

4. STUDY METHODOLOGY

Eric Osthoff

Introduction

This section addresses considerations of study methodology, design, and sample. The first part describes the SSI and USI programs and the national reform context in which NSF's systemic initiatives strategy originated. The second part provides an overview of the Texas SSI and the three USI initiatives—San Antonio, El Paso, and Dallas—that receive major attention in later chapters. The description of the Texas SSI and USIs describes changes in policy, infrastructure, professional development, and instructional practice in each initiative so that readers may appreciate the nature and extent of the reform activity associated with findings on student achievement impacts that are presented in the later sections. Specific comments are included on the ways and degree to which available implementation data for Texas SIs meet the needs of the analytic models presented in later sections. The third part of this section describes the TAAS mathematics database received from the Texas Education Agency (TEA) and methods used to clean the database to prepare it for each planned analysis.

The Emergence of Standards-Based Systemic Reform

In 1991, the National Science Foundation (NSF) assumed a leadership position in national efforts to enable public schools to educate children more effectively and equitably in mathematics, science, and technology by launching the Statewide Systemic Initiatives (SSIs)—the first in a series of closely related programs, including the Urban Systemic Initiatives (USIs) launched in 1993, Rural Systemic Initiatives (RSIs) in 1994, and the Urban Systemic Program in 1999.

All of NSF's Systemic Initiatives (SIs) have reflected the NSF's desire to merge two distinct forces evident in the national educational reform arena. The first force—standards-based reform—rose to prominence with the 1989 publication of the National Council of Teachers of Mathematics' (NCTM) *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The NCTM *Standards* appeared amidst a spate of reports (e.g., *A Nation At Risk*, 1983) by blue-ribbon panels that advocated more demanding educational standards generally. What distinguished the NCTM *Standards* from many previous reports is that it went beyond calling for simply more of the same—more courses for high school graduation, more school days per year, more homework—to advocate higher quality curricular content, instruction, and assessment. NCTM further argued that high quality content could be effectively taught to virtually all students. The remarkable success of the NCTM *Standards* in helping focus the national reform debate simultaneously on equity and quality content had the effect of stimulating the professional organizations of most other subject areas to put forth national standards of their own (McLeod, Stake, Schappelle, Mellissinos, & Gierl, 1996).

The second force the NSF tapped into with the SIs was a rapidly growing systemic reform movement. At the movement's inception, during the mid-1980s, proponents of systemic

reform appeared primarily concerned with the *alignment* of standards, practices, and assessments, where alignment was broadly understood as coordinating policy and practices across multiple dimensions of the educational system to make them work in concert instead of at cross-purposes. But policymakers and reformers had curriculum standards clearly in view by the time Smith and O’Day (1991) put forth what is generally considered to be the first comprehensive proposal for systemic school reform. In the view of Smith and O’Day, systemic school reform would be most effective by making curriculum standards (or standards-based curriculum frameworks) the cornerstone of the instructional guidance system. Based on curriculum standards that give adequate attention to considerations of curricular quality (e.g., conceptual understanding, knowledge construction through analysis and synthesis of real life problems), state policymakers and others could reconfigure other components of the instructional guidance system (e.g., assessment, professional development) to increase the ability of districts, schools, and teachers to effectively engage all students in high quality curricula.

Merging these two national trends yielded *standards-based systemic reform*, an idea that was being actively pursued by only a relatively small number of educational policymakers and scholars in 1991 when the NSF committed to making it the centerpiece of its new series of reform initiatives.

Guiding Principles of Systemic Reform Initiatives

General parameters of the Systemic Reform initiatives were contained in the Request for Proposals that the Human Resources Directorate of the National Science Foundation used to launch the SSI program in 1990. In that document, the NSF encouraged prospective grantees to outline a coherent strategy for reforming all aspects of the state’s K-16 system for mathematics, science, and engineering education. Among the components of the system to be considered were: (1) organizational structure and decision making; (2) provision and allocation of resources; (3) recruitment and preparation of teachers and college faculty; (4) retention and continuing professional development of teachers and other professional personnel; (5) curriculum content and learning goals; (6) delivery of instruction, including the use of educational technology; (7) facilities and equipment; (8) articulation within the system; and, (9) accountability systems.

By 1995, based primarily on feedback from the SSIs, NSF presented what have become known as six “drivers” of systemic reform. The drivers represented NSF’s effort to provide more specific guidance regarding the aspects of educational systems and the relationships among system components that were expected to be of special significance in systemic reform. Soon after introducing the drivers, the NSF began requiring reform managers to include in progress reports to NSF descriptions of how their initiatives were addressing each of the drivers.

Below are listed the six drivers (in italics), plus clarifying parenthetical remarks.¹

1. *Rigorous, standards-based instruction for all students, and the curriculum, professional development, and assessment systems to support that instruction.* (High-quality science and mathematics education for all students is one of the key goals of the NSF’s Systemic Initiatives programs and one of the strategies to accomplish this

¹ The NSF Drivers evolved over time. The version provided here appeared in 1995.

- is through professional development, as well as improved assessment systems. The SIs seek changes in classroom practice by introducing new types of curriculum and instruction to increase students' conceptual understanding of mathematics and scientific concepts).
2. *A unified set of policies that facilitate and enable driver 1.* (SI programs urge states and districts to align their policies with rigorous academic standards. This driver emphasizes the need for systems to move toward aligned, standards-based policies for curriculum, instruction, and student assessment, as well as teacher preparation, licensing, and recertification).
 3. *A unified application of all resources to facilitate and enable driver 1.* (This driver stresses the need for initiatives to develop the resources and capacity to teach high-quality science and mathematics to all students. As stated above, since the NSF funding for the SIs is limited, it is necessary for initiatives to leverage funds to promote standards-based reform activities).
 4. *Mobilization of the full community of stakeholders.* (This driver encourages initiatives to build both professional and public support for reform. The support of stakeholders such as parents, teachers, businesses, and higher education must be mobilized to encourage improvements in the teaching and learning of mathematics, science, and technology education).
 5. *Increased student attainment in science, mathematics, and technology.* (The SIs are specifically designed to increase student achievement in mathematics and science, they therefore place a premium on achievement data that enables system managers and others to assess the impact of achievement growth system wide).
 6. *Reduction in attainment differences between those students traditionally underserved and their peers.* (Another goal of the SI program is a significant reduction in the performance gaps between underserved groups, such as women and minorities, and white males. The SIs support the strategy of moving away from a focus on basic skills for lower-achieving students to higher academic standards for all students).

Additional background on the SSI program, including NSF strategies developed during the SSI program that were carried over into the USI program, is available in Webb, Kane, Kaufman, and Yang (2001) and in Zucker, Shields, Adelman, Corcoran, and Goertz (1998).

Urban Systemic Initiatives Program Roll-Out and Scope

From 1993 to 1999, the NSF funded 22 USIs in five cohorts, as shown below:

Cohort 93	Cohort 94	Cohort 95	Cohort 97	Cohort 99
<ul style="list-style-type: none">• Baltimore• Chicago• Dallas• Detroit• El Paso• Miami-Dade• New York• Phoenix	<ul style="list-style-type: none">• Cleveland• Columbus• Fresno• Los Angeles• Memphis• New Orleans• Philadelphia	<ul style="list-style-type: none">• Milwaukee• St. Louis• San Antonio• San Diego	<ul style="list-style-type: none">• Atlanta• Jacksonville	<ul style="list-style-type: none">• Houston

Many USIs were subsequently refunded after five years as part of the USP. Other cities that had not been USIs also joined the USP. The current study includes only USIs and, for reasons noted later, focuses only on a subset of three Texas USIs—Dallas, El Paso, and San Antonio.

The NSF established several requirements for USI funding. First, USI districts had to serve at least 40,000 students in a single urban area. In many cases, collaboratives of districts serving a common urban area were permitted to join together to propose and implement a single USI. Where this occurred, at least one district in the collaborative was required to meet the 40,000-student requirement. Second, USI districts had to include a high percentage of traditionally underserved students, including minority and economically disadvantaged students. Districts were also required to commit to cost-sharing of 20%, using local funds. Qualifying proposals were reviewed for merit.

The typical funding level for USIs was approximately \$1 million to \$2 million dollars per year, based on the number of students enrolled in the area served by a USI. The Houston USI, one of the largest, has a five-year (1999-2004) budget of \$15 million dollars, including cost-sharing funds and funds leveraged from other federal sources such as Title I, and Eisenhower Professional Development funds.

The overall scale of the USI program is indicated by summary descriptive statistics for School Year 1998-99, the year in which the greatest number of USIs were at or near full implementation.² At that time, USI districts served 4.5 million students, 41% of whom were Black, 38% Hispanic, 15% White, and 6% Asian/Pacific Islander. Sixty-nine percent of students in USI districts received free/reduced-price lunch, 12% were Limited English Proficient, 13% were in Special Education, and there was a 31% mobility rate. These students attended 5,559 schools (3,710 elementary, 837 middle, 864 high, and 148 ungraded), with more than 158,000 teachers of mathematics and science (126,018 elementary, 17,143 mathematics in grades 6-12,

² The following statistics are from Kim, J., Crasco, L., Smith, R., Johnson, G., Karantonis, A., and Leavitt, D. (April 2001), *Academic Excellence for All Urban Students*. Norwood, MA: Systemic Research, Inc. Author.

and 15,002 science in grades 6-12). Seventy-five percent of these teachers had participated in some form of professional development. Professional development accounted for 49% of USI expenditures. USIs as a whole used NSF funding to leverage an additional \$547 million for standards-based mathematics and science reform in FY 1999, including \$166 million from districts/schools, \$148 million from states, \$134 million from Title 1, and \$24 million from the Eisenhower program.

USIs conducted activities to change many aspects of K-12 mathematics and science teaching and learning. Six percent of USI funds were used to supplement funding from other sources for the purpose of purchasing new standards-based curricula. High-quality standards-based mathematics curricula implemented by many districts and schools included Everyday Mathematics (grades K-6), Investigations in Numbers, Data, and Space (grades K-5), Mathland (grades K-6), Connected Mathematics Project (grades 6-8), and Interactive Mathematics Program (grades 9-12). New standards-based science curricula included the Full Option Science System (grades K-6), Insights (grades K-6), Science and Technology for Children (grades K-6), and the Science Education for Public Understanding Program (grades 4-9). Many of these curricula were developed during the 1990s with major funding from the NSF. Broad implementation of these curricula required extensive professional development for thousands of elementary and middle school teachers. Additional information on specific reform activities affecting grades K-8 mathematics and science is included in descriptions of the Texas SSI and USIs later in this section.

Although the main focus of entirely new curriculum packages has been at the elementary and middle school levels, major changes have also been carried out at the high school level as part of standards-based systemic reform. A major goal for high schools has been the elimination of low-level mathematics courses that emphasize arithmetic and that have served as academic dead-ends for students, especially for disadvantaged and minority students, who have historically been tracked into such courses in disproportionate numbers. Many USI districts have simultaneously eliminated such courses and implemented new graduation standards. Frequently, districts have gone beyond mandating that students merely take additional credits to requiring that students take more demanding courses (e.g., Algebra I, Geometry, science laboratory courses). Likewise, sustained efforts have been made to increase the number of minority and disadvantaged students taking Advanced Placement (AP) courses. An impressive statistic relative to increased enrollments in higher-level courses is that, for Cohort '93 USIs, the number of students in grades 9-12 taking mathematics courses beyond Algebra I increased from 58% in 1993-94 to 80% in 1998-99. At the same time, as USIs increased overall levels of higher-level course taking, they significantly reduced the disparity in higher-level mathematics and science course-taking between minority and White students.

