A Longitudinal Assessment of Reading Achievement: Evidence from the Harvard/UTD Texas Schools Project

John F. Kain and Daniel M. O'Brien*

The Green Center for the Study of Science and Society
The University of Texas at Dallas

July 15, 1999 (Revised)

Prepared for presentation at "Achieving Universal Literacy." Conference sponsored by the Cecil and Ida Green Center for the Study of Science and Society, The University of Texas at Dallas, April 2-4, 1998.

We are grateful to our team of programmer/research assistants, Karen Chaddick, Ankur Gupta and Jaison George for their crucial contributions to this paper and to Jean Kingsley for her editorial assistance and help in manuscript preparation. We also thank the Spencer Foundation and Smith Richardson Foundation for their financial support.

A Longitudinal Assessment of Reading Achievement: Evidence from the UTD Texas Schools Project

Introduction

In January 1996, Texas Governor George W. Bush announced his reading initiative, setting forth the "goal of having all students reading on grade level by the end of grade 3 and continuing to read on grade level throughout their schooling" (Moses, circa 1996). There is no disputing the fact that large numbers of children in Texas public schools are not acquiring the basic literacy skills that they will need to function effectively in an increasingly complex and technological society. This paper documents the extent of reading deficiencies among Texas public school students and, in particular, quantifies the large and persistent gaps between high income Anglo children and other groups, particularly low income, Limited English Proficient (LEP), learning disabled and various ethnic and racial minorities.

The paper also describes the history of standardized testing in Texas and the development of the Texas Education Agency's (TEA's) highly regarded accountability system. Initiated in 1993, this evolving, complex system is meant to assess the performance of individual districts and campuses (TEA 1994, 1997a). One important problem any such system faces is how to deal with the huge differences in socio-economic status among districts and campuses. If a consistent finding exists about student achievement, it is that the characteristics of a child's parents and family are of paramount importance. While we do not dispute the usefulness of TEA's current accountability system, its treatment of district and campus differences in socio-economic composition could clearly be improved.

At the same time, research on student achievement and other questions relating to the performance of students, teachers and schools may be an even more valuable use of the Texas Assessment of Academic Skills (TAAS) and Public Education Information Management System (PEIMS). One important contribution of this paper is its suggested extensions and modifications of TEA's data collection, which would greatly increase the value of these data for accountability and for research on educational policy issues. Since a number of states either have or are developing similar data bases, these suggestions presumably have widespread relevance.

The History of Testing in Texas

The first extensive achievement test data for Texas came from U. S. Army tests of 103,500 white recruits and 19,000 Negroes during World War I. As these data clearly demonstrated, the performance of Texas's education system lagged far behind the average for the nation as a whole. A comparison of state medians for white recruits provided by Bond (1924), for example, revealed that Texas ranked 36th out of 45 states included in the analysis. The 10 that ranked below Texas were all southern or border states. If Negro recruits educated in the South's separate and unequal black schools had been included, these comparisons would have been even more unfavorable to these states.[1]

The earliest use of standardized tests by a Texas State agency was the administration of the ACT to a sample of Texas high school students in 1968 by the Governor's Committee on Public School Education.[2] Two years later, TEA conducted a follow up study, the Texas Achievement Appraisal Study, which tested 69,000 high school students. In the following year (1972), TEA carried out its first large-scale state assessment using criterion-referenced tests, when it tested 22,000 sixth grade students in reading and mathematics. Then in 1975, TEA used the Texas Career Education Measurement Series (CEMS) to examine 26,000 students in 16 areas of career education, and in 1978 it gave 94,000 grade six and eleven students tests in reading and mathematics. Finally, in 1979, it tested 50,000 Grade 5, 8 and 11 students in consumer and career education.

The Texas Assessment of Basic Skills (TABS), a criterion-referenced test in reading, writing and mathematics, given in spring 1980, could reasonably be considered the origin of TEA's accountability system. This test, which was given to 433,000 students in grades five and nine, broke new ground by making campus and district results

---

[1] The Army's Alpha data were also used as the dependent variable in one of, if not the, earliest educational production function studies. Using an index of the quality of state school systems developed by Leonard P. Ayers, Bond obtained a correlation between this school quality index and a state's median Alpha score for white recruits of 0.74, indicating that this school quality index explained nearly half of the variation in Alpha scores.

[2] We are indebted to Keith Cruse, Senior Director, TEA for providing us with this history of early testing in Texas.

available to the public. During 1981-1985, TABS was extended to all students in grades 3, 5 and 9; approximately 675,000 students were tested in each of these years.

In 1986, six years after TABS was first implemented, TEA replaced it with the Texas Educational Assessment of Minimum Skills (TEAMS), which it gave to all students in grades 1, 3, 5, 7, 9 and 11. Roughly 1.4 million students per year were tested during this period. Ferguson's (1991) widely cited study of district variations in student achievement in Texas was based on TEAMS. In a departure from past practice and a precursor to future policies, all students were required to pass the TEAMS tenth grade exit level to graduate from high school. The steady expansion of testing in Texas in the previous decade and the decision to make passing the exit TEAMS a requirement for high school graduation may have contributed to what appears to be a steady growth in student achievement during this period (Texas State Board of Education 1990).

Both TABS and TEAMS assessed minimum skills. In 1991, TEA introduced a more demanding, criterion-referenced test, TAAS, which it initially gave in the fall semester to students in grades 3, 5, 7, and 11. The Texas State Board of Education (1991, p. 15) observed that the new tests "measure reading, writing, and mathematics at a more rigorous level." The 1992-93 school year was a transitional year for TAAS. TEA gave the test to some grades in the fall and some in the spring. Since then, it has given TAAS to students in grades 3-8 plus the 10^{th} grade in the spring (TEA, 1997b, p. 2).

In 1990 the state legislature "directed the State Board of Education to adopt one nationally recognized norm-referenced test to be administered annually to students in grades 4, 6, 8 and 10" (State Board of Education and the Texas Education Agency, 1995, p. 20). State law also required TEA to obtain national norm-referenced results for all of the grades that were to be tested by TAAS. TEA responded to these requirements by administering the Norm-Referenced Assessment Program for Texas (NAPT), a variant of the Iowa Test of Basic Skills (ITBS), to all students in grades 3 through 11 in 1992 and 1993.

The April, 1993 administration of NAPT revealed that Texas public school students in grades 3, 4, 5, 10 and 11 performed somewhat above the national average (State Board of Education and the Texas Education Agency, 1995, p. 20). Further analysis indicated that the performance of Texas school children improved between 1992 and 1993 in five of the nine grades tested. With the decision to move TAAS to the spring and to include all students in grades 3-8, TEA discontinued NAPT.

The decision to stop giving NAPT left Texas without a statewide norm reference test that could be used to compare the performance of Texas school children to children in the rest of the country. National Assessment of Educational Progress (NAEP) data permit limited comparisons. Using data from the 1992 NAEP, the State Board of Education and TEA (1995, p. 21) found that Texas fourth-grade students earned an average score of 214 on the NAEP reading assessment, which was slightly above the basic level score. In addition 53 percent scored at or above the assessment's basic level, 20 percent achieved at the proficient level, and 3 percent attained the advanced level.

As originally conceived, TAAS precluded comparisons with students in other states and across grades. In response to the second failing, the Texas legislature in 1992 required that student achievement, as measured by TAAS should "be comparable across grade levels." TEA, as a result, developed "the Texas Learning Index (TLI) ... to provide information about student growth over time" (TEA, 1995, p. 11). TLI aligns "the passing standards at grades 3 through 8 in reading and mathematics ... with the passing standard of the exit level tests."[3]

Table 1 shows mean TLI scores for 1994 through 1997, as well as 1994-97 gains, for all students, and for various categories of students, taking the tests in grades 3,4,5 and 10. These data suggest Texas students have made significant gains in reading achievement over the past four years, as TLI scores improved for all 24 grade/categories.[4] The smallest improvement was 0.09 of a standard deviation (1.3 TLI points) for Anglos in the third grade; the largest was 0.45 of a standard deviation (6.7 TLI points) for African Americans in the tenth grade. Even with this progress, the mean 1994 exit TLI for African Americans remains nearly 0.6 of a standard deviation (8.4 TLI points) below the mean Anglo score.

[3] The TAAS tests administered in 1994 serve as a baseline for the TLI. For each year and each subject TEA determines the number of questions that a student must answer correctly in order to have achieved the same level of achievement as the 1994 passing score. For each test the z score based on the number of questions which the student answers correctly is transformed by assigning a TLI score of 70 to the number of right answers required to pass, then centering TLI on this score with a standard deviation of 15 (TEA, 1997a).

