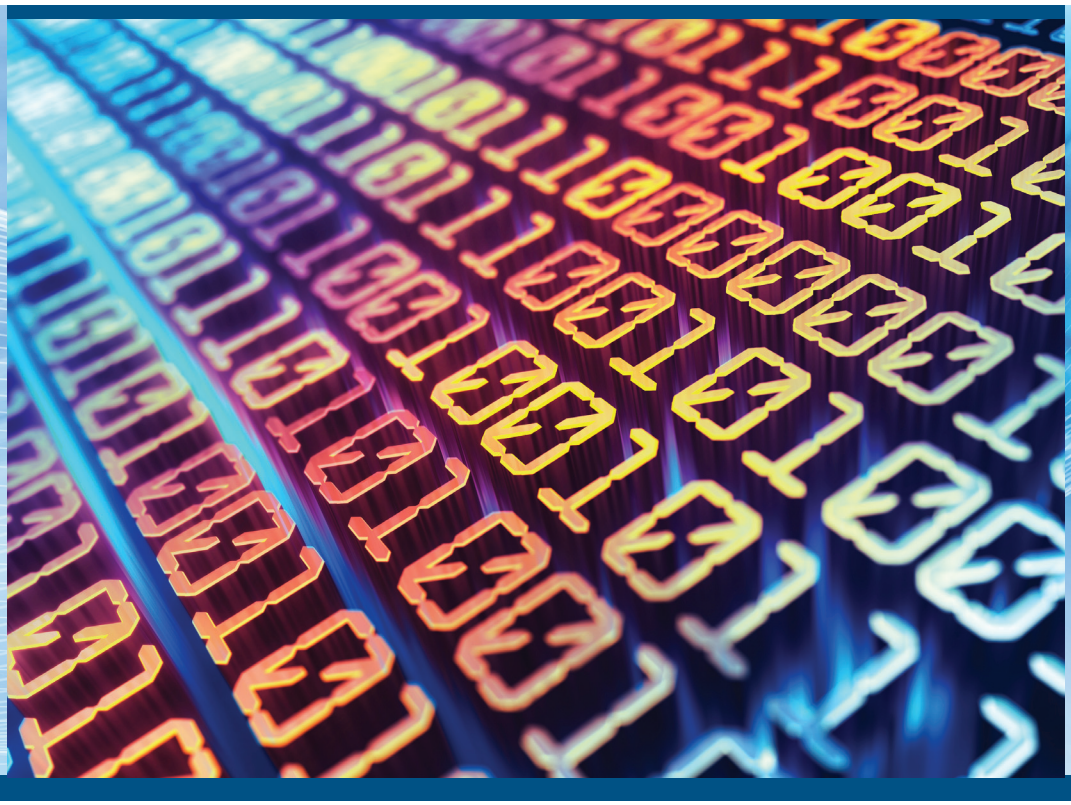

Predicting the Outcome: Business Analytics for Education

Sponsored by

IBM

1105 MEDIA Education Group

3 Steps for Getting Started with Predictive Analytics

1. Just do it. Don't be stymied by the accuracy of your data, says Stephen Gold, vice president of worldwide marketing for IBM's SPSS division. "We have never yet encountered a situation where a customer has perfect data. There is always going to be variability in information." As he notes, the technology is far more forgiving today in how it can work with data "that's less than perfect. That's OK. I think too many institutions worry about trying to correct the information they have and improve the quality of that content, and they're not getting the benefit from what they already have."

2. Don't worry about what you're looking for. Traditional statistics required a hypothesis: You made a statement and then proved or disproved it. In the new world of data mining and predictive analytics, you have the ability to go ahead and mine the data for insights and allow the information to "talk to you" about the findings. "Take the data you have for whatever period it represents, point the technology against that data, and see what comes out," Gold insists.

3. Turn the results into action. "The best insights, the best predictions won't facilitate a better outcome if they're not actually applied," Gold observes. He advises organizations to think about how they're going to act on the findings they uncover "to transform their institution, their district, their school, into a stronger performer with better outcomes."