Linking Student Achievement to Specific Systemic Initiatives

The preceding provides a general background of USIs. More specific information about the design and implementation of individual systemic initiatives is necessary to isolate possible causal links between reform activities and student achievement patterns. It is the exploration of causal links that warrants the attribution of initiative effects on student achievement patterns to specific factors.

In the present study, we did not have the resources necessary to collect first-hand data for assessing issues of attribution. Instead, reliance was placed on information provided by the SIs themselves. Discerning readers of the initiative descriptions in this chapter will find that the information SSIs and USIs were able to provide contains many gaps and is often collected or summarized in ways that reduce the level of confidence with which one may attribute changes in student achievement to specific reform activities. We believe readers require examples of the kind of information the SIs have been gathering on reform implementation so they may appreciate that more needs to be done before the technical potential of the analytic models presented in this study may be fully realized.

Readers should recognize that the preceding summaries of SSI and USI reform activities are selective and do not constitute a basis for summary evaluations of any or all systemic initiatives. Our focus was on using available data to assess SI effects and to identify areas where additional data would increase researchers' capacity to address issues of attribution. In surveying types of assessment data available nationally, it became apparent that Texas was the only state that had 1) multiple SIs (e.g., an SSI and several USIs), 2) an assessment that had been in place and relatively stable in content for more than a year or two³, and 3) tested a majority of students in all grades from 3 to 8 every year.⁴ All of these factors made possible the type of value-added and longitudinal analyses we sought to do.

Using Texas data restricted the purview of the study in important ways. First, Texas assesses science only in grade 8, so science performance in Texas cannot be studied using longitudinal or value-added models. Second, the Texas Assessment of Academic Skills (TAAS) mathematics assessment covers only grades 3-8, so we did not gather information about reform activities targeting the high school grades. Readers should understand that many USIs have invested heavily in improving high school mathematics instruction and learning and that the activities we identify as affecting grades 3 to 8 are often far from the total reform effort undertaken at a given site. Third, although there has been a USI in Houston since 1998-99, we treated Houston as a non-USI because we found no evidence that reform activities by the USI in Houston were initiated in time to have any potential influence on TAAS scores in Spring, 2000. Fourth, we made a decision to do the Texas analysis just once with the database for the English version of the TAAS mathematics assessment. We did not repeat the analysis of the data a second time for the Spanish version of the TAAS mathematics assessment.

The other segment of the system this study cannot address is the substantial proportion of Texas students that are excluded from our analyses for the four reasons noted below. Section 5 (Thorn) discusses in greater detail three groups of students who are not captured in the English-language version of TAAS mathematics assessments. Group 1 includes students categorized as

³ The success of systemic reform during the 1990s resulted in many states replacing traditional with standards-based assessments. If these new assessment systems remain in place for a several years or more, it will substantially increase the number of states that might be examined using the analytic models featured in the present study.

⁴ Texas also has a TAAS assessment that is used as a high school exit examination. Students have their first opportunity to sit for the exit examination in grade 10. This examination is not designed to measure content covered in advanced mathematics courses (including Advanced Placement courses), such as are typically taken by college-bound students during the last two years of high school. Given that SIs generally dedicate many of their resources to increasing enrollment and completion levels in such advanced courses, we decided it would not be fruitful to use the state exit examination to assess SI impacts at the high school level.

Limited English Proficient (LEP). In Texas, this is primarily students whose native language is Spanish. LEP students in Texas who are Spanish-proficient take a Spanish-language version of the TAAS, but the Texas Education Agency does not combine data from the Spanish and English TAAS assessments in a single database. In fact, it was not until 1999 that the state began using Spanish-language TAAS results for accountability purposes. In 1999, the state included Spanish-language TAAS results for mathematics and reading at grades 3 and 4 in the measures used for accountability purposes. In 2000, the use of Spanish-language TAAS test results for state accountability purposes was expanded to include mathematics, reading, and writing for grades 3-6. However, although the state began combining TAAS Spanish-language and English-language mathematics test results for accountability purposes in 1999, the database used for this study contains only data for English-language test takers for all years examined.

Group 2 includes students who are categorized as Special Education students. Especially in earlier years, Special Education students were routinely excluded from TAAS assessments. Group 3 consists of highly mobile students. As indicated in Section 5, some students who would otherwise take the TAAS do not due to their absence from school during the testing period. Finally, Group 4 consists of a group of students who are present at the time of testing but who do not participate (or if tested, their scores are not incorporated in the TAAS *Accountability Subset*—the part of the TAAS database used for school accountability). These are students who entered school too late in the year, or moved in and out too much, to have received enough of their instruction at the testing school to accurately reflect the impact of the school's instruction. For example, in a typical year, 1996-97, the TEA exempted students from the TAAS accountability subset if they were not enrolled in a given school for six weeks or more of the school year.

Section 5 (Thorn) contains detailed information on inclusion rates of student demographic subpopulations for the English-language version of the TAAS mathematics assessment from 1994 to 2000. That analysis shows that the inclusion rate for the English TAAS mathematics assessments rose overall during the period studied. Part of this change was due to the fact that the TEA steadily reduced the number of years during which schools were allowed to classify native Spanish-speaking students as LEP for TAAS purposes. Nonetheless, as Table 4.1 shows, even as late as 1999-2000, students in the three demographic subgroups identified above were excluded at a considerable rate from the part of the TAAS database used for accountability. Furthermore, when broken out by race, African American and Hispanic students are much more likely than White students to be excluded from the TAAS testing and the TAAS-based accountability system.

Table 4.1

Number and Percentage of Enrolled Students in Demographic Subpopulations Excluded From Accountability Subset for the English Version of TAAS Assessments for Grades 3-8 and 10, by Ethnicity: 1999-2000

TAAS Inclusion/Exclusion Category (Grades 3-8, 10)	STATEWIDE	African-American	Hispanic	White
# Enrolled as of Oct 29, 1999	2,108,287	na	na	na
# Tested	1,902,683	260,921	709,722	870,876
% Tested	90.2%	86.6%	87.8%	93.4%
Account. Subset (#)	1,803,541	244,349	673,346	831,714
Account. Subset (%)	85.5%	81.1%	83.3%	89.2%
Mobile Subset(#)	97,033	16,270	34,759	38,229
Mobile Subset(%)	4.6%	5.4%	4.3%	4.1%
# Not Tested	206,722	40,373	98,617	61,539
% Not Tested	9.8%	13.4%	12.2%	6.6%
Absent (#)	12,656	2,410	5,658	5,594
Absent(%)	0.6%	0.8%	0.7%	0.6%
ARD Exempt(#)	149,768	34,950	62,242	51,283
ARD Exempt(%)	7.1%	11.6%	7.7%	5.5%
LEP Exempt(#)	27,422	603	22,633	932
LEP Exempt(%)	1.3%	0.2%	2.8%	0.1%
Other (#)	16,875	2,712	8,083	4,662
Other(%)	0.8%	0.9%	1.0%	0.5%
Total Answer Docs	2,109,405	301,294	808,339	932,415

Source: Adapted from <http://www.tea.state.tx.us/perfreport/aeis/2000/state.html>

Two explanatory notes about the numbers in Table 4.1 are in order before turning to substantive matters. First, as noted in the AEIS (Academic Excellence Indicator System) Glossary (<http://www.tea.state.tx.us/perfreport/aeis/2001/glossary.html#taaspart>) on the TEA website, ARD refers to the Admission, Review, and Dismissal committee that determines the education plan for every student in Special Education. Students who are given ARD exempt status by this committee are not required to sit for the TAAS. Table 4.1 indicates the number and percentage of Texas students in grades 3-8, and 10 who did not take the TAAS and were therefore excluded from school, district, and state accountability subsets in 1999-2000.

Second, the totals for enrollment and TAAS mathematics participation in 1999-2000 are unequal primarily because enrollment figures are established in October, whereas TAAS testing occurs in the Spring of the year. It is also possible for students who move during the testing period to sit for the examination more than one time, thereby increasing the total number of answer documents. The combined row totals for African American, Hispanic, and White do not equal the Statewide row numbers because Statewide numbers include other groups (e.g., Asian

American; Native American) that have been omitted from the Table to focus on the three racial/ethnic groups that represent the overwhelming majority of students in the state.

Table 4.1 shows that African American students in grades 3-8 and 10 in 1999-2000 were more than twice as likely (13.4%) as White students (6.6%) to not be tested on the English version TAAS assessments, with Hispanic students were almost twice as likely (12.2%) as Whites to not be tested. African American students who did complete the test were almost twice as likely (18.9%), and Hispanic students were more than half again (16.7%) as likely, as White students (10.8%) to be excluded from the TAAS accountability subset. A given African American student was more than twice as likely (11.6%) and a Hispanic student was nearly half again as likely (7.7%) as a White student (5.5%) to be excluded from the TAAS accountability subset due to ARD (Special Education) classification. This means that the TAAS accountability subset for any given year exhibits systematic demographic biases that are related to the student population in Texas as a whole. Readers should refer to Section 5 (Thorn) for observations on how test inclusion/exclusion practices affect the ability of researchers or system managers to evaluate the impact of instruction on the entire student population or whole subpopulations.⁵

There is an additional number of Texas students that does not show up in the TEA's database for the English version of the TAAS Mathematics Assessment in grades 3-8, during the 1994-2000 period, students who were excluded from the longitudinal and value-added analyses described in Sections 7 (Gamoran) and 8 (Meyer) because there was only one test score for the student for the years studied. Longitudinal and value-added analyses require at least two data points per student.

To review the main caveats, in reverse order, we ask readers to bear in mind that not all Texas students are included in English version TAAS mathematics tests or our TAAS-mathematics database, that this study is restricted to reform activities affecting only mathematics teaching and learning, and that this study does not address reform activities, policies, or student performance beyond grade 8.

Texas SSI and USIs

The following overviews of the Texas SSI and the USIs in San Antonio, El Paso, and Dallas focus on two dimensions of Systemic Initiatives: SI Policy and Infrastructure (including important changes in policy from the pre-SI period), and SI standards-based instructional reform. Within each dimension, we attend to both the *breadth* and *depth* of a state or district system for mathematics education and reform. Breadth pertains to the "who" of an educational system and reform initiative: How many students, teachers, classrooms, schools, districts are affected by a given policy or activity? Also, among those affected, how are they distributed with regard to region, ethnicity, gender, economic disadvantage, or other demographic criteria? Depth refers to

⁵ Also see E. Fuller (November, 2000) on the Charles A. Dana Center website: <http://www.utdanacenter.org/products/speced1.pdf> Fuller finds that the percentage of Special Education students exempted from school accountability subsets rose at the same time that a new state policy prompted the inclusion of greater numbers of Special Education students in TAAS testing. Fuller also reports that schools appeared on average to include relatively high scoring Special Education students in their accountability subsets, while giving accountability exemptions to Special Education students who scored lower on average.

the “what” and “how” of reform, including factors that influence an actor’s inclination to embrace the goals, values, or behaviors encouraged by reform; resources and social supports provided to enable reform-engaged actors to succeed; and the quality of data gathered about the observable effects of reform activities—especially effects on teaching and learning in the system’s classrooms.⁶

It is important to understand a basic assumption we made about how the dimensions of systemic initiatives, including breadth and depth, relate to issues of attribution. Statewide and Urban Systemic Initiatives are complex, dynamic endeavors in which multiple strands of reform activities proceed simultaneously. These multiple activities may run on different timetables, with the effects of each activity working their way through the system at different rates before manifesting any influence on student performance. Given the emphasis on innovation in the areas of curriculum, professional development, and reform management in SIs, there is considerable uncertainty regarding the implementation and effects of individual activities or of whole reform programs. One result of the high degree of innovation is that the impact of a given reform activity on student achievement does not always reflect the level of funding or SI staff effort that reform managers dedicate to it. Also, aspects of the educational system not altered by an SI can change in ways that affect student achievement, and factors exogenous to the state and district educational systems (e.g., economic trends, housing patterns, federal education policies) can affect student achievement in a given jurisdiction quite apart from any purposeful reform that may be underway.

