5. MOBILITY, ATTRITION, AND EXCLUSION FACTORS REFLECTED IN TLI MATHEMATICS SCORES: PATTERNS OF TEST INCLUSION AND EXCLUSION

Christopher A. Thorn

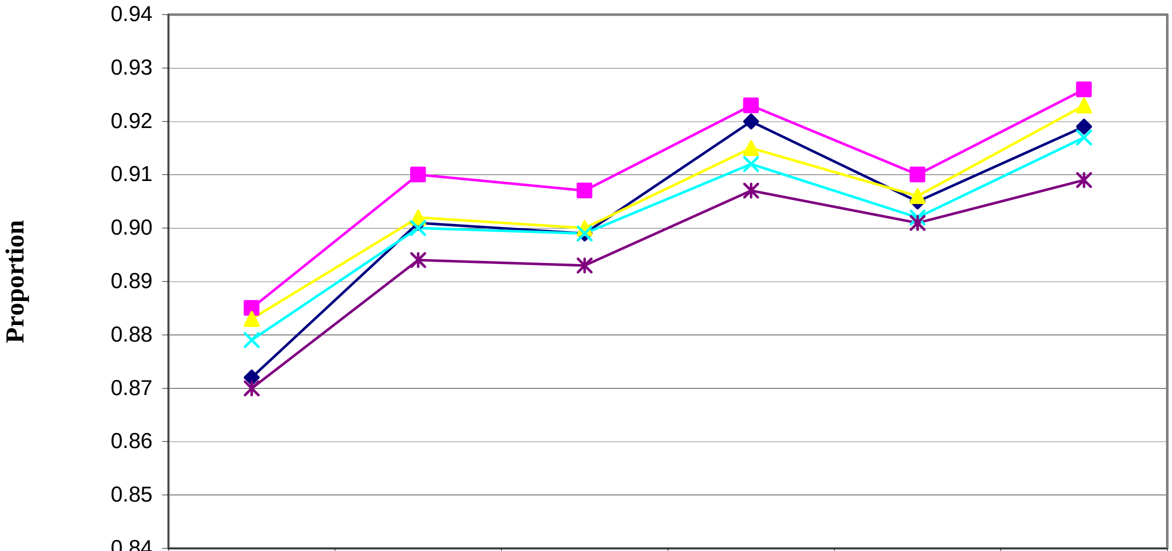
Executive Summary

Student attrition and exclusion are difficult phenomena to study since they are identified by missing data—the student not tested. This study examines the characteristics of students who were tested in one grade within the Texas accountability system and then were not tested in the following year. The important findings of this study are 1) that test inclusions rates went up steadily over time and 2) that student who were not tested in the follow-up year scored much lower, on average, than did the students tested in both years. This means that while the proportion of test takers went up, the worst performing students were the most likely not to be tested in the following year. There is also some concern that individual schools within a single district may have different inclusion policies. Table F shows a number of schools with a significant impact on the probability of being tested while controlling for test scores and individual student demographics. School-level inclusion rates should be examined to find both exemplary and troubling inclusion policies.

Introduction

One of the important ancillary concerns in any study of student achievement is the question of who is and is not tested. This paper explores which groups of children were most likely to remain in the testing system over any given two-year period in grades 3 through 8 across the state of Texas, 1994-1999. For example, the analysis computes the proportion of grade 3 students tested in 1994 who were also tested in grade 4 in 1995. The number of grade 3 students is used as the denominator. The data set used for this study is a 25% random sample of all possible grade combinations from 1994 through 2000. The entire population of valid test scores is approximately 3.5 million individual two-year comparisons. The resulting sample contains 880,181 observations. Figure 5.1 provides an overview of the proportion of children tested across all years and grade comparison cohorts during our study.

Figure 5.1. Proportion of students tested in any two-year comparison.



