

**The UTD Texas Schools Microdata Panel (TSMP):
Its History, Use and Ways to Improve State Collection of
Public School Data**

John F. Kain

The Cecil and Ida Green Chair for the Study of Society and
Director of the Cecil and Ida Green Center for the Study of Science and Society
The University of Texas at Dallas

Prepared for meeting on:

The Secretary's Forum on Research and Value-Added Assessment Data

Sponsored by:

United States Department of Education

and

Education Leaders Council

Washington, D.C.

December 5, 2001

The UTD Texas Schools Microdata Panel (TSMP): Its History and Use

I. Introduction and Overview

This meeting is evidence of the increased interest among officials responsible for developing and overseeing the nation's decentralized system of public and private schools in improving and making better use of the administrative data on individual students and schools collected by state departments of education. A growing number of states have developed, or are currently developing, state-level databases that contain extensive information on individual students, teachers, campuses and districts. In several states these databases include the results for individual students of statewide-standardized achievement tests. Frequently, these data are used in efforts to assess the performance of individual campuses, districts and, less-often, of individual teachers.

Texas has been a leader in this movement. It has: (1) engaged in extensive statewide testing of public school students since 1980, (2) implemented a sophisticated and extensive Public Education Information Management System (PEIMS) to better utilize its massive data collection efforts in 1990, (3) made extensive district and campus level data describing district and campus resources and student performance available to policy makers and the public through its innovative Academic Excellence Indicator System (AEIS) system; and (5) implemented an ambitious accountability system based on PEIMS and AEIS data .

This paper describes an effort I began in 1992, while I was a visiting professor at the University of Texas at Dallas (UTD), to obtain individual data on Texas public school students and teachers for research on the effect of school quality on minority student achievement and to create a panel database for research on a wide range of questions of relevance to educational policy.¹ Prior to accepting a permanent appointment at UTD in

¹ While I use the first person in discussing TSMP, it should be clearly understood that a large number of individuals contributed to its creation and continued maintenance. These include legions of undergraduate and graduate research assistants at both Harvard and UTD as well as several colleagues at Harvard, UTD and elsewhere who provided highly useful advice. I would be especially amiss, however, if I did not single out the contributions of two individuals. Kraig Singleton worked for me full-time for two years and did the heavy lifting in the early efforts to create a panel database from TEA's numerous and disparate annual files. Kraig, who wrote his undergraduate senior honors thesis under my direction at Harvard before taking a job as a programmer and analysis at Charles River Associates, agreed to take a lower paying job at Harvard to help me analyze the data I had obtained from TEA and to create the first version of TSMP. Kraig returned to the Kennedy School for a Masters in Public Policy at the same time I moved to Dallas and UTD. As soon as I arrived at UTD, I began searching for someone with strong programming skills who could do some of the work Kraig had done. I had incredibly good luck. George Farkas, who was then a Professor at UTD's School of Social Sciences and an old friend, recommended Dan O'Brien, who he identified as one of the best graduate students he had encountered at UTD and who had the required quantitative skills and aptitude. Dan had recently returned to school to pursue a Ph. D. after a highly successful 25-year career first as a programmer and then as a manager in a series of increasing responsible positions in the data processing services industry. Dan hit the ground running and soon became a full-partner in both the further development of TSMP and in the research based on it. After Dan completed his Ph. D., I persuaded him to become Associate Director of the UTD Texas Schools Project. Last year UTD's School of Social Sciences offered Dan a tenure track position as an Assistant Professor of Political Economy, which he accepted. This was both bad news and good news. He would have less time for our joint research; but he would be

the spring of 1997, I was the Henry Lee Professor of Economics and Professor of Afro-American Studies at Harvard where the then Harvard/UTD Texas Schools Project was housed. The project is now an integral part of UTD's Cecil and Ida Green Center for the Study of Science and Society (Green Center).

The UTD Texas Schools Microdata Panel (TSMP), currently includes individual data for more than 11 million students and more than 400,000 public school teachers and administrators for the period 1990 to the present. While TSMP initially included only Texas Education Agency (TEA) and State Board of Educator Certification (SBEC) data on public school students and teachers, it has subsequently been expanded to include data for college and university students and limited, but highly valuable, data on significant numbers of Texas private high school students.

The use of common, encrypted student identifiers enables us to track individual students over time, while maintaining the confidentiality of these sensitive data. We hope to be able to add quarterly employment and earnings data to TSMP in the near future and, using the same encrypted identifiers, to link them to the education data that are already included in our database. While we refer to TSMP as a database, we use the term in the generic sense rather than as denoting a relational database. Instead, TSMP consists of a large number of flat files with a few common identifiers or keys. In using TSMP data for specific analyses we create working files that combine data from large numbers of TSMP files.

The sections that follow provide an overview of the development, use and potential improvements in TSMP. Section II provides a history of the UTD Texas Schools Project and the development of TSMP, including fairly detailed descriptions of the data included in both the original (Phase I) version and the current version. This section also identifies sources of funding for this work. Were it not for the generosity of a small number of private foundations and the support of first Harvard and then UTD, there would be no TSMP. Section III presents a brief description of the research that has thus far been completed using TSMP and indicates where more detailed information about these studies may be obtained.