The factors identified above complicate efforts to gather and evaluate evidence of causal links between reform activities and student outcomes. A highly consequential assumption we make to gain analytic purchase amidst complexity is that the main effects of SIs on student achievement will percolate through and reflect back to changes in classroom teaching and learning. It follows from this that anything that can be done to track how reform activities affect the classes students take, the instruction given in relevant classrooms, or student responses to instructional change will help support the attribution of achievement effects to the reform. In cases where direct measures of classroom practice are not possible, the next best thing is data showing that there were reform activities that were relatively broad and deep and that could reasonably be expected to have translated into changed classroom practice. Attribution of effects based on such circumstantial evidence must be regarded as tentative, even when supported by anecdotal evidence from system insiders. The value of recognizing reform activities that were possibly, as well as probably, or probably not, influential is two-fold. First, it provides clues about what especially to look for when synthesizing other existing studies of systemic reform initiatives. Second, and quite importantly, it helps clarify additional types of data that deserve high priority in future research intended to address attribution issues.

⁶ For more information on our framework for describing the breadth and depth of systemic initiatives, see W. H. Clune (1998, December), *Toward a Theory of Systemic Reform: The Case of Nine NSF Statewide Systemic Initiatives* (Research Monograph No. 16). Madison: University of Wisconsin, National Institute for Science Education.

*The Texas SSI*⁷

Practically speaking, the Texas SSI has had two phases. Phase I, which ran from Fall, 1994 (following Spring TAAS testing for the first year included in our TAAS mathematics data set) to Fall, 1998, is by far the most important for our purposes. The fairly close match between the Phase I SSI and the years included in our TAAS mathematics data set is important because it means that Phase I had the potential to affect districts, schools, and student performance for TAAS years 1995-1999, and perhaps beyond.

Phase II ensued shortly after the end of Phase I and remained active as of the writing of this report. We decided not to expend limited resources collecting and analyzing information on Phase II SSI activities because we thought it was improbable that any new activity undertaken by the SSI in Phase II would have exerted substantial effects on student achievement statewide by Spring, 2000. In other words, most of the SSI impacts on student achievement for the period studied were probably already in the pipeline at the time Phase I concluded.⁸

It is important to understand that in this study we are not attempting to measure the direct effects of the Texas SSI on student performance. To do so would require comparing student performance in Texas to that of other states.⁹ The importance of the Texas SSI for the present study is that it provides a context for the Texas USIs, which are of central interest. Understanding the context provided by the SSI puts one in a better position to appreciate the aspects of the grades K through 8 mathematics education system on which USIs chose to focus their resources. For example, as noted below, the Phase I Texas SSI focused heavily on policy-making. This included the creation of new state curriculum standards, plus ostensibly aligned state assessments and accountability systems. Rather than expending limited resources to rework policy issues that had just been recently settled at the state level, USIs tended to focus on building district infrastructure and bolstering professional development for disseminating innovative curricula and instructional approaches.

Phase I SSI: Policy and infrastructure. Phase I of the Texas SSI focused primarily on policy advocacy and infrastructure building. Texas outpaced most other states during the 1990s in the breadth of new curriculum policies established and the speed with which policies were implemented. Both the breadth and speed of policymaking in Texas were enhanced by the fact that Texas had previously engaged in far-reaching basic skills reform that occasioned policy

⁷ We do not attempt to distinguish between the SSI per se and other state organizations (e.g., the state education agency) that obviously play a major role in setting policy for and implementing mathematics education. We have reports that SSI staff were involved in the process of formulating each of the major state policies we discuss. But we do not have data that would permit allocating responsibility or credit for any given feature of state policy to a given organization or agency. Our primary interest is the policies and reform activities themselves. Attempting to disentangle the separate contributions of different organizations to Texas state policy would require considerable research effort and might have little impact on the level of confidence with which one attributes achievement changes to a given policy or reform activity.

⁸ Texas actually had an SSI operating from 1992 to October, 1994, but the NSF terminated it and transferred SSI funding to a second organization, which redesigned and relaunched the initiative. We do not discuss the original Texas SSI that operated from 1992 through the first three quarters of 1994 because we have no evidence that it resulted in any concrete activity that could have potentially affected student TAAS mathematics scores statewide in grades 3 through 8.

⁹ See Webb, Kane, Kaufman, and Yang (2001) for an assessment of SSI effects on student achievement.

changes in many of the same areas that commanded attention in the SSI program. This meant the SSI did not have to start from scratch, but needed only to substitute more ambitious standards-based curricula, assessments, and accountability policies for those that were already in place but oriented to basic skills achievement.

Three strands of state policy were given priority during Phase I of the SSI. First, the state curriculum standards (known as the Texas Essential Elements, prior to the 1998-99 school year) were revised to reflect the increased national emphasis on conceptual understanding and application of knowledge to real-life problems in mathematics. The updated Texas mathematics standards (called the Texas Essential Knowledge and Skills, or TEKS) focused solely on content. Issues of pedagogy were considered too problematic, politically speaking, for the state to risk authorizing innovative pedagogical practices of the kind then being recommended by national organizations such as the National Council of Teachers of Mathematics. Although the then-new TEKS did not tell teachers how to teach to the new standards, the standards themselves did demonstrate to Texas teachers, administrators, and the public how the state wanted mathematics content to place greater emphasis on more challenging problems and concepts, while focusing less on memorization of basic facts.

A second leading policy instrument that was revised during Phase I was the state assessment, known as the Texas Assessment of Academic Skills (TAAS). Since 1995, the TAAS mathematics assessment has moved toward increased emphasis on word problems and on items intended to be more reflective of the national focus on conceptual understanding and the application of knowledge to real-life problems. The biggest changes in the content of TAAS came in 1998-99, when the TAAS was updated to incorporate the TEKS version of state mathematics standards.

In principle, the TAAS is given each year to every Texas student in grades 3 through 8. A similar assessment, which functions as a high school exit examination, is also given starting in grade 10. Annual administration of the same test is a feature of the Texas assessment system that was critical in our choosing Texas as the focus for this study. All other things being equal, more analytic models can be used with data that involve testing on at least an annual basis.¹⁰

The term “alignment” is widely used to characterize the extent to which two or more policy tools or aspects of instructional practice are in accord. For example, in his discussion of the relationship between content standards and assessments, Webb (April, 1997) defines alignment as the degree to which expectations and assessments are in agreement and serve in conjunction with each other to guide the system. Though the extent of agreement or overlap between the TEKS and TAAS specifications is considerable, there is one important area of difference. As noted in Section 6 (Bolt), TAAS test specifications call for items corresponding to numerous objectives (approximately 12-13 per grade level) organized into three domains (i.e., *Concepts, Operations, and Problem Solving*).¹¹ Each objective is also composed of

¹⁰ See Section 8 (Meyer) for an example of an analytic model that is possible under Texas’ annual testing program that would not be feasible in many other states.

¹¹ See, e.g., <http://www.tea.state.tx.us/student.assessment/resources/guides/interpretive/appendix.pdf> on the TEA website for a breakdown of current objectives by domain.

subobjectives. For example, Objective 2 (from the *Concepts* domain) for grade 8 includes the following subobjectives¹²

- Recognize and use rational number properties and inverse operations.
- Determine missing elements in patterns.
- Identify ordered pairs and solution sets in one and two dimensions.
- Apply ratio and proportion.
- Use exponents and properties of exponents.
- Evaluate variables and expressions (formulas).
- Solve simple equations involving integers, decimals, and fractions.

Although some subobjectives may be more difficult than others, the state of Texas, like many other states, does not test all objectives on each test form, nor does the state test all objectives at the same frequency across all test forms combined over time. The analysis conducted by Meyer (Section 8) indicates that item types on which students score consistently high are disproportionately represented in TAAS tests. In the absence of detailed psychometric information on the TAAS, or technically sound studies of alignment between the TEKS and TAAS, analysts should understand that the TEKS and TAAS may fall short of full mutual alignment.¹³

The heaviest lifting in the Texas policy is done by the accountability system. It is the accountability system that contains symbolic and concrete incentives and sanctions that make the TAAS and TEKS consequential for schools, districts, and students. In the Texas accountability system, each school and district receives an annual rating based on dropout and attendance rates and the level of proficiency of students as measured by TAAS. In 1999, for example, the standard accountability ratings, from highest to lowest, were *Exemplary*, *Recognized*, *Academically Acceptable*, and *Academically Unacceptable* (district)/*Low-Performing* (school). A small number of exemplary schools have frequently received minor additional funding.¹⁴ Low-performing schools may be—and at times actually are—under closer state scrutiny, even intensive state intervention in school management. Principals of low-performing schools are frequently removed, and superintendents of Academically Unacceptable districts also run a relatively high risk of removal. However, although the penalties for being designated a low-performing school are thought by many to be powerful motivators, the state has set the performance bar low, thereby precluding any realistic prospect of state sanctions for the vast majority of schools. During the mid- to late-1990s, a school was classified as low-performing

¹² This list of subobjectives is typical of ones made publicly available on the TEA website. Publicly available lists of subobjectives do not necessarily include all subobjectives, provide a full description of any given subobjective, or reflect order of difficulty, although this later type of information is integral to test specification and design.

¹³ An alignment study of TAAS and TEKS over time would provide useful information for assessing student performance patterns, including the degree to which changes in student performance on specific item types corresponds with aspects of content that receive special emphasis in SIs. Examples of technically viable alignment studies of this kind are found in Webb (April 1997); Porter and Smithson (December 2001); and Blank, Porter, and Smithson (July 2001).

¹⁴ For example, in 1997-98, performance awards averaging \$2,430 were given to a mere 823 of approximately 6,300 schools statewide. This equaled approximately \$4.12 per student in award schools. We have heard but not investigated reports that principals and superintendents were given substantial cash bonuses for increased TAAS achievement in some years between 1990 and 1995.

only when fewer than 35% of its students or its lowest-performing demographic subgroup (as defined below) failed to meet minimal expectations on the TAAS. In 1996, for example, only 67 of approximately 6,300 Texas schools were classified as low-performing schools.

Perhaps more important than the concrete incentives is the fact that school and district performance is constantly subject to public review. Information on school and district performance is widely disseminated by the TEA and media. This gives parents, the business community, and others information they can use to advocate for improvement when schools or districts fall short of other stakeholders' expectations. In Texas, as elsewhere, school and district academic achievement attracts considerable public attention for many reasons, not the least of which is the widespread belief that school/district academic achievement affects real estate values and the local economy generally.