	93/94 – 94/95	94/95 – 95/96	95/96 – 96/97	96/97 – 97/98	97/98 – 98/99	98/99 – 99/00
3 to 4	0.872	0.901	0.899	0.920	0.905	0.919
4 to 5	0.885	0.910	0.907	0.923	0.910	0.926
5 to 6	0.883	0.902	0.900	0.915	0.906	0.923
6 to 7	0.879	0.900	0.899	0.912	0.902	0.917
7 to 8	0.870	0.894	0.893	0.907	0.901	0.909

School Year Comparisons

The general story told by the data in Figure 5.1 is that the proportion of students included in the accountability system from year to year has fluctuated, but that the general trend has been toward greater inclusion—in all grades. On a statewide level, rising inclusion suggests that districts and schools are doing a better job of accommodating students with limited English proficiency, some disabilities, and highly mobile students. Rising expectations for student inclusion, the introduction of Spanish language testing in the late 1990s, and increasing pressure from the state to test the total student population may have all combined to increase the proportion of students tested. The parallel movement at all grades—whether up or down — suggests that changes in these years relate to state-wide initiatives such as the introduction of

Spanish language testing or other global policy. The problem with this approach to understanding the dimensionality of inclusion in the testing system is that it does not tell the story of subgroups within the population. This paper explores the differences in participation across ethnic, gender, and regional groupings.

The “failure to observe” a student test score in the second year of a two-year comparison is both a question of student attrition—leaving the educational system—and of who is included in the tested population.

- Student attrition differs from student mobility in that attrition refers to the movement of children and their families out of the district or state.
- Mobility, in this case, would refer to movement of students between schools within the testing system. While this differentiation is important when studying assessment data, the reasons behind student moves are identical. Families move for a variety of school-external reasons (jobs, divorce, new home, etc.), as well as for school-related reasons (perceived problems in current school, reputation of new school, access to parochial education, etc.).
- Exclusion of an enrolled student from the testing system is the other reason a student may not appear in the second year of a two-year comparison. Reasons for exclusion are numerous and include student characteristics such as limited English proficiency, disabilities that make traditional testing too difficult, chronic absence, etc.

There are also important concerns about the counterproductive incentive structure that some accountability systems manifest in practice. Systems that reward districts, schools, or individual teachers based on the proportion of children above a certain test-cut point can encourage schools to neglect students who are far below that point in favor of children who are just below and who might be helped over the bar in the short term. Conversely, accountability systems that focus on student progress encourage participation of all students, since schools receive "credit" for the progress students make from any starting point—no matter how low. These latter types of systems are rare, however, because many consider that they place inadequate emphasis on achieving a minimum level of performance.

From a test-exclusion point of view, reliance on measures such as average test score (or proficiency score that is based on test scores) does not recognize students or schools who have made large educational gains but have not yet achieved high levels of proficiency. Accountability systems that do not include rewards for "student added-value," or other measures of individual level progress provide the strong incentives to exclude students. Teachers and principals operating under non-value added systems often feel that they are being penalized for their students’ demographic characteristics rather than any educational efforts. The logic in these situations seems to be that students are either surreptitiously included in officially excluded groups—LEP or Special Education categories exempted from testing—or they are, in the most egregious cases, sent off on field trips on testing days or are encouraged to be absent. While the latter abuses are probably quite rare, the use of exempted group membership to remove low performing students from the accountability population can be a tempting option. One particular shortcoming of this data set is the lack of any data that would be necessary to distinguish between attrition and exclusion. The following t-tests are one type of circumstantial evidence to suggest that the failure to include some *leavers*[1] in year two of a test cohort is related to their academic ability.

Table 5.1

T-Test for TLI Score Differences Between Students Tested in any Two-Year Grade Comparison Group

	Test 1	Test 2	Difference
Mean	76.7293	78.5782	1.8489
Std. Err.	0.0150	0.0139	0.0094
(N)	794752	794752	794752

Note: Test 1 is pre-test and Test 2 is post-test.

Table 5.1 shows an average TLI increase of 1.85 points for any two-year comparison across all grades and all year cohorts within the sample. This improvement varies across individual years and grades but reflects the general upward trend of the aggregate results identified by others who examined the statewide data. Table 5.2, on the other hand, points to the difference in average pre-test TLI score for *leavers*—students who are tested in a pre-test but who are not tested in the following post-test. There is a disturbingly large difference between the pre-test scores of *leavers* and those students who remain in the system. *Leavers* score, on average, 7.61 points below their counterparts.

