are crafting strategy amid an economic downturn and market uncertainty. Download this free *InformationWeek* Report at: informationweek.com/1205/report_economy.htm

See all our Reports at informationweekreports.com