[4] This finding has been disputed by a recently completed independent assessment of the difficulty level of TAAS in various years. Sandra Stotsky (1998), in a recently completed analysis of the fourth, eighth and tenth grade TAAS reading tests, found "that the tests given from 1995 through 1998 were not of comparable difficulty to each other in any of the grade levels tested." She concluded that "The 1995 tests are longer and more difficult than the 1998 tests at all grade levels."

Governor Bush's goal that all children should be reading at grade level by the third grade could be interpreted as requiring that students pass the third grade TAAS. TLI indexes for 1997 suggest substantial progress during 1994-97. At the same time, they also demonstrate that a major challenge remains in bringing all children to grade level in reading.

Using TSMP Data to Analyze Reading Performance

The multivariate analyses presented in this paper are based on TAAS tests for grades 3-7 given to students who are members of a single cohort during 1990-1996. The TAAS and other data used in the analyses are from the Texas Schools Microdata Panel (TSMP), a multi-year panel data base that has been developed by the UTD Texas Schools Project. TSMP includes up to eight years of data for five cohorts of students attending Texas public schools. It begins in the 1990-91 school year because TEA implemented PEIMS in that year. As long as these students attend Texas public schools we are able to follow them as they move from one school to another.

The cohort used in this analysis is the middle (Cohort 3) of the five cohorts currently included in TSMP. Eight tests, six TAAS and two NAPT, were given to this cohort during 1990-97. The analyses in this paper are based on data for five of the six TAAS tests. TSMP includes extensive student data starting in the first grade, but, as noted previously, statewide tests in Texas are not given until the third grade.

Cohort 3 includes 422,863 unique students. Of these, we have one or more tests for 350,974 of them and a non-zero score for 345,038. To maximize the number of students with multiple tests, we center the cohort on grade 3, the first year standardized tests are given. While 64 percent of the students in Cohort 3 were present for the entire seven year period included in this analysis, significant numbers were absent from the sample for one or more years. Students may be absent because they had not yet moved to Texas or were attending private schools or because they had moved out of Texas or had transferred to private schools. An unknown, but apparently small, percentage of missing students may not be missing at all, but instead have more than one ID. These students appear in the panel as two or more students.

Because TLI is unavailable before 1994 and cannot be used to compare TAAS and NAPT results, we use z scores for the multivariate analyses. 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 and the standard deviation 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 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.[5]

The analysis is complicated by the fact that third grade Spanish speaking LEP

[5] The z scores used in this paper differ from those reported previously by Kain and Singleton (1996) and Kain and O'Brien (1998). In those papers we excluded all students that were "exempt" from the TEA accountability system or had zero scores. For this paper, we include all students with positive scores, even if they were exempted from the State's accountability system. This decision adds just over four percent more students to the analysis.

students in 1992 were generally able to take a Spanish language version of TAAS. In that year, more than 14,000 third grade students took the Spanish TAAS; all but 91 were Hispanic. The scores of these Spanish language test takers are not included in the overall means used in calculating individual z scores for students taking the English language TAAS. Third grade Spanish TAAS z scores, based on the mean for all students taking the Spanish TAAS, are included as explanatory variables in the fourth grade value added regressions.

Assessing the Reading Performance of LEP and Special Education Students

Significant fractions of LEP and special education students are exempt from TAAS for accountability purposes. Approximately 2.7 percent of students received exemptions from TAAS in spring 1997 because of limited English proficiency and between 5.7 and 6.2 percent received special education exemptions. The 1997 fraction of all students with LEP exemptions is considerably lower than in earlier years because a Spanish language version of TAAS was given in grades 3-6 in that year. In prior years, only a third grade version of Spanish TAAS was available.

We attempt to deal with the problems of assessing LEP and special education students by defining three sub-samples: ever LEP, ever special education and neither. The ever LEP and ever special education sub-samples are not mutually exclusive. The ever LEP sub-sample includes students who were classified as special education in one or more years while the ever special education sub-sample includes students who were classified as LEP in one or more years. In contrast, there is no overlap between the neither sub-sample and the first two sub-samples. The number of observations for the neither sub-samples vary from 152,178 (grade 4) to 157,955 (grade 6). The ever LEP sub-samples vary from 25,954 (grade 4) to 34,272 (grade 7). The ever special education sub-samples vary from 27,746 (grade 3) to 32,098 (grade 7).

The fractions of students who are LEP or special education for each test and grade are shown in columns 4 and 5 of Table 2. On average, they are very similar, as evidenced by the bottom row which displays averages across all tests; the LEP average is 10.6 and the special education average is 11.6. The LEP fraction tends to decline somewhat as grade level increases. This decline reflects the tendency for students who were initially classified as LEP to lose this designation as their English Language skills improve. In comparing third grade NAPT and TAAS rates, it should be recalled that TAAS was given in the fall while NAPT was given in the spring. In addition, there was no Spanish language NAPT. The percentage of students who are designated special education exhibits little or no trend. The relatively low percentage (9.7 percent) for the third grade TAAS is probably attributable to the fact that it was given in the fall, while the third grade NAPT was administered in the spring.

Columns 6-8 give mean z scores for each of the seven tests given to Cohort 3 students by category. As noted previously, these scores are relative to the mean of all students with non-zero scores. They demonstrate that even when large numbers of students are excluded from the calculations, the differences between LEP and special education students and those who were neither LEP nor special education are very large. As the means across all seven tests in the last row indicate, the unweighted mean z score for all students that are neither is .15. In contrast, students who were ever LEP scored nearly one standard deviation below the average student and students who were ever special education scored about 0.6 of a standard deviation below the average student. The frequency distributions of z scores for sixth graders belonging to these three groups are shown in Figure 1.

The last three columns in Table 2 give the fractions of each of the three student groups that were either not scored or had a reading score of zero. The large difference between grade 3 and subsequent grades for LEP reflects the availability of a Spanish language TAAS in the third grade. The higher no-score rate for the third grade NAPT than for the third grade TAAS is similarly attributable to the Spanish language TAAS.

Special education no-score rates are much higher than the LEP rates for all seven tests. The means across all tests were 28 percent for LEP students and 42.8 percent for special education students. Comparing the tests, the NAPT fractions without scores greater than zero were also higher than the same statistic for TAAS in both years. This difference, 57.8 percent for the fourth grade TAAS and 31.6 percent for the fourth grade NAPT, no doubt reflects differences in policies about whom should be tested. NAPT's higher testing rate in both years is reflected in lower mean scores for NAPT in both years.

The last column in Table 2 indicates that very few students who are neither LEP nor special education fail to take and obtain a score greater than zero on these tests. The overall no-score rate is only 2.4 percent. There is, once again, a difference between NAPT and TAAS; NAPT rates are always lower.

Policymakers in Texas are committed to including more LEP and special education students in the state's accountability system. Recent legislation requires TEA "to develop and propose a system for evaluating the progress of students eligible for exemption for the statewide assessment program under current law" (TEA, 1996a, p. iii). In 1997, TEA added a Spanish TAAS for Spanish speaking LEP students in grades 4-6 who otherwise might have been exempted from TAAS because of limited English proficiency (TEA, 1998). TEA is hard at work developing a "multi-grade level set of TAAS tests using items from several grade level tests" that could be used for special education students whose performance is too far off grade level to take the regular TAAS (TEA, 1996a).

Trends in Mean Reading Scores by Race/Ethnic Group

Regardless of which test (NAPT vs. TAAS) or grade is used, the mean z scores in Table 3 are remarkably consistent across race/ethnic groups. Starting with the averages for all tests in the last row, Asian Americans, whose scores were .49 of a standard deviation above the mean for all students, had the highest reading scores overall. Anglos, with an average z score for all tests of .41 did second best, followed by Native Americans with an average z score of .16. The overall average z scores for Hispanics and African Americans at -.10 and -.30 were the lowest. In considering both Hispanic and Asian American scores, it should be kept in mind that all students who were ever LEP are excluded. If not, Anglo reading scores would have been much closer to Asian American ones and African American scores would have been much closer to those of Hispanics. Indeed, as estimates reported in Kain (1998) demonstrate, when LEP students with scores greater than zero are included African American scores for this cohort exceed or are equal to Hispanic scores on five of the seven tests.

Multivariate Analyses of Reading Achievement

For the multivariate analyses reported in this paper, we estimate two kinds of equations. So-called level equations are estimated for the third grade, which is currently the earliest statewide testing of Texas public school students, while what economists refer to as "value added" or gains equations are estimated for grades four through seven. It should again be noted that the tests administered to fourth through seventh graders were given in the spring while the third grade TAAS was given in the fall. The third grade TAAS thus actually measures achievement at the end of the second grade minus any losses that occurred over the summer that may have varied by income level (Heyns 1978; Entwisle and Alexander 1992, 1994, 1997; O'Brien 1999a). Except for grade differences the same dependent variable, reading z scores, is used in both the level and value added regressions.