Table 5.2

T-Test for TLI Score Differences in Year One Between Students Who Were and Were Not Tested in Year Two of Any Grade Comparison

	Not Tested	Tested	Difference
Mean	69.116	76.729	-7.613
Std. Err.	0.059	0.015	0.050
(N)	85429	794752	

Note: Not Tested if no post-test and Tested if post-test exists.

The difficulty with analyzing *leavers* from the Texas accountability system is that, while we have good demographic data on the tested students, we do not know what happened to students who were not tested. The only way we would be able differentiate between true attrition—students exiting the Texas public education system—and test exclusion would be to combine the state testing data, which is at the individual student level and includes a unique ID, with the yearly enrollment data in each district in order to identify who has left the district and who is still enrolled. It would also be useful to have limited English proficiency status and Special Education designations to identify students moving in and out of those groups. This would allow for an analysis of the varying policies of inclusion, on both the district and school level.

[1] The term *leavers* is used in this paper to refer to this combination group of attrited and excluded students.

Given the absence of these data, the conclusions that one can draw from any study of the probability of students being tested must be interpreted with this fact in mind. However, even without these detailed data, one can still identify groups that have greater and lesser proportions of *leavers*.

The following pages include sets of tables that address some of the differences in the probability of being tested across student groups from samples drawn at three different levels.[2] The first group is a 25% sample of students across the state. The second group compares students in National Science Foundation-funded Urban Systemic Initiatives [USI] (primarily large urban central cities) to students in similar urban districts that did not participate in the USI program. Finally, other members of this research team examined the impact of the USI on the Dallas school district. Data for Dallas are presented as an example of differences within a large USI district and to provide additional context for readers of the accompanying reports.

Table 5.3 reports the effect of changes in the independent variables[3] as the change in the probability that a student is tested.[4] It is important to remember that the results are conditional on all other factors remaining the same. In the case of the results in Table 3, all other independent variables are set to their mean values and only the predictors in question are allowed to change. The base probability of the model that any individual student will be tested in a subsequent year is 0.911.[5] One of the appeals of this particular approach of presentation is that the probabilities in Table 3 are additive. For example, In the case of Non-White vs. White students, when controlling for all other factors, the probability of white students being tested in year two goes up by 0.089 to near certainty (0.911 + 0.089 = 1.000). At the same time, students on free- or reduced-lunch are 0.029 less likely to be tested in year two. Males are also 0.011 less likely than females to be tested in year two.

---

[2] These analyses were greatly aided by the statistics software Clarify© (Tomz, Wittenberg, & King, 2001), which is used with the statistical package Stata. Clarify allows analysts to estimate any of a wide range of statistical models and then explore changes in the predicted value of the dependent variable when one or more of the independent variable values are changed. Clarify reports the effect (or first difference) of the change in predictors on the dependent variable. The results also include a confidence interval for this predicted difference.

[3] Refer to Appendix Tables A and B for detailed descriptives and probit results.

[4] The statistical technique used to analyze these data was probit analysis. Probit analysis is an appropriate choice for predicting the likelihood of binary outcomes. In this paper, that outcome is the probability of being tested given a number of categorical, ordinal, and continuous predictor variables. The Clarify package was then used to simulate the impact of individual predictors in terms of the difference from the base probability while holding all other predictors constant.

[5] This model—see Appendix Tables A and B—calculates a base probability that includes proportional representation in all years and grades and has all predictor variable set to their sample means.

Table 5.3
Difference in Probability of Being Tested in Year Two, Conditional on Category of Interest—Statewide Data

Category of Interest	Mean[6]	Std. Err.	95% Conf. Interval	
White	0.089	0.001	0.086	0.091
African American	0.031	0.001	0.029	0.032
Hispanic	0.034	0.001	0.032	0.035
Disadvantaged[7]	-0.029	0.001	-0.030	-0.027
Urban	-0.096	0.003	-0.102	-0.089
Male	-0.011	0.001	-0.013	-0.010
Grade 8[8]	-0.007	0.001	-0.009	-0.005

Note: Positive means indicate increasing probability moving from first to second category of interest

This means that disadvantaged, African American males living in an urban setting have a probability of being tested equal to 0.806—approximately 0.20 below the comparison white group and over 0.10 under the sample average.[9] Another way to express that difference is that a male, African American student on free lunch status is –0.105 less likely to be tested than a female, non-African American, non-economically disadvantaged student living in a rural community. These numbers do not tell the story by themselves. What they suggest is that there are significant differences across major subgroups in the sample that are not adequately understood if one only examines simple descriptive statistics.

