

**Comparisons of NAEP Mathematics Score Trends and Patterns for SSI and Non-SSI States**

**Norman L. Webb**

**(nlwebb@facstaff.wisc.edu)**

**Wisconsin Center for Education Research**

**University of Wisconsin, Madison**

**Jung-Ho Yang**

**(jyang@skku.ac.kr)**

**SungKyunKwan University**

A paper presented at the Annual Meeting  
of the American Educational Research Association  
New Orleans, LA, April 2002

The research reported herein was supported by a grant from the National Science Foundation to the Study of the Impact of Statewide Systemic Initiatives (REC-9874171) at the Wisconsin Center for Education Research. The opinions expressed herein do not necessarily reflect the position, policy, or endorsement of NSF, the Wisconsin Center for Education Research, or the University of Wisconsin-Madison.

## Comparisons of NAEP Mathematics Score Trends and Patterns for SSI and Non-SSI States

### Abstract

This study assesses the impact of Statewide Systemic Initiatives (SSI) on student mathematics performance over the time period from 1992 to 2000. Using the State NAEP mathematics data, we focus our analysis on grades 4 and 8 data. The study focuses on three research questions: (1) Are there evident patterns in mathematics achievement in SSI states as compared to non-SSI states? (2) Do SSI states show a greater reduction in the mathematics achievement gap between ethnic minorities and majorities than do non-SSI states? And, (3) How do the changes in performance of the SSI and non-SSI cohort groups differ? The SSI program represents a commitment of over \$400 million by NSF, yet the impact of the program on the learning of students in the SSI states is still unclear. This study investigated the impact of the SSI program on student achievement and attempts to identify those elements that are crucial in designing, implementing, evaluating, and supporting effective statewide systemic reform. In response to question 1, both SSI states (N=14) and non-SSI states (N=13) gained in mathematics composite scores from 1992 to 2000. At grade 8, SSI states gained slightly more than non-SSI states, so that in 2000 there were essentially no differences between the two groups. At grade 4, there were no differences in the mathematics composite scores for any of the three testing times. In response to question 2, the biggest decrease in gaps between White and Black students in SSI states for both grades 4 and 8 was in 1996 compared to 1992. Then in 2000, the gap between White and Black students in SSI states grew in grade 8 and remained the same in grade 4. In contrast, for the non-SSI states, the gap between White and Black students increased in 1996 from 1992 in grade 8 and then stayed the same in 2000. At grade 4 for the non-SSI states, the gap between White and Black students varied by no more than one scale point over the three testing times. The gap between grade 8 White and Hispanic students in SSI states decreased between 1996 and 2000. The gap between grade 8 White and Hispanic students in non-SSI overall between 1992 and 2000 decreased about the same as for SSI states. At grade 4, there was little fluctuation in the gap between White and Hispanic students for either the SSI and non-SSI states. In response to question 3, when considering the same cohort of students and their change in performance from grade 4 in 1992 to grade 8 in 1996, Black students in SSI states made higher gains in two topics—number and operations and algebra and functions—than did White students. Over this same period for a comparable group in non-SSI states, the gain by White students was greater than for Black students. This was not the case for the cohort of grade 4 students in 1996 to grade 8 students in 2000. Over these four years, White students gained more than Black students in both of these topics. Overall, the NAEP data supports the finding of a small improvement by SSI states compared to non-SSI from 1992 to 1996. However, the effects by ethnic groups differed in time, with Black students showing greater improvement between 1992 and 1996 and Hispanic students showing greater improvement between 1996 and 2000.

## Comparisons of NAEP Mathematics Score Trends and Patterns for SSI and Non-SSI States

### Introduction

In 1990, the National Science Foundation (NSF) called for systemic reform in K-12 mathematics and science education in the United States. The Statewide Systemic Initiatives (SSI) Program that embodied this approach to reform was “a major effort by the National Science Foundation to encourage improvements in science, mathematics, and technology education through comprehensive systemic changes in the education systems of the states” (National Science Foundation, 2001). Offered to the states on a competitive basis, the NSF initiatives challenged states to develop clear and ambitious goals for student learning and to create new curricula and instructional materials. The basic elements of this systemic reform are focused on curriculum frameworks, instructional materials and curricula, inservice professional development, preservice professional development, student assessments and accountability, school site autonomy and restructuring, supportive services from districts and the state, and standards-based curricula. The central thesis of systemic reform has been described as follows:

Systemic reformers can bring about a greater degree of alignment of policies of instructional guidance around new standards of learning, thereby producing widespread and substantial gains in the quality of teaching and learning for all students throughout the area affected by the policies. (Clune, 1998, p. 2)

The Statewide Systemic Initiatives represent a commitment of over \$400 million by NSF. The Goals 2000: Educate America Act (1994) and the Improving America’s Schools Act of 1994 (authorized under provisions of the Elementary and Secondary Education Act of 1965) also call for systemic reform. Further, state and local commitments to systemic education reform during this period represent a financial and resources investment many times the federal amount. After the award of the first Statewide Systemic Initiative grants in 1991, 25 states and Puerto Rico made efforts, aided in part by SSI grants, to improve students’ learning of challenging mathematics. These states, and many other states that have engaged in other reform initiatives, have progressed through initial stages of change and are now reaching more advanced stages of reform. In 2000, SSI states had had seven to nine years to produce change. This is adequate time over which to expect some evidence of improved student learning. The study examines the extent to which the SSI effort has effected measurable changes in student achievement, using State National Assessment of Educational Progress (NAEP) data (Webb, Kane, Kaufman, & Yang, 2001).

The State NAEP data of 1990, 1992, 1996, and 2000 provide the only instance in which the same achievement measure has been used at two or more points in time with nearly all of the SSI states and Puerto Rico, 22 of 26 states, as well as with a number of non-SSI states. In evaluating the impact of NSF’s SSI Program, the central research question addressed by this study is: “What differences were evident in mathematics achievement as measured by the NAEP between SSI states and non-SSI states over the period 1992-2000?” In order to do achievement trends analyses of SSI states and non-SSI states in the three test years, we needed to use only those states that participated in all three State NAEP tests. This includes 27 states, 14 SSI states and 13 non-SSI states.

## Background

### The National Science Foundation's Statewide Systemic Initiatives Program

In 1990, NSF instituted a new Directorate for Education and Human Resources to promote and enhance the vitality of mathematics and science education in this country (Webb et al., 2001). In order to achieve a national impact, it adopted an approach that would address entire systems of mathematics and science education, rather than isolated components such as curriculum, professional development, or pedagogy (Clune, 1998; Knapp, 1997; Zucker, Shields, Adelman, Corcoran, & Goertz, 1998). In the strategy it developed for expecting large-scale change, NSF required that the Statewide Systemic Initiatives adhere to high, explicit local and national standards for teaching and learning, such as the newly released National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics* (1989). NSF's objective was to encourage states to seek systemic change in pedagogy, including "hands-on" and "inquiry-based" education, which would relieve students of the unproductive burden of rote learning (Westat & McKenzie Consortium, 1998). Finally, the agency strongly supported newly implemented methods of monitoring by student achievement that were designed to measure students' learning of challenging content.

Beginning in 1991, NSF awarded cooperative agreements to states that proposed initiatives directed toward achieving its vision of systemic reform. NSF granted each successful state applicant up to \$10 million over five years. It recognized that this level of funding was small compared to states' education budgets, but the agency expected these funds to be used as a catalyst for generating the other resources needed to bring about large-scale change in student learning. A total of 26 grants was awarded in three cohorts (Webb et al., 2001):

Figure 1. Three funding cycles of NSF's SSI awards.