TEA, SBEC, and the Texas Higher Education Coordinating Board (Coordinating Board), the principal contributors of data currently included in TSMP, have done a remarkable job of data collection and quality control. Administrators and staff from these organizations, moreover, have spent hundreds of hours educating us about their data and preparing extracts for our use.² While the data collected by these organizations are of

less likely to leave UTD for a teaching position at another university. Fortunately, we were able to recruit T. Robert (Bob) Harris, an experienced statistician and academic researcher, to work full-time on the project. Bob has taken on many of Dan's responsibilities for managing our small army of programmers and overseeing various aspects of database management and development.

² Large numbers of TEA, SBEC and Coordinating Board employees made valuable contributions to our work. I want to particularly acknowledge TEA's Belinda Dyer, Darlene Gouge, Shannon Houssan, Keith Kruse, Cherry Kugle, Nina Taylor and Nolan Wood. At SBEC, James Nelson, the organizations first director gave me valuable advice about the timing of our data requests, Pamela Tackett, who was James Nelson's deputy and is currently SBEC's head, and Dan Junell, SBEC's general counsel, were then very helpful in responding to our requests and Stephanie Korcheck did the required programming. Finally, at the Coordinating Board, I would like to acknowledge the continuing support of David Gardner, Susan Brown, Kathy Benson, and particularly of Ken Dalley, who has had the unenviable responsibility of

uniformly high quality and quite extensive, they are not perfect. Since most participants in this meeting are primarily concerned with elementary and secondary education, the final section of this paper considers TEA's data collection activities and suggests ways in which its immensely valuable databases might be made more useful both to policy makers and researchers.

II. History of the UTD Texas Schools Project and TSMP

Developing TSMP has been an integral part of the UTD Texas Schools Project, since its inception. The goals of the larger project are to obtain a better understanding of the determinants of student performance with the long-term objective of providing a knowledge/research base to improve the performance of public schools. The project's initial focus was on elementary education, but a large Andrew W. Mellon Foundation grant to the Green Center in March 1999 enabled us to add research on higher education. Throughout the project's history I have tried to obtain external funding for development of TSMP with little success. In general funds for data acquisition and database development were only available as part of larger research projects focusing on specific questions.

While the lack of dedicated funding for database development has been a disadvantage, nonetheless, we have had considerable success. As I discuss in greater detail below, TSMP currently includes individual data for more than 11 million individuals who attended Texas elementary and secondary schools and/or Texas colleges and universities between 1990 and 2001, as well as extensive data on public school teachers.

The Spencer Foundation, which provided nearly \$400,000 of funding for Phase I of TSMP's development, supported the collection of the first eight years of Texas public school data, and the difficult and time-consuming effort of combining TEA's disparate and unlinked annual data to create an integrated panel database. Spencer also funded the project's first substantive focus, an investigation of the impact of increased minority access to suburban schools on the academic performance of minority, and especially black, children.

Phase I of TSMP included enrollment and attendance data, as well as standardized test scores, for more than two million Texas public elementary and secondary students for up to 10 years, 1990-1999. TSMP begins in the 1989-90 school year because TEA implemented PEIMS in that year. In each subsequent year, TEA has improved the quality and extent of these data. Students included in the Phase I belonged to five cohorts; members of the youngest cohort were in Pre-K during the 1989-90 school year and the oldest were in third grade in the same year. The grades and school years for these original five cohorts are identified with S in Figure 1. In addition to student data, Phase I included individual data for all Texas public school teachers for the same nine-year period, including their scores on teacher certification tests. Teacher certification test data were initially provided by TEA, but after the teacher certification function was

figuring out how to fit our massive data requests into the work program of an already overburdened data processing group.

Figure 1. TSMP Data By Funding Source: Spencer (S), Mellon (M), Smith-Richardson (SR) by School Year and Grade

School Years	Primary and Secondary School Grades														College				
	PK	K	1	2	3	4	5	6	7	8	9	10	11	12	Fr	So	Jr	Sr	Yr5
1990	S	S	S	S	S	M	M	M	M	M	M	M	M	M	M	M	M	M	M
1991	SR	S	S	S	S	S	M	M	M	M	M	M	M	M	M	M	M	M	M
1992	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M	M	M	M	M	M
1993	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M	M	M	M	M
1994	SR	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M	M	M	M
1994	SR	SR	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M	M	M
1996	SR	SR	SR	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M	M
1997	SR	SR	SR	SR	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M	M
1998	SR	SR	SR	SR	SR	SR	SR	SR	S	S	S	S	S	M	M	M	M	M	M
1999	SR	SR	SR	SR	SR	SR	SR	SR	SR	M	M	M	M	M	M	M	M	M	M
2000	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR	M	M	M	M	M	M	M	M	M
2001	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR	M	M	M	M	M	M	M	M
Data Source	Texas Education Agency														Coordinating Board				

moved to a new agency, we obtained them from SBEC. Use of encrypted identifiers enabled us to follow individual students and teachers over time. For all grades and years we were able to link teachers to individual students at the campus, grade and program (bilingual, ESL [English as a Second Language], special education, gifted and talented and vocational education) level.

In March 1999, The Andrew W. Mellon Foundation provided the UTD Texas Schools Project with a \$1.25 million grant to support research on minority access to higher education and on the impacts of *Hopwood vs. Texas* and the Texas Top 10 Percent Law on minority enrollment in Texas public colleges and universities, and particularly its most selective public universities. In *Hopwood vs. Texas*, the Fifth Circuit Court barred the use of race in college admission and financial aid decisions. The Top 10 Percent Law, signed into law by then Governor George W. Bush on May 20, 1997 (26 months after the Hopwood Decision), provides automatic admission of all students whose grade point averages placed them in the top 10 percent of the graduating class of any accredited Texas high school.