Table 5.4
Difference in Probability of Being Tested in Year Two Conditional on Category of Interest—Large Urban Central Cities

Category of Interest	Mean	Std. Err.	95% Conf. Interval	
White	0.114	0.003	0.109	0.121
African American	0.034	0.002	0.031	0.037
Hispanic	0.035	0.002	0.032	0.038
Disadvantaged	-0.020	0.002	-0.023	-0.017
USI District	0.001	0.001	-0.001	0.004
Male	-0.014	0.001	-0.017	-0.012
Grade 8	-0.008	0.002	-0.012	-0.005

Note: Positive means indicate increasing probability moving from first to second category of interest

Table 5.4 compares the impact of being in a large, urban, central city and whether or not one is part of an Urban Systemic Initiative. The base predicted probability of an average student

[6] Amount of above or below average probability of being tested.
[7] Coded as a proportion based on the number of times a student appeared in the data set with a free and reduced-cost lunch indicator.
[8] Relative to grade 3.
[9] 0.911 + 0.031 – 0.029 – 0.096 – 0.011 = 0.806

being tested in any year two is 0.888. The predicted differences in Table 5.3 do not differ a great deal between the Tables 5.3 and 5.4. The important line is the one that indicates whether a district is a member of an Urban Systemic Initiative or not. First, the effect size is small (0.001). In addition, the confidence interval ranges from –0.001 to 0.003 meaning that the difference is statistically insignificant as well.

The following table illustrates some of the same results for a single Urban Systemic Initiative District—Dallas. Here we see that the positive effect of being white continues to rise, as does the negative effect of being economically disadvantaged. One addition to this table that parallels the comparison of USI to non-USI districts in Table 5.5 is the inclusion of two schools—one with the largest positive effect on the probability of a student being tested and one with the largest negative effect from the probit model,[10] which includes all schools. This, again, is the sort of result that is not available from the analysis of descriptive statistics. In this model, School L has a large negative effect on the probability of being tested—more than twice the effect of being Hispanic, African American, on being economically disadvantaged. At the same time, the effect of being in School A is relatively large and positive.

Table 5.5
Difference in Probability of Being Tested in Year Two Conditional on Category of Interest—Dallas

Category of Interest	**Mean**	**Std. Err.**	**95% Conf. Interval**	
White	0.129	0.007	0.114	0.141
African American	0.059	0.005	0.050	0.068
Hispanic	0.029	0.005	0.019	0.038
Disadvantaged	-0.042	0.005	-0.051	-0.032
Male	-0.014	0.003	-0.020	-0.008
Grade 8	-0.035	0.010	-0.055	-0.016
School L	-0.159	0.067	-0.305	-0.036
School A	0.061	0.011	0.038	0.079

Note: Positive means indicate increasing probability moving from first to second category of interest

Given that these two schools represent the two extremes in effect size, a reasonable analyst would conclude that both of these schools should be examined more closely to determine what observed effect they seem to have on *leavers* and whether this effect is based on some unmeasured characteristic. While the large negative effect attributed to School L may suggest that this school is somehow failing to include all students in the TAAS mathematics tests, the evidence is not clear enough to indicate why. The important conclusions that can be drawn from this analysis are the following:

- Test inclusion rates have varied across years, but show a general upward trend in all grades.
- The negative impact of being in an urban setting and being economically disadvantaged is pervasive and has relatively large effect sizes.

[10] See Table F in the Appendix.

- School effects in the Dallas analysis are either equal to or are larger than the effect size of any other predictor variable. This parallels findings in student assessment studies that show that school and classroom effects dominate any value-added or growth modeling.
- There was no discernable USI effect on the probability of being tested. While USI projects were not focused on including more students in the tested population, but the logic behind the model assumed that a push for test-based accountability would encourage USI districts to improve their inclusion of students exposed to USI programs.

