Getting FERPA Right: Encouraging Data Use While Protecting Student Privacy

Chrys Dougherty

Chrys Dougherty, Ph.D. is a senior research scientist at ACT and the National Center for Educational Achievement

1. Introduction

The creation of large statewide education databases offers an unparalleled opportunity to improve our information about effective schools, programs, practices, and reforms. This opportunity is at risk, however, because of excessive restrictions on access to data based on concerns about student privacy. In many places, privacy has been used as a justification to restrict many types of research, data mining, and data analysis that depend on access to statewide data.

Like their peers in medicine, counseling, law, and accounting, educators have an obligation to protect the privacy of their charges. However, as in medicine, an appropriate balance must be struck between the need to protect individual privacy and the equally compelling mission to use data and research to improve outcomes for students. Because all uses of data contain some small risk that the data will be improperly disclosed, the key to privacy policy is to create arrangements whereby those risks are minimized while the large benefits from use of the data for analysis, research, and the improvement of schools and student learning can still be realized. It would be an error, while focusing on privacy risks, to overlook the even greater risk to which we subject millions of students if we fail to improve their education.

Section 2 of this chapter provides an overview of the federal Family Educational Rights and Privacy Act (FERPA) privacy law. Section 3 briefly describes the U.S. educational and policy environment at the time of FERPA's enactment in 1974. Section 4 describes changes since then, including the increased focus on school accountability and the development of statewide longitudinal student data systems. Section 5 describes the research and analysis opportunities that have been created by these data systems, while Section 6 explains how federal privacy law can be interpreted or amended to take full advantage of these opportunities while continuing to safeguard privacy. Section 7 offers reasons why policymakers may assign greater weight to small privacy risks than to large data use benefits. Section 8 concludes with recommendations in three areas: interpreting appropriately privacy law, strengthening research and data analysis using longitudinal student data, and helping the policy world do a better job of balancing privacy risks and data analysis benefits.

2. FERPA Fundamentals

The Family Educational Rights and Privacy Act was passed by Congress and signed into law by President Gerald Ford in August 1974. Known as the "Buckley Amendment" after the law's principal sponsor, Senator James Buckley of New York, the law gave parents oversight of their children's educational records.[1]

At the time the law was passed, the Watergate scandal was current news and concerns about abuses of government power and invasions of privacy were ubiquitous. Senator Buckley and others were concerned that allegations about students ("Johnny is a troublemaker") were being placed in those pupils' file folders and later inappropriately used against them—without parents being able to view the information, challenge its accuracy, or prevent its unwanted release. Senator Buckley stated that the new law was intended to counter "frequent, even systematic violations of the privacy of students and parents by the schools…and the unauthorized, inappropriate release of personal data to various individuals and organizations."[2]

FERPA guaranteed parents three specific rights with regard to their children's education records. The first was the right to inspect and review the accuracy of the record. Second was the right to challenge the accuracy of the record at a hearing, at which time the parent could ask that inaccurate material be corrected or removed. Third was the right to prevent personally identifiable

information on the student from being disclosed to any third party without the parent's written consent.[3]

The records in question are those "maintained by an educational agency or institution or by a person acting for such agency or institution" in cases where the agency or institution receives "(federal) funds under any applicable program."[4] In 1974, when FERPA was enacted, the agency or institution in question was almost always a local school district, and the "educational records" were paper documents maintained in file folders in the school or district office.

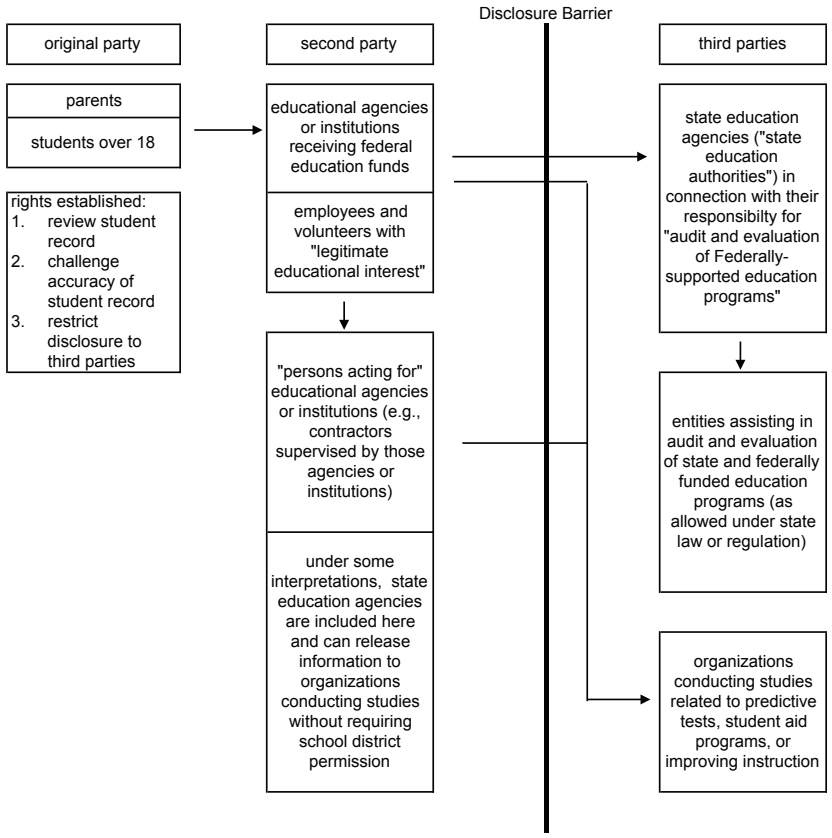
The law's authors understood that the privacy of student records must be balanced with other public priorities, such as the ability of schools to educate students and the ability of law enforcement officials to maintain public safety. With that in mind, FERPA made certain parties eligible to receive personally identifiable student information without parental consent, including:

- Teachers and other school officials who have been determined by the educational agency to have "legitimate educational interests" in the student;
- "Authorized representatives of the Comptroller General of the United States, the (U.S.) Secretary (of Education), or State educational authorities…in connection with the audit and evaluation of Federally-supported education programs, or in connection with the enforcement of the Federal legal requirements which relate to such programs"; and
- "Organizations conducting studies for, or on behalf of, educational agencies or institutions for the purpose of developing, validating, or administering predictive tests, administering student aid programs, and improving instruction…"[5]

Third parties "outside the educational agency or institution" that received confidential information under these provisions could not in turn pass on ("redisclose") the information to other third parties without written parental consent.

