Using TEA Annual Data to Develop a Multi-Year Panel Database: Lessons Learned and Suggested Additions and Improvements to TEA's Data Collection

John F. Kain

The Cecil and Ida Green Center for the Study of Science and Society

The University of Texas at Dallas

May 11, 1998 (Revised)

Preliminary Draft: Do Not Cite Without Permission of the Author

Prepared for the 12th Annual Texas Assessment Conference, February 15-18, Renaissance Austin Hotel, Austin Texas

---

* The author is grateful to Eric Hanushek, Steve Rivkin, and particularly Daniel M. O'Brien for their helpful comments, to Ankur Gupta, Karen Chaddick, Nathanael Altice and especially Jaison George for their excellent research assistance, and to B. Jean Kingsley for editorial and manuscript preparation. He also thanks the Spencer Foundation and the Smith Richardson Foundation for their financial support.

Using TEA Annual Data to Develop a Multi-Year Panel Database: Lessons Learned and Suggested Enhancements of TEA's Data Collection

Introduction

This paper has four parts. First, it provides a brief description of the UTD Texas Schools Project and the Texas Schools Microdata Panel (TSMP). TSMP combines annual data from Texas Education Agency's (TEA) Public Education Information Management System (PEIMS) and standardized test data into a multi-year panel database. Second, it discusses and briefly assesses the various potential uses of data on Texas public schools collected and maintained by the TEA. Third, based on our experience in developing and using TSMP, it suggests a number of additions to TEA's data collection efforts that could greatly increase the usefulness of these data at modest cost. Finally, it briefly describes our fledgling efforts to supplement TSMP with local data obtained from individual school districts.

The UTD Texas Schools Project and the Texas Schools Microdata Panel (TSMP)

The UTD Texas Schools Project is a multi-year research project whose goals 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.

John F. Kain, Cecil and Ida Green Chair for the Study of Science and Society at the University of Texas at Dallas (UTD) and Director of UTD's Green Center for the Study of Science and Society initiated the UTD Texas Schools Project in 1992 when he was a Visiting Professor at UTD. Prior to accepting a permanent UTD appointment in spring 1997, Kain was the Henry Lee Professor of Economics and Professor of Afro-American Studies at Harvard where the project was previously housed. The project is now housed at the Green Center.

The project's primary focus to date has been the creation of the Texas Schools Microdata Panel (TSMP). This multi-year panel database already includes eight years of linked micro data for more than two million students attending Texas public schools. TSMP will be used for research on a large number of important questions that are either poorly understood or for which there has been, heretofore, little or no research or even systematic information.

TSMP includes individual student, teacher, district and campus data.[1] The student data include enrollment, attendance and standardized test records of five cohorts of students. As Table 1 reveals, the members of the youngest cohort were in Pre-K during the 1989-90 school year while members of the oldest were in third grade in the same year. TSMP begins in the 1990-91 school year because TEA implemented PEIMS in that year. In each subsequent year, TEA has improved the quality and extent of these data. The letter and number designations in the columns labeled Test/Grade in Table 1 identify particular standardized tests by type of test, Norm-referenced Assessment Program for Texas (NAPT) and Texas Assessment of Academic Skills (TAAS), and grade. Thus, N-5 under Cohort 1 refers to the fifth grade NAPT, while T-7 under Cohort 1 refers to the seventh grade TAAS.

In addition to student data, TSMP includes individual data for all Texas public school teachers for the same seven-year period. Currently we are able to link these teacher data to individual students at the campus, grade and program [bilingual, ESL (English as a Second Language), special education, gifted and talented] levels. As we discuss in greater detail below, in the future, we hope to link individual students to their specific teachers. Even without this valuable extension, we are able to complete educational histories of individual students for as long as they attend Texas public

[1] PEIMS is a yearly relational database and TEA makes no effort to link these data across years. To construct TSMP, we had to combine annual PEIMS teacher and student data with TAAS, NAPT, and various teacher certification tests that are not part of PEIMS, and link these data across years. As Appendix Table A-1 reveals, to create TSMP we had to combine data from more than 140 individual student files and more than 110 individual teacher files, as well as campus level data from TEA's AEIS files, block group data from the 1990 Census and district level data from the School District Data Book CD ROMs. The number of individual records included in TSMP exceeds 80 million.

Table 1. Total Students and Standardized Tests Included in the Texas Schools Microdata Panel by Cohort, Grade and Test (Eight Years of Data)

Year	Total Students (Enrollment)	Sem	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 5	
			Gr	Test/ Grade	Gr	Test/ Grade	Gr	Test/ Grade	Gr	Test/ Grade	Gr	Test/ Grade
89-90		F	3		2		1		K		PK	
89-90	1,161,358	S	3		2		1		K		PK	
90-91	1,505,551	F	4		3	T-3	2		1		K	
90-91	1,391,735	S	4		3		2		1		K	
91-92	1,420,295	F	5		4		3	T-3	2		1	
91-92		S	5	N-5	4	N-4	3	N-3	2		1	
92-93	1,415,593	F	6		5		4		3	T-3	2	
92-93		S	6	N-6	5	N-5	4	N-4 & T-4	3	N-3	2	
93-94	1,428,908	F	7		6		5		4		3	
93-94		S	7	T-7	6	T-6	5	T-5	4	T-4	3	T-3
94-95	1,438,632	F	8		7		6		5		4	
94-95		S	8	T-8	7	T-7	6	T-6	5	T-5	4	T-4
95-96	1,459,220	F	9		8		7		6		5	
95-96		S	9		8	T-8	7	T-7	6	T-6	5	T-5
96-97		F	10		9		8		7		6	
96-97		S	10	T-10	9		8	T-8	7	T-7	6	T-6

schools. Skillful use of these data should enable us to more accurately and effectively assess the performance of Texas schools than can be done with the fragmentary data that are currently available. These data should also allow us to better understand the causes of low student performance. If the required funding can be obtained, we will continue to follow individual students belonging to the current five cohorts until they have completed high school or dropped out, as well as add additional cohorts. The availability of data for more recent cohorts will enable us to assess the effectiveness of various ongoing school reform efforts, such as TEA’s accountability system and Governor Bush’s reading initiative.