1991 Cohort Group ( $N = 10$ )	1992 Cohort Group ( $N = 11$ )	1993 Cohort Group ( $N = 5$ )
Connecticut, Delaware, Florida, Louisiana, Montana, Nebraska, North Carolina, Ohio, Rhode Island, South Dakota	California, Georgia, Kentucky, Maine, Massachusetts, Michigan, New Mexico, Texas, Vermont, Virginia, the Commonwealth of Puerto Rico	Arkansas, Colorado, New Jersey, New York, South Carolina.

States varied in the strategies they adopted to achieve systemic reform. Nearly all states concentrated on mathematics and science. Eleven focused on grades K through 16. Another six focused on grades K through 12. Other states concentrated their initiatives on the middle or primary grades. Only the Montana SSI addressed primarily high school. Eighty percent of the SSIs developed a strategy for supporting teacher professional development and approximately 90 percent had a strategy for creating an infrastructure for capacity building, the two most common approaches to change (Zucker et al., 1998). Other strategies identified by SRI International's evaluation included developing, disseminating, or adopting instructional materials (13 SSIs), supporting model schools (7 SSIs), aligning state policy (16 SSIs), funding local systemic initiatives (9 SSIs), advocating reforms in the education and preparation of teachers (13 SSIs),

and mobilizing public and professional opinion (14 SSIs). This group of SSI states, as well as many non-SSI states that have sought to improve student achievement in mathematics and science, have progressed through the initial stages of reform and are now reaching more advanced stages (Clune, 1998; Clune, Osthoff, & White, in press; Webb, Century, Darvila, Heck, & LeMahieu, in press).

After the first phase of NSF funding, a large question remained about the actual impact of the SSIs on student learning. A number of evaluation reports have been published on the SSI program (Barley & Jenness, 1995; Corcoran, Shields, & Zucker, 1998; Horizon Research, Inc., 1995; Inverness Research Associates, 1995; Laguarda, 1998; Shields, Corcoran, & Zucker, 1994; Shields, Marsh, & Adelman, 1998; Zucker & Shields, 1995; Zucker, Shields, Adelman, & Powell, 1995). Nearly all of the evaluations have focused primarily on generating formative and descriptive information. Very few studies addressed the question of impact on student achievement specifically. In fact, some even questioned whether five years was enough time for any state to mount an effort that would be large enough to impact student learning (St. John, 1999). At best, it was thought a state could engage in capacity building, but the sequence of changes envisioned—from teacher knowledge to classroom practices to student learning—would not, in the time covered by the SSI program, develop on a scale adequate to influence student achievement levels. In addition, states were engaged in other reform efforts besides those funded by NSF, including development of new accountability measures, raising graduation requirements, grade-to-grade promotion, and the alignment of state curriculum standards and assessments.

This research studied the possible impact of the SSI program on student achievement and gleaned the lessons that could be learned about designing, implementing, evaluating, and supporting statewide systemic reform from the State NAEP data on achievement.

### **Research Questions**

In this paper, we present the results of our study to assess the impact of Statewide Systemic Initiatives (SSI) on student mathematics performance over the time period from 1992 to 2000. Using the State NAEP mathematics data, we focus our analysis on grades 4 and 8 data from 1992, 1996, and 2000. In order to obtain complete data for NSF's funding to each state, we needed to limit our analysis to the 27 states that participated all three years in NAEP. We examined the data in terms of three research questions:

- Are there evident patterns in mathematics achievement in SSI states as compared to non-SSI states?
- Do SSI states show a greater reduction in the mathematics achievement gap between ethnic minority and majority than do non-SSI states?
- How do the changes in performance of the SSI and non-SSI cohort groups that were 4th graders in 1992 and 8th graders in 1996 differ?

## Method

Beginning in 1990, the State NAEP provided uniform data for a number of states. Mathematics achievement data along with teacher, student, and school policy information were available for grade 8 in 1990 and grades 4 and 8 in 1992, 1996, and 2000 (Allen, Jenkins, Kulick, & Zelenak, 1997).<sup>2</sup> This time span proved fortuitous for evaluating the impact of an SSI during its implementation. 1990 data and, for some states, 1992 data served as a baseline, providing information about the status of mathematics achievement and related practices just prior to the beginning of the SSI program. Further information available from the 1996 and 2000 NAEP tests allowed for an initial study of the impact of an SSI on both students' mathematics performance and related policies and procedures.

The overall strategy for determining SSI impact involved comparison of SSI and non-SSI states on a variety of variables. Three types of comparisons were made: status at all three years; 3-point trend analysis using data points at 1992, 1996, and 2000, and 2-point trends using data at 1992 and 1996. Because of the voluntary nature of the State NAEP, each of the comparisons used a different sample of states. In addition to examining differences in means for SSI and non-SSI states, where sample sizes permitted, achievement levels of minority and majority groups were compared to assess the extent of the gap-closing purpose of the National Science Foundation in creating the SSI program.

The use of the State NAEP data enabled researchers to explore three types of variables: 1) cognitive achievement as measured by several types of mathematics questions; 2) demographic analysis; and, 3) policy and practice indicators based on teacher, student, or principal questionnaires. In this study, we focused on only mathematics achievement as the dependent variable.

### The Basic Descriptive Analysis

The State NAEP database is complex because of the sampling procedures employed in collecting the data, the weighting procedures, and the way in which the data are structured to compute an estimate of the error in the findings reported (Allen et al., 1997; Grissmer, Flanagan, Kawata, & Williamson, 2000; Webb et al., 2001). In the SSI impact study (Webb et al., 2001), we developed a set of procedures for analyzing the NAEP assessment data. In the present study, we employ similar descriptive trend analyses to examine the gains of students in SSI states compared to students in non-SSI states during three test years of the State NAEP data.

Based on the two groups of states—14 SSI states and 13 non-SSI states—that participated in the State NAEP in the three assessment years, descriptive studies of trends in average scale scores over 1992, 1996, and 2000 and cohort growth in average scale scores from grade 4 (1992) to grade 8 (1996) were done for the total group; these studies also addressed ethnic breakdowns for composite scores, subtopic scores (i.e., number and operations, measurement, geometry, data analysis, and algebra and functions), and the gaps between the different groups (Barton & Coley, 1998; Grissmer et al., 2000).

In summary, SSI states and non-SSI states were compared in a number of ways in order to detect and describe improved student achievement in mathematics over time and to identify differences:

- average mathematics proficiency by year;
- differences in average mathematics proficiency between years for two grade levels, 4 and 8;
- differences in average mathematics proficiency between years for the cohort group; and,
- differences in average scores by content subscales.