Figure 1 provides a schematic diagram of how these disclosure provisions work.[6] Parents (and the students themselves, when they turn 18) are viewed as having specific rights with respect to student information maintained by educational agencies and institutions. Prominent among these rights is the

Figure 1



ability to restrict access to the information by third parties—entities other than "educational agencies or institutions" or their employees or contractors. This establishes a "disclosure barrier" with third parties on the far side of the barrier: information on identifiable students may only cross that barrier with the parent's signed consent, or with the meeting of one or more alternative conditions specified in the law ("FERPA exceptions"). These exceptions include the transfer to "authorized representatives" of state education agencies ("state education authorities") and the U.S. Department of Education, and the release of information to organizations conducting studies related to predictive tests, student aid programs, and improving instruction. These third parties can share the information with another third party only if the sharing is viewed as part of the initial disclosure, not a redisclosure, and the other party also qualifies under the rules allowing for that initial disclosure.[7]

When FERPA was enacted, state education agencies did not collect student-level data, and alternative governance structures, such as charter schools, charter management organizations, and charter school sponsors did not exist. Nor had any state proposed comprehensive arrangements to gather data on individual students from early childhood through K-12 and into higher education and the workforce.[8] Therefore, it was not necessary to clarify which, if any, of these data-maintaining entities should be classified under the law as "educational agencies or institutions or persons acting for such agency or institution," and which as outside third parties.[9]

To summarize, the issue of who is an "educational agency or institution or person acting on behalf of such an institution" is critical under FERPA, because the law places no constraints on how such an entity can make use of the data, so long as the data are accurate and subject to parental inspection and correction, and the user has a "legitimate educational interest" in the information. The constraints come in as soon as the data are to be released to third parties, and therefore must cross the disclosure barrier in Figure 1. The data can only be released to third parties with the parent's written consent or under specific circumstances described by the law (e.g., to organizations conducting studies on behalf of the educational agency or institution). Third parties, in turn, cannot share ("redisclose") the data with other third parties without parents' written consent. So the law is restrictive with respect to the use of data by third parties, but not by educational agencies or their contractors — "person(s) acting on (their) behalf."[10]

3. The Education Environment in 1974

In 1974, the U.S. was in a period of educational stagnation. Test scores were falling not only on the SAT but also on standardized tests such as the Iowa Tests of Basic Skills, declining by more than could be explained by changes in the composition of the test-taking population.[11] Education did not attract the level of interest from policymakers or the public that it did in earlier or later decades. Anecdotally, the counterculture and various educational fads, such as the "open classroom," were having a negative influence on the ability of schools to educate students.[12]

The federal policy focus was almost exclusively on funding and rules — for example, paying for programs for specific populations, monitoring to make sure that federal dollars were being spent on exactly the right students, and

ensuring that schools had proper procedures in place to involve parents on committees. Both what and how much the funded students were learning were not treated as a matter of equal urgency. In the language of economics, the emphasis was on inputs, not on productivity.

Research on "effective schools" was in its infancy. The 1965 Coleman Report was widely misinterpreted as implying that students' socioeconomic status was the only relevant factor determining educational outcomes, so that schools don't make much of a difference. The 1969 Westinghouse Report suggested that Head Start didn't seem to make much difference either. The 1970s equivalent of the "standards movement" was the call for "minimum competency," asking students to demonstrate sixth-grade performance by the time they graduate from high school. Even these minimal standards were widely regarded as unfair and unrealistic for many students.[13]

Although record numbers of students were enrolling in higher education, the policy emphasis at the time on "equity" did not translate into a call for achievement gaps to be closed or for the majority of disadvantaged students to be academically prepared for college or other postsecondary learning opportunities. Rather the emphasis was on "access" to higher education regardless of whether students were actually prepared to succeed once they enrolled. The 1970s were a heyday of the "shopping mall high school" and of corresponding elementary and middle school practices based on the idea that only a minority of students were cut out for challenging academic content.[14]

This toxic combination of low expectations and a focus on rules over results meant that there was little pressure on students or schools to improve performance and little demand for research or public information on school effectiveness. Thus, the need to create better arrangements for using student data while safeguarding privacy rights was not a salient issue in 1974.

4. Major Developments Since 1974

After 1974, and particularly after the publication of *A Nation at Risk* in 1983, the education policy environment in the United States changed dramatically in ways that placed a strong emphasis on collecting and using data to improve schools and student learning. The emergence of data on international comparisons led to a widespread understanding that American students were underperforming relative to their peers overseas. In addition, the availability

of data on race- and income-based achievement gaps from the National Assessment of Educational Progress (NAEP) helped to focus policy leaders on the adverse implications of failing to educate poor and minority students. Since the 1980s, the emphasis has increasingly shifted from the amount of resources to whether those resources are making a difference.

Other developments influenced the supply of education information and how that information is stored and analyzed. In 1994, the Improving America's Schools Act introduced federal standards-based testing requirements, and several states went beyond the federal requirements for standards-based testing in the 1990s.[15] Around the same time, a number of states began producing "school report cards" with student test results. The increasing collection and publication of school performance information in the 1980s and 1990s was, in part, based on a realization that it is difficult to sustain an effort to improve performance if there are large costs to doing so while actual performance is hidden from view.

At the same time, the expansion of magnet programs, open enrollment, charter schools, and other "school choice" arrangements has made school performance information more valuable both to parents choosing schools and to the policymakers seeking to evaluate those reforms. The issue of consumer and public information on school performance was notably absent from the original FERPA policy discussion.[16] Policymakers have also become more interested in keeping up with the performance of highly mobile students and of students as they cross institutional boundaries, such as between K-12 and higher education. In addition, the expansion of online education and of dual credit programs means that students are more likely to be enrolled in multiple educational institutions at once.

From the point of view of this discussion, the most important change was the development by states of longitudinal student data systems with the ability to follow students over time and across multiple databases. The first statewide student information systems were created in Delaware in 1985, Texas in 1990, and Florida in 1992. By 2001, seven states (Arkansas, Delaware, Florida, Louisiana, Minnesota, Mississippi, and Texas) could match student-level test and enrollment records over time.