The principal difference between these two types of equations is that the value added equations include one or more prior test scores among their explanatory variables. For students who were not in the sample or did not take TAAS in the previous year, we predict the prior year's score from the prior year's level equations. Since these scores take account of each student's prior achievement level, the value added equations measure the achievement gains during the year. The remaining explanatory variables quantify net effects, holding constant the influence of the prior test score(s) on the current reading score.

Value added equations are usually preferable to level equations because they quantify each starting point and control both for prior educational experiences and the prior effects of family background on achievement. Level equations are estimated for the third grade because there is no prior test score for all students. The third grade level equations are included to provide something of a baseline. They quantify the difference in achievement levels by race/ethnicity, household income and other variables that affect student achievement. In the value added equations estimated for grades 4 – 7, the effect of these and various omitted variables are embodied in the lagged or prior test scores.

Both the level and value added equations are estimated with campus fixed effects. Use of campus fixed effects, a differencing procedure, is equivalent to including a dummy variable for each campus. The procedure purges the parameter estimates for individual students of any campus level influences. We estimated both the level and value added equations without fixed effects and found the individual parameter estimates and R^2s are very similar to those obtained in the fixed effects equations. The equations estimated without fixed effects may be obtained from the authors.

The Level Equations

Table 4 contains level regressions for third grade students for the ever LEP, ever special education and neither sub-samples. The ever LEP equations in Table 4 are limited to Asian American and Hispanic students, who are 97 percent of all ever LEP students. The fraction of explained variance (coefficient of determination or R^2) varies between 19 percent for the neither equation and 27 percent for the ever LEP equation. Table 4 also reports the number of campuses represented in the regression. Ever LEP students are more geographically concentrated than those students who are neither LEP nor special education. Evidence of this geographic concentration is provided by the fact that ever LEP students are found at only two-thirds as many campuses as students who are neither LEP nor special education.

The first twelve explanatory variables quantify the joint effects of race/ethnicity and household income/poverty on individual reading scores. The household income variables are based on the eligibility of individual students for free or reduced price lunches under the federal school lunch program. Eligibility is based on federal definitions of the poverty level and thus depends on both family income and family size. To receive a free lunch, a child must be a member of a family whose income is less than 135 percent of the poverty level for its size. Similarly, to receive a reduced-price lunch, family income must be between 135 and 185 percent of the poverty level. Students whose families receive AFDC benefits or who participate in a number of other poverty programs are also eligible for a free lunch. In subsequent discussions, we use the terminology high, low and very low income households in referring to these categories.

The 12 income/race/ethnicity dummy variables in Table 4 are obtained by multiplying the three household, zero-one, income variables times the zero-one race/ethnicity dummy variables. Because there are relatively few observations for Native Americans and Asian Americans, we combine their low and very low income categories to create a single low income category. Thus, there are only two income/race/ethnicity variables for Native Americans and Asian Americans. Even though we calculate three income variables in calculating income/race interaction variables for the remaining three race/ethnic groups, the equations include only two income/race variables for Anglos. This is because one of the dummy variables must be omitted in order to estimate the equations. In this and similar cases, the omitted dummy becomes the base case and the coefficients of the remaining categorical variables are interpreted relative to it. The base case for these variables is high income Anglos.

The relatively small sample sizes for Native American and Asian American students also affect the statistical significance of the coefficient estimates obtained for the income/race/ethnicity variables. Of the nine coefficients for Asian Americans or Native Americans, only three have t statistics larger than 1.8. Two are for the neither equation, which has much larger samples, and one is in the ever LEP equations.

Turning to the neither equations, there is no difference, holding the effects of all other variables constant, in the third grade reading performance of high income Asian American, Native American or Anglo children. Income, however, has a pronounced effect. The third grade reading scores of Anglo children from low income families are 0.14 of a standard deviation lower than Anglo children from high income families and the difference for Anglo children from very low income families is -0.23 standard deviations. This same pattern exists for Anglo children who were ever special education and the relative magnitudes are even larger.

African Americans consistently have the largest negative income/race coefficients and they exhibit the same kind of income gradient. The results for Hispanics indicate that they consistently outperform African Americans in both the neither and ever-special education equations. In assessing the neither results, it should be kept in mind that Hispanic children with limited English proficiency are not included.

The next five variables are Census zip code variables that we include in an effort to provide additional controls for differences in family background. As impressive as TEA's data are, they provide much less information about the family background of each student than would be desirable. This feature of PEIMS/TAAS somewhat reduces their value for either research on student achievement or for the closely related task of assessing the performance of individual districts, campuses and teachers. While we would by no means argue that the sparse family background information for individual students invalidates TEA's accountability system, it clearly makes it vulnerable to criticism and more difficult to defend. The same can be said for teacher assessment schemes, such as that being developed by the Dallas public schools, that rely on PEIMS/TAAS or similar data at the district level (Jordon, Mendro and Weerasinghe, 1997).

We follow Ferguson and Ladd (1996) and Webster et.al. (1997) in using aggregate variables from the 1990 census as proxies for individual family background variables. Hanushek, Jackson and Kain (1974) demonstrated nearly 25 years ago that while micro data are always preferable, aggregate data are usually better than no data at all. The Census variables used in this paper should be interpreted as crude estimates of the probability that each student has a particular trait. Again, following Ferguson and Ladd, we assign Census zip code data for each student's campus to individual students. Since we have up to seven years for most students, we use data for the year/campus that is closest to 1990 in all regressions. These zip code variables are thus traits that remain with students when they move to other campuses. This variation and the fact that, as discussed below, the variables are also race/ethnicity specific enables us to include these campus level variables in campus fixed effects equations.

We assign individual African American, Anglo, Asian American, Hispanic and Native American students the zip code value of each race/ethnic specific variable. Two types of zip code variables are used. The first is the proportion of families that are female-headed. The second describes the proportion of zip code adult residents that were college graduates, had some college or were high school graduates.

The means of the zip code variables vary greatly by race ethnicity. The mean percentage of female-headed households was thus only 4.7 percent for Anglos as compared to 22.4 percent for Native Americans and 18.3 percent for African Americans. There are also large race/ethnic differences in the zip code percentages of college graduates. Among third graders, Asian Americans, have the highest rates of college graduation (40.3 percent), while Anglos (22.9 percent) come in a distant second. In interpreting this result, it should be kept in mind that Asian Americans who were ever LEP are not included in this sample of students who were neither LEP nor special education. At the same time, the zip code estimates are for all families or persons.

In spite of their primitive nature, the family background proxies used in these and subsequent value added regressions have substantial predictive power. The coefficient of the proxy for female-headed households is negative for all three equations and is statistically different from zero in the neither and ever LEP equations, although the negative coefficient in the ever LEP equation is implausibly large. Six of the eight coefficients of the percent college graduate and some college variables are positive and four are large and statistically significant. The t statistics for both variables in the ever LEP equation, however, are vanishingly small .

The college graduate and some college proxies thus play an important role in both these equations and in the value added ones that we present subsequently. The household income/school lunch variables appear to do an adequate job of quantifying the socio-economic status of low and very low income households. They are much less adequate for high income households because the variation in the incomes and other measures of socioeconomic status for this category is simply much greater. While not ideal, the zip code proportions of college graduates and some college provide some differentiation for households belonging to this highly heterogeneous category of households.

The coefficient estimates obtained for the zip code variables are obviously, and unsurprisingly, much less consistent and have much lower statistical reliability than those for the individual income/race/ethnicity variables. Nonetheless, they provide further evidence of the importance of family background and strongly suggest the wisdom of routinely collecting these and other family background variables when parents register their children for school.

The next five variables are all specific to individual students. Gender, which takes the value one for boys and zero for girls, is highly significant statistically in all three equations and confirms the widely held view that, holding everything else constant, third grade girls are better readers than boys. The age variable is also highly significant statistically and is negative in all three equations. The effects, however, are small except in the special education equation.

The next four variables, which are dummy variables for LEP and special education, do not appear in the neither equation because all ever LEP and ever special education students are excluded. Both ever special education and special education in the current year (special education now) are included in the LEP equation; the sign of ever special education is negative and its coefficient is -0.63. When the small positive coefficient is subtracted from it, it appears that the reading scores of LEP students who were ever special education are more than 0.6 of a standard deviation lower than otherwise identical LEP students who were never LEP. Ever LEP and LEPNOW also have offsetting signs in the ever special education equation. In this case the current year variable is both larger and more significant statistically. Combining these two coefficients, special education students who were ever LEP have third grade reading scores that are 0.12 of a standard deviation below an otherwise identical third grade special education student who was never LEP.