I originally envisioned following the original five cohorts into college and then into the labor force. For those that did not enroll in higher education or who worked during school, I would obtain employment data for them as soon as they became employed. The Mellon grant enabled us to accelerate my earlier research plan by providing funding to add TEA data for nine cohorts of older students to TSMP and by obtaining Coordinating Board data for all individuals enrolled in Texas public colleges and universities from 1990 to the present. The Texas Workforce Commission (TWC) has also agreed to provide employment and earnings data for all Texas residents for the period 1990 – 2010. These data will be added to TSMP.

The Mellon grant enabled us to add nine cohorts and 57 years/grades of TEA’s individual student data to the 45 grades/years of data that were included in the Phase I version of TSMP. The Ms in Figure 1 identify years/grades of Mellon funded data

collection including both public school data obtained from TEA data and college and university data obtained from the Coordinating Board. The Coordinating Board also provided ACT and SAT scores and related socioeconomic data. Finally, TEA provided individual data and scores for all persons that took the GED in Texas since 1990.

During March 2000, the Green Center received a \$443,000 grant from the Smith Richardson Foundation for research on charter schools and parental choice. Our research plan for the charter school study would have been impossible without TSMP. These data allow us to compare the performance of individual students before and after they enroll in a charter school. TSMP panel data thus enable us to use the past performance of students to separate student effects from school effects. This will permit us to observe the pattern of achievement growth for a student and determine whether a change from a conventional public school to a charter or other alternative public school results in a change in performance. The approach is similar to the one we used in our earlier Smith-Richardson funded research on special education (Hanushek, Kain and Rivkin, August 1998 and September 1998). Since the youngest students in the original five cohorts were ninth graders in 1999-2000 and there were no Texas open enrollment charters before 1997, we had to add additional cohorts to TSMP. Thus, we have used funds included in the Smith Richardson grant to add 10 cohorts of younger students to TSMP for the charter school study. The years/grades of TEA data funded by this grant are identified by the SRs in Figure 1.

TSMP Today

As we noted previously, TSMP currently includes up to 12 years of individual data for approximately 11 million persons who attended Texas elementary and secondary schools during 1990-2001 and/or Texas colleges and universities during the same period. “Year” here and throughout the paper refers to the academic year; i.e. 1990 is the 1989-1990 academic year.

Figure 2 provides a broad overview of TSMP’s current structure and the primary data sources. The Texas Education Agency (TEA) is the principal supplier of elementary and secondary school data. TEA’s data are shown in the first panel, where the Es and Hs indicate the years for which we have obtained numerous enrollment, attendance and program files and where the Ts identify years/grades for which we have both these data and statewide-standardized tests. The most numerous of these standardized tests are the Texas Assessment of Academic Skills (TAAS), which has been given to most students in grades 3-8 since 1991 and the Norm-reference Assessment Program for Texas (NAPT), which was given to all students in grades 3-8 in 1992 and 1993. We have seven years/grades of the NAPT test as well as one year/grade of data for an earlier test, the Texas Educational Assessment of Minimum Skills (TEAMS). The Ds identify years/grades in which standardized tests were given, but are not yet included in TSMP because of delays in obtaining them from TEA. The Hs identify high school years/grades without statewide-standardized tests. In addition to the various enrollment, attendance and program files for elementary school students, there are a number of other types of TEA data for high school students. TEA also supplied data for more than 600,000 individuals who took one or more GED exams during 1991-2000. Of these individuals

Figure 2. Types of Data Currently Included in the UTD Texas Schools Microdata Panel (TSMP)

School Year Ending In	Primary and Secondary School Grades														College				
	PK	K	1	2	3	4	5	6	7	8	9	10	11	12	Fr	So	Jr	Sr	Yr5
1990	E	E	E	E	E	E	E	E	E	E	H	H	H	H	C	C	C	C	C
1991	E	E	E	E	T	E	D	E	D	E	D	D	H	H	C	C	C	C	C
1992	E	E	E	E	T	N	D	E	D	E	D	D	H	H	C	C	C	C	C
1993	E	E	E	E	T	T	N	N	D	D	H	D	H	H	C	C	C	C	C
1994	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
1995	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
1996	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
1997	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
1998	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
1999	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
2000	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C
2001	E	E	E	E	T	T	T	T	T	T	H	T	H	H	C	C	C	C	C

Notes: E identifies elementary school years/grades for which the enrollment, attendance and program files listed in Table 1 are included in TSMP, identifies years/grades for which the same data represented by E are included in TSMP plus state-wide standardized tests are included in TSMP; D has the same meaning as T except that TEA has not as yet supplied the standardized tests; H identifies high school years/grades for which for which the enrollment, attendance and program data listed in Table 1 are included in TSMP and T and D have the same meaning as for grades Pre-K through 8. C identifies years and grade levels for which the Coordinating Board files listed in Table 1 are included in TSMP.

more than 484,000, many of them dropouts from Texas public schools, were awarded GEDs.