Many other states acquired student-level test data, but could not match test records for the same students across different grades and years.[17] These non-longitudinal datasets were mainly useful for reporting "snapshot"

statistics about student performance levels in a given grade and year. Comparisons of the performance of last year's third graders and this year's fourth graders approximated a measure of average student growth only if student mobility was low. These databases were even less helpful in following students across levels — elementary, middle, high school, and higher education — or in tracking student transfers in order to produce better measures of dropout rates.

The enactment of the No Child Left Behind Act in early 2002, with its requirement of reporting test score data disaggregated by student characteristics, greatly accelerated the development of statewide longitudinal data systems. That was because *accurate* disaggregation of students depends on having such a system.[18] Congress's appropriation of funds to provide grants to states to develop longitudinal student information systems also helped to accelerate the development of these systems. As of 2007, 27 states had received Statewide Longitudinal Data System (SLDS) grants, and every state was working on developing such a system.[19] In that year, four states reported having all of the "ten essential elements" of a robust longitudinal data system described by the Data Quality Campaign, giving those states the ability to follow students across enrollment, demographic, program participation, state test, course completion, dropout, graduation, college readiness test, and college enrollment databases.[20]

These systems were facilitated by the revolution in electronic data collection storage, transfer, and analytic capabilities. The creation of the internet, the lower cost of computers that can handle large data sets, and the increasing user-friendliness of database management and statistical software have made the collection of data by states and the use of the data by third-party analysts much easier and less costly.

The expansion of internet-accessible computer databases has increasingly transformed the student privacy issue into one of computer security: protecting student records from identity theft and the ability of malicious individuals to steal poorly protected data. For example, one federal report stated that over a nine-month period in 2005, 93 documented breaches of computer security occurred involving personal information from education records such as Social Security numbers (SSNs), credit card information, and dates of birth.[21] Almost half of these breaches occurred in colleges and universities.[22] Since there is no legal reporting requirement for data security violations, the total number of such breaches may have been greater. In addition, every news

report of a privacy breach occurring in another industry—whether missing Veterans Administration laptops or stolen credit card records—accentuates these concerns.

In 1974, breaches of privacy mainly consisted of school district officials voluntarily or carelessly releasing information contained in paper files, and the law's emphasis was on schools and districts having policies in place to prevent such releases. The creation of large government databases of any kind was a concern among privacy advocates in 1974.[23] At the time, however, these concerns were more about misuse of the information by government officials, not theft of records by outside individuals.[24]

In addition, the increased public reporting of school results since NCLB has led to concerns that individual student results might inadvertently be "leaked" in these reports. For example, reporting the test passing rate of all 50 students in a grade and of the 49 white students would make it possible to identify whether the remaining one African American student passed or failed the test. This has led to policies of masking (not reporting) results for "small cells" (small student groups) in public reports.[25] While these small student groups do indeed need to be removed from public reports, many behind-the-scenes data investigations require their inclusion in the underlying analysis.[26]

5. Opportunities Created by Statewide Longitudinal Student Data Systems

The creation of statewide longitudinal data systems has multiplied the opportunities to address questions of importance to educators, parents, and policymakers. The fact that the databases are *longitudinal* means that they can be used to address questions about student growth; school, teacher, or program effectiveness; and whether students are "on track" to later success. Getting clear answers to these questions into the hands of educators—while helping them understand what those answers imply in terms of taking action and changing adult behavior—is critical for the goal of improving schools.

The fact that the databases are *statewide* means that they can answer questions that are far better addressed with records on as many students and schools as possible. The questions that these databases can help address may be organized into two main categories as follows:

1. Statewide longitudinal student databases can serve the role of large epidemiological databases in medicine—making it possible to look for patterns in large numbers of individuals over time and predict what is likely to happen to students if certain actions are or are not taken.[27] The availability of statewide data increases analysts' ability to address questions such as:
   - To what extent do students who are academically prepared when they leave elementary school remain "on track" in middle and high school?
   - To what extent do students catch up later if they leave elementary school poorly prepared?
   - What variables are most closely associated with the odds that a student will drop out?
   - How are student course-taking patterns and course grades related to success on college readiness exams and the need for remediation in college?
   - How well does the workplace reward different student achievement levels and educational degrees and certifications, and how is that changing over time?[28]
   - How does the answer to each of these questions vary across student populations in different schools, districts, and regions in the state?

2. Statewide longitudinal student databases can be used to widen the search for the most effective schools, teachers, programs, and policies—making it possible to learn systematically from "What Works." As educators and policymakers pursue information on what is working well and where, they will want to know answers to questions such as:
   - Are your local schools as effective as the best in the state serving similar student populations?
   - How good are the charter schools in your community, and how do charter schools compare with traditional public schools in your community and statewide?

- Are some schools especially good at catching up academically behind students? Are these different from the schools that do the best job with academically advanced students?
- Which schools and programs work best for English language learners and other at-risk student groups?
- Which types of preschool interventions produce the best results for students in elementary school, and in general, which interventions lead to the greatest student success in the next higher level of education?
- How often is school improvement in one subject accomplished at the expense of performance in other subjects?
- What will it take to double the percentage of low-income students reaching college and career readiness benchmarks?
- How well are the state's teacher preparation programs preparing teachers?[29]
- What will it take to attract highly effective teachers to the high-poverty schools in your community and region?

These questions and many others like them have three things in common. First, they cannot be answered well without longitudinal student data. In many cases, this means the use of confidential student data to which FERPA applies.[30] Second, they are best answered by gathering information from as many schools and school systems as possible: hence the advantage of accessing statewide student databases, not just the data from a single school or district. Third, involving third-party data analysts is likely to greatly accelerate the rate at which these questions are addressed.[31]

Efforts to bring outside resources to bear on research questions using statewide data began shortly after the first statewide longitudinal databases were developed. In 1992, Harvard economics professor John F. Kain established the Texas Schools Project (TSP) to take advantage of Texas' statewide data. TSP began studying the achievement of minority students in residentially integrated suburban school districts, and moved on to address issues such as teacher quality, teacher incentives, and charter school effectiveness.[32]

The longitudinal Texas data were also used to identify effective schools and design innovative school reports. In late 1998, the nonprofit organization Just

for the Kids began releasing school reports on the web comparing achievement in each Texas public school with that in the highest performing schools in the state serving equally or more disadvantaged student populations. Though the statistics in these reports were aggregate data and did not reveal individual student information, they were built from longitudinally-matched individual student data. Examples of these longitudinal statistics include "the percent of students meeting academic growth benchmarks," "the percent of below passing eighth grade students who later met college readiness benchmarks in high school," and "the proficiency rate of students who were continuously enrolled in the same school for three years or more."

