Student Data Mining System Raises Privacy Concerns

P20-W initiative will track students throughout academic careers

By Dan Way

March 26, 2015

RALEIGH — North Carolina public schools are developing a multimillion-dollar student data mining system intended to compile and analyze reams of information to improve educational outcomes. But critics say it poses a "creepy" potential to engineer the work force and easily could fall prey to a variety of "malicious" abuses.

Known as the P-20W system, the program captures student data from pre-K through graduate school and follows individuals into their work years.

The Department of Public Instruction is collaborating with the UNC system, North Carolina Community Colleges System, North Carolina Independent Colleges and Universities, the state Department of Health and Human Services, and the Department of Commerce to gather, manage, and analyze the information.

"We're hoping that through this system we can have a better understanding of the outcomes that come from kids going to different schools in North Carolina, and what happens to them after they graduate," said Lou Fabrizio, DPI director of data, research, and federal policy.

"We are required to report certain things to the U.S. Department of Education, all in aggregate form, in terms of things like how many kids … graduate from high school and enroll in a community college or a university within 16 or 18 months after their high school graduation, and how many kids actually complete one year of higher ed or community college within two years of graduating," Fabrizio said.

The system still is being developed as part of an initial $3.4 million grant, and the state expects to seek a one-year extension from the U.S. Department of Education in the coming weeks. No full-time employee is assigned to the program, but about 10 employees do some amount of work on it.

“The children don’t fill out any information. This is all information that we’re collecting from the parents,” Fabrizio said.

Much of the data was being collected prior to the P-20W initiative through the Power School Student Information Management System. At the end of each year, information collected in that system transfers to the Common Education Data Analysis and Reporting System, a longitudinal database that would become part of the P-20 system when it becomes operational.

The General Assembly in the last session passed Senate Bill 815, “which has to do with student data privacy, and every fall the superintendents are supposed to send notifications home to the parents telling them that we do have safeguards in place for any data,” Fabrizio said.

The system will be housed in the Office of State Information Technology Services with built-in safeguards protecting the confidentiality of student and family information, Fabrizio said. Officials in that office referred questions back to DPI.

“If you like privacy, and you have concerns about how people might abuse collective data, then it’s creepy. And I think people who find it creepy have legitimate concerns,” said Neal McCluskey, associate director of the Washington, D.C.-based Cato Institute’s Center for Educational Freedom.

“You do have a problem of hubris among researchers who think that because they have been able to statistically control for various factors that they can reach universal conclusions about how that can be used to educate people,” McCluskey said.

That makes it easier for politicians to cherry-pick data in developing public policy for work force outcomes, “and then the biggest danger is that people, politicians or experts, try to use this to engineer society,” McCluskey said.

The P-20W system is another concern layered atop the National Education Data Model established by the federal government to collect more than 400 data points.

“You can see how this could be seriously abused and dangerously used to say potentially we encourage people not to [practice] a particular religion, or will start saying how children should be raised by their parents, legislating it and saying not to do what the research suggests would be criminal abuse or something like that,” McCluskey said.

And with recent reports of data breaches hitting everything from Target stores to the national health care exchanges, privacy security is a constant concern.

“These inventive hackers or people who steal data are malicious people who seem to find a way to get data that people think is secure,” McCluskey said.

“I would say probably that most North Carolinians have no idea the extent to which these databases are being brought together by P-20W,” and most parents don’t know they can opt out of the data collection, said Terry Stoops, director of research and education studies at the John

Locke Foundation. He is a member of the Longitudinal Data System Committee overseeing development of P-20W.

“Some of it’s definitely benign,” such as test scores and attendance records, Stoops said. And he can envision ways in which the information can be analyzed for beneficial and legitimate purposes to enhance the delivery of education and its outcomes.

“But from the K-12 perspective, there’s data being collected on behavioral issues, [the] number of times a student’s been suspended or put into detention, and that sort of thing I think parents would object to,” he said.

Medical records such as vaccinations, primary care physicians, and medications taken “are areas … where most parents would draw the line, saying the school doesn’t have any right to maintain any data about what kind of medications a student is taking,” Stoops said.

The state has received an Early Learning Challenge Grant that will be used to collect information on K-3 students that eventually will be incorporated into P-20W.

“This is going to be a much more problematic set of data because it’s going to include social and emotional skills, and teachers will have to assess whether a student can function well in a group, is aware of their own emotional state, etc.,” Stoops said. “We’re talking about very young children.”

With a growing database of family, medical, educational, law enforcement involvement, and other information comes the risk that “bad people who are looking for information on someone to use against them for whatever reason” could leak the records from inside, or hack into it from outside, Stoops said.

Like McCluskey, he worries how far officials may go in demanding public policy to comport with interpretations of the research.

“I think the expectation is that [the main users will be] university-based researchers, doctoral students writing dissertations, individuals, professors and researchers from universities, as well as administrators from the departments that may be looking for ways to improve their services,” Stoops said.

While they are not supposed to have access to the databases, “What would prevent a private business, or the state, or federal government from accessing the data to use for various purposes, whether it be to market goods or target voters?” Stoops said.

Since the federal government provides so much of the funding, he asked, “Will you have a point where the federal government says, ‘Because we pay for it, we are then entitled to it’?”