After the students included in TSMP graduate from high school or drop out we hope to continue following those who attend Texas public colleges and universities into and through college, and both those who attend college and those who do not into the labor force. For dropouts, those graduating from high school and not going to college and those attending college, we hope to obtain earnings data from the Social Security Administration. While we have not yet begun discussions (the oldest students in TSMP are currently in the 10th grade), we plan to meet with representatives of the coordinating board about obtaining courses of study and grades for students included in TSMP. These data would enable us to study the relationships between the test scores of elementary-secondary school and their performance in college. Similarly, we would be able to use earnings data obtained from the Social Security Administration to determine the impact of elementary, secondary and college performance on earnings. There are well established procedures for obtaining earnings data for research purposes and a fairly large number of researchers have obtained and used these data to study the determinants of earnings. None has had the wealth of data on student achievement that would be available from TSMP.

While the TEA data are of unprecedented quality and extent, important gaps remain. Therefore, as time and funding permit, we plan to augment TSMP with information obtained from individual school districts. We have already held discussions with officials in 12 districts in the Dallas, Fort Worth, and Corpus Christi PMSAs. Two of them have already provided some data and three others have agreed to participate. We are continuing discussions with the remaining seven districts, have definite plans to meet with representatives of six other districts during 1998 and anticipate adding still others to the list. Data from TSMP will be used to examine a number of specific educational issues. As we add years and cohorts plus district specific data, TSMP's analytical usefulness and power will be greatly increased.

We are already engaged in research on two important areas, the impact of increased African American access to suburban schools on the performance of these children, and an assessment of special education. The Spencer Foundation has provided two grants totaling nearly $400,000 for this research. Spencer's grants supported the collection of data for the first five cohorts and eight years of data, as well as the difficult and time consuming effort of creating TSMP from TEA's disparate and unlinked annual data. 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 African-American, children.

In fall 1996 the Smith Richardson Foundation provided an additional $200,000 in funding for what we anticipate will be the first of several studies that will use data from TSMP for research on a variety of educational policy issues. In this study, Eric A. Hanushek (Rochester University) and Stephen Rivkin (Amherst College) will join Kain and other Green Center analysts in research on special education programs in Texas. Special education, which is the most rapidly growing segment of public education, has been subject to very little systematic analysis.

Early findings of the UTD Texas School Project are available in Kain and Singleton (1996). This paper, "Equality of Educational Opportunity Revisited" was published in the May/June, 1996 issue of the *New England Economic Review.* Originally prepared for a Boston Federal Reserve Bank conference held in Fall 1995, it compares the UTD Texas Schools Project to the Coleman Report published more than 30 years ago. It begins by documenting the dramatic declines that have occurred in the extent of racial/ethnic segregation in Texas public schools since the Coleman report was published. It also demonstrates, however, that the large gaps in mean achievement identified by Coleman et al. (1966) persist. Tables 2 and 3, which give mean z scores by race/ethnicity and grade for reading and math for 25 tests given to members of the five cohorts included in TSMP, provide further evidence of these differences.

Z scores are simply the ratio of the deviation of the number of correct answers for each student from the mean number of correct answers, for all students with meaningful scores, to the standard deviations of all students' scores. Use of z scores makes comparisons across tests with different numbers of questions possible and sidesteps most questions relating to norm referencing or to the differential level of difficulty of tests given in different years to different grades. The z score for each student indicates how well he/she did on a particular test relative to the average performance of all students taking the same test in the same year.

Table 2. Mean Reading z scores and Percent with Scores by ethnicity and Grade

Cohort	Year	Test	Grade	Native American	Asian	African American	Hispanic	Anglo
1	92	napt	5	0.03	0.47	-0.26	-0.40	0.36
1	93	napt	6	-0.04	0.46	-0.24	-0.29	0.30
1	94	taas	7	-0.06	0.35	-0.45	-0.35	0.32
1	95	taas	8	-0.04	0.34	-0.41	-0.34	0.31
2	91	taas	3	0.07	0.26	-0.28	-0.31	0.27
2	92	napt	4	0.00	0.35	-0.39	-0.36	0.35
2	93	napt	5	-0.01	0.29	-0.34	-0.39	0.34
2	94	taas	6	0.01	0.42	-0.42	-0.33	0.31
2	95	taas	7	0.05	0.39	-0.41	-0.34	0.32
2	96	taas	8	0.03	0.31	-0.41	-0.36	0.33
3	92	taas	3	0.02	0.31	-0.28	-0.31	0.28
3	92	napt	3	0.08	0.23	-0.36	-0.37	0.36
3	93	taas	4	-0.06	0.34	-0.45	-0.36	0.33
3	93	napt	4	0.02	0.26	-0.40	-0.40	0.37
3	94	taas	5	-0.03	0.41	-0.45	-0.30	0.29
3	95	taas	6	0.05	0.37	-0.44	-0.33	0.32
3	96	taas	7	0.05	0.32	-0.36	-0.32	0.30
4	93	taas	3	-0.02	0.30	-0.27	-0.32	0.29
4	93	napt	3	-0.08	0.22	-0.37	-0.41	0.37
4	94	taas	4	-0.07	0.39	-0.47	-0.27	0.27
4	95	taas	5	-0.03	0.42	-0.44	-0.27	0.26
4	96	taas	6	0.02	0.41	-0.40	-0.37	0.34
5	94	taas	3	-0.06	0.33	-0.43	-0.28	0.26
5	95	taas	4	-0.06	0.34	-0.49	-0.26	0.27
5	96	taas	5	0.01	0.35	-0.40	-0.27	0.25
Average All Tests				-0.01	0.35	-0.38	-0.33	0.31

While there are some differences among the 25 tests included in Tables 2 and 3, the relative magnitudes of the z scores by race/ethnic group are remarkably consistent across cohorts, grades and tests (NAPT vs. TAAS). As these data demonstrate, African

American and Hispanic students consistently have the lowest mean z scores in both reading and math, and Anglos (non-Hispanic, white) and Asians consistently have the highest. The mean scores for the relatively small number of Native American students are very close to the average of all students.