Encrypted identifiers enable us to link the several types of data/files included in TSMP both to each other and over time. We have worked closely with TEA and other agencies and organizations that have provided data for inclusion in TSMP to develop procedures to insure their confidentiality. Our files include no names and the various IDs are encrypted before they are sent to the Green Center, where they are converted to STATA format and are included in TSMP.³ These data were never meant to be used to create a panel database. As a result, while each year's data are remarkably complete and error free, there are, nonetheless, frequent omissions, inconsistencies and errors in the encrypted personal identifiers we must rely on to link the files. The most important of these identifiers are the encrypted Social Security Numbers (SSNs), which is the principal information we use to link the millions of individual records produced by the agencies, organizations, and divisions that have provided data for TSMP

The right most panel in Figure 2 shows the years/grades of data on college students we have obtained from the Coordinating Board. Most of these data are for students attending Texas public colleges and universities, although in certain years we are able to identify about half of all Texas residents attending Texas private colleges and universities from financial aid data. In addition, we have been able to use Coordinating Board enrollment data to identify the origin-institution of significant numbers of Texas

³ STATA is the data processing and statistical package we use for virtually all of our analyses.

high school graduates who transferred to a Texas public college or university after attending a Texas private or an out-of-state college or university. The Coordinating Board, with the College Board and ACT's approval, has provided, or will provide, ACT and SAT data for Texas residents who took these tests during 1991-2001. These data include test scores, extensive information on family background, student interests, high-school records and activities and codes identifying the colleges and universities where test takers have sent scores. Finally, the Coordinating Board has provided us with three years of financial aid data (1997-1999) for students attending both public and private schools and has agreed to make them available for subsequent years. As noted previously, we plan to add employment and earnings data for all Texas residents for the same period to TSMP. Officials at TWC assured us more than a year ago that we would be able to obtain these data for all covered Texas residents, but we have yet to receive any data or even written approval.

Table 1 provides a summary of the data included in the current version of TSMP. The data currently available are in 579 separate files, which provide individual information for more than 11 million persons. The number of records is 430 million. With a few exceptions we have obtained all of TEA's standardized tests for 1990 through 2001 and all but a few of the files from PEIMS for 1990-2000. We expect to receive the remaining available data we have requested from TEA within the next few months. The Coordinating Board has provided all of their files listed in Table 2 for Fall 1990 through Spring 2001. Each public university reports enrollments and related data to the Coordinating Board three times a year (spring, fall and summer), while community colleges have four reporting periods (spring and fall semesters plus two summer reporting periods). They provide Texas Assessment of Skills Program (TASP) files to the Coordinating Board once a year. As noted earlier, we have also thus far obtained three years of financial aid data and have requested these files for 2000 and 2001.

III. Research Based on TSMP

The panel nature of TSMP, the availability of multiple cohorts and large sample sizes have enabled us to provide persuasive evidence about a number of questions that are highly relevant to improving public education, that previously have been resistant to educational research. Attachment A, which lists 29 working papers and two Ph. D. dissertations by researchers associated with the UTD Texas Schools project, illustrates the range of research that has been completed using TSMP data. Seventeen of the working papers deal with research on grades Pre-K through 8. The first nine deal primarily with questions related to the minority-white achievement gap. The O'Brien-Ware paper, which is not based on TSMP data, describes value-added regression models O'Brien estimated to help the Fort Worth Interdependent School District assess a large-scale direct-instruction reading program the District implemented in the hope of improving what it regarded as unacceptably low levels of reading performance by many of its children.

The next three working papers describe research on special education funded by the Smith Richardson Foundation. These are followed by seven papers by Hanushek,

Table 1. Summary Statistics for Data Included in the Current Version of TSMP

File Sources and Types	Number of Variables (1999)	Years in Data Base			Number of	
		From	To	years	Years X Files	Records
Coordinating Board (THECB)						
Four-Year						
Student Report	26	1990	2001	12	35	11,483,320
Class Report	20	1990	2000	11	32	2,254,509
End of Semester Report	19	1993	2000	8	22	1,561,546
Faculty Report	31	1990	2000	11	22	532,331
Graduation Report	15	1992	2000	9	9	659,671
Application Report	21	1999	2000	2	2	394,545
Financial Aid	74	1997	1999	3	3	679,033
Two-Year						
Student Report	37	1990	2001	12	46	13,064,424
Class Report	24	1990	2001	11	41	2,212,497
End of Semester Report	22	1994	2000	7	24	1,336,714
Faculty Report	32	1990	2000	11	22	499,657
Graduation Report	15	1992	2000	9	9	316,176
Financial Aid	74	1997	1999	3	3	416,794
TASP Report	51	1990	2000	11	11	10,447,534
ACT & SAT						
ACT	342	1991	2000	10	10	441,335
SAT	472	1991	1999	9	9	787,633
Texas Education Agency (TEA)						
Enrollment	23	1990	2000	11	11	40,346,410
Attendance	8	1993	2000	8	48	188,615,143
TAAS						
Grades 3-8	185	1991	2001	11	52	14,751,140
Exit	131	1994	2001	8	16	3,135,315
End of Course	52	1995	2000	6	18	1,623,165
NAPT	7	1992	1993	2	7	1,963,369
Demographic	8	1993	2000	8	9	36,302,827
Course Completion	11	1993	2000	8	8	89,795,340
Graduation Type	11	1991	1999	9	9	1,575,725
School Leaver	9	1993	1998	6	6	180,408
Special Education Attendance	27	1990	2000	10	10	2,626,706
GED	33	1991	2000	10	10	954,256
State Board of Educator Certification (SBEC)						
Teacher Certification Tests						
<hr/>						
Total	1,780				504	428,957,523

Rivkin and me that are based on other research using TSMP data. None of this research could have been done without access to TSMP. The final Pre-K – Grade 8 paper by Caroline M. Hoxby, a Professor of Economics at Harvard, is based on campus-level tabulations we prepared for her without charge. Hoxby first asked TEA to provide these data. They referred her to us with the suggestion that we would be able to prepare the tabulations that she needed more quickly and more cheaply than they could. In preparing these tabulations we used the same suppression rules as TEA uses for its widely distributed AEIS data, thus maintaining the confidentiality of TSMP data.