A handful of state education agencies have joined the effort to promote the use of third-party research for school improvement. Most notable among those is the Florida Department of Education (FLDOE), which provides on request a list of key areas where research and analysis are needed to improve student learning in Florida's schools. FLDOE invites outside third parties to submit proposals for investigations in these areas using Florida's statewide longitudinal student database.[33] The agency also works with researchers and analysts who propose investigations of other topics. Kansas has developed a partnership with the state's two largest universities and the Kansas Board of Regents to promote research using student data. North Carolina and Texas have set up state-sponsored education research centers to take advantage of the availability of student data in those two states, and Arkansas has shared its data with researchers at the University of Arkansas. These efforts have been viewed by their respective states as complying with state and federal privacy laws and fully addressing the need to safeguard the privacy of student records. However, concerns about federal interpretation of privacy law may be why similar efforts are not happening in more states, despite the fact that a few less timid states have been leading the way.[34]

6. Statewide Longitudinal Student Data Systems and Federal Privacy Law

In general, federal privacy law has placed few barriers in the way of teachers and other school and school system personnel using data on their own students and hiring private contractors to help them with those efforts.[35] However, barriers to the analysis and use of statewide longitudinal data by third parties threaten to hamper the search for answers to questions such as those in the previous section. Here is where getting FERPA (and privacy rights in general) right is most likely to make a difference.

To illustrate what we mean by "getting FERPA right," in this section we set out four questions that a sound student privacy policy would answer in the affirmative.

A. Can third-party analysts obtain statewide longitudinal data for studies or evaluations directly from the state education agency, without having to get the permission of each individual school district and charter school?

The FERPA regulations proposed in March 2008 authorize studies initiated by third-party researchers or data analysts using confidential student data "for, or on behalf of, educational agencies or institutions," without the prior consent of students or their parents, if the analysts conclude an agreement with the educational agency or institution that is the source of the data.[36]

This could be interpreted as meaning that, in a state with 1,000 school districts and hundreds of charter schools, data analysts must conclude a separate agreement with each of these entities for each analytic project in order to gain access to statewide data. The legal argument for this position holds that state education agencies have traditionally not been defined as "educational agencies or institutions," nor are they clearly defined in statute or regulations as "person(s) acting for such agency or institution," since they neither directly educate students nor are voluntarily hired as contractors by the agencies that do. Nor do they operate under the direct control of local education agencies. Therefore, when school districts send student data to a state-managed longitudinal data system, this represents a disclosure to a third-party entity outside the local educational agency or institution. Under FERPA, so the argument goes, such third parties lack the independent authority to make further redisclosures to other third parties without written parental consent. Only if an agreement was concluded between the researchers and each school district whose data is provided in the study, as proposed in the March 2008 draft regulations, would such an arrangement be FERPA compliant. According to this line of reasoning, therefore, the answer to the question above would be "no."

The contrary argument holds that the legal responsibility of a state education agency is, in effect, to "act for" the state's school districts and charter schools, even though it is not a contractor and is not controlled by the school districts.

Thus it should be understood as a separate type of "person acting for such agency or institution" with the authority to make independent decisions, including the ability to conclude its own agreements with third-party data analysts. This interpretation would move the state education agency to the left side of the disclosure barrier in Figure 1, and provide a "yes" answer to the question above.[37]

A second opportunity for analysis of statewide data can arise when states establish formal procedures in state law and/or regulation for outside analysts to be authorized to assist the state education agency in carrying out its responsibility to evaluate teacher, school, and program effectiveness in the state. This is the basis on which the U.S. Department of Education's Family Policy Compliance Office has given a green light to the three education research centers established under Texas law.[38]

B. Can a system be established for approving the use of statewide data for analyses that state education agencies may not want?

There are obvious reasons why state agencies or local school districts should not be expected, much less required, to sponsor and control all longitudinal education data analysis. Consider, for example, an assessment of whether the state agency and local school districts are counting dropouts correctly. States and school systems may be reluctant to commission studies that are likely to find major flaws in their own practices.

Similarly, a risk-averse or politically sensitive state agency may have no desire to approve data analyses that are likely to produce results unpopular with influential constituencies. Rather than having to say "no" to the political hot potatoes, the agency might choose the easier path of not approving any third-party data analysis at all—perhaps citing privacy issues as the reason. Or agency leaders might truly believe that the privacy risks of releasing data to third parties almost always outweigh any potential benefits from analyzing the data.

State law might provide an alternative channel, such as a research review board, for approving data analytic projects. The board would need to have the status under state law or regulation as a "person acting on behalf of educational agencies or institutions" or an "authorized representative in connection with

the evaluation of Federally-supported education programs" but with the independent ability to approve studies. A memorandum of agreement would need to exist between the review board and the state agency operating the longitudinal data system, providing for the release of data by the state agency to the research organization if the project is approved by the review board. Such an arrangement might provide political cover for state agencies not wishing to approve studies directly.

C. Can state early childhood, K-12, and higher education agencies combine the data possessed by each of these agencies into a single database for joint research and analysis purposes?

If the state happens to structure multiple levels of its education system under the control of a single education agency, as Florida does,[39] the answer to this question for the data managed by that agency is an unambiguous "yes." But where the agencies are separate, ambiguity arises. Consider the students currently in K-12 — does the higher education agency have a "legitimate educational interest" in them, even though they are not currently enrolled in any of the state's higher education institutions and some may never enroll? As for the students in higher education, many are former participants of the state's K-12 system, but others are not. None (except for dual enrollment students) are currently enrolled in K-12. Does the state education agency have a legitimate educational interest in those students?[40]

As former Massachusetts Commissioner of Education David Driscoll pointed out in a letter to U.S. Secretary of Education Margaret Spellings, the answer to a question such as this should not depend on the accident of how a state configures its education agencies. A state should be able to combine the data from its preschool, K-12, and higher education agencies into a single database for research and analysis purposes.[41]

D. May non-education state agencies, such as employment or social service agencies, obtain longitudinal student data in order to improve their own services to students or former students?