The Texas system for school and district accountability has an additional feature that was unique to that state for the period studied and that potentially has major implications for educational equity. Texas state law requires that student achievement at the district and school level be disaggregated to compare achievement between students who differ by race (i.e., White, Black, Hispanic, Native American, Asian/Pacific Islander), income (i.e., students who do and do not receive free or reduced-cost lunch), and gender. To the extent that state accountability ratings are consequential to schools and districts, it is highly significant that such *ratings are based on the scores of the lowest-performing demographic subgroup* in a school. Thus, for example, if the low-income students in a school score lower than all other subgroups (e.g., Black, and Hispanic, and male students), then the school's entire rating is the one corresponding to the performance level of the low-income subgroup. This represents a considerably more rigorous accountability standard than would result from, for example, basing ratings on the average score for all students.¹⁵ More importantly, the extra accountability pressure created by disaggregating data is all focused on improving the performance of students in demographic groups that have historically lacked equal access to high-quality instruction. This particular policy of Texas appears to represent a potentially important strategy for ensuring that all students benefit from successful educational reform. Studies concerned primarily with attributing student performance effects to state-level policies and reform initiatives would be well advised to carefully investigate whether and in what ways the TEA's method of data disaggregation affects the student populations receiving highest priority in district and school instructional programs and reform activities.

Individual students may be affected by their TAAS performance. Sections 6 (Bolt) and 8 (Meyer) discuss intricacies of the procedure the TEA uses to calculate the number of items a student must answer correctly in a given year to meet minimum expectations for a given objective, or to pass the entire test. State policy requires schools to provide students with increased instructional time in subject areas that students fail on the TAAS. We are aware of no

¹⁵ An alternative and even more rigorous basis for defining demographic subgroups for accountability in Texas would be to further disaggregate data by income and gender within racial subgroups. This would yield separate scores and ratings based on finer demographic distinctions (e.g., Low-income, White, Male; Low-income, White Female; Low-income Hispanic Male; Low-income Hispanic, female, etc.) Section 5 (Thorn) discusses the additive nature of performance advantages and deficits as these relate to income, race, and gender. This discussion suggests how the overall impression of school and district performance achievement in Texas would be affected by use of the alternative method of disaggregating achievement data.

state policy that prescribes concrete rewards for grades 3 through 8 students who score high on TAAS. The grade 10 TAAS, not included in this study, is used as a high school exit examination.

In addition to attempting to influence state policy, the SSI invested heavily in developing organizational infrastructure, such as professional networks, for supporting standards-based instructional reform in mathematics and science. From the outset, the SSI decided that the large size of the Texas system rendered it impossible to have extensive direct interaction with a significant number of the state's teachers of mathematics and science.¹⁶ Rather than dedicate resources to maximizing direct delivery of professional development to teachers and administrators, the SSI gave top priority to influencing the amount and type of professional development provided by other "intermediate" organizations. Among the intermediary organizations that collaborated with the SSI were the state's 20 regional education service centers, the state's 30 largest school districts, other NSF SIs and CPMSAs (Comprehensive Partnerships for Mathematics and Science Achievement), teacher and administrator professional organizations, and institutions of higher education (including especially, Teacher Preparation Programs). The Texas SSI was also unique in the extent to which it was designed to give support to and draw strength from other federal and state programs affecting standards-based reform. This special relationship was due to the fact that the SSI was located in and managed by the same organization that administered many related federal (i.e., Title I, Goals 2000, Eisenhower Professional Development for Mathematics and Science, Regional Comprehensive Assistance Center Consortium, McKinney Homeless Assistance) and state (Technology Education, Center for Educator Development (for state-funded professional development for mathematics and science teachers, and for implementation of state TEKS)) programs, and other research and curriculum dissemination initiatives in mathematics and science. The presence of so many programs and initiatives related to standards-based mathematics and science reform in a single organization, as well as the organization's ability to manage the programs in ways that were often mutually supportive, was due in no small part to the organization's leadership. This included a charismatic Principal Investigator who had a reputation for being politically astute and well connected and who demonstrated exceptional sophistication about strategic planning and action. It also included other SSI staff members who were well integrated into the state's professional organizations for mathematics, science, and administrators.

The SSI has a number of activities designed to make deep impacts on instruction in relatively small numbers of schools.¹⁷ However, the bulk of SSI resources appear to have gone into several activities designed to reach large numbers of schools and educators.¹⁸ Many of these activities are internet-based, internet-supported, or internet-coordinated. This is true of activities related to the TEXTEAMS initiative, which was designed and managed by the SSI to foster professional development for dissemination and implementation of the state curriculum

¹⁶ In 1997-98, Texas had approximately 7,500 teachers of mathematics at the middle school level, and 10,310 high school mathematics teachers. At this time, there were approximately 105,000 elementary teachers, approximately 22,000 of whom taught mathematics, or science, or both. The SSI also sought to influence the instructional leadership of the state's school principals and assistant principals, numbering approximately 12,000 in 1997-98.

¹⁷ Many of these smaller projects are described on the SSI website. See, for example, <http://www.utdanacenter.org/ssi/pdf/Annual98.pdf> and <http://www.utdanacenter.org/ssi/projects/>.

¹⁸ For example, during 1997 the SSI worked with the Alliance School Network. The network included the 22 largest school districts in the state, serving more than 730,000 of the 3.7 million students in the state. See the SSI website: <http://www.utdanacenter.org/ssi/docs/PER.pdf>

standards. One major piece of the TEXTEAMS initiative is on-going professional development events that are sponsored and coordinated by the SSI yet delivered by professional developers from intermediary agencies. Many such professional development events, called TEXTEAM Institutes, are designed to help teachers acquire standards-based content knowledge, pedagogical knowledge, and strategies for shaping instructional programs to simultaneously incorporate TEKS standards and prepare students for TAAS examinations.¹⁹ TEXTEAMS also has a strand of professional development activities that target school administrators and other potential instructional leaders in schools and districts. TEXTEAMS professional development for administrators and other instructional leaders is similar to the TEXTEAMS teacher activities in the way it attends simultaneously to TEKS curriculum standards and TAAS performance. The main difference is that TEXTEAMS activities for leaders are designed to foster school-wide instructional improvement (based on knowledge of the proper relationship between state curriculum standards, assessments, and instructional practices), whereas teacher-oriented activities focus on classroom-level issues.

Two other areas that have received major SSI resources are facilitating standards-based instructional reform in Title I schools and the standards-based reform of teacher preparation programs. SSI efforts to support instructional change in Title I schools appear to have occasioned much more intensive SSI-school interaction than was typical of, for example, TEXTEAM-based interaction. By the end of 1996, the SSI had established a relationship with more than 90 Title I schools.²⁰ In addition to representing a significant area of SSI activity, the SSI's emphasis on Title I illustrates a point made earlier about the strategic sophistication exhibited by SSI leaders in finding ways to make programs funded by the SSI and other sources work together to support standards-based reform. When Title I was expanded to include mathematics, in addition to reading, under the 1994 federal reauthorization of the program, leaders of the Texas SSI seized an opportunity lost on many other SSIs to coordinate the pressure for standards-based mathematics instructional reform from several directions. Some of the analyses demonstrated in Sections 6, 7, and 8 of this report would be well-suited to a study comparing student performance in the Title I schools that worked closely with the Texas SSI to those that did not.

Unfortunately, there is not much data available to use as a basis for assessing directly the impact of any SSI professional development or technical assistance initiative. A major shortcoming of available data is that they do not clearly identify specific teachers, schools, and districts that participated to any given extent in any or all SSI activities. For example, the Texas SSI reported that, as of March 1997, 55,000 educators, including 34,000 teachers (also including, according to our estimates, approximately 15% of the state's middle school mathematics teachers and 14% of the state's elementary teachers of mathematics, or science, or both) had participated in "up to" three weeks of SSI professional development.²¹ This example, common in SI self-reports, is too vague to be used in rigorous technical analyses, such as those demonstrated in the present study. Examples of data that would facilitate more thorough evaluation studies include information on the school and grade level of SI participants, by SI activity; the number of hours devoted to each activity; and information to allow for matching students with their own

¹⁹ See <http://www.utdanacenter.org/ssi/projects/textteams/about/need.html> for an example of how the SSI incorporates these three considerations in TEXTEAMS activities.

²⁰ See, <http://www.utdanacenter.org/ssi/projects/title1/map.html>

²¹ See the SSIs' 1997 Program Effectiveness Review on their website: <http://www.utdanacenter.org/ssi/docs/PER.pdf>

mathematics teachers. It would also be useful to have systematic information on the content and quality of delivered SI professional development activities.²²

The preceding points about the additional data desired for addressing issues of attribution in SIs pertain specifically to professional development and technical assistance activities. Below, we assess the availability of data on the effects of the SSI on classroom instructional practices. We conclude discussion of the SSI with brief consideration of attribution issues that relate to the broader effects of an entire SSI.

Texas Phase I SSI: Standards-based classroom instruction. Good data on classroom practice can play an important role in evaluating the impact of reform activities on student achievement. As noted earlier, we made the assumption in this study that broad and deep effects of SIs on student achievement would likely be associated with observable changes in the content of classroom instruction. There is substantial evidence that student achievement in Texas has risen during the SSI.²³ If we possessed good data on classroom instruction—assuming such data indicated that the SIs had influenced what content teachers cover, how they present it, or which students have access to reformed, standards-based instruction—it would substantially increase the confidence with which we could attribute achievement gains to initiatives such as SSIs and USIs. As it stands, we are aware of no data for Texas for the period studied that would enable us to confidently render systematic descriptions of relevant changes in classroom instruction in affected classrooms, schools, or districts in Texas. Greater use of tools for monitoring and describing relevant features of instructional content in ways that are amenable to rigorous quantitative analysis, as well as qualitative accounts of practices based on robust purposeful samples, would be a valuable addition to the data on all Texas SIs.²⁴

We began this overview of the Phase I Texas SSI suggesting that the SSI was important in this study because it provided a context for the USIs. We then proceeded to summarize Phase I SSI activities in the areas of policy, the building of infrastructure, and direct professional development and technical assistance to districts, schools, and teachers. We conclude here with general observations about the SSI and its implications for efforts to evaluate the impact of any or all Texas USIs on student TAAS achievement.

First, we have seen that the SSI worked with and through the TEA to effect standards-based policies that applied to all districts, schools, teachers, and students in the state. The available SSI data do not tell us whether these policies had differential effects in terms of local district, school, or classroom circumstances (e.g., the racial or socio-economic composition of the student population, or district participation in a USI). However, given the strong emphasis of

²² The Texas SSI, like others, tried to influence the quality of service of professional development and technical assistance providers primarily by designing somewhat standardized activities and providing common training and support to providers. Data on the quality of enacted professional development or technical assistance are not available. An evaluation of SI professional development would probably require the relatively broad collection of standardized (quantifiable) measures of the content of professional development sessions, plus qualitative inquiry into a purposeful sample of a smaller number of activities.

²³ See Webb, Kane, Kaufman, and Yang (June, 2001), and Grissmer and Flanagan (November 1998).

²⁴ For an example of a viable tool for generating quantifiable descriptions of classroom practice see Blank, Porter, and Smithson (July, 2001), or on the website of the Council of Chief State School Officers, at <http://www.ccsso.org/pdfs/newtools.pdf>

the state assessment and accountability system on elevating the performance of the state's lowest-performing students, it would seem reasonable to expect that the state policies under consideration had greater impact on districts and schools whose students started at the lower end of the achievement distribution. As Sections 6 (Bolt), 7 (Gamoran), and 8 (Meyer) all show, USI districts did indeed have relatively low-achieving students when the SSI and their own USIs began. In cases where it is found that USIs posted greater gains than non-USI districts, it is important to remember that USIs were part of a subset of Texas districts that may have been disproportionately affected by SSI-era policy changes. It would have been helpful in this study to have the benefit of other research addressing the possible differential effects of state policy on districts and schools. Because we lacked such information, the main strategy used to tease out USI effects was to compare USI districts with various control groups consisting of some or all non-USI districts. Readers should understand that the potential indirect effects of state policy may differ depending on the control group used for a given analysis.