Data has become the lifeblood of education. Governments predicate their school funding on it; districts demonstrate the effectiveness of their practices with it; schools prove the merits of their instructional approaches by it; instructors redirect their lesson plans because of it. All of these scenarios address one common question: How do you make better decisions to improve student performance and student retention?

Yet data has also become like the guest who stays too long at the party. At the beginning it's nice to have, but at some point it becomes too much. As personnel and resources in the school environment continue to shrink, people struggle to keep up with the continual flow of data. They just don't have the time to make sense of it, let alone put it to good use.

The challenge is to identify and prioritize the most important data so that educators can take action on it and not shut down when the topic is brought up. What's needed is a way to streamline the use of data so that teachers and administrators can use their time more wisely and get back to their real business: educating students.

Business analytics has already proven its worth in the education field. Leading K-12 districts, and education agencies have discovered how to borrow from the business world the tools and practices that will help them improve school, student, and teacher performance.

Within the education space, the use of analytics has received a lot of attention in the last decade. Business intelligence tools like IBM Cognos drive usability of data for decision making by providing reporting tools, real-time dashboard monitoring, and scorecards that provide a critical view of historical results, trends, and progress against goals. BI tools drive transparency across the institution by making key information easy to access and share. Predictive modeling—known commonly as predictive analysis—complements this historic view by using algorithms to find patterns and hidden associations within the data that may not be immediately apparent from a traditional report. These "golden nuggets" provide much deeper insight into the historical trends by revealing the factors that drove a particular historic outcome—and then predicting what is likely to happen next.

What does this look like in action? As Kirk Kelly, director of Accountability and Testing for Tennessee's Hamilton County Department of Education, explains, "Predictive analytics allows a school or school system to bring in past history and look at the current status of the students. Then based on the information you have, you can make predictions about how well students can do. You can have three looks at what is going on in your environment—past, current, and future. It allows you to be proactive as opposed to being reactive to stuff that may happen."

Predictive analytics also provides a means to evaluate which forms of intervention would likely yield the best outcomes. While it's important to understand which student is at risk and would compromise the school's ability to continue to receive state and federal funding under No Child Left Behind mandates, it's also vital to know what to do about it. Predictive analytics addresses both challenges.

Making Data Work for You

Predictive analytics has long been used in business and research. Financial services companies, for example, use it to calculate a credit score before they issue a loan or credit card. By examining the details of an individual's past history and comparing them to others' performance, the lender gains insight into how likely it is that the applicant will pay back the loan.

In the case of education, schools and districts run the data they have through a specialized application for analysis. The software crunches the data to create a model that generates correlations or connections between seemingly unrelated pieces of data. As Kelly explains, "There's a strong likelihood that if this occurs, that's also going to occur. They're associated; they're related to each other."

Those related pieces of data become the leading indicators, relevant variables that have an impact on an outcome. For example, a district knows through long history and experience

Reporting Results to Your Education Community

Running analytics is really only the first step in successfully changing learning outcomes. The next part of the story is what you do with the results. That means presenting what you've learned to the people who can best make use of it.

The real value of analytics for education surfaces when you can link traditional business intelligence capabilities, such as scorecards and dashboards, to report on the results generated by running predictive models. That reporting may encompass staff, teachers, administrators, parents, and students themselves. Only then can that information be acted on.

IBM's Cognos, also a part of IBM's Business Analytics division, offers a family of products that addresses the "presentation layer" of analytics. Cognos has a number of features for presenting data in ways that help decision-making.

For example, Mobile County Public Schools in Alabama selected Cognos to replace static reporting and enable delivery of customized and dynamic dashboards to thousands of district users with timely reports and measures specifically focused on monitoring the academic lifecycle of each student in a number of areas: class attendance, grades, interventions, and other educational requirements.

that the likelihood of a student failing a grade is tied to multiple variables: absenteeism, tardiness, health records, the availability of hot lunches, parental involvement in back-to-school events, and teacher performance assessments. That's a long list for anybody in a school to stay on top of, let alone manage. Predictive analytics refines that analysis to uncover which of those indicators is the most influential among all of the possible scenarios in order to give it top billing for attention.