This approach to modeling test taking—whatever the reason—is a first step towards understanding what measures one would need to adequately separate the phenomena of mobility, attrition, and test exclusion. With additional data—including language proficiency, Special Education designators, and enrollment status—it would be possible to provide a much finer-grained analysis of inclusion policies that would point to both exemplary and questionable practices on the part of districts and schools.

References

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science, 44*(2), (April): 347-61.

Tomz, M., Wittenberg, J., & King, G. (2001). *CLARIFY: Software for interpreting and presenting statistical results*. Version 2.0. Cambridge, MA: Harvard University, June 1. http://gking.harvard.edu

Appendix

Table A[11]

Descriptive Statistics for Statewide Analysis Sample[12]

Variable	Obs	Mean	Std. Dev.
Pre-Test	880181	75.990	13.945
Post-Test	794752	78.578	12.417
Tested[13]	880181	0.903	0.296
Male	880181	0.500	0.500
Disadvantaged	880181	0.437	0.459
Students in USI Districts	447637	0.248	0.432
- Ethnicity			
White	880181	0.514	0.500
African American	880181	0.135	0.342
Hispanic	880181	0.320	0.466
Other Minority	880181	0.017	0.130
Mixed Ethnicity	880181	0.014	0.116
- Community Context			
Rural or Small Town	880181	0.213	0.410
Urban Fringe	880181	0.283	0.450
Mid-sized Central City	880181	0.180	0.384
Large Central City	880181	0.322	0.467
- School Year Groups			
Year 94-95 |	880181	0.162	0.369
Year 95-96	880181	0.162	0.369
Year 96-97	880181	0.166	0.372
Year 97-98	880181	0.170	0.376
Year 98-99	880181	0.173	0.378
Year 99-00	880181	0.167	0.373
- Grade Comparisons			
Grade 3 to 4	880181	0.189	0.391
Grade 4 to 5	880181	0.193	0.395
Grade 5 to 6	880181	0.200	0.400
Grade 6 to 7	880181	0.209	0.407
Grade 7 to 8	880181	0.209	0.407

[11] Table A includes the descriptive statistics for the 25% state-wide sample of student test cohorts.
[12] Pre-test and post-test of mathematics TLI score, which ranges from 0 to 93. Disadvantaged is expressed as the proportion of occurrences of the free- and reduced-cost lunch indicator for each student across all years in which that student is in the data set. All other variables are dichotomous variables coded 0 or 1 where 1 = True. This means that they represent the within-category proportion of cases for which that subcategory is True.
[13] Students tested in both pre- and post-test.

Table B
Probit Model Predicting Who Is Tested in Year Two for Statewide Analysis[14]

Tested – Dependent	Coef.	Std. Err.	P>|z|	95% Conf. Interval	
Pre-Test	0.017	0.000	0.000	0.017	0.017
Male	-0.071	0.004	0.000	-0.079	-0.064
Disadvantaged	-0.176	0.005	0.000	-0.186	-0.166
- Ethnicity					
African American	0.213	0.006	0.000	0.201	0.226
Hispanic	0.221	0.005	0.000	0.210	0.231
Other Minority	0.164	0.016	0.000	0.132	0.196
Mixed Ethnicity	0.248	0.017	0.000	0.214	0.282
- Community Context					
Urban Fringe	-0.106	0.006	0.000	-0.117	-0.095
Mid-sized Central City	-0.141	0.006	0.000	-0.153	-0.129
Large Central	-0.266	0.006	0.000	-0.276	-0.255
- School Year Groups					
Year 95-96	0.082	0.006	0.000	0.069	0.094
Year 96-97	0.018	0.006	0.004	0.006	0.031
Year 97-98	0.077	0.007	0.000	0.064	0.090
Year 98-99	-0.004	0.006	0.570	-0.016	0.009
Year 99-00	0.045	0.007	0.000	0.032	0.059
- Grade Comparisons					
Grade 4 to 5	0.020	0.006	0.001	0.008	0.032
Grade 5 to 6	-0.035	0.006	0.000	-0.047	-0.023
Grade 6 to 7	-0.032	0.006	0.000	-0.044	-0.020
Grade 7 to 8	-0.046	0.006	0.000	-0.058	-0.034
_constant	0.177	0.012	0.000	0.153	0.201