Similar issues arise when we turn to potential impacts of the SSI's efforts to support district and school professional development and technical assistance delivered by intermediary organizations. We know that this activity focused primarily on the 22 largest districts in the state, as of 1997. This presumably included all Texas districts that were already, or would soon become, USIs. But we do not know how evenly the activities were implemented and whether impacts may have differed systematically in some way in USI districts. We also do not know how evenly SSI-sponsored activity was distributed across schools within affected districts. Better information on district and school participation in SSI-sponsored activities would aid efforts to determine the extent to which student performance differences in USI districts and schools might be attributable to the USIs themselves, as compared to the greater-than-usual effects of SSI policy and reform activities.

An additional wrinkle related to selecting control groups for Texas USIs arises with the 1998 inception of the Texas Rural Systemic Initiative (TRSI). The TRSI somewhat counters the SSI emphasis on large districts by lending extra support to systemic reform activities in the state's most impoverished rural districts. The TRSI was established too recently to have broadly affected student TAAS scores prior to school year 2000 in districts that were given low priority by the SSI and USIs.²⁵ However, future studies may benefit from considering districts participating in the TRSI as an additional comparison group for USIs.

Evidence of standards-based reform in classroom practices is the third area we considered in summarizing the Phase I Texas SSI. Knowing the relative level of participation of teachers in the SSI-related activities and being able to match teachers with their students would substantially increase researchers' ability to identify possible SSI effects of student achievement. Better documentation of actual changes in instruction in SSI-impacted classrooms would permit analysts to go yet further in linking SSI activities to student achievement.

²⁵ The TRSI began working with 20 rural districts in 1998, and added 40 more in 1999. See, <http://www.texasrsi.org/>

The Texas USIs

The focus turns now to the Texas USIs. Our initial plan was to compile equally systematic descriptions of each of the three Texas USIs that had operated for several years or more during the period of time included in the TAAS mathematics data set—Dallas, El Paso, and San Antonio. Such descriptions would have included USI implementation data, down to at least the school level. With good implementation data, we believed it would be feasible to use a framework developed for a previous study²⁶ to devise quantified ratings of reform breadth, depth, and strength in the three USIs. However, as noted previously, the El Paso USI was too overburdened by research studies already underway in its districts to participate in our study. We then turned to Dallas, where we received the necessary permissions. The Dallas central office and USI staff provided us with much appreciated cooperation and sound advice about where we might find useful data. Examination of the preliminary data indicated that implementation data for the Dallas USI was fairly thin. Given that an equally systematic comparison of implementation in all three USIs was no longer feasible, we decided to focus on a single USI. Rather than invest limited resources and time in exploring access and preliminary data in a third USI, we opted for getting as much analytic value as possible out of the Dallas data. We subsequently devoted limited project resources to digging for additional data in Dallas, even making modest attempts to gather original data on USI implementation in the district. To provide some perspective on variation in design across USIs, we note major components of the San Antonio and El Paso USIs before proceeding to a more detailed description of Dallas.

San Antonio USI. Nine districts (Alamo Heights, East Central, South San Antonio, Edgewood, Harlandale, Judson, North East, San Antonio, and Northside) participated in the San Antonio USI (SAUSI). SAUSI received more than \$15 million dollars in NSF funding from September 1996 to September 2001. In 2001, the initiative received \$6 million dollars of additional funding for a three-year Urban Systemic Program.

As of November, 1999, the districts included in the San Antonio USI enrolled approximately 235,000 students in grades K-12 (elementary students, middle school students, and high school students) in 363 schools. At that time, the student population in San Antonio USI districts was approximately 62% Hispanic, 9% Black, and 28% White. Fifty-nine percent of all students were economically disadvantaged. Eight percent were in bilingual or English-as-a-Second Language (ESL) programs²⁷

SAUSI leadership was provided by USI staff members in collaboration with superintendents from the nine participating districts. SAUSI staff leadership was stable during the SAUSI. From 1993, when SAUSI planning commenced, to 1999-2000, the nine participating districts collectively experienced 23 changes in superintendents.

²⁶ See, Clune (December, 1998).

²⁷ For comparison purposes, the state of Texas in SY 1999-2000 had almost 4 million students in grades K-12. There were 7,219 schools in 1,041 districts. Statewide, 43% of students were White, 40% were Hispanic, 14% were Black, and 3% were from other racial or ethnic categories. Forty nine percent of students were economically disadvantaged. Twelve percent were in bilingual or ESL. See the TEA website:

<http://www.tea.state.tx.us/perfreport/snapshot/2000/state.html>

Demographic statistics for the USIs are based on information available at the TEA website:

<http://www.tea.state.tx.us/perfreport/snapshot/2000/pdf/snap004.pdf>

*SI policy and infrastructure.*²⁸ The SAUSI districts collaborated to formulate consistent policies for mathematics and science education across the greater San Antonio area. In the Summer of 1994, participating superintendents identified eight areas they would focus on to lend greater policy support to standards-based reform activities:

1. Helping all students meet state graduation standards in mathematics and science.
2. Integrating mathematics and science content with each other and with other content areas.
3. Focusing on mathematics and science in state-mandated Campus Improvement Plans, including explicit planning for curriculum development, professional development, and accountability.
4. Increasing mathematics and science professional development in all nine districts.
5. Forming a cross-district Policy Development Committee to encourage shared policy supports for reform.
6. Enlisting business partners and the city of San Antonio to support systemic change.
7. Forming a network among administrators to support reform.
8. Increasing high quality, integrated science at the elementary level.

These commitments led to concrete decisions and activities affecting reform, including provision of infrastructure support for reform; for example, the USI employed four mentor teachers. Mentor teachers were on loan from participating districts to the USI. Mentor teachers helped schools integrate standards-based curriculum and instructional practices into their Campus Improvement Plans and implement new curricula. Mentor teachers logged 8,370 hours in 4,336 school visits in 1999-2000. This included a heavy focus on supporting implementation of a new middle school mathematics curriculum.

Another related example of infrastructure building was that the nine districts in combination added 23 new district curriculum specialists for mathematics and science over the course of the USI. District curriculum specialists were approximately evenly divided between mathematics and science. Most specialists had functioned previously as SAUSI mentor teachers.

The SAUSI collaborated systematically and substantively with the city and area business leaders. The San Antonio mayor supported the SAUSI and saw it as an integral part of the area's economic development plan. In 1999-2000, the city funded two positions to increase collaboration between the SAUSI and the city and business. The two people hired with city funds were housed in and supervised by the SAUSI. Area businesses provided substantial numbers of tutors for mathematics and science and important public support for reform.

The SAUSI worked with local institutions of higher education to increase the amount and quality of professional development. In collaboration with the University of Texas at San Antonio, the USI designed sequences of standards-based mathematics and science courses. One sequence targeting middle school teachers, for example, included courses in Algebra, Geometry, and Probability and Statistics.

²⁸ Information on SAUSI design and implementation comes from the San Antonio USI Annual Report: 1999-2000 (San Antonio USI, undated), and the initiative's April 1995 USI proposal to the NSF.

SI standards-based instructional reform. Available information does not summarize all teacher participation in SSI-sponsored or standards-based professional development, but it does list numerous major strands of activity. As is common with SI self-reports, the information provided speaks more to the breadth than the depth of implementation activities. In 1999-2000 the USI coordinated 121 study groups involving 821 teachers. The SAUSI provided study group leaders with two days of training covering group interaction skills, goal setting, and portfolio development. Groups also received some assistance from USI mentor teachers and were given stipends for meeting during non-school time. Groups submitted portfolios to the USI at the end of their projects to demonstrate that they had achieved their goals. A survey of study group participants reported by the USI indicates that almost all teachers saw the activity as affecting their teaching and that it increased their level of collaboration with other teachers. More than 75% of teachers indicated that the study groups were more effective than traditional professional development activities.

Summer institutes represented a major strategy for SAUSI professional development. In Summer of 2000, 403 teachers participated in six institutes. One institute that targeted elementary and middle schools was a five-day workshop that addressed state and national standards, standards-based curricula, equity, assessment, and equity. A second five-day content-oriented advanced institute was conducted for people who had attended the introductory institute in previous summers. The SAUSI sponsored TEXTEAMS institutes (see *Texas SSI, Policy and Infrastructure*, above) for an unspecified number of K-12 teachers. Additional workshops affecting elementary and middle schools focused on the integration of technology into mathematics and science.

Data on school- or classroom-level impact of SAUSI activities is thin. After four years, 269 of approximately 360 in participating districts had become USI participating schools. At a minimum, schools seeking “participating” status in the USI needed to send a campus leadership team to a two-week professional development institute on designing campus improvement plans with a standards-based emphasis. The initiative reported that all but one of the participating schools implemented at least some standards-based mathematics by the end of the fourth year of the USI. These schools employed 11,140 teachers of mathematics and science serving almost 213,000 students. However, we do not know how uniform the level of intensity of curriculum implementation was across or within schools.

The SAUSI is not atypical in juxtaposing summaries of professional development activities with data showing increasing TAAS scores for affected districts, without presenting systematic, methodologically sound analysis to demonstrate causality between the SI and improved student achievement. As discussed in greater detail in the summaries of the El Paso and Dallas USIs, attributing student achievement gains to SIs requires better data on the intensity of SI implementation at the school level. It would be desirable for implementation measures to include representative data on standards-based instructional practices in SI classrooms, as well as district-wide or school-level summaries of teachers’ participation in SI professional development activities.

El Paso USI. The El Paso USI (EPUSI) included three districts (El Paso, Socorro, and Yesleta). The USI was funded by the NSF from September, 1994, through August, 1999. The initiative was subsequently awarded an Urban Systemic Program award, which is still active.

As of November, 1999, the three districts included in the El Paso USI enrolled approximately 135,000 students in grades K-12 (elementary students, middle school students, and high school students) in 168 schools. At that time, the student population in the El Paso USI districts was approximately 84% Hispanic, 3.6% Black, and 12.5% White. Sixty-nine percent of all students were economically disadvantaged. Twenty-three percent were in Bilingual or ESL programs.

The EPUSI enjoyed stable leadership, especially among USI staff housed at the University of Texas at El Paso.

*SI policy and infrastructure.*²⁹ As in other Texas USIs, El Paso expended a considerable amount of its mathematics reform efforts at the high school level. The EPUSI districts eliminated remedial mathematics courses, increasing enrollment in upper level courses, implemented the SIMMS IM (*Systemic Initiative for Montana Mathematics and Science-Integrated Mathematics*) curricula developed by the Montana SSI, and creating end-of-course examinations as mandated by the state.

The El Paso USI engaged in limited policy activity affecting grades 3 through 8. Most of the activity amplified state policy but did so in a way designed to emphasize the more rigorous and challenging aspects of state policy. By 1997, the EPUSI had developed local standards—*El Paso Standards for Academic Excellence*—to extend and clarify national and state standards. The local standards were disseminated to all classrooms, as well as to parents and community groups. At the level of informal policy, EPUSI mentors (see below) worked with teachers to conduct performance assessments to expand teachers' understanding of instructional effectiveness on content not covered on the TAAS. Local textbook adoption was modified to take advantage of changes in state policy that added standards-based curricula (e.g., SIMMS, IMP, CMP) to others that could be purchased with state funds. Finally, EPUSI mentors implemented a new state-mandated teacher evaluation instrument in ways that put special emphasis on standards-based instructional practices and systematic planning for standards-based professional development.