## **Results**

### **Trends in Average Scale Scores During the 1990, 1992, and 1996 Test Years**

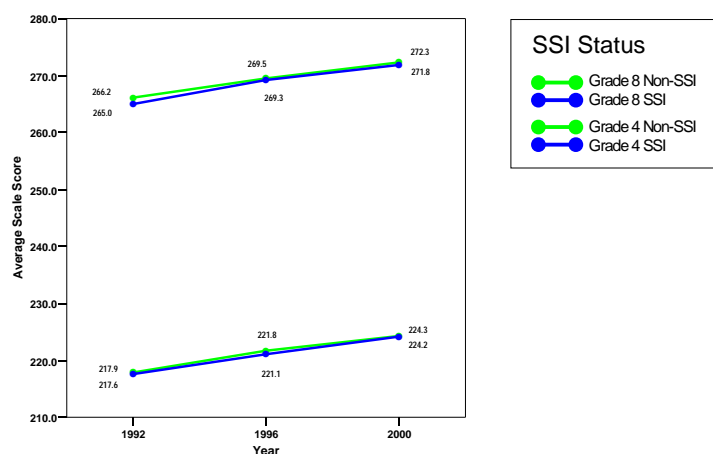
#### ***Composite Scores***

In mathematics in both the SSI states and non-SSI states, students showed continuing increases in average scale scores from 1992 to 2000 for grades 4 and 8. Overall, the average performance across the 14 SSI states was slightly lower than the average for the 13 non-SSI states across all three years. But, the initial gap between SSI states and non-SSI states narrowed by 2000.

The average scale score for grade 8 mathematics from 1992 to 2000 showed a 6.8-point increase for the 14 SSI states and a 6.1-point increase for the 13 non-SSI states (Figure 2). The average increase in the SSI states was slightly higher than in the non-SSI states, by 0.7 points. In 1992, just after the initiation of the SSI program, those states that were to become SSI states scored, on average, lower than non-SSI states on the NAEP grade 8 mathematics test. In 1992, the 14 SSI states averaged 1.2 points less than the 13 non-SSI states. In 1996, the difference was narrowed about 0.2 points, and in 2000 it was slightly less, around 0.5 points.

Similarly in grade 4, both the SSI states and non-SSI states made performance gains in average scale scores from 1992 to 2000. In 1992, the SSI average was 0.3 points lower than the non-SSI average. The performance gap was reduced by 0.1 point in 2000.

Figure 2. Trends in average scale scores, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states).



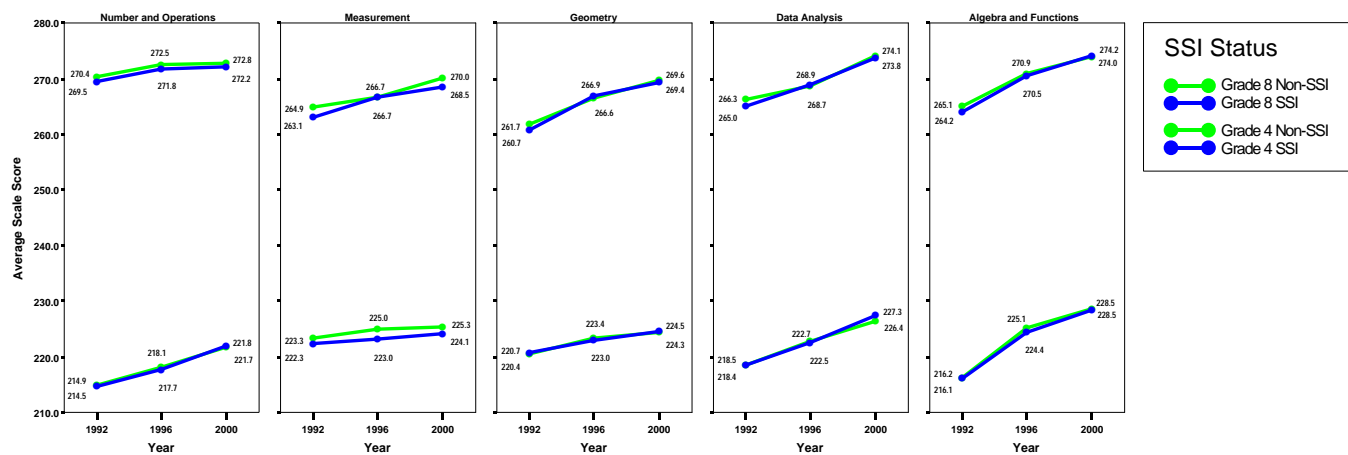
### Subtopic Scores

Very few differences were observed in the pattern of achievement among the five mathematics topics tested by NAEP in the three testing years and between SSI and non-SSI states (Figure 3).

On each of the five mathematics topics, SSI states had average scale scores below those of the non-SSI states for both grade 4 and grade 8 in 1992, but the gaps were significantly narrowed less than 1 point in 1996 and 2000. Overall, students in both SSI states and non-SSI states increased their mathematics performance on each of the five mathematics topics. The greatest gains at both grade 4 and grade 8 levels were in algebra and functions. The smallest gain was in measurement. Grade 8 students in SSI states gained slightly more than grade 8 students in non-SSI states on four of the five subscales and grade 4 students in SSI states gained more on almost all five subscales. However, the grade 8 gaps between SSI and non-SSI states in measurement decreased in 1996, but widened in 2000.



Figure 3. Trends in average scale scores on content strands, by SSI status: Trend Group 92-00 grade 8 and 92-96 grade 4 (14 SSI and 13 non-SSI states).

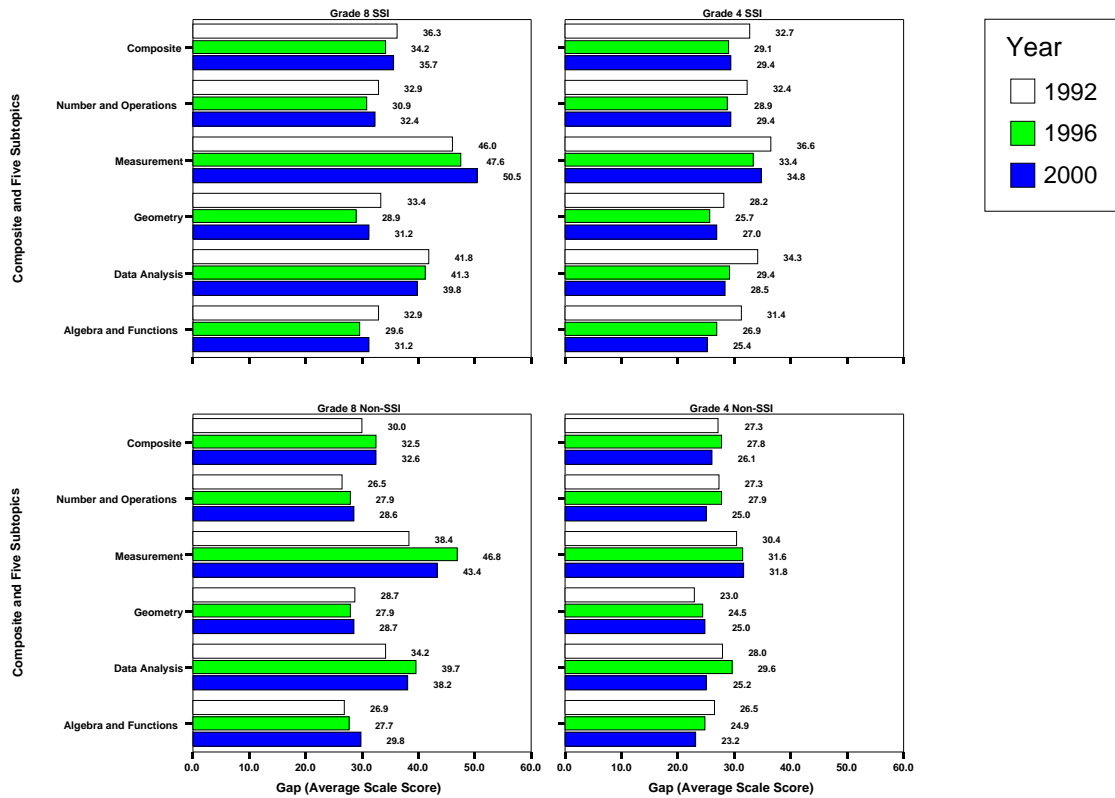


### Gaps Between Ethnic Groups

There were differences in the composite score and in the five content strand scale scores for White and Black students, but the gaps between the two groups remained (Figure 4). Although White students got higher composite scores than Black students and scored higher in the five content strands in both grades, there were no consistent patterns across the six scale scores among SSI and non-SSI states. In grade 8, SSI states slightly reduced the scoring gaps between White and Black students across the three assessment times, especially in 1996. The gaps in non-SSI states increased in 1996 from 1992, but decreased in 2000. For example, in data analysis, the gap increased 5.5 points in 1996, but decreased by 1.5 points in 2000, although SSI gaps narrowed or remained stable.

