

Executive Summary

A multidisciplinary team worked over a year to achieve the *Systemic Initiatives: Student Achievement Analysis Study*'s main goal of developing an analytic framework for studying the degree to which systemic reform contributes to improved student achievement and other outcomes. In pursuing this goal, we sought to provide information to the National Science Foundation that will enable NSF, its education constituencies, and the education research community address the following questions:

1. How can the data submitted to NSF by systemic initiatives (SIs) be used to evaluate systemic reform?
2. How does the precision of analysis depend upon the qualities of student assessment data?
3. What statistical models best fit the data linking systemic initiatives to student achievement?
4. What are the lessons learned about the kind of databases and analyses that are most effective for evaluating and understanding systemic reform?

To demonstrate our approach to developing analytic frameworks, we analyzed data from the Texas Assessment of Academic Skills (TAAS) for grades 3 through 8 from 1994 through 2000 and compared Urban Systemic Initiative (USI) districts with other districts in the state.

We identified a number of desirable features of data and databases that are needed to study the impact of the systemic initiatives and other large-scale reforms:

1. Data that describe for the USI, or other initiative, the participation level of teachers by school;
2. An identifiable control group;
3. Testing of students in consecutive years;
4. Vertically scaled scores over grades;
5. An assessment that measures the full range of student knowledge without ceiling effects;
6. Assessment data linked with student, school, and district demographic and program data;
7. Alignment of assessments with district standards and USI goals; and,
8. A means of determining student attrition rates in the population and the selective exclusion of students from testing.

We identified three general areas that can influence the precision of analyses of student achievement data in studying the impact of systemic reform that need to be considered: (1) the extent to which teachers, schools, and districts participated in the systemic initiative over time; (2) what students were excluded from the testing and analyses; and, (3) the standard error of measurement in the assessment instruments.

There is no one best model for analyzing the link between systemic initiatives and student achievement. Each model is based on specific assumptions made necessary by the incompleteness of available data or other constraints. We developed three approaches, each by a different researcher. All three models provide information about students' growth over time and compare the performance by students in USI districts with those in other districts. In this respect, the results from one model serve as a replication of those from the other models. However, each of the three researchers made different assumptions about growth that influenced which groups of students were tracked over time. Bolt examined changes in school means on TAAS at a given grade level (e.g., grade 5 in 1994, 1995, 1996, etc.). He assumed that this approach would more effectively control for teacher effects because the same teachers are more likely to teach the same grade in successive years. Within-grade analysis also is more comparable with the nature of TAAS and the TLI scores that are equated within grades. His model implies that the variation among different cohorts of students (e.g., 4th grade students in 1995 compared with 4th grade students in 1996) is due to program effects, rather than to other factors.

Gamoran used nearly all of the students in the database to estimate the growth intercepts and slope. In this very robust model, students with any two scores, even those whose scores are not for consecutive years, can be used to estimate the parameters. To develop estimates for a model with a quadratic term requires students with four data points in order to estimate the three parameters *intercept*, *slope* (linear term), and *changed* (quadratic term). Students with fewer than four data points contribute to estimating the lower-order terms. Thus, Gamoran's growth model included students who left or entered the system during the period investigated. He also included students who changed schools within the system, state, or district. Students who were retained in a grade and had two scores for the same grade were deleted from the database used in the analysis.

Meyer's value-added analysis examined students' performance in a grade by taking into consideration achievement from the year before. His analyses included only students who had test scores for two consecutive years (e.g., grades 3 and 4, grades 4 and 5, etc.). This was a more restrictive requirement than that used by the other two researchers. The advantage of this approach is that greater precision could be given to improved student performance that can be attributed to a school year. The other two researchers computed the intercept term, or the initial performance at grade 3, that was used to compare the starting points among districts, but their models computed the difference between any two grades using equations developed to fit all of the points over the seven years, rather than just between two years.

Based on the three analytic models, we drew the following conclusions:

1. Texas Assessment of Academic Skills (TAAS) scores improved from 1994 to 2000 for all groups. Annual gain scores by Blacks and Hispanic students relative to White students improved over time. Annual gain scores for economically disadvantaged remained constant over time.
2. USI districts began lower, but raised scores faster than non-USI districts:

- A. There is some evidence that USI districts improved faster than non-USI districts, but this may be due to the scoring metric used (e.g., TLI rather than Rasch). Meyer found positive USI effects over time for most grade levels. Bolt found a positive USI effect at grade 7.
 - B. Texas USI scores started below those of non-USI districts and produced smaller annual gains than non-USIs in 1994.
 - C. There is no evidence that USI districts lost ground compared to non-USI districts from 1994-2000.
3. Achievement gaps between minority and White students and between advantaged and disadvantaged students narrowed statewide on the TLI. However, this finding was highly dependent on the scaling metric used. There was much less narrowing of gaps when the Rasch scale, which is more sensitive to gains at the extreme ends of achievement, was used rather than the TLI. There was some evidence from one model that the gap between White students and Black students narrowed more in USI districts than in the contrast districts.
 4. There is no difference in the rate at which achievement gaps are narrowing in USI and non-USI districts.
 5. An increasing proportion of students were tested over time.
 - A. TAAS attrition rates in students not being tested went down over time.
 - B. There was no difference in the TAAS attrition rate between USI districts and large urban districts that are not USIs.
 6. There are large differences in TAAS attrition by demographic group.

Demonstrating the impact of large-scale reform is immensely complex. This project has revealed the reasons for much of this complexity and has demonstrated specific analytic techniques that can be used to study the growth in student learning over time, given this complexity. Being restricted to only using existing data, we were unable to overcome major design flaws and the lack of data on the independent variables. Even with these deficiencies, the models produced some evidence of the improvement in student learning by districts with USIs compared to other districts. The analytic models used in this study have wide applicability in studying large-scale reform. The most important implication of this study is to inform the design of the evaluations of large-scale reform efforts so that the necessary data will in future be available to more effectively measure the impact of such interventions on student learning.