Building infrastructure for standards-based reform received high priority in the EPUSI. The initiative was managed by the El Paso Collaborative for Academic Excellence, housed at the University of Texas at El Paso.³⁰ The director of the collaborative, who served as co-principal investigator of the EPUSI with the superintendents of the three participating districts, played an important role in bringing about a high level of collaboration among the districts and the

²⁹The primary source of information about the design and implementation of the El Paso USI comes from the El Paso Urban Systemic Initiative End of Grant Report: September 1, 1994, through August 31, 1999, with additional information coming from the El Paso USI Program Effectiveness Review reports for 1997, 1998, and 1999. These documents were supplied to us by the NSF. Some of the information is also available on the EPUSI website: <http://www.epcae.org/>

³⁰ The website for the El Paso Collaborative for Academic Excellence is found at: <http://www.epcae.org/>

University. For example, the University became an important source of standards-based professional development for EPUSI districts. The University also worked closely with the EPUSI in redesigning its teacher preparation program to increase the number of certified mathematics and science teachers and the readiness of beginning teachers to engage in standards-based instruction.

Within the three participating K-12 districts, the USI helped engender and sustain a common commitment to improving student achievement and closing the academic gap among historically underserved student populations and others. The USI helped districts become more strategic and systemic in support of standards-based reform. Districts increased commitment to a shared vision of standards-based teaching and learning-informed district professional development programs, coordination of funding streams, the selection of curricula, and planning for new accountability systems.

The EPUSI also facilitated increased networking among the districts' principal actors. The initiative fostered standards-based instructional leadership among principals. The USI used summer institutes and on-going monthly seminars to help administrators improve their ability to recruit and select good teachers, monitor classroom practices, prioritize professional development activities, encourage collegiality and reflective interaction among staff, and use data to create and sustain urgency for change—all in the context of a standards-based vision. Likewise, the USI encouraged networking and collaboration among teachers who were involved in implementing new standards-based curricula.

Another important EPUSI strategy for building infrastructure was the use of a cadre of USI mentors. We have no information as to the number of mentors who were active, or their level of contact at the school or classroom level. Isolating the possible effects of mentors from USI reports is difficult because reports of USI impacts on schools and teachers do not differentiate between activities conducted by mentors and ones carried out by other actors (e.g., external professional development providers). The reports we have suggest that many of the USI activities that resulted in deep, intensive engagement of school staff in standards-based reform were facilitated by USI mentors. Mentors appear to have provided technical assistance as well as professional development. They helped teachers implement new curricula and develop personal professional development plans, and they assisted schools in disaggregating TAAS data for school improvement planning purposes. Mentors also helped EPUSI internal evaluators collect information to monitor and assess the status of reform implementation.³¹

SI standards-based instructional reform. EPUSI reports provide information on overall teacher participation in USI-sponsored professional development. The *El Paso Urban Systemic Initiative End of Grant Report* (undated) indicates that the initiative affected 4,700 (or 83%) of the teachers of mathematics in participating districts. The initiative reports that as of 1998, 71% of mathematics and science teachers had participated in some USI professional development, including 44% who had participated in two to four weeks of such activity. At least some of the professional development involved groups of teachers and administrators from schools rather than individuals. For example, 17 school teams (each consisting of five teachers and a principal)

³¹ The EPUSI Performance Effectiveness Review for 1997 indicates, for example, that 40 USI mentors made 150 classroom observations in 1997.

received 60 hours of professional development on integrating technology into mathematics and science. None of the EPUSI documents available report professional development participation rates at the school level. In the discussion of the Dallas USI below, we note a similar situation there and note analytic limitations stemming from the absence of such information. Related analysis and discussion is found in Section 7 (Gamoran).

The EPUSI also provides some information on the breadth and depth of the impact of EPUSI activities on students and classroom practice. The initiative reportedly reached 146 of 163 schools by 1999. These schools served 115,000 (84%) of all students in participating districts. In another measure, the USI reports that approximately one-third of all students in 1998 had classes with teachers who had received “intensive” USI mentor support. The 1998 Performance Effectiveness Review for the initiative indicates that by Year 4 of the USI, the Connected Mathematics Program curriculum was being used to some extent in all middle schools, Math Investigations was in 41 of 102 middle schools, and 37 additional elementary schools were using some other standards-based curricula.

The EPUSI presents one set of data that represents a rare attempt by an SI to simultaneously summarize aspects of the breadth and depth of SI implementation. The data, shown in Table 4.2 below speak generally to EPUSI impacts on classroom practice at several levels of intensity.

Table 4.2
Implementation of Standards-Based Mathematics Curricula in the El Paso USI

	Intensive Level		Moderate Level		Minimal Level		Nonexistent Level		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Elementary	24	23	42	41	32	31	5	5	103	63
Middle	14	44	5	16	10	32	3	9	32	21
High	8	29	5	18	13	46	2	7	28	17
Total	46	28	52	32	55	34	10	6	163	100

Source: The El Paso Urban Systemic Initiative End of Grant Report, p. 6.

These data represent ratings made by USI mentors using a common rubric. Schools characterized by “Implementation of standards-based curriculum and instructional strategies with reflection and adaptation by 60% or more of the faculty” were deemed to be implementing intensively. A score of “moderate” was given to schools where raters discerned “A strong core of faculty or faculty teams (12% to below 60%) implementing standards-based curriculum and instructional strategies”.³² Lower ratings were given to schools with lower levels of standards-based activity.

Though the rubric used may not have been ideally suited to analyses demonstrated in this report, the data in Table 4.2 suggest that the EPUSI did actually gather implementation data at the school level that could in principle be used to assess the relationship between an important dimension of SI implementation strength and student achievement growth. Our understanding of

³² El Paso Urban Systemic Initiative End of Grant Report: September 1, 1994, through August 31, 1999 (undated), (p. 5).

SI impacts generally would be enhanced by moving further along this analytic path down which the El Paso USI started.

Dallas USI. The Phase I Dallas USI (DUSI) was first awarded NSF funding in 1993. Actual implementation activities occurred from Fall, 1994, through Fall, 1999. The initiative continued as an Urban Systemic Program site in 2000, with a five-year NSF award of \$11.5 million. The Dallas USI included only schools in the Dallas Independent School District (DISD).

As of November, 1999, the DISD enrolled approximately 160,000 students in grades K-12 (101,000 elementary students, 23,000 middle school students, and 36,100 high school students) in 218 schools. At that time, the DISD student population was approximately 52% Hispanic, 38% Black, 8.5% White, 1.49% Asian, and .41% American Indian. Seventy-three percent of all Dallas students were economically disadvantaged. Thirty-two percent were in Bilingual or ESL programs.

The Dallas district as a whole and the DUSI itself experienced frequent changes in leadership during the years encompassed by the DUSI. There were five changes in the district superintendency. Since district superintendents are required by the NSF to act as principal investigators within USIs, top leadership for the DUSI changed each time a new superintendent came to the district. Though district superintendents sometimes take an active role in formulating the basic design and strategy of their USI, daily administration of the initiatives generally goes to co-principal investigators and USI project directors. Co-principal investigators of USIs rarely devote all of their time to the USI. Many co-principal investigators do little more than participate in occasional planning meetings and draw no salary from the initiative. USI project directors are typically assigned full-time to their USI. In Dallas, the USI project directorship also turned over five times from 1994-99. As the following descriptions of SI Policy and Infrastructure, and Standards-Based Instructional reform indicate, frequent changes in leadership for the DUSI were mirrored by equally frequent shifts in USI design and implementation activities. Consequently, it was not until Summer, 1998, that the USI implemented an activity that both appeared to have reasonable potential to affect standards-based teaching and learning in grades 3-8, and for which there existed usable data on teacher and school participation. The Summer, 1998, activities, described below in *SI Standards-Based Instructional Reform*, were the basis for one analysis demonstrated by Gamoran in Section 7.

Although the DUSI may have taken longer than usual to stabilize, many SIs go through an initial period of adjustment where substantial modifications are made to the initiative's underlying theory of action and theory of change. While it is beyond the scope of this study, it appears it might be worthwhile to investigate whether there is a clear relationship between the amount of time USIs take, on average, to coalesce around a coherent, sustained set of activities, and the amount of turnover in superintendents/principal investigators in participating districts. A related question is whether change in USI project directors generally accompanies a change in superintendents. Also, if there were cases where USI project directorship was more stable than the corresponding district superintendency, did USIs with more stable directorships tend to

undergo less frequent or less extensive change in implementation strategies or activities over the span of the initiative?³³

*SI policy and infrastructure.*³⁴ During the five years of the USI, DISD implemented several significant policies affecting mathematics and science that had considerable potential to affect mathematics and science teaching and learning, but not primarily at the grade levels (i.e., grades 3 through 8) included in this study. For example, in 1994 the district simultaneously eliminated remedial mathematics and science courses and began requiring all students to take a mathematics and a science course every year. This primarily affected high school students who otherwise would have been subject to a state policy requiring only two years each of mathematics and science. Another example is a 1996 policy that provided a \$1,000 stipend (e.g., a salary supplement) for certified mathematics or science teachers. DISD internal evaluators indicated that this stipend probably contributed to reducing retirement rates among mathematics and science certified teachers. This was especially significant because a disproportionate share of the district's certified mathematics and science teachers had reached or were closing fast on retirement eligibility. This policy would have had little impact on elementary schools because almost no teachers at that level are certified in mathematics (or science). The policy would have mainly impacted the composition of the district's high school mathematics (and science) teacher cadre, but would also have affected middle schools because many teachers of mathematics at that level also have subject area certification. Although this policy probably resulted in maximizing the proportion of certified mathematics teachers at the middle school level, we would have no way of assessing the likely implications of that factor vis a vis standards-based reform. We are aware of no studies that examine whether mathematics certified teachers are more receptive to standards-based reform, nor studies that determine whether engagement in reform among certified teachers is correlated with teacher age or number of years of teaching experience.

Other district policies adopted during the DUSI had greater potential to affect mathematics at all grade levels. One important policy was district adoption of new standards-based curricular materials. The main mathematics curricula for grades K-8 that were endorsed, encouraged, or mandated by the district during the DUSI were Everyday Mathematics (for grades K-5), and Connected Mathematics (for grades 6, 7, 8). The district also distributed conventional as well as graphing calculators more widely. What is known about the extent and rate of dissemination of these new curricula and materials during the DUSI is related below in the discussion of *SI Standards-Based Instructional Reform*.

Another important district policy was a requirement that all teachers engage in at least 49 hours of professional development annually. Teachers were routinely given paid professional leave and release from classroom duties to pursue professional development opportunities. For the first two to three years of the DUSI, professional development was driven primarily by individual teacher preference. Many teachers chose topics unrelated to mathematics and science,

³³ In this regard, it is interesting to note, for example, that the Dallas USP, which followed and built upon the USI, was entering what has become several years of stable leadership with a clear, sustained implementation strategy at the time we visited the district for the present study. It appears that Dallas might be a productive site for future studies of USP effects.