Success Story No. 1: Facing Down the Dropout Rate

Kelly is a long-time user of IBM SPSS tools, including Statistics and Modeler. A little over two years ago, Kelly and his team of analysts began to look at why so many students were dropping out. In 2009, about 19 percent of the county's students left school before graduating. "We were having 700 students drop out a year," Kelly says. "We started to look at what feature jumps out that's a high correlation with these students dropping out."

Using IBM SPSS Modeler, the team pulled data from student information systems, examined academics, and looked at age, behavior, and attendance.

"What we tend to think is some students walk up and announce, 'I'm dropping out.' Others may disappear, and then you have bad attendance. When you pull out the attendance piece and you begin to look at it, the No. 1 thing that jumps out is the age factor. It's highly correlated with all these students who were dropping out."

Kelly's staff discovered that even before attendance problems surfaced, if a student was one, two, or especially three years older than this or her peers, there was a high likelihood of that student leaving school before graduation. Next, the researchers began to analyze where students were getting off track. "That's when we began looking at it by grade level and discovered that the problems surfaced at entry grade levels, at transitions," he notes.

The first transition would often actually occur when a child began kindergarten with or without some pre-K program and was held back in kindergarten or first grade. The next transition would occur in moving from fifth grade to sixth, elementary school to middle school. The final transition happened in the move from ninth to 10th grade. Difficulties would surface in ninth grade and manifest in attendance problems during 10th grade.

"If a kid got retained twice before they got to high school, there was a strong likelihood of them dropping out. If they come to high school at age 16 or 17, they're not going to stay until they're 20," explains Kelly.

In response to that finding, Hamilton County held a citywide summit that brought community leaders together to discuss what could be done to address the problem. "We made it a K-12 initiative, not just a high school initiative," he says.

Kelly's team developed an on-track indicator to identify those children that needed extra help or support to make sure they could move from grade to grade with their classmates. By 2010, just a year later, the dropout rate had plummeted from 19 percent to 13.5 percent for the district. "And we're looking to keep that momentum rolling right through 2011," Kelly adds.

Success Story No. 2: Understanding Early Literacy

Nicole Catapano is the coordinator for data analysis at the Washington-Saratoga-Warren-Hamilton-Essex Board of Cooperative Educational Services (BOCES) in New York State. This BOCES serves 31 school districts with about 44,000 kids in a five-county region. Catapano and her small team of data analysts have been working with one district specifically on its early literacy data. As she explains, "A common goal people have is that they want students reading at grade level by third grade. That's when our large-scale assessments start statewide."

The district faced the challenge of generating a lot of data through assessments, but without any good way to organize the results. Schools would experience analysis paralysis wondering how to use that data to identify gaps in their teaching. "Folks would get this huge spreadsheet of data points about a student and have to figure out, 'Where do I start?'" Catapano notes. "Teachers or administrators might have hypotheses about which variables are the most important. Our office would go into these meetings and say, 'Look, we can pit these variables against each other and see which ones are really making the difference or not making the difference.'"

District leaders told Catapano, "Whatever you can do, please do it. They gave us carte blanche: 'You tell us what you need from us, and you've got it.'"

Other Uses for Analytics in Education

Although analytics is finding its home in education as a predictor of student outcomes, it's also finding additional uses tied to regulatory and compliance requirements. The mission of schools is clearly to deliver education, but financial and operational requirements also play a major part in ensuring their success.

Analytics for education can help a district evaluate the educational value of a given program against other similar offerings. In the case of professional development, it can look at the gains and growth the students are making and tie that back to instructor training to identify which programs are worth investing in and which ones should be limited.

As an example, recently Hamilton County performed a controlled study of the use of virtual schools among four of its brick-and-mortar schools. Director of Accountability and Testing Kirk Kelly and his team discovered that students who received both regular and virtual school instruction in mathematics at the seventh-grade level had an average gain of three points in student achievement, "That's got some implications," Kelly notes. "We could find that a virtual school is a great supplement to students who may be struggling or for all students across the board."