The next two variables, years retained in grade and years double promoted, as expected, have opposite signs in all three equations. Of the two, years retained has a larger effect on reading and much higher statistical reliability. The size of the years retained coefficient is particularly large in the neither equation. It indicates that the reading scores of 3^{rd} graders who have been retained in grade are more than 0.5 of a standard deviation below those of other neither 3^{rd} graders holding constant the effects of all other variables. It is tempting to interpret this result as a condemnation of Governor Bush's proposal to end social promotions. It is probably more correct, however, to interpret it as signaling serious learning or behavioral problems for a small number of students. The size of this coefficient for the two remaining groups is smaller, but still significant. The size of the double promote coefficients are similar in all three equations, ranging from 0.08 to 0.11.

The coefficient of the days absent variable is highly significant in all three equations and its value, -0.01, is the same in all three equations. The standard deviation of average days absent is about six days; evaluating this figure using –0.01 indicates a one standard deviation increase in days absent produces roughly 0.06 of a standard deviation decrease in the reading score.

The next five variables are all measures of student mobility between the end of the 2nd grade and the beginning of 3rd grade. While these variables also appear in the value added equations, their definitions are slightly different for this analysis. The reason is that the mobility variables for grades four through seven are calculated from attendance data and identify moves made between the sixth attendance period in the previous year and the first attendance period in the current year. Most of these moves occur during the summer. Since individual attendance data were not included in PEIMS until 1993, the mobility variables included in the third grade level equations are based on enrollment data. In contrast to attendance data, which provide a record for every student even if they were in attendance for a single day, enrollment data are a snapshot and only include students if they enrolled at a particular campus on the particular day the enrollment data are compiled. We also include dummy variables for the number of years each student is in the sample (not shown).

The coefficients of all five mobility variables are negative in all three equations. Since the omitted category, which is used as the reference point, is no move, these results indicate that for third graders, each type of mobility have an adverse effect on achievement. These estimates suggest that district-to-district moves have a small, but statistically significant negative impact on achievement and that the impact is somewhat larger for ever LEP and ever special education students than for students that were neither LEP or special education. In the case of ever special education students, a district-to-district move reduces their reading score by 0.09 of a standard deviation, relative to otherwise identical third grade special education students.

We distinguish between two kinds of within district campus-to-campus moves. Voluntary moves most often arise from within district residential moves, but they less often may result from parents' dissatisfaction with a particular campus. Transfers are situations where all of the students attending a particular campus move en masse to another campus. An example would be graduation from elementary school and entry into middle school. Because of the almost unlimited combinations of grades found in Texas schools and because the grades provided at a particular campus may differ across programs, transfers are difficult to define. The resulting estimates suggest that both types of within district moves negatively affect student achievement, although for the third grade, at least, the effects of transfers are quite small and are not statistically different from zero. Voluntary within district moves have a much larger negative impact on achievement.

The variable "into sample" identifies students who were not present in TSMP in the previous year during the fall enrollment snapshot when these students were in the third grade, but were present for both the fall third grade snapshot and for TAAS. In nearly all cases these would be students who moved from other states or transferred from private schools. Out of sample, in contrast, are students who were in TSMP fall snapshot in the previous year, were not present in fall enrollment snapshot in the current year, but had reappeared by the time TAAS was given in the spring. The coefficients of both of these mobility variables are negative and highly significant statistically in all four equations.

Percent no score, which is calculated separately for each of the three samples, is the fraction of students who are enrolled in particular district at the time the exam is given, but do not have a positive score on the reading portion of TAAS. This variable is meant to account for differences in district policy toward testing, and particularly differences in the testing of exempt LEP and special education students. Of the three coefficients, only the one in the ever special education equation is statistically different from zero. It indicates, as might be expected, that testing fewer special education students increases average reading scores at that campus even when the effects of all other included variables are held constant.

The Value Added Equations

The specifications of and explanatory variables included in the twelve value added equations are nearly identical to those used in the level equations. The four equations for never LEP and never special education students are shown in Table 5. The addition of prior reading scores is the most important difference between these equations and the third grade level equations presented above. The huge t statistics for the previous year's English language TAAS reading score in all four equations speaks to the importance of including them. The smaller t statistic obtained for this variable in the fourth grade equation may reflect the fact that the third grade TAAS was given in the fall. As a result 19 months, instead of 12 months, elapsed between the third and fourth grade tests. Even with the difference in timing and other factors, the coefficient of the prior English reading score in all four equations varies in the narrow range from 0.61 to 0.64.

The fourth grade equations include two prior reading scores because some students took the Spanish rather than English language version of TAAS in the previous year. It is not surprising that the ever LEP and ever special education samples include children who took the Spanish language TAAS. It is somewhat more surprising that the neither sub-sample does. Admittedly, the number is relatively small, 200. This is consistent with our experience - TSMP includes at least one of everything; in this case it includes 200. While the coefficient for the Spanish TAAS is not significantly different from zero (t = 1.5), it is positive and about half as large as the coefficient for the prior English language TAAS.

The third explanatory variable, Peer Read, is the mean prior year reading score of “other” never LEP, never special education students enrolled at each students campus/grade. “Other” refers to the fact that a particular student’s score is excluded in calculating his or her mean peer reading score. The Peer Read variable does not appear in the ever LEP and ever special education equations. The reason is that, for these students, the value of the variable is the same for all students at a particular grade/campus. This means it is a campus/grade variable for these students, precluding its use in campus fixed effects equations. The coefficients of the Peer Read variable are highly significant in all four equations, although the effects are relatively small, ranging from 0.12 in the fourth grade to 0.02 in the fifth.

The mean values of the variables included in the fifth grade neither regressions can be used to assess the magnitude of this peer effect. One indicator is obtained by multiplying the difference in the means of Peer Read for African Americans and Anglos (-0.81- .321 = -.402) times the peer coefficient, 0.12. This calculation yields the result that the fourth grade reading score for a representative African American student would be about -.05 of a standard deviation below that of an otherwise identical Anglo student because of the lower average reading scores of the other children enrolled at his campus/grade. This result, which is qualitatively consistent with findings by Kain and O'Brien (1998) using a somewhat different approach to the same question, would be smaller for the remaining grades.

The second block of explanatory variables gives coefficient estimates for the twelve household income/race/ethnicity dummies, which are incredibly consistent across both income/race categories and grades. First, all but five of the 48 coefficients are negative and only one of the positive coefficients, high income Asian American fifth graders, is significantly different from zero. It indicates that the reading scores of grade five, high income Asian American students (excluding those that were ever LEP) exceeded those of high income Anglos by 0.06 of a standard deviation. Second, there is only one case where the income gradient within race/ethnicity groups is inconsistent. Even in this case, the difference was small and not statistically significant. Students from lower income families consistently have lower scores and lower gains from year to year.

The magnitudes of the individual coefficients across race/ethnicity categories are equally consistent and conform to the pattern of overall means by race/ethnicity in Table 2. Relative to high income whites (the omitted category) African Americans consistently have the largest negative values. The largest negative value, -0.44 is obtained for very low income black children in grade 4 and the coefficients for high income African Americans in the same grade were nearly as negative. Of the five race/ethnicity groups Hispanics consistently had the next largest negative coefficients (in absolute value).

As in the third grade level equations, the family background proxies used in the neither value added regressions have substantial predictive power. The most consistent of the three education variables, moreover, is again the proportion college graduates. All four coefficients are positive with values ranging from 0.32 (grade 3) to 0.18 (grade 5) and the smallest t statistic is 3.5, which also is for the same grade. All four of the some college coefficients are also positive, although generally much smaller than those for college graduates and all but the grade seven coefficient is statistically significant. The female-headed percent of families with school age children has the expected negative sign in three of the four equations, but only one of the positive coefficients is statistically different from zero at the five percent level.