MacGregor Stephenson, a Ph. D. candidate in Education at Texas A&M used TSMP data for his Ph.D. dissertation. In this case, the Coordinating Board referred Stephenson to us. They concluded that they could not supply him with the data he needed for his research, but suggested he contact me to see if we could help. After signing the same confidentiality agreement that we require of all persons working with TSMP data, MacGregor spent several months at UTD using Green Center computers to complete doing analyses based on TSMP data. I agreed to serve on his thesis committee and several Green Center staff helped him understand and use STATA and TSMP data. Even though names and other identifying information have been removed from our TSMP files, they, nonetheless, can be used only on Green Center secure computers.

The final eight working papers focus on high schools and higher education. All but one of these papers, “Ethnic and Racial Differences in Graduation, Dropout Rates and Course Completion for Students Attending Texas Public High Schools,” are part of our Mellon funded research on minority access to higher education. I prepared this paper at the request of Hardy Murphy for the Texas Commission on a Representative Student Body. Murphy, who is currently Superintendent of Evanston, Illinois’ Elementary School District, was an Associate Superintendent of the Fort Worth Independent School District and a member of the Commission.

IV. Suggested Additions Improvements in TEA’s Data Collection

As good as PEIMS and standardized test data included in TSMP are for assessing and monitoring public schools in Texas, their usefulness could be greatly improved at modest cost. The following are a number of suggestions for augmenting and improving TEA’s data collection. I recognize that an effort by TEA and other state departments of education to collect these data might be viewed as intrusive and be quite controversial. Nonetheless, developing fair and accurate accountability systems, obtaining a more complete understanding of the determinants of student achievement and improving the state’s capacity to assess the performance of its public schools would greatly benefit from the collection of some or all of the data described below.

Testing in the Early Grades

The most expensive, but most important, improvement would be statewide standardized testing of first and second grade, and possibly kindergarten, students. Many

districts routinely test first and second grade students, but there is an urgent need for statewide testing in these grades. These data would strengthen the state's accountability system and provide districts and campuses with the information they need to assess students and for timely intervention. Identifying low student, campus and district performance at the end of the third grade is simply too late. Texas requires school districts to annually administer a reading skills assessment to K-2 students, but allow them to use a variety of instruments and TEA does not obtain scores for individual children. TEA has also been developing tests that can be used to assess the gains of large numbers of special education students, who heretofore have been exempt from both testing and the accountability system. It appears they agree with my view that all but a tiny fraction of both special education and Limited English Proficient (LEP) students should be tested in each year and be included in the accountability system.

Insofar as districts already test in the early grades, the net cost of a statewide-standardized testing will be proportionately less. If many districts object to early testing, the state should follow the federal government's example and make the testing of students in the early grades voluntary. At the same time, they should pay the entire cost of these early tests to persuade more districts to do early testing and require them to use a common test.

Family Background, Pre-School Experiences and Remedial Programs

Research on the determinants of student performance consistently identifies family background as the most important determinant of student achievement. The fairness and adequacy of TEA's accountability system and the usefulness of PEIMS and TAAS data for research on the determinants of student achievement would be greatly increased at low cost by the collection of additional data on the characteristics of students' parents and their families. Listed below are data that might be collected by districts at the time of enrollment and included in their fall and summer submission of demographic data to TEA. They are listed in order of importance and cost-effectiveness.

- A. Mother's education (years of school completed) and, ideally, father's education.
- B. A family ID for each student. The best ID would be the mother's social security number. If some parents are reluctant to provide their social security numbers, a state supplied family ID might be used for them.
- C. Enumeration of all household members by sex, age and relation, including all adults, school age children and other children who are living at home.
- D. Prior schools attended (country, state/province, district and campus) for children enrolling in Texas public schools after kindergarten. This includes both new Texas residents and students transferring from private schools.
- E. The scores obtained by LEP children on English proficiency tests.
- F. Information on after school and summer remediation programs.

- G. Information on pre-school attendance. Number of years of pre-school attendance and their names/identities and descriptive information about other less formal forms of out-of-home child care experiences.
- H. Current street address and months/years at current residence.
- I. Course grades and/or grade point averages for high school students. PEIMS and TSMP already include lists of all courses taken by high school students and information on whether they were successfully completed.

The case for obtaining parent's, and particularly mother's, years of schooling completed is overwhelming. These data are routinely collected in the U. S. Census and the Current Population Survey and by a few school districts. There may be some opposition, but questions about years of schooling are generally considered less intrusive than other measures of socio-economic status such as family income.