If state education agencies can provide data to third parties for the purpose of evaluation or analysis, then other state agencies should be able to qualify as third parties as long as they sign an appropriate interagency agreement. However, to qualify for a FERPA data-sharing exception to parental consent, analyses conducted by those other agencies must be intended to "improve instruction" or "evaluate Federally-supported education programs." Thus it is not clear that an analysis addressing practices that improve a student's educational outcomes indirectly would qualify. In addition, if the study is strictly to improve the agencies' own services, without any anticipated impact on the students' educational success, such a study would clearly not qualify for a FERPA data-sharing exception.

Federal privacy law may need to be amended so that third-party studies intended to improve students' educational outcomes, but not necessarily through the mechanism of improving instruction or evaluating educational programs, qualify for a FERPA exception to parental consent. Alternatively, "education programs" in this context might be defined broadly to include any program that is likely to affect educational outcomes.

7. Weighing Privacy Risks Against Data Analysis Benefits

Even if each of the four questions above is answered affirmatively by federal policy, state privacy laws in some states impose restrictions beyond those established by the federal law. For example, as of fall 2007, four states—Connecticut, Ohio, New Hampshire, and Wisconsin—forbade sharing of student records between K-12 and higher education institutions.[42] Ohio prohibits reporting of student names or social security numbers to the state education agency.[43]

Unless state policies are also open to the use of data, states can use privacy laws or pure risk aversion to avoid sharing data. For balanced policies to become the norm, it is necessary not only to get federal privacy policy right, but also to establish the right policy climate in each state. States, after all, are the keepers of the data.

As we saw in Section 5, only a handful of states have adopted policies that allow or encourage third-party analysis with statewide longitudinal student data. This could imply that policymakers in most states assign greater weight to the risks of data analysis than to its benefits.

Reasons why policymakers may see the risks of data analysis as greater than the benefits include:

- *Policymakers and the broader public are more easily motivated by fear than by lost opportunity. Privacy issues are easily framed in terms of fear.* Research in economics and psychology has documented the human tendency to "loss avoidance"—to giving greater weight to possible losses than to potential gains.[44] Similar research has documented that human beings do a poor job of weighing the risk of relatively infrequent but salient events.[45]
- *Privacy violations have clearly defined victims, whereas the beneficiaries of research and data analysis are a large and ill-defined group.* Breaches of privacy and thefts of student records happen to specific individuals, whereas it is harder to identify the beneficiaries of a piece of analysis that contributes, for example, to the overall understanding of teaching mathematics. Public officials face clear political consequences when individuals suffer losses of which they are readily made aware, but are likely to receive fewer political benefits when the advantages of a policy are spread out over many individuals who do not know that they have benefited.[46]
- *Because other types of databases are frequent targets of identity thieves, policymakers may overestimate the privacy risk from databases created for education data analysis.* Statewide databases created for research and analysis can be made more secure and less target-rich (i.e., with statistics less useful to identity thieves) than is the average database maintained by a school district or college.[47]
- *The benefits from education data analysis are little understood by policymakers and the public.* Education research and data analysis lack the dramatic examples that medical research has of diseases cured and lives saved. In addition, because the widespread use of data by teachers and school administrators is relatively new, and analysis using statewide longitudinal data has simply not been available in the past, educational practitioners themselves are just beginning to learn of the benefits of data analysis and use. Some educators' complaints about data ("too much testing") may have been more loudly heard than testimonials from other educators about the benefits of data.

- *The culture and folklore of education emphasize the special talents of individual teachers over accumulated research and professional knowledge as a source of teacher effectiveness.* In medicine, we usually think of the effectiveness of doctors as mainly due to accumulated medical research and professional knowledge acquired through training and experience, rather than to individual doctors having "the right stuff" or the inborn personal talent of a great doctor. Yet in education, our culture tends to think of teacher ability as an inborn talent or "magic spark" whose expression is as likely as not to be hindered by encouraging teachers to follow research-based practices.
- *The relative shortage of independent education data analysis may have adversely affected policymakers' and the public's perception of the value and credibility of education data.* When much of the data story is "spun" by school district officials, when the public hears about "teaching to the test" and manipulations of dropout rates, and when there is little independent information or transparency about what is actually going on, much of the public comes to mistrust education data and to heed the voices in education that say that educational measures and indicators don't carry much meaning.
- *Powerful interest groups in education are not comfortable with the transparency that widespread third-party data analysis could bring.* State and federal accountability systems have made many educators uncomfortable by taking away some of their control over the story on how their educational institutions are performing. Yet the limited information and analysis provided in most accountability systems leaves plenty of room to avoid transparency. For example, most such systems do not provide clear answers to the questions posed in Section 5. Third-party data analysis, coupled with investigations into educational practices, could make school systems more transparent.

Strategies to help policymakers find an appropriate balance between privacy risk and data analysis benefits must take these issues into account. Some of these strategies are discussed in the following section.

8. Promoting Data Analysis and Use While Protecting Privacy

Below are recommendations aimed at encouraging the use of data to improve education outcomes, organized under four headings: 1) making the

necessary interpretations of or amendments to privacy law; 2) taking steps to reassure the public that privacy risks are being minimized; 3) strengthening and expanding analysis using longitudinal student databases; and 4) helping policymakers see the benefits of this analysis.

1. *Make necessary interpretations or amendments to privacy law.* To ensure that federal laws and regulations do not pose an unreasonable barrier to data analysis conducted with adequate attention to the protection of student privacy, make sure that federal privacy policy provides a "yes" answer to the four questions in Section 6, whether through regulatory interpretation or statutory amendments. Where necessary, state privacy laws should also be made consistent with these requirements. To summarize, regulation and/or legislation should clarify that:

    a. State education agencies can conclude agreements with and provide confidential student data directly to third-party analysts without having to receive local school district approval of the planned analysis.

    b. States may establish additional entities, such as education research centers or education research review boards, with the same authority as the state education agency to approve third-party data analysis projects.

    c. States may establish a comprehensive longitudinal research database spanning all levels of the education system (early childhood, K-12, and higher education), which can be accessed for analysis intended to evaluate programs and improve instruction and student outcomes at any or all of these levels.

    d. State employment and social service agencies may gain access to confidential student data under the same conditions as other third-party analysts.