³⁴ Unless otherwise noted, information on DUSI implementation activities is from the *1994 – 1999 Summary Report* (Dallas Public Schools, November 30, 2000) submitted to the NSF by the DUSI.

or activities that lacked a clear focus on intellectually challenging standards-based work for all students. In Year 4, the DUSI exerted much stronger guidance in the area of mathematics and science professional development. Instead of merely cataloging professional development opportunities and letting teachers take it from there, the DUSI organized a summer professional development institute for mathematics, science, and technology. The nature of the 20 mathematics workshops offered and levels of teachers participation are discussed in *SI Standards-Based Instructional Reform* below. Participation data from this summer institute proved to be the best available for demonstrating one of the analyses conducted by Gamoran in Section 7.

The DUSI undertook various activities to bolster infrastructure for standards-based reform. Most of these activities were short-lived, even though they were originally intended to persist for most or all of the DUSI years. The first activity was a project to develop a CD-based curriculum guidance resource to support school and classroom efforts to design and implement standards-based instruction. The original plan called for extensive staff development to enable teachers to fully utilize the resource. After several years of work that yielded an Alpha and Beta version, the CD the project was terminated. The decision to abandon the project was due in part to the district's increasing appreciation of the NSF's strong support and rationale for using nationally developed and field tested standards-based curricula. District leaders for mathematics and science realized that massive resources would be needed to make all the materials on the CD of equally high quality from a content and pedagogical standpoint.

The second activity was a push to disseminate Total Quality Management (TQM).³⁵ In Year 1 of the USI, TQM training was conducted with school principals only. In Year 2, the program was expanded to 67 schools, in Year 3 to 99 schools, and in Year 4 to 215 campuses. Starting in Year 2, the district provided TQM training to Campus Instructional Leadership Teams (see below), whose members were then expected to repeat the training with their respective school staffs. In 1997-98, the district arranged the school calendar to permit schools to work with teachers for two to three days before the start of school in the fall to engage in TQM-based school improvement planning. The focus in TQM workshops was on using data at the school level to improve instruction. No data are available regarding the extent to which standards-based teaching and learning, or mathematics and science were emphasized as the specific target for instructional improvement, either in TQM workshops or in school-level uses of data. Internal evaluators at the Dallas Independent School District indicate that the main importance of the TQM training proved in retrospect to be that it helped prepare principals assume greater instructional leadership during the DUSI years, when central administration was in constant turmoil and decision-making became increasingly decentralized to newly created subdistricts and schools.

The DUSI tried several strategies for increasing the pool of subject-area expert teachers to support school- and classroom-level standards-based instructional reform. One strategy was to foster a strong emphasis on mathematics and science by guiding Campus Instructional Leadership Teams (CILTs) and other instructional improvement activities in subdistricts and schools. CILTs were charged with using data to pursue instructional improvement at the sub-

³⁵ Some information on TQM training in Dallas during the USI comes from the Dallas Urban Systemic Initiative 1998-99 Annual Report (Dallas Public School District, undated).

district and school level. CILTs were comprised primarily of administrators and department chairs. The DUSI attempted to increase the emphasis CILTs gave to standards-based mathematics and science reform by hiring USI specialists (experienced mathematics and science teachers), who were placed in at least seven of nine subdistrict offices. In addition to providing CILTs with information about new directions in district mathematics and science curricula and instruction, USI specialists provided additional resources and guidance to teachers who showed special interest in reform.

The use of USI specialists to assist CILTs in subdistricts appears to have lasted one year before giving way to a strategy of using the USI specialists to cultivate cadres of teacher leaders in mathematics and science at the school level. The theory of action for this change was that the DUSI would identify teachers who had already shown strong commitment to standards-based instruction, increase their knowledge and authority through relatively intensive, focused professional development, then encourage them to share their perspectives, knowledge, and practices with other teachers in their school. No data were available on the breadth or depth of contact between USI specialists and classroom teachers or schools in relation to these activities.

For 1998-99, the DUSI modified its infrastructure building strategy once again. Plans were made to create positions for three or more USI Lead Teachers in each of the nine DISD sub-districts. This was expected to permit more intensive support for the efforts of schools and teachers to implement the new standards-based curricular materials being promoted by district and USI leadership. However, the district encountered budget shortfalls part way into the year and determined it no longer had the non-USI funds that would have been needed to fully staff the positions.

In January, 2001, we conducted on-site interviews and surveys with seven DUSI staff members who had worked in the various specialist roles that existed in the DUSI during years 3 through 5. By the time of the on-site data collection we had determined that otherwise available data on breadth and depth of school impacts of DUSI infrastructure building activities were too thin to demonstrate the analytic models featured in Sections 7 (Gamoran) and 8 (Meyer) as fully as we would have liked. We sought limited original data consisting of indirect measures to supplement the sparse direct measures otherwise available. The instruments we developed were designed to yield numerical ratings of the relative intensity of various aspects of USI implementation at the school level. The hope was that the specialists still on staff with the DUSI, as a group, would provide ratings that could be used to extend planned analyses. Unfortunately, the effort yielded no usable data because only a few of the specialists on staff at the time of our visit happened to be among those who had been previously focused on DUSI activities targeting mathematics in grades 3 through 8. It is possible that surveys we used for this purpose would have produced usable data had it been possible to administer it to all the DUSI specialists on staff in 1998-99 when the number of specialists peaked and the potential for school-level USI impacts was probably at its height. Our efforts to collect original data based on retrospective reports underscores how important it is for the overall evaluation plan of fast-moving reforms, such as the SIs, to have on-going data collection to capture important implementation activities while they are underway and not to rely heavily on collecting implementation data retrospectively.

Another DUSI strategy for building infrastructure consisted of forming partnerships with other formal and informal educational organizations. The DUSI listed at least 41 partnerships with businesses, museums, foundations, higher education, and the Texas SSI. The following summary of the relationship of the DUSI and the Texas SSI is provided in the *1994-1999 Summary Report* (Dallas Public Schools, November, 2000):

- Texas SSI provided updates and clarification of new state policies, assessments and timelines
- Shared promising practices and latest MST research
- Hosted TEKS for Leaders conference,
- Investigated distance learning opportunities between sites, and
- Shared resources in the area of professional development. (p. 7)

We are aware of no data on the breadth and depth of partnership activities or the impact of these activities on schools or classrooms.

SI standards-based instructional reform. The following excerpt from the DUSI 1994-1999 *Summary Report* (2000) is illustrative of available information on the breadth and depth of standards-based mathematics and science instructional practices in the district and the impact of the DUSI, in particular, on such practices.

Based on classroom observations, interviews with specialists, replacement cost purchases, and teacher survey results, the District estimates that standards-based programs are in almost every campus. But realistically, approximately one third of elementary schools, half of middle schools in mathematics, more than three fourths of middle schools in science, two thirds of high schools in science, and no high schools in mathematics have standards-based (SB) curricular programs in at least one third of their classes. (p. 10)

Reports do not indicate which schools have standards-based classrooms, the quality of standards-based instruction that occurs, or to what extent the prevalence or quality of standards-based instruction in the district is associated with USI activities. For example, although we know the Everyday Mathematics curriculum was implemented by an increasingly large subset of district elementary schools during the last two to three years of the DUSI, there are no data as to which schools implemented the curriculum, how many teachers in those schools used it, how the USI facilitated school-level implementation, how instruction changed, or whether different subpopulations of students had equal access to the new curriculum.

1998 Summer Institute. During Summer, 1998, the DUSI coordinated a professional development institute for mathematics, science, and technology.³⁶ Several DUSI and DISD staff indicated that the 1998 Summer Institute was the broadest and possibly most intensive Phase I DUSI activity.³⁷ The Institute included 38 different workshops for mathematics, science, and

³⁶ Participation data for the 1998 Summer Institute are based on records supplied to us by DISD and DUSI staff.

³⁷ Isolating the effects of the 1998 summer workshops is complicated by the fact that significant numbers of teachers received potentially influential professional development through the DUSI in other years, and many teachers pursued professional development in other settings (e.g., institutions of higher education) during each year of the

technology. The 38 workshops had a combined 552 attendees. The mathematics workshops—of which there were 20—had 417 attendees. Of the 20 mathematics workshops, eight met seven hours per day for five consecutive days. Four met for three days. Teachers received a \$300 stipend for attending five-day workshops, and \$180 for three-day sessions. Twelve of the 20 mathematics workshops 12 targeted elementary and middle schools. Among elementary and middle school mathematics workshops, six focused on Connected Mathematics Project, three on Everyday Mathematics, one on the Interactive Mathematics Program, and two on using technology in mathematics instruction.³⁸

Section 7 (Gamoran) demonstrates the use of the data on teacher participation in the 1998 Summer Institute workshops to investigate possible SI effects on student achievement. As Gamoran notes, one should exercise caution in attributing causality to correlations between teacher participation in a single professional development event and student outcomes. Such attribution must be tempered by other information about the participants, including overall levels of professional development and evidence showing that the professional development activity did indeed affect classroom instructional practices.

Next, we turn to a description of the TAAS mathematics data supplied to us by the Texas Education Agency, including steps taken to prepare the data for use in the analytic models presented in subsequent sections of this report.

TAAS Mathematics Data Set, Data Cleaning, and Data Management

The TAAS mathematics data provided by the TEA needed substantial cleaning and preparation before being usable for longitudinal and value-added analyses. A description of the data received from the TEA, as well as the cleaned TAAS mathematics data set used for many of our analyses, follows.

Structure and Characteristics of TAAS Mathematics Data Set

Analyses for the current study are based on data provided by the Texas Education Agency. The original data set supplied by the agency included 10,459,927 records of the 3,584,290 grade 3 through 8 students who took the English version of the TAAS mathematics examination from Spring, 1994, through Spring, 2000. This represents an average of 2.918 records per student. Table 4.3 shows the information included in each record in the TAAS mathematics data set.³⁹

USI. Studies of SI effects would benefit from more detailed information on professional development participation down to the school level if not the classroom level, or grade level, within schools.

³⁸ Technology workshops were conducted by the T³ (Teachers Teaching with Technology) organization. See, <http://www.t3ww.org/t3/index.html>

³⁹ The original TAAS data set was organized by grade and year. We assume the grade level for a given student in a given year matches the grade level of the TAAS examination taken by the student.

Table 4.3

The Field Names and their Descriptions in the TAAS Mathematics Data Set

Field Name	Description
YEAR	Year of Test Administration
DISTRICT	Texas Education Agency District Number
CAMPUS	Texas Education Agency Campus Number
DNAME	District Name
CNAME	Campus Name
SEX	Sex of Student
DISADVG	Economically Disadvantaged (i.e., receives free or reduced-price lunch)
ETHNICT	Ethnicity of Student
ITEMRSP	Individual Item Responses (1=CORRECT 0=INCORRECT) for the entire test
STUID	Student ID Number, Scrambled (to assure anonymity)
TLIMTH	Texas Learning Index Score for Student for the TAAS Mathematics Test.

Characteristics of the Original Data Set and Need for Data Cleaning

All large data sets have holes and inconsistencies. Such flaws are more or less problematic, depending on the analytic models to be applied to the data and the software used to run the models.

An important difference between many of the analyses for the present study—namely, the analyses reported in Sections 7 and 8 and analyses typically reported by the TEA, on their website and elsewhere—concerns the unit of analysis used. TEA academic performance reports typically involve *trend analyses* for which *schools, districts, demographic subpopulations* (e.g., female students, Black students, etc.), or *grade level* are treated as the unit of analysis. In contrast, two analyses in the current study (see Sections 7 (Gamoran) and 8 (Meyer)) use *student* as the unit of analysis. Furthermore, these analyses require student-level data that are *longitudinal*. Certain problems in the data set that may reasonably be ignored in analyses conducted by the TEA must be redressed in order to do for longitudinal, student-level analysis.