As the mean test scores in Tables 2 and 3 reveal, African American mean z scores for all 25 tests were –0.38 in reading and –0.40 for math. The Hispanic mean z score for reading, which is –0.33 for the 25 tests in Table 2, is slightly higher than the average score for African Americans. Nonetheless, the Hispanic reading scores are lower than those for blacks in eight of the 25 tests; these lower scores are concentrated in the earlier grades in each cohort suggesting many Hispanic students are able to overcome their initial disadvantage in English proficiency. Hispanic scores in math are higher than African American scores for all 25 tests. As we discuss in greater detail below these African American – Hispanic differences in mean z scores need to be interpreted with some care because much higher fractions of Hispanic than African American students are excused from taking these tests. These exempt students would in nearly all cases have low scores; if these low scores were included, the relative position of Hispanics would worsen in all tests.

The data in Tables 2 and 3 also support the widely held perception that Asian Americans are currently America's highest performing students. They have the highest mean scores on both tests in every grade and, again, conforming to the stereotype, their advantage is particularly large in math. The mean reading z score of Asian Americans for all 25 tests, of 0.35, is not very different from the mean score for Anglos, which is 0.31. In the case of math, however, the mean Asian score of 0.53 far exceeds the Anglo mean score of 0.30.

As noted above, significant fractions of the students who took the tests, reported in Tables 2 and 3, did not have their scores reported, either because they missed the test or because their tests were not scored. As the data in Table 4 reveal, the percentages of students without reading scores vary widely across racial/ethnic groups and grades.

Table 3. Mean Math z scores and Percent with Scores by ethnicity and Grade

Cohort	Year	Test	Grade	Native American	Asian	African American	Hispanic	Anglo
1	92	napt	5	0.03	0.69	-0.41	-0.33	0.34
1	93	napt	6	-0.07	0.48	-0.46	-0.34	0.34
1	94	taas	7	-0.08	0.59	-0.51	-0.32	0.31
1	95	taas	8	-0.05	0.63	-0.54	-0.39	0.36
2	91	taas	3	0.06	0.36	-0.39	-0.32	0.31
2	92	napt	4	-0.01	0.55	-0.46	-0.33	0.33
2	93	napt	5	-0.09	0.54	-0.42	-0.35	0.33
2	94	taas	6	-0.10	0.58	-0.51	-0.28	0.30
2	95	taas	7	-0.01	0.63	-0.57	-0.37	0.36
2	96	taas	8	-0.02	0.60	-0.50	-0.32	0.32
3	92	taas	3	-0.03	0.44	-0.35	-0.30	0.29
3	92	napt	3	0.05	0.46	-0.40	-0.31	0.31
3	93	taas	4	-0.01	0.53	-0.51	-0.29	0.30
3	93	napt	4	0.00	0.48	-0.41	-0.34	0.32
3	94	taas	5	-0.04	0.58	-0.51	-0.25	0.26
3	95	taas	6	0.02	0.56	-0.53	-0.33	0.34
3	96	taas	7	0.00	0.57	-0.51	-0.32	0.33
4	93	taas	3	0.00	0.41	-0.40	-0.26	0.28
4	93	napt	3	-0.12	0.42	-0.40	-0.33	0.31
4	94	taas	4	-0.08	0.57	-0.51	-0.23	0.25
4	95	taas	5	-0.02	0.54	-0.54	-0.22	0.26
4	96	taas	6	-0.04	0.53	-0.45	-0.28	0.29
5	94	taas	3	-0.10	0.46	-0.51	-0.26	0.27
5	95	taas	4	-0.07	0.52	-0.55	-0.24	0.26
5	96	taas	5	-0.02	0.53	-0.54	-0.20	0.24
Average All Tests				-0.03	0.53	-0.48	-0.30	0.30

Hispanics, many of whom arrive at school with limited English language skills, have the highest no-score rates for both tests in every grade. As these data indicate, these rates vary from a high of 24 percent for 5th grade reading to a low of 15 percent for 7th grade reading and math. Most of the Hispanic students without TAAS scores have been excused from taking TAAS because of an LEP (Limited English Proficiency) exemption.

If these excused LEP students were required to take the English language versions of TAAS, the mean scores for Hispanics would be much lower.

No-score rates for Native Americans, Asian Americans and African Americans are similar and substantially below Hispanic rates, while those for Anglos (non-Hispanic whites) on both tests and in every grade are much lower than those for any of the remaining four groups. The higher no-score rates of Hispanics and Asians, of course, reflect the large number of LEP excuses given to them. Hardly any Anglos or African Americans are LEP; in this cohort only one percent of blacks and 0.9 percent of Anglos were ever classified as LEP. The fractions for Hispanics and Asians are 47 percent and 43 percent. Finally, five percent of Native Americans were classified as ever LEP. In spite of the fact that the LEP fractions for Asians are nearly as high as Hispanic LEP fractions, their no-score rates are much lower. Asian no-score rates for the reading test, however, are considerably higher than those of Anglos, a fact that should be kept in mind when assessing the mean scores in Tables 2 and 3. The rates of non-test taking by race/ethnicity for the math test are similar to those shown in Table 4 for the reading tests.

While the evidence presented in Kain and Singleton (1996) on the large and persistent differences in test scores by race/ethnic group are important, the paper's principal contribution was in quantifying the extensive within district variations in selected school inputs that exist for campuses of varying racial/ethnic and family income composition. In contrast to Coleman et al.'s finding of no significant difference, Kain and Singleton (1996) found substantial within-district variations for four types of school inputs: teacher test scores, years of education, experience and class size (student-teacher ratios).