The coefficients for gender are negative in all but the sixth grade equation. We have no good explanation for this result or for the huge negative and highly significant coefficient for the seventh grade equation. Given the overall consistency of these results, it is tempting to conclude that the sixth grade test was biased against girls and that the test makers decided to make up for it by writing a seventh grade test that favored girls in the following year. The age variable is highly significant statistically and negative in all four equations. The effects, however, are small. The next two variables, years retained in grade and years double promoted, as in the level equations have opposite signs. Of the two, years retained in grade has a larger effect on reading and much higher statistical reliability. Except for grade 4, when its coefficient is nearly a third of a standard deviation for each year retained, the coefficient is about half this amount. Of course, the earlier effects do not disappear; they are embodied in the lagged reading scores. The effects of double promotions, 0.18 for the fourth grade and 0.02 in the seventh, decay over time. Again, however, the earlier effects are presumably embodied in the prior test scores in the later years. As the results in Table 7 clearly show, absenteeism has a significant effect on student achievement. These effects are highly significant statistically, and the magnitudes are fairly small. The standard deviation of average days absent is about 5.2 days; evaluating this figure using the coefficients in Table 7 indicates a one standard deviation increase in days absent produces between 0.03 and 0.04 of a standard deviation decrease in reading scores.

As promised the value added equations have one more mobility variable than the third grade level equations. The added variable, within year moves, varies between zero and six. While students could move more frequently, we allow only one move per attendance period. The reason is the difficulty of distinguishing within period moves from concurrent attendance at more than one campus. The coefficient of the number of within year moves is highly significant in all four equations, with values of between -0.04 (grade 5) and –0.08 (grade 6).

Of the between year moves, between district moves have no effect in the fourth grade, but they have a significant negative affect on student performance thereafter. Both types of within district moves also negatively affect student achievement and, somewhat surprisingly, transfers appear to have a larger negative effect than voluntary campus-to-campus moves. We are uncertain about how to interpret this result. The last two mobility variables, which are based on attendance periods, differ slightly from those included in the level equations. For the value added equations the Into Sample variable identifies students that were not in the sample during the sixth attendance period in the previous year, but took TAAS in the current year. Out of Sample students, in contrast, were in TSMP during either the sixth attendance period in the previous year or the first attendance period in the current year, but took TAAS in the current year.

In third grade level equations, the coefficient estimates for the percent no score variable are positive. In the value added equations, the coefficient was positive in the fourth grade and negative in grades 5-7. Not too much should be made of this difference, however, as the t statistic is small. The negative coefficient for grades 5-7 indicates that individual reading scores are somewhat lower in districts that test a smaller percentage of their students.

LEP and Special Education

Tables 6 and 7 present the value added regressions for four ever LEP and four ever special education samples. With a couple of exceptions, the explanatory variables are the same as those included in the neither equations. The sample sizes, though much smaller than for the neither equations, are still sizeable, with more than 25,000 students in each year/grade. As with the level equations, the ever LEP equations are estimated for only Asian and Hispanic students. The coefficients of determination (R^2) for all eight equations vary between 0.42 (fourth grade ever LEP) and 0.59 (seventh grade ever special education). The variances explained by the ever special education equations are always larger than for the comparable ever LEP equations.

In all eight equations, prior reading scores (both English and Spanish) strongly affect current reading scores. Even with samples that are about one-sixth as large as those for the neither equations, the t statistics are huge. All but one of the English test coefficients have t statistics that are greater than 100 and the exception is 78. Even the Spanish test coefficients in the fourth grade equations have t's of 33 and 13. The coefficient estimates for English tests, moreover, range from .45 for the fourth grade LEP to .70 for the seventh grade special education equation. The coefficients for the Spanish TAAS are .44 (ever LEP) and .55 (ever special education). The latter is actually larger than the coefficient of the lagged English score in the same equation.

The income/race/ethnicity variables generally show the same pattern as those found in the neither estimates, albeit with less statistical significance. For ever LEP students, all other 4^{th} grade students have lower gains than the omitted category, high income Asian students. However, many of the coefficients are not statistically significant. This may be due to several factors, including the small number of ever LEP students who are not eligible for free or reduced rate lunch. For ever special education students, gains for Asian high income students are similar to those for high income Anglos students. The familiar pattern of negative and statistically significant coefficients for low income and minority students is evident. Low income and minority special education students have smaller year to year gains than high income Anglo students in each year/grade.

Only three of the twelve schooling proxies are statistically different from zero at the five percent line in the four special education equations and two of the twelve in the ever LEP equations. The ineffectiveness of the years of schooling proxies in the ever LEP equations is hardly surprising. Many, if not most, of these parents would have been educated in their home countries. Translating their years of schooling to the U.S. context is problematic. In addition, the zip code variables refer to all Native American, Asian

American, African American, Hispanic or Anglo residents of the zip code rather than only LEP residents. Only one of the female-headed coefficients (seventh grade ever special education) are statistically significant, although all are negative. The results for the remaining variables, with few exceptions, are similar to those previously discussed for the neither equations. Indeed, the similarities between the equations for the three groups are remarkable.

Suggestions for Improving TEA's Data Collection

While PEIMS and TAAS are exceedingly valuable tools for assessing the overall performance of Texas school children and of individual districts and campuses, as well as for research on the determinants of student achievement, they are far from perfect. Their usefulness for both accountability and research could be greatly improved at modest cost. Based on our experience in developing TSMP and using PEIMS/TAAS data for research on student achievement, we offer the following suggestions for augmenting and improving them and for creating similar data bases in other states. Many of the data we identify for inclusion are already collected by individual districts or could easily and cheaply be added to their routine data collection activities.

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

TEA obtains prior school attended for all students transferring from another Texas public school. Inexplicably, they do not obtain these data for students transferring from private schools or from out of the state or out of the country. These data would be of great value in linking the records of individual students over time and could be used to answer a number of important questions about the movement of students between private and public schools and the impact of migration on student achievement.

The age and relationship of all family members should be routinely collected annually when children enroll in school. Districts should also be required to provide the scores LEP children obtain on the English Proficiency tests that are used in determining their assignments to bilingual, ESL and regular programs and to determine whether particular children should be moved to regular classrooms. These scores would greatly increase TEA's capacity to monitor programs for LEP children and would be of great value in understanding of the persistent achievement gaps of LEP children.

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. TEA (1998, 1-21) reports that "districts must offer intensive instruction in either reading, writing, mathematics, or a combination of these subjects to between 21 percent and 35 percent of the students tested at each grade level in Grades 3-8." Yet PEIMS contains no information on any of this activity, and we know of no other systematic reporting of information that describes the extent and nature of these special instructional activities in either the aggregate or for individual students. 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 importance of pre-school experience in laying the groundwork for kindergarten and beyond. TEA obtains little or no information about the pre-school experiences of Texas school children, and it appears that individual districts know little more. Texas and other states should at minimum collect, for each child, the pre school name and months and hours of attendance as well as more generic descriptions of the nature and extent of more informal forms of out-of-home child care.

Street addresses are routinely obtained by individual districts and could easily be added to PEIMS. These data could be used to identify residential moves, to assign Census block group data to individual students for additional family background controls and to identify siblings when a family ID is not collected.

If TEA collected and included links that identified each student's specific teachers in its PEIMS data base, it would be possible over several years to obtain incontrovertible evidence about the relative contributions of individual teachers and campuses to individual student achievement. Ideally, PEIMS student records would identify 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. Of these, the most important is research on the determinants of student achievement. The lessons learned from this research would provide valuable guidance to educators in Texas and elsewhere in their efforts to devise cost-effective ways of improving student performance. Kain (1999) provides a more detailed discussion of the use of standardized tests to assess teacher effectiveness.

TEA collects and maintains a great deal of information about individual students and teachers and the performance of individual students 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 (1997) 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.

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. The following example for reading instruction is but one case where states should obtain systematic information on curriculum and instructional practices.

The recent National Academy of Sciences report on reading and a large number of earlier studies and reports have 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 differences in reading gains, and particularly if the results of early tests were obtained by TEA. 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 and the burden need not be great. Specifically, elementary school teachers who teach reading should be required to each year 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.

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.

Summary and Conclusions

To achieve universal literacy, educators, policy makers and researchers will need to work together to develop tools to assess each child's reading and writing abilities, to determine the best methods of instruction and to devise systems to evaluate the effectiveness of alternative instructional methods. This paper addresses three issues that are central to achieving universal literacy: the use of assessments, analyses of the contribution of individual and school characteristics to reading achievement and improvements that might be made to TEA's TAAS/PEIMS system.

The paper begins with a review of the history of assessment by the Texas Education Agency. TEA's use of these tools and related reforms appear to have produced steady improvements in the performance of Texas public school children in both reading and other subjects. Given Texas' apparent success, it may provide useful lessons for other states that have already begun, or are considering, statewide testing of public school students and the use of these data to assess the performance of districts and campuses.

There are large differences in the performance of learning disabled and limited English proficient children and children that were never classified as LEP or learning disabled. The mean reading scores of fifth grade students who were neither LEP nor special education, for example, were 0.72 standard deviations higher than fifth grade students who were ever LEP and were 0.65 standard deviations higher than fifth grade students who were ever assigned to special education classes. If the large numbers of LEP and special education students who were exempt from TAAS, or did not have positive scores, were somehow included the differences would be much greater.