Gaps in grade 4 scale scores between White and Black students showed a sharply contrasting pattern for SSI and non-SSI states. SSI states reduced the gap from 1992 to 2000, but in non-SSI states the gaps between White and Black students remained in all content strands except algebra and functions. Whereas the gap between White and Black students declined slightly in both grades 8 and 4 in SSI states, in non-SSI states the gap increased or remained constant at both grades.

Figure 4. Differences in average scale scores between White and Black students, by SSI Status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).

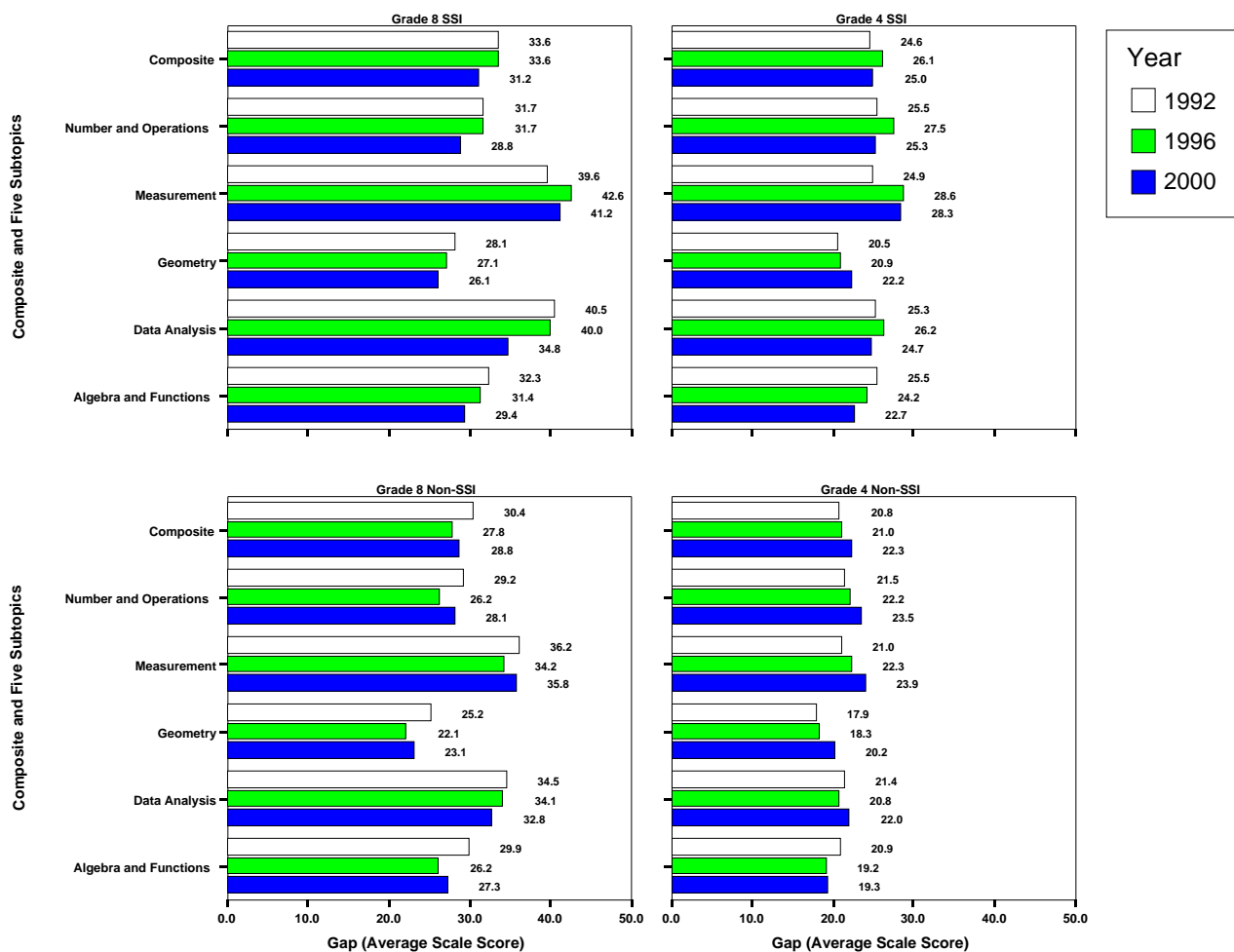


\* Due to the insufficient sample size of these subgroups, results are based on 12 SSI states and 8 non-SSI states.

Regarding score differences of White and Hispanic students, there were different trends for students in grades 4 and 8 (Figure 5). In grade 8, the gaps between these groups in SSI states and non-SSI states narrowed between 1992 and 2000. However, the score gaps for grade 4 White and Hispanic students widened from 1992 to 2000 except algebra and functions. At grade 8 on the measurement content strand, a different pattern of score-gap change emerged: The gap in SSI states dropped in 2000 after a 3-point increase in 1996, but the score gap in non-SSI states increased after a 2-point decrease in 1996.

The gaps between White and Hispanic students in non-SSI states were smaller than those in SSI states in the composite and in the five content strand scale scores.

Figure 5. Differences in average scale scores of White and Hispanic students, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).



\* Due to the insufficient sample size of these subgroups, results are based on 11 SSI states and 12 non-SSI states.

### Cohort Growth in Average Scale Scores from Grade 4 (1992) to Grade 8 (1996)

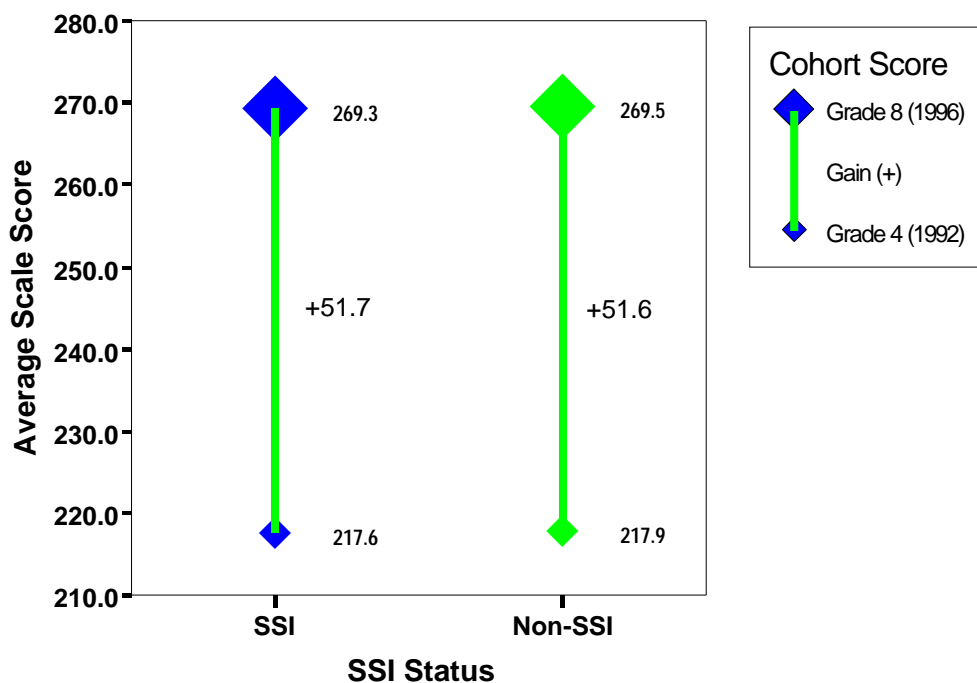
This performance comparison of cohort growth based on two grade levels (grade 4 and grade 8) allowed us to track achievement growth (see Appendix for results of cohort growth from 1996 to 2000).