A family ID for each student would enable researchers to exploit the common family background of siblings in efforts to identify the respective contributions of parents/families, schools, classrooms and communities to individual student achievement and other aspects of behavior. It appears that relatively few districts currently identify families on their student databases, but it would not be difficult and would be of great value. The best measure would probably be the mother's social security number because, in cases of separation or divorce, children usually stay with their mother.

TEA obtains the prior school attended for all students attending Texas public schools. Regrettably, they do not obtain these data for students transferring from private schools, from other states or from other countries. Adding this information to PEIMS would presumably be fairly easy, would be of great value in linking the records of individual students over time and would be useful in answering a number of important issues relating to the movement of students between private and public schools and the impact of migration on student achievement.

While a family ID number would enable analysts to identify siblings currently attending Texas public schools, it would not provide information on children who are either too young to attend school, are enrolled in private schools, have dropped out or have already graduated from high school. Similarly, it provides no information about other important aspects of family structure, and particularly about whether the child comes from a one or two-parent family. Collecting these data would be more difficult than simply obtaining the mother's social security number, but they would be of enormous value in answering a large number of questions related to student achievement and the performance of individual schools and teachers.

Districts should also be required to supply TEA with the scores LEP children obtain on the English Proficiency tests that are used in determining their assignments to bilingual, ESL and regular programs and their scores on subsequent tests used to determine whether particular children should be transitioned to regular classrooms. TEA currently has limited capacity to monitor these practices. Addition of these scores as regular PEIMS data items would be of great value in developing an improved understanding of the persistent achievement gaps of LEP children. If TEA mandated the specific tests to be used for these purposes, it would increase the utility of these data.

Even if TEA decides to continue allowing districts to use different tests, the scores should be included in PEIMS with information that identifies which tests were used.

Individual students, particularly low achieving ones, currently receive a variety of extra services. These include pullout programs during the school day, after school programs and summer programs. Yet PEIMS includes no information on the participation of individual students in these activities, even though they are frequently funded by state or federal money and usually are meant to increase student achievement. One explanation may be the difficulty of describing them in an economical, yet meaningful, manner. Nonetheless, a serious effort should be made to collect data for individual students on the types and quantities of remedial instructional services they receive within the school day, before school, after school, on weekends and during school vacations.

There is a growing awareness of the importance of pre-school experience. TEA currently obtains little or no information about the pre-school experience of Texas school children, and it is my impression that districts know little more. Pre-school may be particularly important for children who come from immigrant and low-income families. While obtaining relevant information on these programs is likely to be much more difficult than the collection of the other information listed above, TEA should, nonetheless, make a serious commitment to developing a workable scheme for collecting these data. At minimum school names as well as months and hours of attendance should be obtained for each pre-school attended as well as more generic descriptions of more informal forms of out-of-home child care.

In comparison with the two previous items, street addresses are routinely obtained by individual districts and could easily be added to PEIMS. If these data were included in PEIMS it would be possible to obtain additional and stronger controls for family background and neighborhood characteristics, *albeit* at some expense for geo-coding, from secondary sources such as Census block and block-group data. Added family background controls and information on the extent of residential, as opposed to school mobility, would help strengthen the various kinds of analyses discussed previously, that employ standardized test data to evaluate individual campuses, districts or teachers as well as research on student achievement and the efficacy of various programs and policies. If parents' or mothers' years of schooling and data on family composition were added to PEIMS, the benefits of adding street addresses to PEIMS would be significantly reduced, although there would remain important questions relating to residential mobility and its impact on achievement.

PEIMS currently includes the individual courses taken by high school students and information on whether they passed or failed these courses. Course grades, or at minimum end of year grade point averages and class ranks, should be added. These data would permit TEA to evaluate the consistency of grading standards across districts and the connection between course grades and student performance on TAAS and other standardized tests. The importance of these data is likely to be increased as policymakers begin to come to grips with the implications of legislation that would discontinue social promotions; class-rank data, moreover, would be useful in assessing the Top 10 Percent Law.

Student Teacher Links

TSMP is a nearly ideal tool to assess teacher effectiveness and to attempt to obtain a better understanding of which teacher characteristics contribute to this effectiveness. It could also be used to explore the feasibility and desirability of using student performance or gains on standardized achievement tests as part of a teacher assessment program. The qualification, nearly, refers to the fact that TEA does not currently include information on who teaches particular students in its PEIMS data base. While I do not know all of the reasons for this omission, concerns by individual teachers and their professional organizations that this information might be used to assess the performance of individual teachers is obviously an important part of the story.

While there is considerable opposition to including student-teacher links in PEIMS, the usefulness of standardized test data collected by TEA would be greatly increased if PEIMS student records identified each student's teacher(s), along with the subject taught and contact hours. While these data might ultimately be used for teacher assessment, evaluating teacher performance is but one of many uses. Of these the most important is research on the determinants of student achievement. The lessons learned from this research would provide valuable guidance to Texas educators in their efforts to improve teacher training, teacher effectiveness and ultimately to improve student performance.

If PEIMS included links that identified each student's specific teachers it would be possible over several years to obtain incontrovertible evidence about the relative contributions of individual teachers and campuses to student achievement. The above phrase "over several years" deserves further emphasis. Even if sophisticated methods and strong data are used, it would be a mistake to rely on a single year's performance. On the other hand, if the same teacher's students perform much better for a number of years (after adjusting for differences in prior student achievement and family background), it would be reasonable to believe that there is something about the teacher's educational practices, preparation or other attributes that account for his/her consistently higher performance.