The equations presented in the paper quantify the effects of individual student characteristics, such as gender, family income, race and household income on reading performance. The results are remarkably consistent across years and category. The results obtained for the grades 4-7 value added equations provide strong evidence of the cumulative nature of education. The previous year's reading score has a huge and statistically significant effect on reading performance in the following year.

In the case of students that were neither LEP nor special education, the rankings by ethnic group are consistent across all five equations (third grade level equations plus grades 4-7 value added equations. African Americans and Hispanics, holding constant the effects of all other variables included in the equations, including differences in school inputs, perform much less well than Asian Americans and Anglos. In all pairwise comparisons, moreover, Afican American scores are below those for Hispanics. In interpreting this result, it should be kept in mind that the large number of Hispanics who were ever LEP are not included in the neither LEP or special education sample. If they were, the African American vs. Hispanic differences would be much smaller.

Within race/ethnic groups the effect of household income is incredibly consistent. Comparing African Americans included in the third grade neither LEP nor special education equation to Anglos in the same income category, the scores of high income African Americans are 0.30 of a standard deviation below those of high income Anglos. Similarly, the difference between low income African Americans and low income Anglos is 0.20 of a standard deviation and the difference between very low income African Americans and very low income Anglos is 0.29 of a standard deviation. The analyses also indicate that all types of mobility, within year, district-to-district between year moves and two kinds of between year within-district campus-to-campus moves all have large negative impacts on reading performance.

In spite of their high quality, the TAAS/PEIMS data include less information on family background than would be desirable. Therefore, the regressions also include family background proxies obtained from 1990 Census data for each campus' zip code. While these variables are less precise than individual versions of the same variables would be, they nonetheless have fairly large and statistical effects on individual reading scores. These results provide strong evidence of the need to obtain more complete family background data.

While the statistical analyses included in this paper are highly informative and confirm earlier findings about the importance of family background on the performance of individual children, they also reveal gaps in TEA's impressive TAAS/PEIMS system. Therefore, the final section of the paper focuses on possible improvements that would both strengthen the state's accountability system and increase the power of these data as a tool to assess educational policies and practices.

The most expensive, but most important, of the several proposed improvements would be statewide standardized testing in the early grades. The econometric analyses in this paper and earlier research on the determinants of student performance consistently identify family background as the most important determinant of student achievement. These findings provide further support for the view that TEA and other state departments of education should require districts to collect and provide more extensive data on the characteristics of students' families. We assign the highest priority to mother's education (years of school completed). In addition, we urge TEA to begin including a family ID for each student in PEIMS and begin obtaining the name and location of the prior school attended for students who transfer from schools in other states and from private schools. Finally, we urge TEA to obtain and include in PEIMS an enumeration of all household members by sex, age and relation, the scores of LEP children on English proficiency tests, and information on after school and summer remediation programs and on pre-school attendance to PEIMS.

In addition to earlier testing and more family background data, we also feel strongly that statewide data bases should identify each student's teachers along with the subject(s) taught and contact hours in PEIMS. The emphasis to date on collecting these data has been on teacher assessment. An even more important use is research on student achievement. It is widely believed that "teachers matter," but the systematic evidence, though growing, is still limited. Data that link individual students could be used to begin to see how much difference individual teachers make and ultimately if combined with information on teacher characteristics and their educational practices could have a huge impact on public education.

In many respects the most serious gap in TEA's data collection is its failure to collect systematic information on curriculum and educational practices. Campuses and classrooms remain black boxes. The absence of information on curriculum and educational practices may account for the widespread failure of educational productin function studies to obtain positive and statistically significant relationships between student achievement and various school input measures, such as class size, teacher education, and teacher experience.

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. We contend, however, that developing such instruments would not be impossible and these data would be far more valuable than other information that is currently collected. The paper uses reading instruction to provide an example of the kinds of curricular and teaching practices data that should be collected from individual teachers during each school year.

References

Bond, Horace M. 1924. "What the Army 'Intelligence Tests' Measured," *Crisis* (July). Reprinted in Russell Jacoby and Naomi Glauberman (ed.), *The Bell Curve Debate: History, Documents, Opinions.* New York: Times Books, 1995: 583-598.

Entwisle, Doris R. and Karl L. Alexander. 1992. "Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School." *American Sociological Review*, Vol. 57 (February: 72-84).

Entwisle, Doris R. and Karl L. Alexander. 1994. "Winter Setback: The racial Composition of Schools and Learning to Read." *American Sociological Review*, Vol. 59 (June): 446-460.

Entwisle, Doris R., Karl L. Alexander, and Linda Steffel Olson. 1997. *Children, Schools, & Inequality*. Boulder, Colorado: Westview Press.

Ferguson, Ronald P. 1991. "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal on Legislation* 28 2 (Summer): 456-465.

Ferguson, Ronald P. and Helen F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools," in Helen F. Ladd, *Holding Schools Accountable*. Washington, D.C.: The Brookings Institution: 265-298.

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. 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 Jackson and John F. Kain. 1974. "Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy," Political Methodology 1, No. 1 (Winter).

Heyns, Barbara. 1978. *Summer learning and the effects of schooling*. New York: Academic Press.

Hoxby, Caroline Minter. 1996. "The Effects of Private School Vouchers on Schools and Students," in Helen F. Ladd, *Holding Schools Accountable*. Washington, D.C.: The Brooking Institution: 177-208.

Jordan, Heather R., Robert L. Mendro and Dash Weerasinghe. 1997. "Teacher Effects on Longitudinal Student Achievement, A Report on Research in Progress," Presented at the CREATE Annual Meeting, July, Indianapolis, IN.

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

Collection," Paper presented at the 12th Annual Texas Assessment Conference, February 15-18, 1998, Austin, Texas.

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, Chicago, Illinois, January 5.

Moses, Mike. Circa 1996. "Forward to Beginning Reading Instruction: Components and Features of a Research-Based Reading Program, Texas Reading Initiative," Texas Education Agency. 1996.

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.

O'Brien, Daniel M. 1999. "Family and School Effects on the Cognitive Growth of Minority and Disadvantaged Elementary Students," Paper prepared for the meetings of the American Education Finance Association, Seattle, March 18-20.

State Board of Education and the Texas Education Agency. 1995. "Texas Public School Education: Making the Grade, 1992-94," A Biennial Report to the 74th Texas Legislature (Spring).

Stotsky, Sandra. 1999. "Analysis of the Texas Reading Tests, Grades 4, 8 and 10, 1995-1998." Independent report prepared for the Tax Research Foundation, Houston, Texas (November).TEA (Texas Education Agency). 1993. *Snapshot '92: 1991-92 School District Profiles*. January, Publication Number GE3 410 03.

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).

TEA (Texas Education Agency). 1995. *Snapshot '94: 1993-94 School District Profiles*, (winter).

TEA (Texas Education Agency). 1996a. "Assessment System for Special Education Students Exempted from the Texas Assessment Program: A Report to the 75th Texas Legislature from the Texas Education Agency."

TEA (Texas Education Agency). 1996b. "Assessment System for Limited English Proficient Students Exempted from the Texas Assessment Program: A Report to the 75th Texas Legislature from the Texas Education Agency."

TEA (Texas Education Agency). 1997a. "Texas Student Assessment Program Technical Digest for the Academic Year 1995-1996."

TEA (Texas Education Agency). 1997b. "Expanding the Scope of the Texas Public Accountability System," *Policy Research*, Report Number 9, June.

TEA (Texas Education Agency). 1998. "1997 Interim Report on Texas Public Schools: A Report to the 75^{th} Texas Legislature." www.tea.state.tx.us, March 31, 1998.

Texas State Board of Education. 1991. "Quality, Equity, Accountability: Long-Range Plan for Public Education 1991-1995," Austin, Texas, March.

Webster, William J., Robert L. Mendro, Timothy H. Orsak and Dash Weerasinghe. 1997. "Comparison of the Results Produced by Selected Regression and Hierarchical Linear Models in the Estimation of School and Teacher Effect," Presented at the annual meeting of the American Educational Research Association (AERA), Chicago Illinois, (March 24-28).