#### Composite Scores

Both SSI and non-SSI states had substantial cohort growth from grade 4 in 1992 to grade 8 in 1996 (Figure 6). Students in SSI states scored slightly higher than those in non-SSI states in the two assessment years; for example, the grade 4 scale score in SSI states was 217.6, compared to 217.9 points in non-SSI states. After four years, grade 8 students in SSI states scored 269.3

points and their counterparts in non-SSI states scored 269.5. The cohort growth for SSI states and non-SSI states was 51.7 points and 51.6 points, respectively.

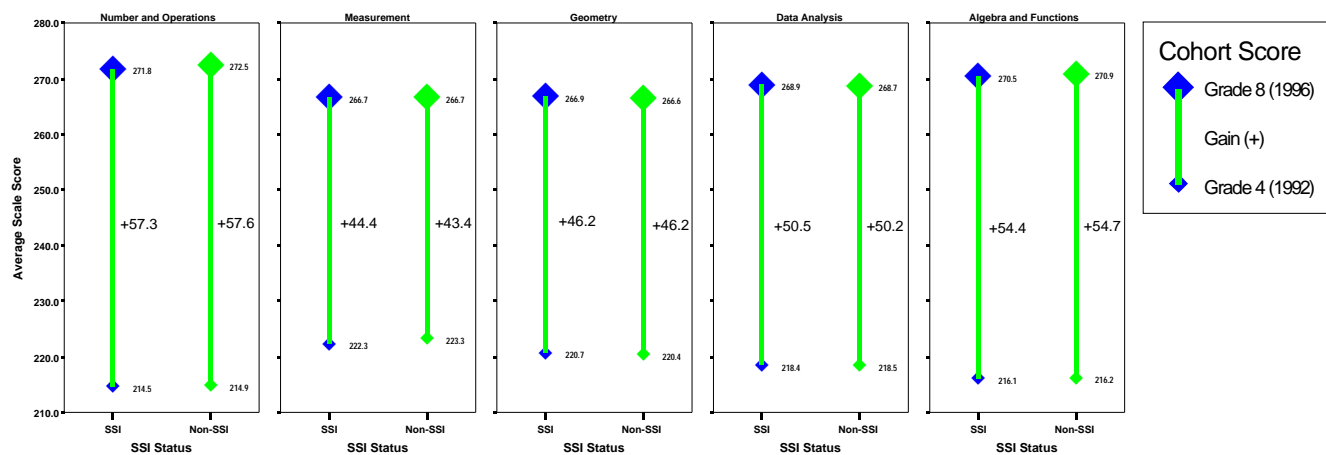
*Figure 6.* Cohort growth in average scale scores from 1996 to 2000, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states).



### *Subtopic Scores*

Both SSI and non-SSI states showed similar pictures of cohort growth in the five mathematics content strands over the four years. Students made a gain from grade 4 in 1992 to grade 8 in 1996. In all five content strand scale scores, students in non-SSI states showed the same cohort gains as students in SSI states; also note that the non-SSI grade 4 students started at a slightly higher point than grade 4 students in the SSI states. The results show some variations of cohort growth across the five content strands (Figure 7). Cohort students in both SSI and non-SSI states made greater gains in number and operations (57 points), algebra and functions (54 points), and data analysis (50 points) than in geometry (46 points) and measurement (up to 44 points).

Figure 7. Cohort growth in average scale scores on content strands, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states).

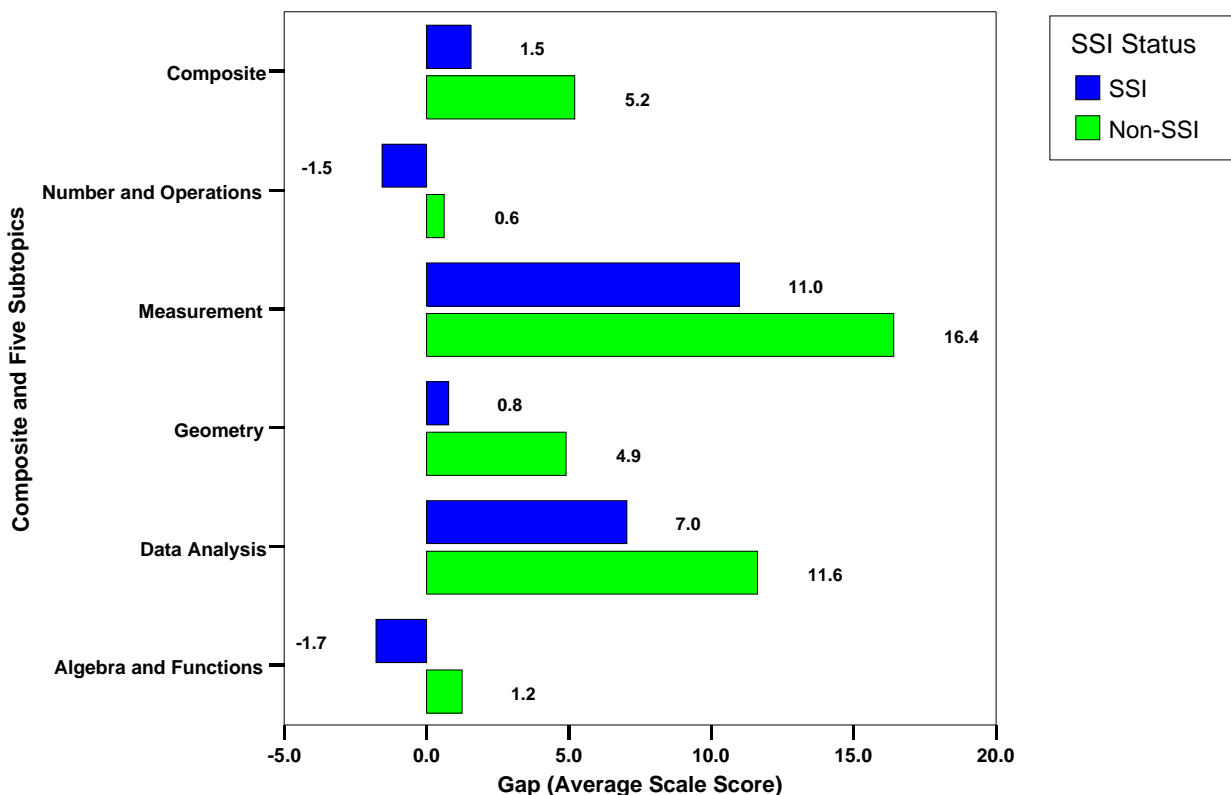


### Gaps Between Ethnic Groups

The results for 1992-1996 cohort growth differences indicate that SSI states were successful in reducing the gap in cohort growth between White and Black students (Figure 8). In the composite and in the five content strand scale scores, cohort growth gaps in SSI states were smaller than those in non-SSI states. The biggest score difference in cohort growth between SSI and non-SSI states was noted in measurement. Non-SSI states had a 16.4-point difference, while SSI states had an 11.0-point difference.