The district shoveled data to the BOCES analysts in multiple formats—Excel spreadsheets and text and comma-delimited—culled from multiple sources, including student information systems, classroom reading and math assessments, demographic systems, and attendance and tardiness programs.

The BOCES' tool of choice for processing that data, IBM SPSS Modeler, comes from IBM Software Group's Business Analytics Portfolio. Catapano has been a long-time user of SPSS tools—particularly SPSS Statistics—beginning in her days as a college student performing statistical modeling on a mainframe. That software has traditionally offered multiple statistical analysis features accessible by pull-down menus and a syntax language. Now she's become a major fan of SPSS Modeler, a graphical data mining application, which she was introduced to in 2009.

"SPSS Statistics used to be cumbersome when you had multiple data fields," she says. "Things would get a little unwieldy when you tried to pull everything together—especially when we're serving so many school districts with a very small staff. So when IBM SPSS Modeler came out, it allowed us to really ramp up the ability to do predictive analytics on a much larger scale for a lot more people with a lot more approaches. Instead of looking at one or two types of research questions, now we could look at a dozen at a time and not bat an eye.

"Modeler allows us to pull in data from a variety of sources. The data still stays where it is, in that original format; but the Modeler software allows us to do our merging, clean it up, and recode variables without having to change that original data set."

The speed with which that work can now be done, she explains, is important, since the bulk of time consumed by any research project is in the data management—"making sure the data is clean, making sure the model contains the variables in the format we want it to contain. When we get to the point where we're running the statistical analysis, that's the shortest part of the whole process."

In the case of early literacy for third graders, Catapano's team identified the variables that are having the biggest impact, as well as those that are secondary and tertiary predictors. As a hypothetical example, if gender were the biggest predictor, Modeler would do additional breakouts to uncover the next level of predictors within each of those categories—such as special education status for females or the beginning-of-year fluency benchmark results for males. That entire model would form a decision tree, with each category of variable breaking out into ever smaller branches.

The process, she points out, "allows us to see which variables really don't have impact once you have them pitted against each other. We can really start to rule out variables that are part of our data set but are not really adding anything to our decision making."

After the initial analytics work was completed, Catapano's staff communicated with school and district administrators to broaden their understanding of the data. The teachers were the ultimate recipients of the work. "We streamlined that huge, crazy spreadsheet to [focus on] the key variables they need to pay attention to," Catapano explains. Administrators were able to say to teachers, "Here's the student's profile. We know this is the biggest indicator that's the problem area. This is what their predicted results are. This is the population of students we really need to target some more reading instruction to."

Once a given predictive model is built in IBM SPSS Modeler, new data can be added without having to start from the beginning. As the longitudinal data set grows, the findings become even more compelling because time proves their veracity. Now a full cohort of students has passed through, and the analytics work is in its second year. What teachers have discovered is that those folks who took that data and made targeted interventions have seen definite improvements over the ones who didn't.

"Action speaks louder than words," Catapano declares. "They're starting to see some of the results. Now they're already looking this year at being able to build the programs for those early literacy components so that when they get to their mid-year benchmarks, which were the No. 1 predictor, they know how students will do."

She adds that teachers were able to have conversations over the summer that weren't possible before, when data was fed to them on the fly and they'd have to do their analysis in the midst of the academic year. "Now they can say, 'We've got a classroom full of kids, and we know these five kids—based on their profiles—are at risk for these reasons.' It's beginning to take on a life of its own. They're using it on a regular basis. They're asking for the data and understanding it. They're making interventions based on it. They're seeing results from it. So now the cycle is continuing, and it's exciting. That's really generating the buzz.

Educator-Friendly Analytics

Although these two individuals—Catapano and Kelly—both have long histories in statistical analysis, which can certainly amplify the impact of analytics for education, significant advances in the technology have lately put those tools into the hands of non-data experts too. The operations of applications such as IBM SPSS Modeler are much more visual and the functions to perform data modeling have evolved so that the user's ability to interact with the software is far easier than in the past.

IBM SPSS Modeler lets the user quickly identify and source disparate data from existing systems, without the need for a new data warehouse. Once the data is pulled in, the visual drag-and-drop interface lets the user view the information collected and apply filters. For example, the loaded data may include transfer students who have begun attending the school mid-cycle. To prevent them from skewing results, they can quickly be identified and filtered out of the model.