The probit results presented in Table B show statistically significant results for all predictors – with the exception of the 1998-1999 school year. This difference may be explained by the introduction of the Spanish language math test during this period. Some students at the low end of English proficiency may have been moved back into Spanish language testing when it became available. Moving down the list of predictors, it may be confusing to see that the effects for minority students are all positive – when anecdotal accounts would lead one to expect that minority students would be a more mobile population. This can be explained by noting that analysis controls for ethnicity, economically disadvantaged status, and community context (expressed in terms of urbanicity). The fact that the majority of minority students are poor and live in an urban setting means that the overall negative effect of being poor and living in an urban setting outweigh the positive effect of being a member of a minority.

[14] Reference categories are White, Rural, Year 94-95, and Grade Cohort 3 to 4

Table C[15]

Descriptive Statistics for USI/Non-USI Analysis Sample

Variable	Obs	Mean	Std. Dev.
Pre-Test	283826	74.263	14.771
Post-Test	250064	76.969	13.330
Tested	283826	0.881	0.324
Male	283826	0.496	0.500
Disadvantaged	283826	0.516	0.456
Students in USI Districts	266827	0.409	0.492
- Ethnicity			
White	283826	0.325	0.468
African American	283826	0.219	0.413
Hispanic	283826	0.414	0.493
Other Minority	283826	0.026	0.159
Mixed Ethnicity	283826	0.016	0.126
- School Year Groups			
Year 94-95 |	283826	0.167	0.373
Year 95-96	283826	0.161	0.367
Year 96-97	283826	0.162	0.368
Year 97-98	283826	0.166	0.372
Year 98-99	283826	0.173	0.379
Year 99-00	283826	0.171	0.376
- Grade Comparisons			
Grade 3 to 4	283826	0.188	0.390
Grade 4 to 5	283826	0.192	0.394
Grade 5 to 6	283826	0.199	0.399
Grade 6 to 7	283826	0.209	0.407
Grade 7 to 8	283826	0.212	0.409

[15] Table C includes the descriptive statistics for the large, urban central city subsample of student test cohorts.

Table D

Probit Model Predicting Who Is Tested in Year Two for USI/Non-USI Analysis[16]

Tested – Dependent	Coef.	Std. Err.	P>|z|	95% Conf. Interval	
Pre-Test	0.014	0.000	0.000	0.014	0.015
Male	-0.073	0.006	0.000	-0.086	-0.061
Disadvantaged	-0.095	0.009	0.000	-0.113	-0.078
USI District	0.008	0.007	0.256	-0.006	0.021
- Ethnicity					
African American	0.194	0.010	0.000	0.174	0.214
Hispanic	0.190	0.009	0.000	0.171	0.208
Other Minority	0.208	0.023	0.000	0.163	0.253
Mixed Ethnicity	0.315	0.029	0.000	0.259	0.371
- School Year Groups					
Year 95-96	0.151	0.011	0.000	0.130	0.172
Year 96-97	0.084	0.011	0.000	0.063	0.105
Year 97-98	0.226	0.011	0.000	0.204	0.248
Year 98-99	0.202	0.011	0.000	0.181	0.224
Year 99-00	0.214	0.011	0.000	0.192	0.236
- Grade Comparisons					
Grade 4 to 5	0.025	0.011	0.018	0.004	0.046
Grade 5 to 6	-0.083	0.010	0.000	-0.103	-0.063
Grade 6 to 7	-0.056	0.010	0.000	-0.076	-0.036
Grade 7 to 8	-0.044	0.010	0.000	-0.064	-0.024
_constant	-0.028	0.020	0.153	-0.068	0.011

Table D presents the state wide model applied to a comparison of large, urban central city school districts. There are three interesting findings one can draw from these data. First, for the traditional demographic variables the sign and magnitude of the effects is similar to the state wide model. This indicates that the effects of race/ethnicity, gender, and income are systemic, societal effects that hold across settings. Second, the effect of being a USI district is small and statistically insignificant. Third, insignificant effect of the 1998-99 school year does not repeat itself within urban districts. It appears that whatever happened to affect test inclusion in that year did not play out in urban districts.

[16] Reference categories are White, Rural, Year 94-95, and Grade Cohort 3 to 4.