2. *Take steps to reassure the public that privacy risks are being minimized.* To provide assurance that reasonable precautions are being taken to reduce privacy risks, state agency officials can:

a. Implement a system of data security audits that are applied to every repository of statewide student data and on a spot-check basis to the databases maintained by local education agencies.

b. Delete key variables that are useful to identity thieves from databases provided to outside analysts. These variables, such as student names and social security numbers, are important for the state agency itself to collect in order to match records correctly across multiple databases. But once the matching is done, an appropriate alternative student identifier may be attached to each student record and the social security number deleted from the data supplied to third-party analysts. This makes the research databases the state creates relatively useless as targets for identity thieves.

3. *Strengthen and expand analysis using longitudinal statewide and cross-state student databases.* To encourage third-party data analysis, not just allow it, state and federal policy and private philanthropy can do the following:

   a. Continue to fund the development of statewide longitudinal student data systems with the ten essential elements recommended by the Data Quality Campaign.[48]

   b. Increase state, federal, and private funding to promote data analysis using statewide and multistate longitudinal student databases.[49]

   c. As a bolder policy, establish a multistate or national repository of student data combining the contents of the longitudinal data systems of multiple states. This might be done with the support of private philanthropy if the federal government finds it too politically difficult to sponsor such a repository.[50]

4. *Help policymakers, educators, and the attentive public see the benefits of analysis using longitudinal student data.*

For policymakers and other audiences to keep the benefits of analysis using longitudinal data in mind, they must be continually reminded of these benefits. This can be done if data analysts, funders, and advocates do the following:

a. Remind policymakers and their staffs, educators, and other audiences of questions that cannot be answered well without longitudinal student data. This should include questions that have come from these audiences.

b. Encourage states to publish data tables derived from the analysis of longitudinal student data (e.g., achievement and academic growth statistics disaggregated by the students' prior academic performance; test scores disaggregated by the prior school the student attended; longitudinal graduation rates; and higher education enrollment and success rates tied back to students' high school). These statistics can help make educators and the public aware of what can be done with the data.

c. Present examples of progress that has been made in answering important questions using statewide longitudinal student data. Describe the decisions that allowed the data analysis to happen. Discuss the implications of the analysis for policy and educational practice, keeping the language accessible to non-technical policymakers, educators, and other laypersons.

d. Work with educational practitioners to help them use the knowledge generated by the data analysis. Where possible, document where this knowledge was used to improve outcomes for students. Work with school and school system leaders to bring these examples in front of policymakers.

In conclusion, *any* use of data entails some small incremental risk of a breach of student privacy. If the sole goal were to minimize privacy risk, there would be no use of data at all. On the other hand, the risk from the appropriate use of data for third-party analysis can be held to a minimum, while the potential benefits from such uses of data are large. While working to protect students' privacy rights, policymakers must keep in mind the value of appropriately used data to answer important questions about student progress, teacher quality, and school effectiveness—to help students and schools get better. State and federal privacy law must do its job but must not become an obstacle to improving schools and student learning.

Bibliography

Buckley, James. Joint Statement in Explanation of the Buckley/Pell Amendment. *Congressional Record — Senate*. Vol. 120, December 13, 1974.

Buckley, James. Address to the Legislative Conference of the National Congress of Parents and Teachers, March 12, 1975. *Congressional Record*, Volume 121, S7974, May 13, 1975.

Copperman, Paul. *The Literacy Hoax*. New York: Morrow Quill, 1980.

Dougherty, Chrys. "States Must Improve Data for Adequate Yearly Progress," *Education Assessment Insider,* August, 2002.

Driscoll, David. Letter to Margaret Spellings, Secretary of Education. August 30, 2007. http://www.dataqualitycampaign.org/files/Publications-Driscoll_FERPA_letter_to_Spellings-080107%20final.pdf

Finn, Chester. *We Must Take Charge: Our Schools and Our Future*. New York: Free Press, 1992.

———. *Troublemaker: A Personal History of School Reform since Sputnik*. Princeton, N.J.: Princeton UP, 2008.

Hansen, William D. Key Policy Letter Signed by the Education Secretary. January 20, 2003. http://www.ed.gov/print/policy/gen/guid/secletter/030130.html

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5:1, 1991.

"Legislative History of Major FERPA Provisions." U.S. Department of Education, 2002. http://www.ed.gov/policy/gen/guid/fpco/ferpa/leg-history.html

Lerner, Barbara. "Good News about American Education." *Commentary,* March, 1991.

Lockwood. J.R. and D.F. McCaffrey. "On the Use of Value-Added Assessment to Evaluate Teacher Preparation Programs." Manuscript submitted for publication, 2007.

Mellor, Lynn, Mary Lummus-Robinson, Veronica Brinson, and Chrys Dougherty. "Linking Teacher Preparation Programs to Student Achievement in Texas." Manuscript submitted for publication, 2008.

Pfeiffer, Jay. Email communication with Chrys Dougherty, March 4, 2008.

Pipho, Chris. "Assessment—High Stakes. Seven Lessons Learned from Minimum Competency Testing." Policy Brief, Texas Schools Project. The Education Commission of the States, 2002. http://www.utdallas.edu/research/tsp/Index.htm

Powell, Arthur G, Eleanor Farrar and David Cohen. *The Shopping Mall High School: Winners and Losers in the Educational Marketplace*. Boston: Houghton-Mifflin, 1985.

Proposed Rules. *Federal Register* vol. 73, no. 57. U. S. Department of Education, March 24, 2008.

Rothstein, Richard. *The Way We Were? The Myths and Realities of America's Student Achievement*. New York: Century Foundation Press, 1998.

The Statewide Longitudinal Data System Grant Program. Institute of Education Sciences, National Center for Education Statistics. U. S. Department of Education, 2007. http://nces.ed.gov/Programs/SLDS/

Winnick, Steven Y., Scott R. Palmer and Arthur Coleman. "State Longitudinal Data Systems and Student Privacy Protections under the Family Educational Rights and Privacy Act." 2006. http://www.hklaw.com/content/whitepapers/StateLongitudinalsDataSystems2.pdf

Endnotes

1 The oversight passes to students when they turn 18. However, as stated in the law and reemphasized in the March 2008 proposed regulations, parents still have access to the academic records of their dependent over-18 children without those children's written permission.

2 See Buckley, Address to the Legislative Conference. Then, as now, concerns about possible abuses of government databases and potential invasions of privacy resonated with both the political left and right, with the left being more concerned about abuses by law enforcement and intelligence agencies and the right being suspicious of government power in general. Senator Buckley himself was a member of New York's Conservative Party and the brother of famed conservative commentator William F. Buckley, Jr.