TABLE 1. Mean TLI (Texas Learning Index Scores by Grade, Race/Ethnicity, LEP and Economic Disadvantaged
(Excludes All Exempt Special Education and LEP)

	1994	1995	1996	1997	97-94	1994	1995	1996	1997	97-94
All Non-Exempt Except Special Education						Hispanic				
Grade 3	78.2	78.0	78.6	79.7	1.5	74.0	73.8	74.7	75.8	1.8
Grade 4	78.4	80.1	79.9	80.9	2.5	74.3	76.5	75.8	77.1	2.8
Grade 5	78.8	79.9	81.6	83.8	5.0	74.2	75.5	77.3	79.6	5.4
Grade 10	77.7	77.8	80.0	82.1	4.4	71.7	71.9	74.3	76.8	5.1
African American						LEP				
Grade 3	71.7	71.5	71.9	74.1	2.4	68.7	69.8	71.9	73.0	4.3
Grade 4	71.2	73.2	72.9	74.7	3.5	68.2	71.0	70.5	71.3	3.1
Grade 5	71.9	72.7	75.0	77.9	6.0	65.4	66.9	69.0	71.3	5.9
Grade 10	71.4	71.1	75.1	78.1	6.7	58.3	58.7	58.7	63.1	4.8
Anglo (Non-Hispanic White)						Economically Disadvantaged				
Grade 3	82.2	82.0	82.7	83.5	1.3	73.2	72.9	73.7	75.1	1.9
Grade 4	82.6	83.9	84.1	84.9	2.3	73.3	75.4	74.7	76.1	2.8
Grade 5	83.2	84.3	85.8	88.0	4.8	73.3	74.5	76.3	78.9	5.6
Grade 10	82.9	82.9	84.6	86.5	3.6	70.5	70.9	73.3	76.0	5.5

Table 2. Mean Reading z Scores for Ever LEP, Ever Special Education and Neither LEP nor Special Education by Grade and Test (English Tests Only) for Students in Cohort 3

			Percent		Mean Z Score			Percent Not Scored or Zero Score		
Year	Test	Grade	LEP	Spec Ed	LEP	Spec Ed	Neither	LEP	Spec Ed	Neither
92	TAAS	3	12.3	9.7	-0.66	-0.37	0.12	9.8	48.5	2.4
92	NAPT	3	10.0	12.5	-0.99	-0.38	0.19	22.5	24.4	1.8
93	TAAS	4	11.1	11.3	-0.91	-0.53	0.11	48.7	57.8	4.1
93	NAPT	4	11.3	11.4	-0.97	-0.69	0.18	22.2	31.6	2.6
94	TAAS	5	10.7	12.0	-0.98	-0.73	0.14	36.9	52.7	2.4
95	TAAS	6	9.9	12.3	-0.98	-0.83	0.17	26.4	44.0	1.7
96	TAAS	7	8.9	12.3	-1.12	-0.94	0.18	29.3	40.8	2.1
Average All Tests			10.6	11.6	-0.94	-0.64	0.15	28.0	42.8	2.4

Table 3. Mean Reading z scores and Percent with Scores by Ethnicity and Grade for Cohort 3 Students Who Are Neither LEP nor Special Education in the Current Year
(English Tests Only where Neither Refers to the Test Year)

Year	Test	Grade	Native American	Asian	African American	Hispanic	Anglo
92	TAAS	3	0.07	0.39	-0.29	-0.14	0.29
92	NAPT	3	0.20	0.46	-0.29	-0.09	0.44
93	TAAS	4	0.14	0.53	-0.27	-0.12	0.47
93	NAPT	4	0.15	0.52	-0.33	-0.12	0.47
94	TAAS	5	0.13	0.53	-0.34	-0.10	0.38
95	TAAS	6	0.18	0.54	-0.33	-0.10	0.43
96	TAAS	7	0.21	0.48	-0.24	-0.06	0.42
			0.16	0.49	-0.30	-0.10	0.41

Table 4. Level Regressions with Campus Fixed Effects for Third Grade Students by Category

Variables	Neither Coef.	Neither t	Ever LEP Asian and Hispanic Coef.	Ever LEP Asian and Hispanic t	Ever Special Education Coef.	Ever Special Education t
Native Amer High	0.00	0.0			-0.56	-1.3
Native Amer Low	-0.25	-2.7			-0.14	-0.5
Asian High	0.00	0.2			0.10	0.9
Asian Low	-0.13	-3.1	-0.26	-4.1	-0.33	-1.8
Black High	-0.30	-21.2			-0.46	-8.7
Black Low	-0.34	-16.2			-0.43	-5.3
Black Very Low	-0.52	-37.4			-0.70	-14.5
Hispanic High	-0.10	-8.2	-0.21	-3.2	-0.24	-5.5
Hispanic Low	-0.20	-11.7	-0.22	-3.3	-0.35	-6.1
Hispanic Very Low	-0.32	-25.4	-0.29	-4.5	-0.40	-9.4
Anglo Low	-0.14	-11.5			-0.17	-4.5
Anglo Very Low	-0.23	-30.9			-0.33	-14.2
College Grad	0.37	5.9	0.07	0.2	0.36	1.7
Some College	0.35	6.2			0.65	3.2
High School Grad	0.04	0.7	-0.01	0.0	-0.09	-0.4
Female Head	-0.15	-2.1	-0.58	-2.0	-0.30	-1.2
Gender	-0.12	-31.7	-0.08	-8.8	-0.05	-3.7
Age	-0.07	-14.9	-0.05	-5.5	-0.27	-22.2
Ever Special Ed			-0.63	-36.8		
Special Ed Now			0.04	1.3		
Ever LEP					0.01	0.3
LEP Now					-0.12	-2.8
Years Retained	-0.51	-51.9	-0.19	-12.0	-0.10	-4.6
Years Dbl Promote	0.11	3.9	0.09	2.6	0.08	1.5
Days Absent	-0.01	-14.2	-0.01	-9.4	-0.01	-7.2
District Move	-0.04	-5.8	-0.01	-0.1	-0.09	-3.6
Campus Transfer	-0.01	-0.8	-0.02	-0.1	-0.04	-0.6
Campus Vol Move	-0.07	-9.4	-0.10	-5.2	-0.08	-3.1
Out of Sample	-0.15	-6.3	-0.21	-3.6	-0.14	-1.5
Into Sample	-0.11	-15.1	-0.05	-2.7	-0.20	-6.7
Percent No Score	0.09	1.0			0.32	3.0
Constant	0.95	18.4	0.52	2.3	2.05	13.2
R Square	0.19		0.27		0.20	
Campuses	3,468		2,161		3,098	
Observations	156,469		33,764		33,670	

Table 5. Value Added Equations With Fixed Effects for Neither LEP nor Special Education Students in Grades 4-7

	Grade 4		Grade 5		Grade 6		Grade 7	
Variable	Coef.	t	Coef.	t	Coef.	t	Coef.	t
English Read	0.62	246.6	0.62	316.6	0.64	314.6	0.61	317.1
Spanish Read	0.30	1.5						
Peer Read	0.12	10.5	0.02	2.8	0.07	8.4	0.04	5.7
Native Amer High	-0.14	-1.8	-0.03	-0.5	-0.09	-1.3	0.03	0.5
Native Amer Low	-0.17	-1.9	-0.22	-2.9	-0.08	-1.0	0.04	0.5
Asian High	0.01	0.6	0.06	3.1	-0.01	-0.7	-0.02	-1.2
Asian Low	-0.11	-2.8	0.05	1.4	-0.12	-3.8	-0.01	-0.2
Black High	-0.32	-24.4	-0.15	-13.6	-0.20	-18.0	-0.09	-8.7
Black Low	-0.39	-20.1	-0.20	-11.9	-0.24	-14.7	-0.13	-8.2
Black Very Low	-0.44	-34.3	-0.27	-24.8	-0.27	-24.2	-0.18	-16.8
Hispanic High	-0.15	-12.7	-0.04	-4.2	-0.10	-10.1	-0.03	-2.9
Hispanic Low	-0.24	-15.1	-0.09	-7.0	-0.15	-10.9	-0.05	-3.9
Hispanic Very Lov	-0.30	-25.6	-0.14	-13.9	-0.21	-20.6	-0.10	-10.2
Anglo Low	-0.09	-8.6	-0.07	-7.4	-0.08	-8.9	-0.05	-5.9
Anglo Very Low	-0.18	-25.4	-0.10	-16.2	-0.12	-20.1	-0.06	-9.2
College Grad	0.32	5.5	0.18	3.5	0.23	4.7	0.20	4.5
Some College	0.13	2.4	0.16	3.5	0.12	2.6	0.03	0.7
HS Grad	-0.10	-1.7	0.05	0.9	0.00	0.0	0.05	1.0
Female Head	-0.13	-1.9	0.04	0.7	-0.05	-0.9	-0.04	-0.7
Gender	-0.03	-9.4	-0.03	-8.8	0.02	6.5	-0.18	-64.2
Age	-0.05	-10.7	-0.08	-20.0	-0.07	-17.7	-0.07	-19.6
Years Retained	-0.32	-31.4	-0.16	-16.4	-0.15	-13.7	-0.16	-15.0
Years Dbl Promot	0.18	5.4	0.09	2.9	0.06	1.8	0.02	0.6
Days Absent	-0.01	-23.3	-0.01	-20.9	-0.01	-25.3	-0.01	-18.3
W/I Year Moves	-0.06	-9.7	-0.03	-6.6	-0.08	-14.4	-0.06	-12.4
District Move	0.00	0.3	-0.02	-3.0	-0.07	-9.0	-0.04	-6.1
Campus Transfer	-0.05	-4.8	-0.06	-7.3	-0.10	-19.4	-0.05	-10.3
Campus Vol Move	-0.03	-4.0	-0.02	-2.7	0.00	0.3	-0.01	-1.3
Out of Sample	0.04	1.8	-0.02	-0.8	-0.06	-2.7	-0.01	-0.4
Into Sample	0.02	2.1	0.01	0.8	-0.07	-5.4	0.02	1.7
Percent No Score	0.16	1.9	-0.09	-1.7	-0.37	-5.8	-0.11	-2.1
Constant	0.66	12.4	1.00	20.3	1.09	20.6	1.17	22.2
R Square	0.47		0.54		0.55		0.54	
Campuses	3,570		3,663		4,025		4,269	
Observations	152,178		156,361		157,955		157,146	