Curriculum and Educational Practices

TEA collects and maintains a great deal of information about individual students and teachers and their performance on standardized tests. Surprisingly, there appears to be little or no systematic collection of information on curriculum and educational practices. Campuses and classrooms remain black boxes. TEA and Texas are not unusual in this respect. The absence of variables describing curriculum and educational practices in educational production function studies may account for the failure of these studies to regularly find positive and statistically significant relationships between student

achievement and various school input measures, such as class size, teacher education, and teacher experience.⁴

One reason these data are not collected is the difficulty of devising useable and reliable questionnaires that would not place too great a burden on already busy administrators and teachers. Developing such instruments would not be impossible, however, and these data would be far more valuable than other information that is currently collected. Described below are two examples of the kinds of data that might be collected from individual teachers during each school year.

Reading Curriculum and Instructional Practices. The National Academy of Sciences Committee report on reading and other recent studies have raised important questions about how reading should be taught, and particularly about the effectiveness of direct phonics instruction vs. whole language (Committee on the Prevention of Reading Difficulties in Young Children, 1998). If these differences are as important as the protagonists in this debate allege, it should be possible to find supporting evidence in TAAS performance. It is clear that the methods used to teach reading differ widely across districts and campuses and even within the same grade and campus. The fact is that no one really knows what methods classroom teachers in Texas currently use.

Information on the methods and materials used could be obtained from individual classroom teachers. Specifically, elementary school reading teachers might be asked to complete a simple form describing their primary approach to reading instruction (whole language vs. phonics) and, in the case of those who use a mixed strategy, the fraction of time they spend using each method. Elementary school teachers might also be asked to indicate how much use they make of reading groups and ability grouping, the amount and types of seatwork they assign and they amount of time devoted to reading instruction. If information on the educational background and training of these teachers and the nature and extent of their training in reading instruction is not already available, these data should be collected as well.

Teachers should also be asked to identify the teaching materials they use. State approved textbooks play an important role in reading instruction in Texas schools, but many teachers supplement them with other materials and some make very little use of prescribed textbooks. Finally, teachers might be asked to indicate the types and amount of homework they assign during a typical week. Data on these and other instructional practices could be used in combination with the data already collected by TEA to determine whether differences in curriculum and instructional practices have any significant impact on student performance. The power of these analyses, of course, would be greatly increased if individual student-teacher links were included in PEIMS.

⁴ Hanushek (1997, p. 141) in a 1997 survey of close to 400 studies of student achievement finds “there is not a strong on consistent relationship between student performance and school resources, at least after variations in family inputs are taken into account. More specifically, he found that in 277 studies that included the teacher-pupil ratio as an explanatory variable only 15 percent reported a positive and statistically significant coefficient. Whether the comparisons are limited to statistically significant coefficients or if both significant and insignificant coefficients are included, the result remains the same. There are about as many studies with negative as with positive coefficients for the teacher-pupil ratio (Hanushek, 1997, p. 144). Similar discouraging results were obtained for teacher education, teacher experience, teacher salary and expenditure per pupil.

Bilingual/ESL Programs. The performance of LEP, children lags badly behind those raised in English-only households. The relationship of their poor performance to the state's bilingual and English as a Second Language programs is unclear. Critics of bilingual education contend that in all too many instances bilingual programs fail to prepare students to function in English and that the children participating in these programs would be better off in regular English language classrooms. Just like reading instruction, it is clear that bilingual programs differ greatly in objectives and practice. Yet there is no systematic information about these differences. Without these data, bilingual and ESL programs cannot be meaningfully evaluated.

To provide these critical data, individual bilingual/ESL teachers might be asked annually to complete a brief questionnaire that describes their instructional goals, and, more importantly, gives the fraction of instructional time they devote to each subject and the fraction of this instruction that is in English. As in the case of the reading instruction example, data on bilingual instructional practices could be used in combination with the testing data already collected by TEA to determine whether differences in curriculum and instructional practices have a significant impact on the performance of students participating in bilingual education programs.

The examples given above are simply illustrative. It would be desirable to enlist the aid of reading specialists and experts in teaching LEP children in developing survey instruments that would not place too much of a burden on individual teachers and yet would capture the essential features of these practices. There is no doubt, moreover, that many other kinds of information could and should be collected on educational practices. These data in combination with an enriched PEIMS/TAAS information system would enable researchers to begin finding out what does and does not work in educating Texas' young people.

Conclusion

TSMP provides an unparalleled opportunity to accomplish meaningful assessments of the effectiveness of school resources, organization, programs, and practices on student achievement.

TSMP, however, is far from perfect. Statistical techniques can help to control the effects of family environment, preparation for school, and other confounding influences, but more effective analysis could be performed through the collection of additional data. Among the most important of these are family characteristics, such as numbers of siblings, mother's education, home address, teacher-student matching, early and special test scores, and additional program participation information. Teaching practice information is also critically important; opening the black box of what really happens in the classroom is critical to program evaluation and improvement.

While some of these data would be difficult to obtain and may require periodic surveys, much additional information is already maintained by school districts. I believe that these data will allow researchers to better identify the causal relationships of school and program practices and policies on student achievement, and that this understanding is a prerequisite to effective school improvement.