3 In this section "third parties" is used as follows: parents (and students over 18) are the first party, education agencies or institutions and persons acting for them are the second party, and all others are third parties.

4 "Applicable programs" are those funded by the U.S. Department of Education (in 1974, the U.S. Office of Education), given that the law is part of the General Education Provisions Act.

5 Other groups given access to confidential student records without parental consent include school accrediting organizations and juvenile justice authorities as permitted by state law, and "in connection with an emergency, (to) appropriate persons if the knowledge is necessary to protect the health and safety of the student or other persons." (In the wake of the Virginia Tech shootings, proposed regulations released in March 2008 clarify that the U.S. Department of Education will not second guess the judgment of the educational institution in making this determination. See *Federal Register*, March 24, 2008, p. 15589.)

6 This diagram is based on the March 2008 proposed regulations, the most recent information available at the time of this writing.

7 Because the law distinguishes between initial disclosures (allowable without parental consent under FERPA exceptions) and redisclosures (allowable only with parental consent), one of the roles of FERPA regulations has been to clarify what is a "disclosure" as compared with a "redisclosure." For example, the March 2008 proposed regulations indicate that when students move from one educational institution to another, having the receiving institution share

information with the sending institution for purpose of records verification is not a "redisclosure" and therefore is permissible without written parental consent.

8 Given the widespread concern—ssome might say paranoia—about the creation of government databases, such an idea might well have been viewed more as a threat than as an opportunity had it been proposed in 1974.

9 For example, if state education agencies or charter school authorizing organizations are regarded as "educational agenc(ies) or institution(s)" or "persons acting for such agency or institution," then they have the same status as school districts under the law and move to the left side of the disclosure barrier in Figure 1. If, on the other hand, they are third-party recipients of student records, then they cannot redisclose those records to other third parties without the consent of parents or of the school districts that are viewed as the primary custodian of the records. See Winnick, Palmer, and Coleman, "State Longitudinal Data Systems."

10 The revised FERPA regulations proposed in March 2008 clarified that a contractor hired by an educational agency or institution, operating under the direction of the educational agency and performing functions that would otherwise be done by agency employees, could send the data to another education agency or a third party on behalf of the agency or institution employing the contractor. In other words, the educational agency approves the release, but the contractor does the actual data transfer. Such a transfer is not treated as a redisclosure.

11 See Rothstein, *The Way We Were?* pp. 58-68.

12 For extensive anecdotal evidence of this from the state of California, see Copperman, *The Literacy Hoax*. Copperman's evidence was taken from before the passage of Proposition 13, so it had nothing to do with how schools were funded, but rather with how they were managed and with the prevailing student ethos of the time.

13 See Lerner, "Good News." The first minimum competency testing laws were passed in four states in 1975 (Pipho 2002).

14 See Powell, Farrar, and Cohen , *The Shopping Mall High School*. The "shopping mall high school" describing high schools that deliver a strong curriculum to the students who want it and a weak curriculum to the poorly prepared and to students wishing to coast through high school with minimum effort.

15 Norm-referenced testing was required for Title I programs beginning in the 1960s, but these testing programs as implemented were vulnerable to "Lake

Wobegon effects," with students in all states or districts performing above average, and were not tied to standards for student performance or academic growth. See Finn, *We Must Take Charge*.

16 Since FERPA was an amendment to the General Education Provisions Act offered on the floor of the Senate and there were no committee hearings, there is little 1974 legislative history on the bill (U.S. Department of Education, 2002). See Buckley, Joint Statement.

17 For example, California began statewide testing in grades 2-11 in 1998, but could not match scores for third grade test-takers with the following year's scores for fourth graders.

18 In the absence of a statewide student information system, disaggregation of data by income and ethnicity as required by NCLB often had to depend on students or teachers filling in the information on test answer sheets, resulting in notoriously unreliable data. See Dougherty, "States Must Improve."

19 See Statewide Longitudinal Data System Grant Program. A total of $115 million has been awarded for these grants. The post-1974 period has also seen the creation of nationwide student databases outside state agencies, such as the National Student Clearinghouse's database of college enrollment and graduation records collected in order to help colleges verify student eligibility for financial aid.

20 For information on the ten essential elements and on states' development of longitudinal data systems with those elements, see www.dataqualitycampaign.org and Aimee Guidera's chapter in this book. The Data Quality Campaign was funded by the Gates Foundation after a number of persons associated with the National Center for Educational Achievement and other organizations, including the author, spoke and wrote in favor of making the development of these systems an important public policy priority.

21 Office of Inspector General, 2006.

22 Office of Inspector General, 2006.

23 Privacy fears about government databases have led to restrictions on some states' ability to maintain student records: for example, Pennsylvania was prohibited from having student records until recently, and Ohio is prohibited by law from sharing the state's K-12 student ID with other agencies, including the state's higher education institutions. The same concerns have blocked the creation of federal student-level databases, such as the one recently proposed for higher education.

24 See Buckley, Address to the Legislative Conference pp. 13990-1.

25 In general, records for enough students must be masked in order to prevent the reader from calculating the results for individual students or very small groups using published data. In the example above, the passing rate for *both* whites and African Americans would need to be masked. It should be noted that small cell rules need to be applied to the denominator but not to numerator in these passing percentages. For example, if one African American student out of 50 African Americans failed the test, the number need not be masked—how can one know which student out of 50 failed? If, however, all 50 either passed or failed, reporting a 100 percent passing or failing rate does in fact convey the results for each tested student.

26 See the discussion in footnote 30 below.

27 A well-known example of a longitudinal medical database used for research was the database created for the Framingham Heart Study, which originally consisted of 5,209 men and women between the ages of 30 and 62 from the town of Framingham, MA. (http://www.framingham.com/heart/). In education, the U.S. Department of Education has created a number of student-level longitudinal databases, including the well-known High School and Beyond (HS&B), National Education Longitudinal Study (NELS), and Early Childhood Longitudinal Surveys (ECLS) cohort data sets. These contain confidential data subject to FERPA but have been made available to researchers under license agreements, a successful example of addressing privacy concerns while maintaining the ability of third parties to access the data. Each of these federally maintained data sets contains information on a national sample of students—but possibly few or no students from a particular school or school system—while the state databases discussed here contain records on all of the students enrolled in a state's public schools.