Table E[17]

Descriptive Statistics for Dallas Analysis Sample

Variable	Obs	Mean	Std. Dev.
Pre-Test	46635	74.584	15.071
Post-Test	41243	77.358	13.454
Tested	46635	0.884	0.320
Male	46635	0.489	0.500
Disadvantaged	46635	0.465	0.453
- Ethnicity			
White	46635	0.362	0.480
African American	46635	0.367	0.482
Hispanic	46635	0.229	0.420
Other Minority	46635	0.034	0.181
Mixed Ethnicity	46635	0.008	0.091
- School Year Groups			
Year 94-95 |	46635	0.156	0.363
Year 95-96	46635	0.160	0.367
Year 96-97	46635	0.154	0.361
Year 97-98	46635	0.170	0.376
Year 98-99	46635	0.181	0.385
Year 99-00	46635	0.178	0.382
- Grade Comparisons			
Grade 3 to 4	46635	0.198	0.398
Grade 4 to 5	46635	0.198	0.398
Grade 5 to 6	46635	0.195	0.396
Grade 6 to 7	46635	0.200	0.400
Grade 7 to 8	46635	0.209	0.407

[17] Table E includes the descriptive statistics for the Dallas school district subsample of student test cohorts.

Table F
Probit Model Predicting Who Is Tested in Year Two for Dallas Analysis[18]

Tested – Dependent	Coef.	Std. Err.	P>|z|	95% Conf. Interval	
Pre-Test	0.015	0.001	0.000	0.014	0.016
Male	-0.077	0.016	0.000	-0.108	-0.046
Disadvantaged	-0.229	0.026	0.000	-0.281	-0.178
- Ethnicity					
African American	0.348	0.030	0.000	0.290	0.406
Hispanic	0.171	0.031	0.000	0.110	0.231
Other Minority	0.302	0.055	0.000	0.194	0.409
Mixed Ethnicity	0.388	0.097	0.000	0.197	0.579
- School Year Groups					
Year 95-96	0.123	0.026	0.000	0.071	0.175
Year 96-97	0.143	0.027	0.000	0.089	0.197
Year 97-98	0.205	0.027	0.000	0.151	0.258
Year 98-99	0.209	0.027	0.000	0.155	0.262
Year 99-00	0.159	0.027	0.000	0.106	0.213
- Grade Comparisons					
Grade 4 to 5	0.030	0.025	0.235	-0.020	0.080
Grade 5 to 6	0.065	0.026	0.013	0.014	0.116
Grade 6 to 7	-0.159	0.027	0.000	-0.211	-0.106
Grade 7 to 8	-0.184	0.051	0.000	-0.285	-0.083
- Individual Schools[19]					
School A	0.463	0.116	0.000	0.236	0.690
School B	0.309	0.078	0.000	0.156	0.461
School C	-0.605	0.122	0.000	-0.845	-0.365
School D	-0.472	0.090	0.000	-0.650	-0.295
School E	-0.438	0.092	0.000	-0.618	-0.258
School F	-0.478	0.087	0.000	-0.647	-0.308
School G	0.268	0.083	0.001	0.107	0.430
School H	0.215	0.067	0.001	0.083	0.346
School I	0.273	0.089	0.002	0.099	0.447
School J	0.320	0.101	0.002	0.121	0.518
School K	-0.278	0.088	0.002	-0.451	-0.106
School L	-0.621	0.207	0.003	-1.027	-0.215
constant	-0.045	0.073	0.540	-0.187	0.098

The Dallas model underlines the importance of examining local application of systemic policies and the variability of school-level effects. A number of differences begin to emerge in the demographic variables – the effect sizes increased for ethnicity and disadvantaged indicators. The negative impact of increasing grade declined relative to the statewide and urban analyses. However, the most interesting finding was the large effects for individual schools ranging from large positive to large negative effects.

[18] Reference categories are White, Rural, Year 94-95, & Grade Cohort 3 to 4.
[19] Schools A-L represent anonymous coding for Dallas schools whose results had a significance level ≤ 0.05. 85; schools with non-significant coefficients were excluded from this Table but were included in all analysis. In addition, schools were only included if there were 50 or more test scores for the school.