The most interesting picture in cohort growth differences was displayed in number and operations and algebra and functions. In SSI states in 1996, Black students in the cohort of students who were in grade 4 in 1992 showed a greater gain in algebra and functions over four years than White students. In this instance, the gap between White and Black students was reversed, with Black students gaining more than White students. Although not statistically significant, the fact that Black students gained more than White students is noteworthy. In non-SSI states, White students gained slightly more than Black students, 0.6 points and 1.2 points, respectively.

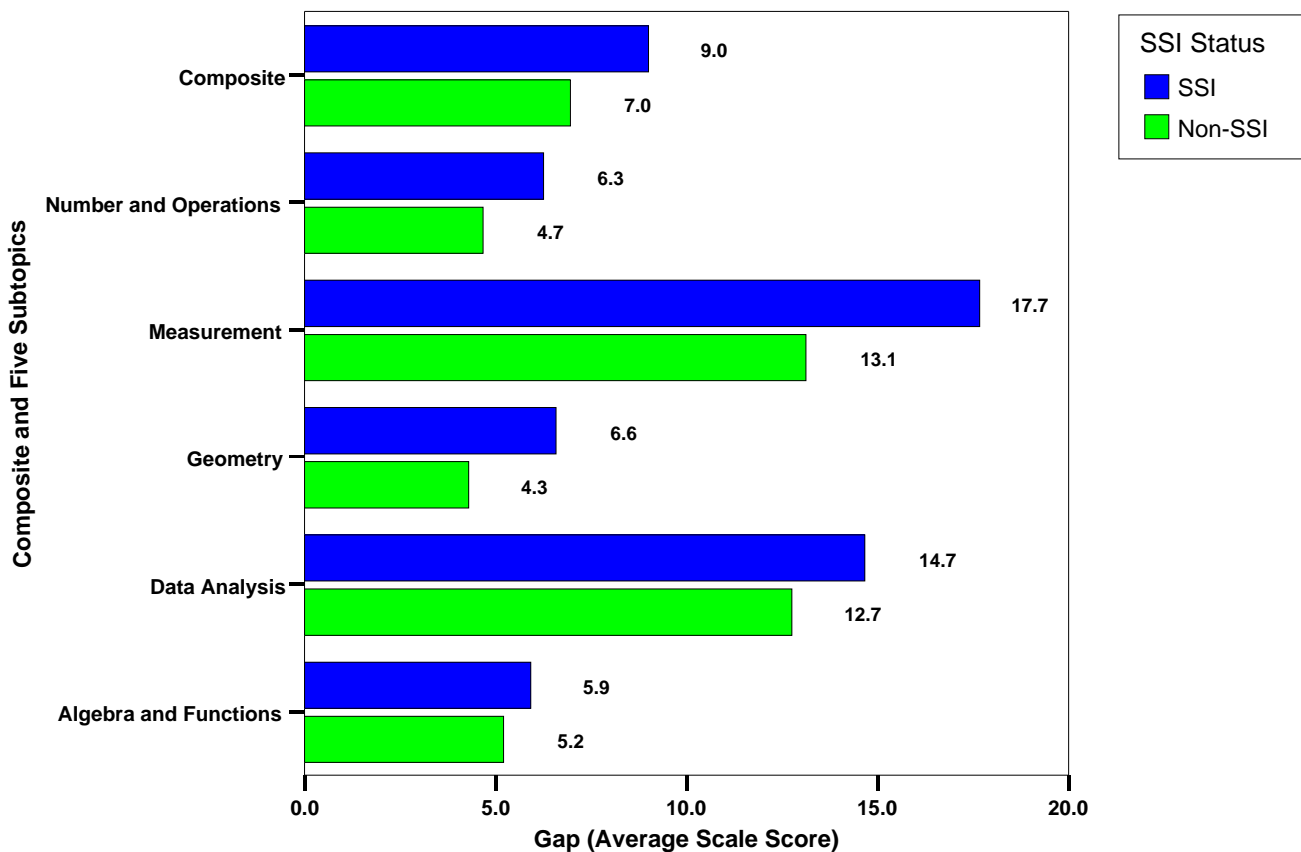
Figure 8. Differences in average scale scores between White and Black students from 1992 to 1996, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).



\* Due to the insufficient sample size of these subgroups, results are based on 12 SSI states and 8 non-SSI states.

In contrast to the patterns of White and Black cohort growth gaps, White and Hispanic cohort students in SSI states did worse in reducing the gap than these groups did in non-SSI states (Figure 9). In five of the six scale scores, the cohort growth differences widened more in SSI states than in non-SSI states: Composite (9.0 points for SSI, 7.0 points for non-SSI); number and operations (6.3 for SSI, 4.7 for non-SSI); measurement (17.7 for SSI, 13.1 for non-SSI), geometry (6.6 for SSI, 4.3 for non-SSI); data analysis (14.7 for SSI, 12.7 for non-SSI); and algebra and functions (5.9 for SSI, 5.2 for non-SSI). However, the difference in measurement was the largest compared to other scales for both SSI and non-SSI states, with Whites outperforming Hispanics. In number and operations, in geometry, and in algebra and functions, the cohort gaps between White and Hispanic students were smaller than in other content strands.

Figure 9. Differences in average scale scores between White and Hispanic students, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).



\* Due to the insufficient sample size of these subgroups, results are based on 11 SSI states and 12 non-SSI states.

### Discussion and Conclusions

This study has presented the results of our evaluation of the impact of the State NAEP mathematics assessments for grade 4 and grade 8 in 1992, 1996, and 2000, and for cohort students of grade 4 in 1992 and grade 8 in 1996 in SSI states and non-SSI states. We have focused our analyses on the 27 states—14 SSI and 13 non-SSI states—that participated in all three assessment years.

### Summary of the Findings

First, the trend differences between SSI and non-SSI states in the composite score and in each content strand were based on the descriptive trend analyses comparing the group means of SSI states and non-SSI states across each assessment year. In general, the results revealed that

substantial student gains in the mathematics composite score and in the five content strands over time were observed for grade 8, grade 4, and the cohort in both SSI and non-SSI states in the assessment years. Considerable improvements were also noted for students by race/ethnicity. However, SSI and non-SSI states showed no clear patterns in the gaps between different ethnic groups across the assessment years. Summaries of performance trends for different subgroups and gaps between Whites and Blacks and between Whites and Hispanics, follow:

- Both SSI and non-SSI states experienced an increase in the average composite scale scores from 1992 to 2000 at grades 4 and 8.
- The pattern in the performance gap between White and Hispanic students was different from the performance gap pattern between White and Black students. At grade 8, the gap between White and Hispanic students improved from 1992 to 2000 on nearly all of the scales for both SSI states and non-SSI states. Only on the measurement scale did the gap increase for both SSI and non-SSI states. However, at grade 4, the performance gap between White and Hispanic students increased on all of the scales for both SSI and non-SSI states, except on the algebra and functions scale for the SSI states.
- Considering the same cohort of students as fourth graders in 1992 and eighth graders in 1996, students in both SSI states and non-SSI states gained the same over the four years, 51.7 and 51.6 respectively.
- The White-Black gap in the gain scores between grade 4 and grade 8 was less in SSI states than in non-SSI states on all six scales. On the number and operations and on the algebra and functions scales, Blacks in SSI states actually gained more between grade 4 and grade 8 than did White students. The greatest difference in the White-Black gap between SSI and non-SSI states was on the measurement scale.
- The White-Hispanic gap in the gain scores between grade 4 and grade 8 was more in SSI states than in non-SSI states on all of the scales except measurement.