In January 1998, researchers associated with the UTD Texas Schools Project presented two papers, based on TSMP, at the Allied Social Sciences meetings in Chicago. Rivkin, Hanushek and Kain (1998) presented a paper, "Teachers, Schools and Academic Achievement," to a session of the Econometrics Society and Kain and O'Brien presented a paper, "How Much Has Moving to the Suburbs Increased African-American Educational Opportunities?" to a session of the American Economics Association.

Table 4. No-Score Rates by Cohort, Year, Test, Grade and Race/Ethnicity

Cohort	Year	Test	Grade	Native American	Asian	African American	Hispanic	Anglo
1	92	napt	5	6.2	5.2	8.4	7.9	4.6
1	93	napt	6	7.3	4.2	9.1	9.1	5.1
1	94	taas	7	11.5	10.4	14.6	19.0	7.9
1	95	taas	8	19.0	12.1	17.1	21.1	9.2
2	91	taas	3	8.9	11.4	9.3	7.8	6.6
2	92	napt	4	7.2	6.1	8.4	9.5	4.5
2	93	napt	5	9.6	5.6	8.6	9.4	4.8
2	94	taas	6	13.9	12.0	13.4	19.7	7.6
2	95	taas	7	15.1	11.4	14.0	18.9	7.4
2	96	taas	8	15.0	11.0	14.3	19.0	7.5
3	92	taas	3	12.3	14.4	9.6	7.9	7.2
3	92	napt	3	10.9	6.6	7.8	10.5	4.4
3	93	taas	4	18.9	15.1	15.4	24.0	10.2
3	93	napt	4	9.1	6.9	9.1	11.5	5.3
3	94	taas	5	15.0	17.0	14.1	24.7	8.7
3	95	taas	6	15.1	13.0	14.2	20.2	7.9
3	96	taas	7	12.5	10.6	12.6	16.5	6.2
4	93	taas	3	11.4	16.3	9.8	10.0	7.4
4	93	napt	3	7.6	6.1	7.9	12.3	4.6
4	94	taas	4	20.7	20.7	17.1	31.0	10.6
4	95	taas	5	16.7	18.6	15.5	26.5	8.7
4	96	taas	6	10.5	12.5	13.2	15.9	6.4
5	94	taas	3	14.5	21.0	13.4	31.1	8.4
5	95	taas	4	17.2	21.0	17.0	32.5	10.3
5	96	taas	5	11.3	16.8	14.1	20.6	6.9
Average All				12.7	12.2	12.3	17.5	7.1

In many respects the most striking finding of Kain and O'Brien (1998) was that in 1996 more than half of all black children enrolled in the 7th grade in the five largest Texas metropolitan areas attended suburban schools. As Table 5 from Kain and O'Brien (1998) reveals, this narrow black majority in suburban schools resulted from high 1990 levels of black suburban residency, rapid growth in the number of blacks attending suburban schools between 1990 and 1996 and a substantial decline in the number of black children in the five largest central city districts. Between 1990 and 1996, black suburban enrollment increased by 27 percent while black central city enrollment declined by 12 percent.

Using standardized test data for a single cohort of students attending Texas public schools between 1990 and 1996, Kain and O’

The concluding section of Kain and O'Brien (1998) presents the results of analyses in which the reading and math scores of individual black students are regressed on both individual characteristics and each of the three school quality measures. These analyses indicate that school quality (based on standardized test data for all students) has a substantial impact on the reading and math scores of individual black students. Combining these estimates with data on the mean differences in school quality for suburban and central city schools suggests that increased access to "better" suburban schools could have a large positive effect on black achievement.

"Teachers, Schools and Academic Achievement," is the first of several joint papers that we expect to produce from a Smith Richardson funded study that will use TSMP to study special education. In contrast to Kain and O'Brien (1998), which used a single cohort for its analysis, this paper uses 3rd, 4th, 5th and 6th grade math and reading scores for two cohorts of students (4th graders in 1993 and 1994) and 3rd, 4th and fifth grade scores for 4th graders in 1995. The use of multiple test scores for the same student provides an excellent way of controlling for student heterogeneity through a difference in differences framework.

While the results of Rivkin, Hanushek and Kain (1998) are preliminary, they are, nonetheless, a starting point in a line of research that is certain to yield important new results. The authors find that "substantial portions of the variation in individual performance between the 4th and 6th grades is the result of stable differences in achievement gains by schools and classroom" (Ibid., p. 28). The paper further concludes that the most significant component of these differences is due to "heterogeneity among individual teachers" (Ibid., p. 28). Exploratory estimates of educational production functions, moreover, indicate that commonly used school input measures capture only a small fraction of these differences. These provisional estimates do indicate, however, that "Class size appears to have a small but significant effect on student performance in fourth and perhaps fifth grade, and the effect appears to be stronger for children in low-income families." Refinements of these analyses, moreover, that are not included in the original version of the paper suggest even larger class size effects, and particularly for low-income children.

Uses of PEIMS/TAAS

There are at least four possible ways in which TAAS has been and can be used to increase the achievement of Texas school children and the performance of individual schools and districts. First, TAAS results for individual students may be used by classroom teachers to identify gaps and deficiencies in the skills and knowledge of individual students or entire classrooms and to develop instructional strategies to correct them. While some systems and teachers currently use TAAS in this way, it is almost certainly inferior to other assessment methods and tools for this purpose. Among its disadvantages is the fact that it is administered in the spring, rather than the fall or winter, when a diagnostic tool to assess individual progress would be more useful. This no doubt explains why TAAS was initially given in the fall and why many individual districts currently administer standardized tests to their students. While TAAS and other standardized tests given during the fall semester undoubtedly have some value as diagnostic tools, they are inferior to other, more individualized, methods of assessment that can be more closely tailored to the needs of individual students and classrooms.

The second use of PEIMS/TAAS data is to evaluate school and district effectiveness. According to Elmore, Ableman and Fuhrman (1996, p. 66), a survey conducted by the Council of Chief State Officers published in 1994 indicated that "twenty-eight states have included school-level test scores in public performance reporting." In Texas and in many other states these campus level scores are widely distributed and are frequently published in local and regional newspapers as soon as they are released. A 1993 survey, moreover, revealed that six states, including Texas, provided monetary rewards for improved student performance (Elmore, Ableman and Fuhrman,1996, p. 67).