Today's analytics users may not understand statistical regression or know what a See5 algorithm is, nor do they have to—it all happens in the background. The user doesn't even have to know which model is best. IBM SPSS Modeler is "smart" enough to create several types of models and recommend which one is the best fit. And forget about p-values and z tests—the predictive insights are expressed in terms that educators understand, like whether a student is at-risk for dropping out or needs a special program.

Fundamentally, it's about gaining insights that help educators make better decisions when it matters most. The key, however, is to turn the math into decisions by taking action on what the predictors reveal. Simply knowing that a student is likely to drop out is not enough—you then have to do something about it. IBM SPSS decision management tools enable this by leveraging predictive models to generate actual recommendations for programs, intervention, or other necessary actions. This broad accessibility to insights is what is driving the tremendous uptake in the use of analytics by the non-statistician.

Moving Into the Realm of What-If

In just the last two years, as the use of analytics has grown in K-12, so too has thinking about how predictive analytics can be applied. This whitepaper has already touched on the "at-risk student" and shown how analytics can help identify those children whose performance isn't meeting standards. But the analytics work doesn't have to end there.

The next phase may well be to examine "under-performers," such as the C students who should be B students. When analytics is used to identify at-risk people, that C student could be overlooked because he or she will most likely pass.

Likewise, analytics can promote the notion of personalized education. Say that student A is very proficient in English, but requires additional time and counseling in math. Conversely, student B may be great in science but not do so well in languages. For those situations, a personalized lesson plan derived from predictive models could allow more tailored course curriculum to be offered to those students, perhaps by extending the time they spend in a particular area or providing insight on what kind of help they need to accelerate their learning.

Advancements in the technology will enable schools to extend the purview of its analysis to improve the quality of the total student population.

Likewise, it can be used to assess various intervention approaches and identify those that are the most effective. One of the biggest expenses for schools is the special accommodations they have to make for small numbers of students. There are dozens, if not hundreds, of programs to address learning gaps in a district, and not all of them are equally effective with different children.

The ability to match intervention with an individual to take corrective action is part of the benefit of predictive modeling. If a student is having difficulty keeping up with reading, for instance, the school may find that one reading program used judiciously is far more effective than liberal application of a proprietary homegrown methodology that has been in place for years.

That's exciting to educators because it allows them to better use the resources they have by capitalizing on what they know will be more effective. For one, they won't be buying useless "gadgets." For another, they can get out of a situation ahead that, if left unchecked, could consume massive resources in reviewing, reporting, analysis, and testing "downstream."

The BOCES' Catapano likens the tools of analytics to vehicles. She talks about Excel as a "decent kind of basic starter car." The typical instructional data management system, which delivers data to teachers to inform their instruction, is a "a sports car or luxury car." IBM SPSS Modeler, she states, "is like having the Space Shuttle. It's taking you to places you probably haven't been before much faster, and it's extremely exciting."

This master analyst predicts that the use of analytics can really change the way schools do business: "How we look at data, how we help students, how we save time, save resources, and have better outcomes along the way. I just see this as something that everyone should be doing at every level in every district."

About Us

About THE Journal

THE Journal is dedicated to informing and educating K-12 senior-level district and school administrators, technologists, and tech-savvy educators within districts, schools, and classrooms to improve and advance the learning process through the use of technology. Launched in 1972, THE Journal was the first magazine to cover education technology.

THE Journal's franchise consists of the monthly print magazine (which is also available in digital format), the web site thejournal.com, six newsletters (District Cloud Computing, K-12 Mobile Classroom, THE News Update, T.H.E. Journal Insider, IT Trends, THE SmartClassroom, and School Security), and targeted list rental opportunities.

To learn more, visit www.thejournal.com.

the JOURNAL
9201 Oakdale Ave.
Suite 101
Chatsworth, CA 91311

(818) 814-5277

IBM
IBM Corporation
1 New Orchard Road
Armonk, New York 10504-1722
United States

(914) 499-1900