## **Conclusion and Future Directions**

Tracking the impact of any large-scale educational reform confronts the researcher with a number of methodological and conceptual issues. In this evaluation study of the impact of NSF's Statewide Systemic Initiatives, our analyses of the State NAEP in 1992, 1996, and 2000 allow us to verify findings from other evaluations and to detect statewide changes consistent with NSF's SSI program. However, the design of the State NAEP does not allow us to attribute observed changes to a specific state program or initiative. NAEP uses a stratified sample of schools within a state and students within those schools to support inferences about the general student population for a state at grades 4 and 8. State NAEP data do not support inferences about individual districts or schools. Most likely, not all of the 2,000 to 3,000 students included in a State's NAEP sample were in schools directly influenced by the state's SSI. In some states, very few sampled students may have directly benefited from the SSI. Information about the scaling-up strategy of each state SSI is needed to determine how likely it is that the observed findings are associated with the systemic initiatives. Given the limitation of the data noted above, our findings regarding the effectiveness of SSI states over non-SSI states in improving mathematics achievement need to be interpreted with care.



Even though our trend analyses of average scale scores suggest that there was evidence in most cases for the differences between SSI and non-SSI states on the overall composite scale and on each of the five content strands, it is unclear whether the differences can be attributable to the relative effectiveness of an SSI. In the descriptive analyses, we did not attempt to determine which SSI-related factors contributed to achievement growth of the SSI states. There are many factors involved in how students learn over time. School structures, home environments, state educational policies, and other factors can affect learning. This will be a primary concern for policymakers intent on increasing student learning at the state level, as well as for researchers trying to identify the relationship between reform initiatives and their effects on students.

Analyzing NAEP data, although very complex, was found to be promising for detecting differences between SSI and non-SSI states. However, the within-group differences among individual states is striking and points to the need to consider individual states, as will be done in further analyses (Webb et al., 2001). The gains by SSI states between 1992 and 1996 reported here could be the beginning of a trend that can only be determined by systemically analyzing the 2000 State NAEP data. In our future study, we will extend current approaches to assessing the effects of state policies and practices related to the SSI program on student mathematics achievement using the State NAEP data of 1992, 1996, and 2000 for grades 4 and 8.

## REFERENCES

- Allen, N. L., Jenkins, F., Kulick, E., & Zelenak, C. A. (1997). *Technical report of the NAEP 1996 State Assessment Program in Mathematics*. Washington, DC: National Center for Education Statistics.
- Barley, Z. A., & Jenness, O. (1995). *The evaluation of the Michigan statewide systemic initiative: Description and sample materials*. Kalamazoo, MI: Western Michigan University, Science and Mathematics Program Improvement Center for Research on At-Risk Students.
- Barton, P. E., & Coley, R. J. (1998). *Growth in school: Achievement gains from the fourth to the eighth grade* (ETS Policy Information Report). Princeton, NJ: Educational Testing
- Clune, W. H. (1998). *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.
- Clune, W. H., Osthoff, E., & White, P. (In preparation). *Theory and practice of systemic reform of mathematics and science education*.
- Corcoran, M. E., Shields, P. M., & Zucker, A. A. (1998). *Evaluation of NSF's Statewide Systemic Initiatives (SSI) Program: The SSIs and professional development for teachers*. Menlo Park, CA: SRI International.
- Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What NAEP state test scores tell us*. Santa Monica, CA: RAND.
- Horizon Research, Inc. (1995, October). *SC-SSI evaluation plan: Plan organized by evaluation questions and related activities*. Chapel Hill, NC: Author.
- Inverness Research Associates. (1995, April). *The New York SSI: Profiling dimensions of the R & D school development process: Inquiry-based math, science, and technology teaching*. Inverness, CA: Author.
- Laguarda, K. G. (1998). *Assessing the SSIs' impacts on student achievement: An imperfect science*. Menlo Park, CA: SRI International.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Science Foundation. (2001). *Statewide Systemic Initiatives*. Retrieved July 10, 2001, from <http://www.her.nsf.gov/esr/programs/ssi>
- Shields, P. M., Marsh, J., & Adelman, N. (1998). *Evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program: The SSIs' impacts on classroom practice*. Menlo Park, CA: SRI International.
- Shields, P. M., Corcoran, M. E., & Zucker, A. A. (1994). *Evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program: First-year report. Volume I: Technical Report*. Arlington, VA: National Science Foundation.

- St. John, M. (1999). Understanding the value of NSF's investments in systemic reform. In N. L. Webb (Ed.), *Evaluation of systemic reform in mathematics and science. Synthesis and Proceedings of the Fourth Annual NISE Forum* (Workshop Report No. 8). Madison: University of Wisconsin, National Institute for Science Education.
- Webb, N. L., Century, J. R., Davila, N., Heck, D. J., & LeMahieu, P. (in press). *Challenges to evaluating systemic reform in mathematics and science*. Madison: University of Wisconsin, National Institute for Science Education.
- Webb, N. L., Kane, J., Kaufman, D., & Yang, J.-H. (2001). *Technical report of the Study of the Impact of the Statewide Systemic Initiatives*. Madison, WI: Wisconsin Center for Education Research.
- Westat & McKenzie Consortium. (1998). *The National Science Foundation's Statewide Systemic Initiatives (SSI) Program: Models of reform of K-12 science and mathematics education*. A pamphlet prepared for the National Science Foundation's Statewide Systemic Initiatives (SSI) Program. Washington, DC: National Science Foundation.
- Zucker, A. A., Shields, P. M., Adelman, N. E., Corcoran, T. B., & Goertz, M. E. (1998). *Statewide Systemic Initiatives Programs: A report on the evaluation of the National Science Foundation's Statewide Systemic Initiatives Program*. Menlo Park, CA: SRI International.
- Zucker, A. A., Shields, P. M., Adelman, N. E., & Powell, J. (1995). *Evaluation of NSF's Statewide Systemic Initiatives (SSI) Program: Second-year report: Cross-cutting themes*. Arlington, VA: National Science Foundation, Directorate for Education and Human Resources.
- Zucker, A. A., & Shields, P. M. (Eds.). (1995). *SSI case studies, Cohort J: Arkansas and New York*. Menlo Park, CA: SRI International.