TEA moved TAAS from the fall to the spring because the principal rationale for giving it was to assess the performance of Texas students overall, but, in particular, to assess the performance of individual districts and campuses. This use is embodied in TEA's Accountability Rating System for Texas Public Schools (TEA, 1994). Using TAAS passing rates for reading, writing and mathematics for non-special education students as well as the previous year's dropout and attendance rates, TEA designates individual campuses as Exemplary, Recognized, Acceptable or Low Performing. To obtain one of the top two ratings, both all students and each of the student groups African American, Hispanic, White and Economically Disadvantaged must meet each standard. Dallas ISD implemented what Clotfelter and Ladd (1996, p. 29) describe as "one of the most complete and sophisticated accountability and incentive programs of any big-city district in the country," for the 1991-92 school year. They observe further that the introduction of DISD's system was "facilitated by the statewide orientation toward testing and accountability, the absence of strong teacher unions that might have opposed it, the presence of competent statisticians and evaluators within the district," and other favorable factors that were particular to Dallas.

The third use of TAAS and PEIMS data is to assess the performance of individual teachers. Empirical studies by Hanushek (1971 and 1972), Murnane (1975) and Murnane and Phillips (1981) indicate that there are systematic and large differences in student achievement across classrooms that are attributable to differences in teacher skill or performance. These studies calculate mean residuals for each classroom/teacher from regression equations that attempt to hold constant the effects of student and family characteristics on individual student achievement. Citing evidence from these studies, Hanushek (1989, p. 48) states that "The findings … are unequivocal: Teachers and schools differ dramatically in their effectiveness. The formal statistical tests employed in these studies confirm that there are striking differences in average gain in student achievement across teachers."

Hanushek's conclusion, if correct, is of major importance. Yet there are two fundamental problems. First, Hanushek (1989, p. 48) adds the qualification that while "an important element of skill is involved in successful teaching … it is currently impossible to measure with any precision any readily identifiable components or elements of this skill." Worse yet, he adds that "it is unclear whether any form of teacher training could be organized to foster high levels of skill in teachers." Second, the evidence concerning teacher effectiveness is less strong than Hanushek implies. His claims are based on three small studies:

(a) a study of 1,000 third grade children in one California school district (Hanushek, 1971);

(b) a study of approximately 900 second and third grade Chapter I students attending 16 inner-city schools in New Haven (Murnane, 1975); and

(c) roughly 800 third, fourth, fifth and sixth grade students enrolled in a federally funded welfare reform experiment in a large Midwestern city (Murnane and Phillips, 1981).

The numbers of teachers and classrooms included in these studies are not provided, but they could not exceed 150 for all three studies combined.

TSMP is a nearly ideal tool to assess Hanushek's claims about 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 mean deviations from standardized achievement test scores 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 database. While we do not know the precise reasons for this omission, our sense is that it reflects, in part, concerns by individual teachers and their professional organizations that this information might in fact be used to assess the performance of individual teachers.

Teachers have reason for concern. According to Cohen (1996, p. 117), "Kentucky is one of the few states in which explicit rewards and punishments for professionals have been attached to student's performance," that this is "probably because the reform there was ordered by a court rather than devised through political bargaining in the legislature or executive branch." Within Texas, DISD has developed and used a sophisticated teacher assessment system that uses both TAAS and ITBS test results for individual students to assess the performance of individual teachers. These assessments, which are part of the previously discussed campus assessment program described by Clotfelter and Ladd, have not been used, thus far, in a formal way to reward or punish individual teachers, although principals are provided with these data for their campuses and apparently make some use of them.

A February 7, 1998, *Dallas Morning News* story provides a brief description of DISD's efforts to assess individual teachers. It reports that Robert Mendro, DISD's Executive Director of Institutional Research, "has been tracking student performance since 1993 to create annual ratings of half, or 4,500, of the district's teachers who teach reading or math to grades one to 12," adding "These educators teach about 95 percent of the 150,000 teachers in Dallas schools" (Wertheimer, 1998, p. 37A). According to the story, Mendro used 16 factors to net out the effects of student characteristics such as gender, race and socioeconomic status on individual student performance.[2] Calculating the mean residuals for each teacher's students, he used these means to assign DISD teachers to five categories, from least to most effective. Finally, he examined the impact of teacher assignments on the achievement of individual students. The same article quoted Acting Superintendent James Hughey as saying he "plans to take the research a step further and recommend that all schools start providing help to the least effective teachers." He gives as his reason the study's finding that "the longer a student has ineffective teachers, the worse the child does in school."

Mendro's findings, as reported by the *Dallas Morning News*, on the effects of assigning different quality teachers on the achievement of individual students are, to say the least, striking. Starting with two groups of students with above average reading scores in first grade, he finds that the average test scores of the group who subsequently had ineffective (category 1) teachers for three years in a row dropped by 41 points (from the 63rd to the 22nd percentiles). The mean scores of the second group of students, who had the highest rated teachers in 1994, average teachers in 1995 and the highest rated ones again in 1996 increased by five points, from the 62nd percentile in the first grade to the 67th percentile in the fourth grade.

The idea of using student achievement data in teacher assessment systems has a long history. The reactions to Mendro's study, as reported in the Dallas Morning News, however, illustrate the difficulty in actually using such a system to reward and punish teachers. Harry Hiscock, President of the Alliance of Texas Educators, was quoted as saying that "he had thought judging teachers on test scores was a closed subject in Dallas." He then elaborated saying, "You can't do it. … you cannot accurately rate teachers based on what students do on tests."

As Cohen (1996, p. 118) points out, "it is very difficult to devise and enact accountability schemes," because it is "difficult to devise new standards, tests, and information systems that are usable, effective, and professionally defensible." A major obstacle to developing accountability is insufficient information on individual students and their families. These data are required to separate the contributions of parents and communities to the performance of children enrolled in particular schools and classrooms from those of individual schools and teachers. The suggestions we offer below about additions to PEIMS are thus critical if the use of TAAS or other standardized tests to assess individual teachers and schools is to be seriously considered.