References

- Committee on the Prevention of Reading Difficulties in Young Children. 1998. *Preventing Reading Difficulties in Young Children*, Washington, DC: National Academy Press.
- Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, Summer, Vol. 19, No. 2: 141-164.
- Hanushek, Eric A., John F. Kain and Steven G. Rivkin. 1998a. "Does Special Education Raise Academic Achievement for Students with Disabilities," August.
- Hanushek, Eric A., John F. Kain and Steven G. Rivkin. 1998b. "Does Special Education Work," September.
- Hanushek, Eric A., John F. Kain and Stephen G. Rivkin. (forthcoming). "Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disability?" *Review of Economics and Statistics*.

Attachment A

UTD Texas Schools Project Publications and Working Papers

Most of the following working papers are, or soon will be, available at:

<http://utdallas.edu/research/greenctr>

Pre-K through Grade 8 Research

John F. Kain and Kraig Singleton., "Equality of Educational Opportunity Revisited." *New England Economic Review*. (May/June), 1996.

John F. Kain and Daniel M. O'Brien, "Has Moving to the Suburbs Increased African American Educational Opportunities?" January 23, 1998.

John F. Kain and Daniel M. O'Brien, "A Longitudinal Assessment of Reading Achievement: Evidence from the UTD Texas Schools Project," April 2-4, 1998.

John F. Kain, "Using TEA Annual Data to Develop a Multi-Year Panel Data Base: Lessons Learned and Suggested Additions and Improvements to TEA's Data Collection," May 11, 1998.

Daniel M. O'Brien, "Family and School Effects on the Cognitive Growth of Minority and Disadvantaged Elementary School Students," November 8-10, 1998.

Sharon Leigh Wrobel and Daniel M. O'Brien, "Assessment of Bilingual Education Programs in a Large Texas School District," October 31, 1998.

John F. Kain, "The Impact of Individual Teachers and Peers on Individual Student Achievement," October 29-31, 1998.

John F. Kain and Daniel M. O'Brien, "Black Suburbanization in Texas Metropolitan Areas and Its Impact on Student Achievement," March 9, 2000

Daniel M. O'Brien and James C. Murdoch, "School Choice in a Large Texas School District," December 2000.

Daniel M. O'Brien and Anne M. Ware, "Implementing Research-Based Reading Programs in the Fort Worth Independent School District," *Journal of Education for Students Placed at Risk*," (Forthcoming).

Research on Special Education

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "Does Special Education Raise Academic Achievement for Students with Disabilities," August 1998.

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "Does Special Education Work," September 1998.

Eric A. Hanushek, John F. Kain and Stephen G. Rivkin, "Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disability? March 2001 (revised). Accepted for publication, *Review of Economics and Statistics* (forthcoming).

Charter Schools, Choice and other Hanushek, Kain and Rivkin Research

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "Do Higher Salaries Buy Better Teachers?" February 8, 1999.

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "The Cost of Switching Schools," April 1999.

Steven G. Rivkin, Eric A. Hanushek and John F. Kain, "Teachers, Schools and Academic Achievement," July 1998 (Revised April 2000).

Eric A. Hanushek, John F. Kain, Jacob M. Markman, and Steven G. Rivkin, "Does Peer Ability Affect Student Achievement?" October 2000, Accepted for publication by the *Journal of Applied Econometrics*.

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "How Much Does School Integration Affect Student Achievement?" November 2, 2000.

Eric A. Hanushek, "Evidence, Politics, and the Class Size Debate," February 2000.

Eric A. Hanushek, John F. Kain and Steven G. Rivkin, "Disruption Versus Tiebout Improvement: The Costs and Benefits of Switching Schools," September 2001.

Eric A. Hanushek, John F. Kain, and Steven G. Rivkin, "Why Public Schools Lose Teachers," October 2001.

Other Pre-K through 8 Research Using TSMP

Caroline M. Hoxby, "Peer Effects in the Classroom: Learning from Gender Variation," July 2000.

High School and Higher Education Research

John F. Kain, "Ethnic and Racial Differences in Graduation, Dropout Rates and Course Completions for Students Attending Texas Public High Schools," June 4, 1998.

John F. Kain and Daniel M. O'Brien, "High School Outcomes and College Decisions of Texas Public School Students," November 2, 2000.

Kain, John F. and Daniel M. O'Brien paper, "High School Outcomes and College Decisions of Texas Public School Students," November 3, 2000.

Kristin Klopfenstein, "Advanced Placement: Do Minorities Have Equal Opportunity?" April 2001.

Kristin Klopfenstein, "Minority Advanced Placement Participation: Is Equal Opportunity Enough?" 2001.

M. Kathleen Thomas, "College Enrollment Demand and the Number of Credit Hours: Differences in Full and Part Time Students," 2001.

M. Kathleen Thomas, "Where College-Bound Texas Students Send Their SAT and ACT Scores: Does Race Matter?" October 2001.

John F. Kain and Daniel M. O'Brien, "Hopwood and the Top 10 Percent Law: How They Have Affected the College Enrollment Decisions of Texas High School Graduates," November 9, 2001.

Ph. D. Dissertations Based on TSMP

Daniel M. O'Brien, "Three Essays on Early Academic Achievement of Minority and Disadvantaged Students," The University of Texas at Dallas, August 1999.

MacGregor Stephenson, "The Impact of High School Curriculum and High School Campus Characteristics on Graduation Rates from Texas Four-Year Public Institutions of Higher Learning," Texas A & M University, November 2001.

December 2001