28 This question requires matching data from the state employment agency with state education data.

29 The issues involved in addressing this question are discussed in Mellor et al., "Linking Teacher Preparation" and Lockwood and McCaffrey, "One the Use of Value-Added Assessment."

30 Some research projects use "de-identified" student-level data from which all information has been removed that could be used to identify an individual student or small group of students. These data sets are not subject to FERPA restrictions on confidential student data. However, completely de-identifying

student-level longitudinal data so that FERPA no longer applies entails removing not only student names, but also information on individual students or small groups of students with combinations of characteristics that might make those students "readily identifiable" by members of their local community if the data were made public. This removal of data on students in "small cells" often creates datasets with too many missing records, especially when the student's school and grade level are part of the record. Because small cells are created by unique combinations of variables, they multiply exponentially with the number of pieces of information collected on each individual student, and multiply further when multiple datasets are combined. Also, students who are different from others in their grade or school tend to end up in small cells. Thus, removing "small cells" to de-identify data tends to limit research to data sets with a) not much information on each student; b) little information on characteristics that might make some students unusual; and c) little matching of students across multiple data sets.

31 Many state education agencies have had difficulty getting funding for enough staff to develop and maintain their longitudinal data systems, let alone conduct multiple investigations, data-mining, and research exercises into what can be learned from all of those data. Small staffs are also unlikely to be able to implement all of the ideas that they and others can devise. Finally, the advantages of having multiple groups and individuals working with different approaches on a variety of educational problems should be apparent. (It is worth noting that an inquisitive school district research and program evaluation staff has the status of a "third party" with regard to statewide data on students in other school systems.)

32 See www.utdallas.edu/research/tsp/Index.htm. In addition, an early example of an organization that worked with a single school district was the Consortium on Chicago School Research, organized in 1990, that conducted analysis of longitudinal student data from the Chicago Public Schools.

33 These areas include effective ways of training teachers, research on teacher effectiveness, effects of retention and promotion policies, performance of charter schools, and the relationship between earlier academic performance and student success in higher grades, college, and the workforce. See Pfeiffer, Email communication.

34 Section 7 of this chapter discusses why timidity over small privacy risks can overwhelm the benefits of data analysis in some policymakers' minds.

35 Up until spring 2008, there were questions about whether a contractor could transfer student records on behalf of an educational agency or institution to a second educational institution (e.g., if the student transfers and the contractor is managing the "sending" education agency's data system). That was clarified as being permissible under FERPA in the proposed regulations released for comment in March 2008, which stated that the contractor, as a "person acting on behalf of an education agency or institution," could disclose records on the agency's behalf. See footnote 8.

36 Previously the agency had interpreted the "for, or on behalf of, educational agencies or institutions" language to mean that the educational agency or institution had to *control* the study.

37 See Winnick, Palmer and Coleman, "State Longitudinal Data Systems."

38 The Family Policy Compliance Office (FPCO) is the federal office that oversees FERPA compliance, interpretation, and enforcement. According to the March 2008 proposed regulations, state law or regulations govern the circumstances under which outside analysts may be given this authorization, as FERPA is silent on the issue. The one restriction is that the state agency still exercises oversight and control over who can access the data and is responsible for the maintenance of student confidentiality.

39 Although Florida combines the management of K-12 and higher education under a single agency, Head Start and private early childhood programs remain under separate management.

40 The new regulations seek to clarify that educational agencies or institutions have a right to records on former students without written parent or student consent—receiving these records will not count as a "disclosure"—but that is described as applying mainly for verifying the identity of former students, and at any rate does not apply to higher education students who never attended the state's public K-12 institutions.

41 See Driscoll, Letter to Margaret Spellings p. 3. Given the May 2008 letter from FPCO approving the educational research centers in Texas, similar centers in other states established under state law and supervised by the state education agency should be able under these guidelines to combine data from multiple agencies even if the governance of the agencies is separate.

42 The source is the fall 2007 Data Quality Campaign annual survey. See http://www.dataqualitycampaign.org/files/element9_survey_responses.pdf

43 Ohio Revised Code, Title 33, Section 3301.0714, paragraph (D)(1). See http://codes.ohio.gov/orc/3301.0714. In Ohio, students are assigned student IDs, called "data verification codes" out of sensitivity to the idea of having a student identifier. This assignment is performed by local school districts or regional technology centers contracting with school districts. Having IDs assigned by local or regional entities increases the odds of different students being assigned the same ID or of students being assigned new IDs when they cross district boundaries, producing errors in longitudinal student data. It also creates greater difficulties in merging in data from outside sources, such as SAT, ACT, or AP scores, employment data, or the already forbidden higher education data.

44 See Kahneman, Knetsch, and Thaler , "Anomolies." This is why the benefits to school improvement have often been more effectively framed in the policy world as protecting students from bad outcomes such as academic failure and low wages. No Child Left Behind was presented in that way when it was enacted in 2002. On the other hand, failure-avoidance in education can lead to an undue emphasis on minimum standards.

45 To remind people of the small size of the privacy risks, organizations that use student data must be willing to document and emphasize the safeguards they have put in place to protect student privacy, and any evidence they have of the small size of the risks to privacy that are created by the databases that they maintain. For example, research databases may have suffered relatively few of the privacy breaches that have affected student databases maintained for other purposes.

46 Another way of saying this is that costs and risks are easy to picture, whereas the possible benefits of something one has never had before are often nebulous and hard to estimate. See Finn, *Troublemaker* p. 76.

47 For example, social security numbers (SSNs) can be used as a match key to help identify which student a record belongs to, but then a different identifier for that student can be assigned to the record (and the SSN removed) before placing the record in the research database. The Florida Department of Education is an example of an agency that follows this practice. These non-SSN student identifiers and other educational variables such as test scores and free and reduced price lunch indicators are of little use to identity thieves, as those pieces of information are not used by lending institutions to verify individuals' identities.

48 See Aimee Guidera's chapter in this volume for a description of the ten essential elements.

49 A modest amount of federal funding for that purpose is currently provided to the Center for the Analysis of Longitudinal Data in Education Research (CALDER), a collaborative of researchers at the Urban Institute, Duke University, Stanford University, the University of Florida, the University of Missouri-Columbia, the University of Texas at Dallas, and the University of Washington.

50 A precedent for this has been the creation of the SchoolDataDirect database funded by the Gates Foundation. As of fall 2008 this database contained no student data and held few statistics built from longitudinal student data, but that could change in the future.