The fourth and final use for PEIMS and statewide-standardized test results is for research on the determinants of student achievement and on the effectiveness of various educational initiatives. Coleman et al. (1966), in fact, used data that are similar to those included in PEIMS/TAAS for their pathbreaking study. This is the use envisioned for TSMP and, not surprisingly, it is what we regard as the most appropriate and potentially most important use of PEIMS/TAAS data.

While TEA uses TAAS to assess individual campuses and districts, there are significant, and largely correctable, weaknesses of these data for this purpose. Indeed, with the possible exception of the use of TAAS to assist in the diagnosis of individual students, all four of the possible uses identified above have the same weaknesses. We now discuss a number of enhancements to the PEIMS/TAAS system and TSMP which would greatly increase their utility for these purposes.

---

[2] Mendro has agreed to provide additional documentation of his work, but it was not available at the time this paper was prepared.

Suggestions for Improvement

While TSMP is a remarkable tool, it is far from perfect. While we do not expect to ever attain perfection, we are making strenuous efforts to improve TSMP in several dimensions. First, we are continuing to add years. Our long-range plan is to follow students in the current five cohorts until they graduate from high school. Thereafter, we hope to obtain data on the college performance of those students who attend Texas public colleges and universities. Finally, in subsequent years, when high school dropouts, high school graduates and those attending college have entered the labor market, we plan to ask the Social Security Administration to supply us with earnings data for them. Obtaining these data will require TEA's cooperation because, while TEA has social security numbers for most of these students, we have only an encoded ID for confidentiality reasons.

As good as TEA's PEIMS and standardized test data 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. Individual districts already collect many of these data. As we discuss in greater detail in the final section, we have already contacted 12 individual districts in an effort to obtain some of these data from them. Negotiating with one district at a time, however, is inefficient. As a practical matter, the high cost of obtaining supplementary data on a district by district basis means that this form of data collection is necessarily limited to the largest districts, something that is a potential source of bias.

A more cost-effective approach to obtaining those data that are already collected by most districts would be to add them to PEIMS. TEA might develop uniform methods of data collection and reporting for these data and require all districts to provide them as part of their regular PEIMS submissions. In addition, as we discuss below, there are other high priority data, not routinely collected by districts currently, that could easily and cheaply be added to the system. These should be considered for inclusion in the PEIMS systems as well. We hasten to add that we fully understand that adding items to PEIMS or making other alterations is not a simple matter.

We also recognize that an effort by TEA to collect some of 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 in the state's capacity to assess the performance of its public schools would all greatly benefit from the collection of some or all of the data described below. The following proposals are listed by type of data, and within type of data by relative importance.

Testing in the Early Grades

The most expensive, but most important, improvement would be statewide standardized testing of first and second graders, and possibly kindergartners. Many districts within the state already routinely test first and second graders, but there is an urgent need for statewide testing. These data would strengthen the state's accountability system and provide districts, campuses and districts with the information they need to assess students and for timely intervention. Learning about both individual and campus low performance at the end of the third grade is simply too late. Both early and later grade tests should include all but a tiny fraction of children, including children in special education and LEP (Limited English Proficiency) programs. Of course, insofar as districts already test in the early grades, the net cost of a statewide testing would be proportionately less. If many districts object to early testing, the state could follow the federal government's example and make the testing of students in the early grades voluntary. At the same time the state should pay the entire cost of these early tests as an inducement for more districts to do early testing and to encourage those that already do to use the same tests.

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. Some of these data are already collected by schools, but are not submitted to TEA. Others would require revision of existing enrollment forms and data processing

As noted previously, research on the determinants of student achievement unambiguously demonstrates the importance of parent's education as a determinant of individual student achievement. 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 that might be good substitutes.

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 to do so and could be of great value. The best measure would probably be the mother's social security number because, in cases of separation or divorce, the children usually stay with the mother.

TEA obtains prior school attended for all students attending Texas public schools. Regrettably, they do not obtain these data for students transferring from private schools or from out of the state or out of the country. Adding this information to PEIMS would presumably be fairly easy and 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 go a long way towards permitting analysts to identify siblings currently attending Texas public schools, it would not provide information on the number and ages of children who either are too young to attend school or have dropped out or 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. Data on pre-school children would similarly be of great value to districts by providing them with data that would help them anticipate enrollment changes.

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 a regular data item for PEIMS would greatly increase TEA's capacity to monitor these programs and would be of great value in developing an improved understanding of the persistent achievement gaps of LEP children. If TEA were to mandate the 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 in terms of pull-out programs during the school day, after school programs and summer school programs. Yet the PEIMS system has no information on any of this activity, even though they are frequently paid for by state or federal money and usually are meant to increase student achievement. One explanation for the absence of data on these programs in PEIMS may be the difficulty of describing them in an economical, yet meaningful, manner. Nonetheless, a serious effort should be made to collect meaningful data for individual students on the types and quantities of remedial instructional services they receive within the school day, before school, after school or during vacations.

There is a growing awareness of the possible importance of pre-school experience in laying the groundwork for kindergarten and beyond. TEA currently obtains little or no information about the pre-school experience of Texas school children and it is our impression that districts generally know little more. These experiences are very likely of particular importance for children who come from immigrant and low-income families.

While obtaining relevant information on pre-school 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 the name and 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, albeit at some expense for geo-coding, from secondary sources such as Census block and 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 mother's years of schooling and data on family composition were added to PEIMS, the benefits for 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 tests. Course grades 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 Governor Bush's proposal to discontinue social promotions.

Student Teacher Links

If any doubts existed about Hanushek's (1997) conclusion that "Teachers apparently differ greatly in effectiveness," Mendro's (1998) findings, if they hold up under scrutiny and are replicated for other situations, would presumably largely eliminate them. The usefulness of standardized test data collected by TEA would be greatly increased if PEIMS student records identified each student's teachers, 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 for them. Of these the most important is much more precise research on the determinants of student achievement. The lessons learned from this research would provide valuable guidance to Texas educators in their efforts to devise cost-effective ways of improving student performance.