Table 6. Value Added Regressions With Fixed Effects for Ever LEP Students in Grades 4-7

	Grade 4		Grade 5		Grade 6		Grade 7	
Variable	Coef.	t	Coef.	t	Coef.	t	Coef.	t
English Read	0.45	78.0	0.64	105.3	0.54	115.6	0.67	142.2
Spanish Read	0.44	32.6						
Asian Low	-0.12	-2.3	-0.17	-3.3	-0.12	-2.9	-0.02	-0.5
Hispanic High	-0.06	-0.7	-0.17	-2.0	-0.03	-0.5	-0.06	-0.9
Hispanic Low	-0.10	-1.2	-0.25	-2.9	0.01	0.1	-0.07	-1.0
Hispanic Very Low	-0.19	-2.3	-0.32	-3.8	-0.11	-1.5	-0.12	-1.8
College Grad	0.60	1.7	0.17	0.5	0.20	0.7	0.16	0.6
Some College	0.35	0.9	0.15	0.4	0.06	0.2	-0.40	-1.2
HS Grad	-0.30	-0.7	0.28	0.6	-0.97	-2.4	0.08	0.2
Grade School	-0.32	-0.5	0.25	0.4	-1.31	-2.3	-0.09	-0.2
Female Head	-0.01	0.0	0.26	0.6	0.29	0.8	-0.21	-0.5
Gender	-0.04	-3.9	-0.03	-2.9	-0.04	-4.8	-0.22	-26.6
Age	-0.05	-4.8	-0.12	-12.2	-0.09	-10.7	-0.10	-12.1
Years Retained	-0.21	-11.0	-0.12	-5.7	-0.13	-6.5	-0.07	-3.5
Years Dbl Promot	0.03	0.6	0.00	0.1	-0.11	-1.8	-0.20	-3.6
Days Absent	-0.01	-12.9	-0.01	-13.1	-0.02	-17.6	-0.01	-13.0
District Move	-0.03	-1.1	0.00	0.1	-0.15	-5.6	-0.01	-0.4
Ever Special Ed	-0.19	-8.5	-0.22	-9.3	-0.17	-7.9	-0.12	-5.6
Special Ed, 93	-0.03	-1.0	-0.35	-10.7	-0.40	-14.8	-0.43	-16.4
W/I Yr Moves	-0.06	-3.6	-0.04	-2.2	-0.06	-3.7	-0.08	-5.5
Campus Transfer	-0.05	-1.3	-0.07	-2.0	-0.15	-9.4	-0.10	-7.2
Campus Vol Move	-0.09	-4.6	-0.02	-1.0	-0.15	-6.3	-0.02	-0.8
Out of Sample	0.11	2.2	-0.18	-3.6	-0.04	-0.9	-0.13	-3.3
Into Sample	-0.06	-2.0	-0.31	-8.2	-0.25	-7.0	-0.38	-11.8
Percent No Score	0.34	3.9	0.33	3.3	0.15	1.6	0.09	1.0
Constant	0.32	1.2	1.38	5.4	1.61	7.2	1.59	7.4
R Square	0.42		0.47		0.48		0.57	
Campuses	2,018		1,702		1,710		1,890	
Observations	25,313		27,754		32,271		33,453	

Table 7. Value Added Equations With Fixed Effects for Ever Special Education Students in Grades 4-7

Variable	Grade 4 Coef.	Grade 4 t	Grade 5 Coef.	Grade 5 t	Grade 6 Coef.	Grade 6 t	Grade 7 Coef.	Grade 7 t
English Read	0.52	103.7	0.65	114.5	0.59	120.6	0.70	137.1
Spanish Read	0.55	13.3						
Native Amer High	-0.20	-0.9	-0.12	-0.5	0.39	1.5	0.00	0.0
Native Amer Low	-0.29	-1.4	-0.42	-1.7	-0.36	-2.1	-0.20	-1.1
Asian High	0.14	1.7	0.07	0.9	0.08	1.1	0.03	0.3
Asian Low	-0.24	-2.1	-0.17	-1.4	-0.08	-0.8	0.06	0.5
Black High	-0.42	-10.5	-0.26	-6.4	-0.28	-7.6	-0.10	-2.7
Black Low	-0.50	-8.0	-0.26	-4.1	-0.37	-6.7	-0.14	-2.4
Black Very Low	-0.49	-13.3	-0.34	-9.0	-0.41	-12.0	-0.20	-5.6
Hispanic High	-0.17	-5.2	-0.09	-2.9	-0.15	-5.0	-0.01	-0.5
Hispanic Low	-0.20	-4.6	-0.16	-3.6	-0.18	-4.7	-0.04	-0.9
Hispanic Very Low	-0.25	-7.8	-0.22	-6.7	-0.29	-9.7	-0.09	-3.0
Anglo Low	-0.14	-5.2	-0.06	-2.3	-0.09	-3.7	-0.06	-2.5
Anglo Very Low	-0.20	-11.7	-0.12	-6.7	-0.17	-10.4	-0.12	-6.9
College Grad	0.00	0.0	0.00	0.5	0.00	1.0	0.00	1.7
Some College	0.01	3.4	0.00	0.8	0.00	2.5	0.00	1.6
HS Grad	0.00	0.1	0.00	1.0	0.00	-0.8	0.00	0.7
Female Head	0.00	-0.1	0.00	-0.7	0.00	-0.3	-0.01	-2.8
Gender	0.02	1.5	0.00	0.1	0.03	3.4	-0.23	-23.1
Age	-0.18	-17.5	-0.23	-22.1	-0.20	-21.2	-0.21	-21.3
Years Retained	-0.15	-8.1	0.03	1.3	0.02	1.0	0.01	0.3
Ever LEP	-0.07	-2.5	-0.01	-0.4	-0.05	-2.0	-0.06	-2.7
LEP93	-0.19	-5.6	-0.32	-9.5	-0.26	-9.1	-0.23	-7.9
Years Dbl Promote	0.04	0.7	-0.20	-2.6	-0.30	-3.9	-0.11	-1.6
Days Absent	-0.01	-6.3	-0.01	-8.3	-0.01	-12.8	-0.01	-7.3
W/I Year Moves	-0.07	-4.5	-0.04	-2.7	-0.08	-5.5	-0.09	-6.9
District Move	-0.04	-1.8	-0.04	-1.7	-0.15	-6.3	-0.02	-1.0
Campus Transfer	-0.08	-2.5	-0.11	-4.2	-0.13	-8.3	-0.09	-6.0
Campus Vol Move	-0.04	-2.1	-0.05	-2.3	-0.05	-1.7	-0.08	-2.5
Out of Sample	-0.17	-2.7	-0.09	-1.4	-0.04	-0.6	-0.09	-1.4
Into Sample	-0.07	-2.4	-0.08	-1.8	-0.27	-6.3	-0.04	-1.0
Percent No Score	0.26	4.3	0.44	8.5	0.23	5.1	0.23	5.1
Constant	1.60	12.2	2.38	16.8	2.48	18.0	2.71	18.2
R Square	0.54		0.57		0.57		0.59	
Campuses	3,055		3,112		3,264		3,394	
Observations	27,746		28,274		31,218		32,098	

Figure 1. Distribution of Reading Z Scores

Grade 6 Students by Category