One problem with the TAAS mathematics data supplied by the TEA was that a significant number of records had missing values for STUID (student ID). In a non-longitudinal analytic model, such as that typically used by the TEA, and as is used in the IRT (Item Response Theory) analysis featured in Section 6 (Bolt) of this report, such records may be mined for the other information they contain. The grade level of the student is always known, so tests with missing student IDs can still be used in non-longitudinal analyses of grade-level averages at the school, district, or state level. Records that contain values for variables such as ethnicity, sex, and economic disadvantage can also be used in analyses addressing these factors. However, in longitudinal analyses, records with missing STUIDs are unusable because it is impossible to distinguish between records that represent unique students and data that actually pertain to a student with records for one or more test years.

The use of a longitudinal analytic model brought other problems to the surface in the original TEA data set that also required data cleaning prior to analysis. For example, evaluating the status of the data against the needs of the analytic model revealed that some students, across years, had missing or inconsistent values for ethnicity, sex, and economic disadvantage. Ultimately, we identified the eight situations shown below as factors that presented problems for at least some planned analyses.

1. The STUID was missing in a record.
2. Ethnicity value was missing in all records for a given STUID.
3. Economically Disadvantaged value was missing in all records for a given STUID.
4. Sex value was missing in all records for a given STUID.
5. For a single STUID, there existed two records at different grade levels in a single year, or a record in a later year for a lower grade level.
6. A STUID was associated with two or more years of data that were inconsistent as to the sex of the student, this occurred in equal proportions of female and male records.
7. A STUID was associated with two or more years of data that were inconsistent as to the sex of the student, with unequal proportions of female and male records.
8. A STUID had inconsistent ethnicity value in its records.

The Data-Cleaning Process and Decision Rules

The first step in the data-cleaning process entailed identifying all records that contained one or more of the eight problems noted above. The number and proportion of problematic records, by grade level, is shown in Table 4.4.

Table 4.4
Total Problem Records by Grade: 1994-2000

Grade	Bad Record	Total Record	Proportion (%)
3	154,174	1,688,113	9.13
4	126,015	1,698,751	7.42
5	108,657	1,729,157	6.28
6	93,476	1,791,027	5.22
7	90,714	1,797,429	5.05
8	96,415	1,755,450	5.49
Total	669,451	10,459,927	6.40

Further investigation showed that decision rules could be devised to rectify data problem types 7 and 8. For problem type 7, where a given student was reported as both male and female an unequal number of times, we changed the least often reported value for sex to match the most reported value. To resolve problem type 8, where a student record contained more than one value for ethnicity, we created a new ethnicity value—Mixed—and assigned that value to all records of students who originally had more than one value for ethnicity. Table 4.5 shows the number and proportion of records, by grade level, that were still faulty in the database after the above cleaning measures.

Table 4.5
Type 7 and 8 Problem Records, by Grade Level: 1994-2000

Grade	Bad Record	Total Record	Proportion (%)
3	22,607	1,688,113	1.34
4	26,587	1,698,751	1.57
5	28,894	1,729,157	1.67
6	29,660	1,791,027	1.66
7	27,206	1,797,429	1.51
8	21,724	1,755,450	1.24
Total	156,678	10,459,927	1.50

Table 4.6 shows the number and proportion of records by grade level that were affected by problem types 1 through 6. Upon examining numerous examples of these problem types, we determined that no reliable decision rules could be devised to clean the affected records. Consequently, the records associated with Table 4.5 were deleted.⁴⁰

Table 4.6
Type 1 to 6 Problem Records, by Grade Level: 1994-2000

Grade	Bad Record	Total Record	Proportion (%)
3	131,567	1,688,113	7.79
4	99,428	1,698,751	5.85
5	79,763	1,729,157	4.61
6	63,816	1,791,027	3.56
7	63,508	1,797,429	3.53
8	74,691	1,755,450	4.25
Total	512,773	10,459,927	4.90

Several additional data cleaning steps were performed with the retained data. The first step concerned the need to modify the format of the STUID field in the version of the database provided by the TEA. To protect the identity of individual students, the TEA had scrambled the original student identification numbers before releasing the database. The scrambled codes contained non-numeric (e.g., the letter “S”), as well as numeric characters. The former were problematic because alphanumeric fields substantially increase the amount of computer memory and processing time needed—a serious issue with large databases such as the one at hand. To ease data manipulation, all STUIDs were converted to a numeric format in which each student was assigned a new, unique number.

Second, in many instances, where there were multiple records for the same student ID number, not every record had recorded values for Sex or Ethnicity. In such cases, when all

⁴⁰ Deleting records with missing values was warranted on pragmatic as well as theoretical grounds. HLM5 (Raudenbush, Bryk, Cheong, & Congdon, 2000), one of the software analysis packages used for longitudinal analyses, does not respond well to missing data.

values reported for a single ID number for sex or ethnicity was consistent, we assigned the reported value to other records with the same ID number.

One final technical note on data cleaning is in order, given that the data manipulations involved in this data-cleaning process are very difficult—some, perhaps impossible—to implement in common statistical software such as SPSS and S-PLUS2000. The work is much more manageable with relational database management systems. All data cleaning for this study was accomplished using the shareware *MySQL* (www.mysql.com), which supports Structured Query Language (SQL) to access databases. The SAS system, version 8 (SAS8), also supports SQL and would therefore be another option.

TAAS Mathematics Data Set

The resulting cleaned data set, named the *TAAS Mathematics Data Set*, is the one used for both of the longitudinal analyses reported here.⁴¹ Table 4.7 compares the composition of the sample of students for the original and cleaned data sets. It shows that the original and cleaned TAAS data sets are very similar.

Summary

This section has provided a context for analyses carried out in the remainder of the report. Included in the discussion has been a review of the origins and purpose of the National Science Foundation's Statewide and Urban Systemic Initiatives, overviews of four Texas systemic initiatives, and a description of the TAAS mathematics data set used by the study to investigate potential SI effects on student achievement in Texas from 1994-2000. A major objective has been to help readers appreciate the intricacies of attributing changes in student achievement to reform activities in the context of complex initiatives such as SSIs and USIs. We have provided numerous examples of SI implementation data, noting to what degree and in what ways available data do or do not enable us to get full benefit from analytic models that have considerable potential to inform our understanding of SI achievement effects. Sections 5 through 8 of this report each present a different analytic model or approach to investigating properties of the TAAS Mathematics Data Set and possible relationships between the TAAS data and Texas SIs. This is followed by a section that synthesizes findings and issues across analyses, plus a final section of conclusions and recommendations.

⁴¹ As noted earlier, the IRT analysis contained in Section 6 of this report used the TEA TAAS data basically as it was, except that it too used recoded student ID numbers to minimize computer memory demands and processing time.

Table 4.7

Comparison of Number of Student Records in Various Subsamples in Original and Cleaned TAAS Mathematics Data Sets for all Years, 1994-2000

Sample Segment	Number of Records in Original TEA Data Set	Average TLI Score (SD) for Original Sample Segment	Number of Records in Cleaned TAAS Math Data Set	Average TLI Score (SD) for Cleaned Sample Segment
Whole sample	10,459,927	76.45 (13.81)	9,947,265	76.52 (13.76)
Grade 3	1,688,113	75.55 (14.27)	1,556,546	75.61 (14.24)
Grade 4	1,698,751	76.91 (13.29)	1,599,322	76.94 (13.25)
Grade 5	1,729,157	78.27 (13.45)	1,649,393	78.30 (13.42)
Grade 6	1,791,027	76.85 (13.40)	1,727,211	76.89 (13.37)
Grade 7	1,797,429	76.02 (14.03)	1,733,920	76.08 (13.99)
Grade 8	1,755,450	75.13 (14.14)	1,680,873	75.27 (14.06)
Female	4,716,030	76.82 (13.39)	4,969,635	76.80 (13.37)
Male	4,724,794	76.24 (14.16)	4,977,630	76.23 (14.13)
Black	1,218,122	69.86 (15.52)	1,334,952	70.04 (15.47)
Hispanic	3,117,962	73.31 (14.62)	3,253,664	73.41 (14.57)
Other	137,847	83.76 (9.63)	175,313	83.62 (9.61)
White	4,948,921	80.03 (11.56)	5,062,714	79.98 (11.58)
Mixed	NA	NA	120,622	76.10 (13.47)
Missing Ethnicity	1,030,705	75.62 (14.09)	NA	NA
Economically Disadvantaged	4,079,063	72.40(14.93)	4,049,481	72.42 (14.92)

References

- Blank, R., Porter, A. C., & Smithson, J. L. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science*. Washington, DC: Council of Chief State School Officers.
- Clune, W. H. (1998, December). *Toward a theory of systemic reform: The case of nine NSF Statewide Systemic Initiatives* (Research Monograph No. 16). Madison: University of Wisconsin, National Institute for Science Education.
- Dallas Public Schools. (2000, November). *1994-1999 Summary Report*. Dallas, TX: Author.
- Grissmer, D. W., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas (microform): Lessons from the states*. Washington, DC: National Education Goals Panel, U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.
- Kim, J., Crasco, L., Smith, R., Johnson, G., Karantonis, A., & Leavitt, D. (2001). *Academic excellence for all urban students*. Norwood, MA: Systemic Research, Inc.
- McLeod, D. G., Stake, R. R., Schappelle, B. P., Mellissinos, M., & Gierl, M. J. (1996). Setting the Standards: NCTM's role in the reform of mathematics education. In S. A. Raizen & E. D. Britton (Eds.), *Bold ventures. Volume 3: Case studies of U.S. innovations in mathematics education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- National Commission on Excellence in Education. (1983). *A nation at risk. The imperative for educational reform*. Washington, DC: U.S. Department of Education.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Porter, A. C., & Smithson, J. L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states—One hundredth yearbook of the National Society for the Study of Education, Part II* (pp. 60-80). Chicago: University of Chicago Press.
- Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws & T. J. Cooney (Eds.), *Perspectives on research on effective mathematics teaching* (pp. 96-113). Hillsdale, NJ: Erlbaum.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, (2000). *HLM 5: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

- Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. H. Fuhman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). Bristol, PA: Falmer.
- Webb, N. L., Kane, J., Kaufman, D., & Yang, J.-H. (2001). *Study of the impact of the Statewide Systemic Initiatives Program*. A Technical Report to the National Science Foundation. Madison: Wisconsin Center for Education Research.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Wisconsin.
- Zucker, A. A., & Shields, P.M. (Eds.). (1998a). *SSI Case Studies, Cohort 1: Connecticut, Delaware, Louisiana, and Montana*. Menlo Park, CA: SRI International.
- Zucker, A. A., Shields, P. M., Adelman, N. E., Corcoran, T. B., & Goertz, M. E. (1998b). *A report on the evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program*. Menlo Park, CA: SRI International.
- Zucker, A. A., & Shields, P. M. (1997). *SSI strategies for reform: Preliminary findings from the evaluation of NSF's SSI Program*. Menlo Park, CA: SRI International.
- Zucker, A. A., & Shields, P. M. (Eds.). (1995, December). *Evaluation of the NSF's SSI Program. Second-year case studies: Connecticut, Delaware, and Montana*. Washington, DC: National Science Foundation.