The empirical evidence supporting Hanushek's claims about the importance of individual teachers is, on close examination, not all that impressive. The numbers of teachers and classrooms included in the three studies identified by Hanushek could not have exceeded 150 for all three studies combined. These numbers pale in comparison with those used by Mendro (1998) in his research and, Mendro's numbers and generality are dwarfed by the more than 300,000 teachers and more than 6,000 campuses included in PEIMS. 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 individual 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 place too much emphasis on a single years performance as measured in this way. On the other hand, if the same teacher's students perform much better (after adjusting for differences in family background) for a number of years, it would be reasonable to believe that there is something about his or her or their educational practices that may be responsible for these consistently higher performances. After better and worse teachers have been identified, the next step would be to make an effort to determine just what makes them better or worse. If these traits or practices can be assessed for prospective teachers or taught, it may be possible to use this information to improve teacher training, classroom practices and teacher 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 the educational production functions surveyed by Hanushek 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.[3]

One reason these data are not collected is no doubt 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.

<u>Reading Curriculum and Instructional Practices</u>. Governor Bush's reading initiative has raised important questions about how reading should be taught, and particularly about the effectiveness of direct phonics instruction vs. whole language. If these differences are as important as the protagonists in this debate allege, it should be possible to find supporting evidence in the TAAS. 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.

---

[3] In a 1997 survey of close to 400 studies of student achievement Hanushek (1997, p. 141) finds "there is not a strong or 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). Similarly discouraging results were obtained for teacher education, teacher experience, teacher salary and expenditure per pupil..

Information on the methods and materials used could be obtained from individual classroom teachers, and the burden need not be great. Specifically, elementary school teachers who teach reading might be asked to complete a simple form describing their primary approach to reading instruction (whole language vs. phonics) or, in the case of those who use a mixed strategy, the fraction of time spent using each method. Elementary school teachers might also be asked to indicate how much use they make of reading groups and ability grouping, and the amount and types of seatwork they assign. 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 at this time 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 in some instances may make very little use of the textbooks themselves. 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 teachers were identified on PEIMS student records.

<u>Bilingual/ESL Programs</u>. The performance of LEP (Limited English Proficiency) children lags badly behind those raised in English-only households. The relationship of their poor performance to the state's bilingual education program is unclear. Critics of bilingual education contend that in all too many instances bilingual programs fail to prepare their 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 knowledge 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, the data on bilingual instructional practices could be used in combination with the 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 which 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.

Contacts with Individual School Districts

An alternative way of obtaining some of the added data identified above is to obtain them from individual districts. During the past year, we have held discussions with officials in 12 individual school districts in the Dallas, Ft. Worth and Corpus Christi PMSAs. We have learned a great deal from these conversations that have increased our understanding of these complex data. Two examples of these many lessons are described below.

each student has unique ID's. Nonetheless, it is fairly clear that their efforts have been less than 100 percent effective, and particularly in the early years. One piece of evidence for this assertion is the fact that every year we find students with identical Alt PID's among the enrollment and attendance files they provide. Try as we would, we were unable to come up with a plausible explanation. Discussions with one of the local analysts we met with provided what, with the benefit of hindsight, is a simple and persuasive explanation. He observed that, in too many times to count, mothers with more than one child give their social security numbers instead of those of their children when they enroll them in school. Careful error checking, first by the individual districts and then by TEA, obviously catches most of this type of error, but it is clear that a few occurrences slip through.

More recently, in analyzing data in anticipation of a visit to an unnamed district (District X), we found that the number of "out-of-sample" 9^{th} grade students was implausibly large. Out-of-sample students are students included in PEIMS in the current year, who were not there in the previous year. The numbers of ninth grade out-of-sample students was so large that we were convinced we had made some kind of a programming mistake. Incidentally, this was the first year in which TSMP included any 9^{th} grade students and they were in cohort 1, one of the two boundary cohorts.

In spite of spending huge amounts of time searching for coding errors, we were unable to come up with a plausible explanation. We decided to ask the PEIMS coordinator at District X, whom we were meeting the next day, if he could explain the bulge. He immediately advanced two explanations, neither of which had occurred to us. First, he pointed out that 9^{th} grade was different than the elementary school grades we were used to working with. Unlike pre-K to 8, where students are either promoted to the next grade or not, 9^{th} graders only become 10^{th} graders when they pass four courses. If they fail just one course they remain 9^{th} graders. Since 9^{th} graders who failed one or more class the previous year, were 9^{th} graders in that year, they were not included in any of the five cohorts that are currently included in TSMP. Thus, they were out of our sample in the previous year and appeared as an unusually large number of out-of-sample 9^{th} grade students in the following year.

Second, he pointed out that large numbers of students transfer from private to public schools in the 9th grade. The reason is that there are many more elementary/middle school seats in private schools than high school seats. If PEIMS included the previous school of these private to public school transfers the extent and nature of these students could be easily determined.

While we value the many insights we have obtained from analysts and other officials at the local districts we have visited, the primary object of our meetings has been to identify the kinds of data they routinely collect and maintain on students and teachers that would improve TSMP. In these discussions, we have emphasized identifying and obtaining three kinds of data. Our highest priority has been early (kindergarten, first and second grade) tests for students in our five cohorts. The reason is obvious. The earliest statewide tests are given in the third grade, but children's experiences in the early grades are important, perhaps decisive.

As we have discussed above, the greatest weakness of PEIMS, TAAS and thus TSMP is the sparse information they provide on individual student characteristics and family background. Currently individual student and family background variables included in PEIMS/TAAS are sex, age and race/ethnicity of individual students, whether these students are LEP (Limited English Proficient) and receive reduced price or free lunches and whether they are enrolled in special programs (ESL, bilingual education, special education, gifted and talented, Chapter 1, and career and technology). As analyses included in Kain and O'Brien (1998) demonstrate, these variables are powerful predictors of student achievement. Obtaining early tests would greatly reduce the severity of problems associated with missing family background data by providing a better measure of the skills children possess when they enter school, but would not eliminate them. As a result, after early tests our priority has been to obtain data that provide better family background measures.

Among family background variables, previous studies have consistently found that parents', and particularly mother's, education is the strongest predictor of children's success in school. Unfortunately, we have thus far found only one district that routinely obtains information on mother’s education as part of the enrollment process and thus far they have refused to provide us with any data. Needless to say, we do not plan to take “no” for an answer, as obtaining these data for even a single school district would enable us to assess the importance of the absence of mother’s education in the larger analysis. Richardson ISD collected these data until a few years ago, but a new manager of their data processing department removed it from the enrollment forms when he found it was never used.

In addition to parents’ education, we have tried to obtain street addresses and data that would permit us to identify siblings. This would enable us to strengthen controls for family background by matching the addresses to Census block and block group records and possibly appraisal district files. If we could identify siblings, moreover, the very large samples available from TEA would enable us to examine the effects of birth order on student achievement and to use family fixed effects to provide better estimates of the contributions of schools, as opposed to families, on individual student achievement. Districts differ in terms of both their capacity and willingness to provide these data, but we are thus far greatly encouraged by our discussions about their possible availability.

We are mindful of the fact that this paper will raise questions about confidentiality. In this regard we would emphasize that we have no interest in the identities of individual students and teachers. The development of TSMP and the multivariate research that is based on it require accurate student and teacher identifiers to create a useable multi-year panel database, but the methods we use mask the identities of individuals.

Conclusion

TSMP, with its five cohort, eight-year panel of individual student, teacher and school data, provides an unparalleled opportunity to accomplish meaningful assessment of the effectiveness of school resources, organization, programs, and practices on student achievement. Initial research results are very promising, and include an analysis of central city and suburban school quality, the effect of teachers and class size, and the changes in patterns of segregation, suburbanization, and 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 this 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

Clotfelter, Charles T. and Helen F. Ladd. 1996. "Recognizing and Rewarding Success in Public Schools, Elmore, Richard F., Charles H. Abelmann, and Susan H. Ruhrman. 1996. "The New Accountability in State Education Reform: From Process to Performance," in Helen F. Ladd (ed.), *Holding Schools Accountable*. Washington, D.C.: The Brookings Institution: 65-98.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data," *The American Economic Review*, vol. 61: 280-88.

Hanushek, Eric A. 1971. *Education and Race: An Analysis of the Educational Production Process*. Cambridge, MA: Heath-Lexington.

Hanushek, Eric A. 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," Journal of Human Resources 14, no. 3 (Fall): 351-88.

Eric A. Hanushek. 1989. "The Impact of Differential Expenditures on School Performance," <u>Educational Researcher</u> (May).

Hanushek, Eric A. and John F. Kain. 1971. "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," in *On Equality of Educational Opportunity*. Edited by Frederick Mosteller and Daniel P. Moynihan. New York: Random House, Inc..

Hanushek, *Eric A. 1981. "Throwing Money at Schools, Journal* of Policy Analysis and Management, Vol. 1, No. 1, 19-41.

Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance," *Educational Researcher* (May).

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.

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

Kain, John F. And Daniel M. O'Brien. 1998. "Has Moving to the Suburbs Increased African American Educational Opportunities," Paper prepared for the meetings of the American Economic Association Meetings, Chicago, Illinois, January 5.

Murnane, Richard. J. 1975. *The Impact of School Resources on the Learning of Inner City Children.* Cambridge, MA: Ballinger.

Murnane, R. J. and B. Phillips, 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research*, 10(1), 83-100.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain. 1997. "Teachers, Schools and Academic Achievement," paper presented at the meetings of the Econometric Society, Chicago, Illinois, January 4, 1998.

Rossell, Christine and Keith Baker. 1996. "The Educational Effectiveness of Bilingual Education," *Research in the Teaching of English* 30 (no. 1): 7-74.

<u>TEA (Texas Education Agency). 1994. *Accountability Manual: The 1994-95 Accountability Rating System for Texas Public Schools and School Districts*, Office of Policy Planning and Evaluation (April).</u>

Wertheimer, Linda K. (1998). "Teachers Rated on Students' Performance: Some Oppose Evaluations Based on Tests," Dallas Morning News, February 7: 37A & 42A.

Table A-1. Data and Files Included in the Texas Schools Panel Data Base (TSPDB)
(Data for the 1990-97 School Years, Eight Years of Data)

File Types	Years	Files	Total Records
Student			
PEIMS Enrollment	8	8	11,248,135
PEIMS Chapter I Enrollment	8	8	4,839,000
PEIMS Special Ed Enrollment	8	8	1,475,460
PEIMS Voced Enrollment	3	3	1,123,375
PEIMS Voced Serv	1	1	466,580
PEIMS Gifted Enrollment	6	6	333,250
PEIMS Summer Demographic	5	5	8,018,816
PEIMS Drop - Pre	2	2	15,773
PEIMS Drop	2	2	15,255
Course Completion	1	1	2,718,419
PEIMS Basic Attendance	4	24	33,905,285
PEIMS Special Ed Attendance	4	24	6,952,295
PEIMS Voced Attendance	4	24	2,169,001
TAAS	6	14	5,133,855
NAPT	2	7	1,963,368
TEAMS	1	1	286,982
Total Student Files/Records		138	80,664,849
Teacher			
PEIMS Staff	8	8	2,462,424
PEIMS Employment	7	7	2,462,424
PEIMS Payroll	7	7	4,017,232
PEIMS Class	7	7	8,076,668
PEIMS Nonclass	7	7	374,719
PEIMS Permit	7	7	83,008
TECAT	5	1	145,711
ExCET	5	64	292,696
TOPT	5	1	4,894
PPST	5	1	54,125
TASP	5	1	32,032
Total Teacher Files/Records		111	18,005,933
Total Student plus Teacher Files/Records		249	98,670,782