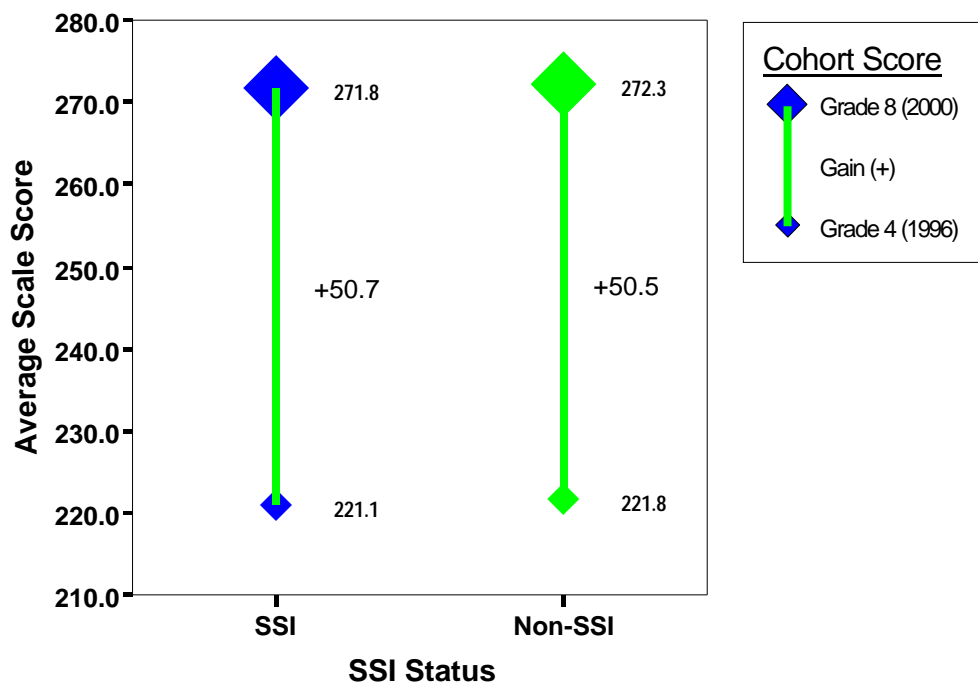
## APPENDIX

### Cohort Growth in Average Scale Scores from Grade 4 (1996) to Grade 8 (2000)

#### Composite Scores

#### Total Group

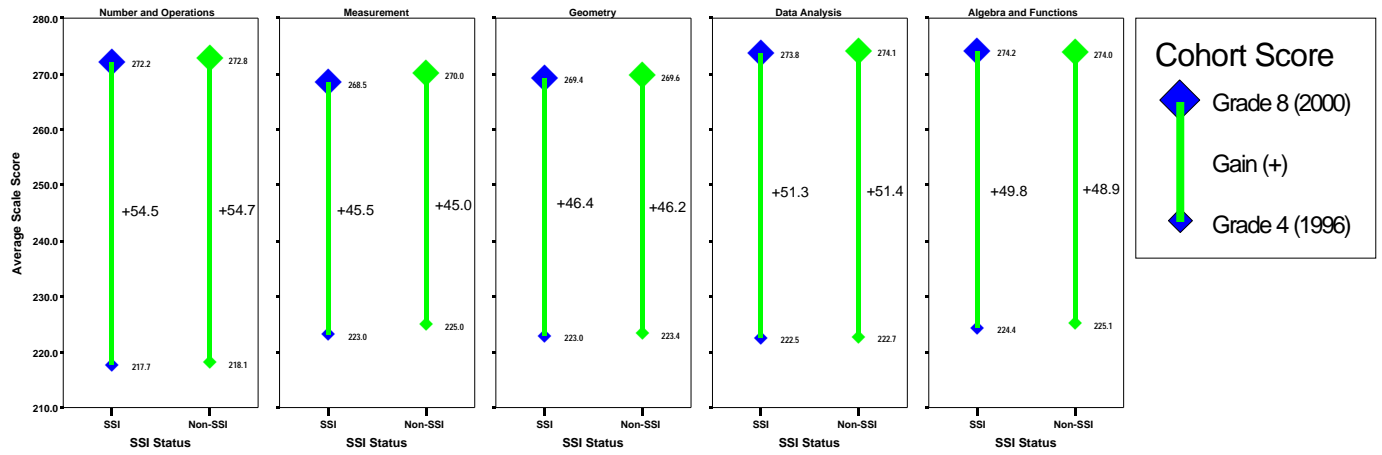
Figure A-1. Cohort growth in average scale scores from 1996 to 2000, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states).



## Subtopic Scores

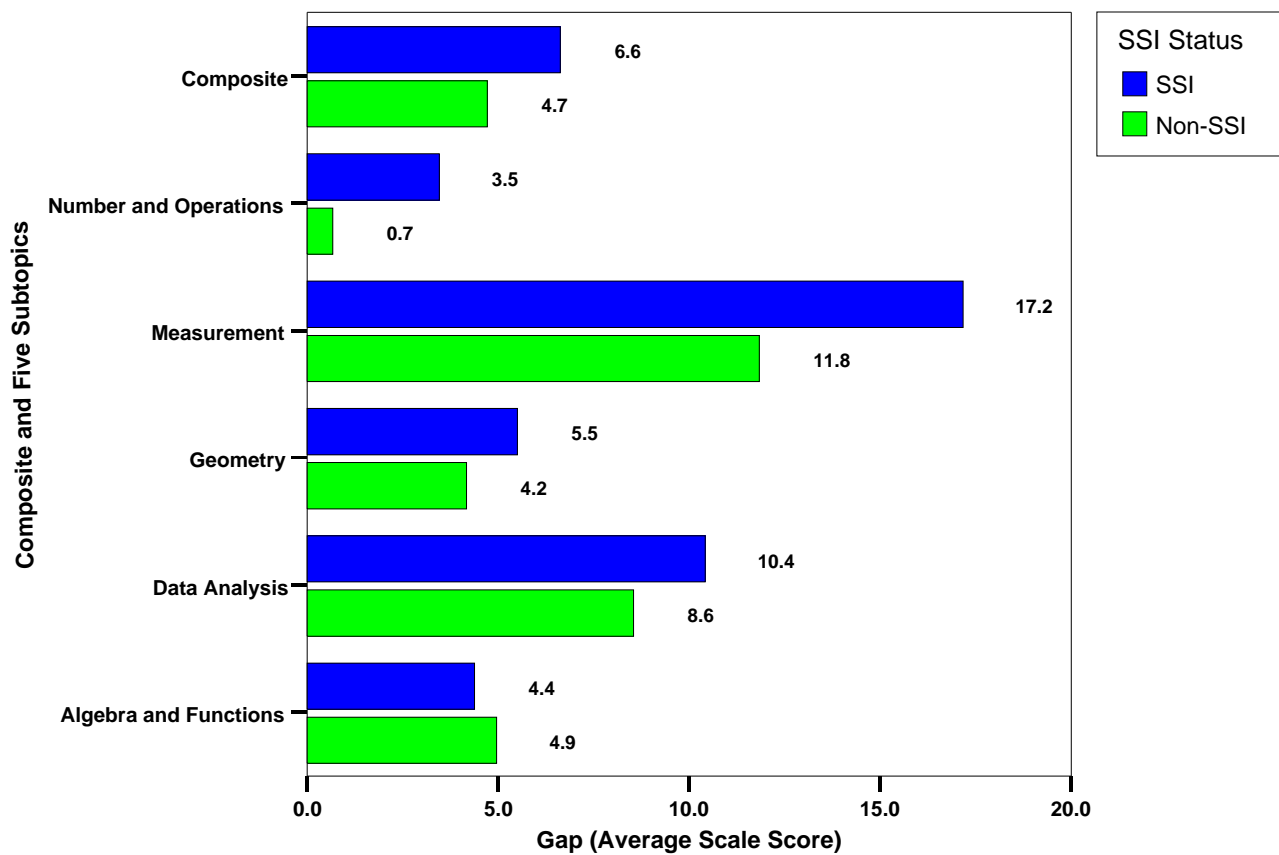
### Total Group

Figure A-2. Cohort growth in average scale scores on content strands, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states).



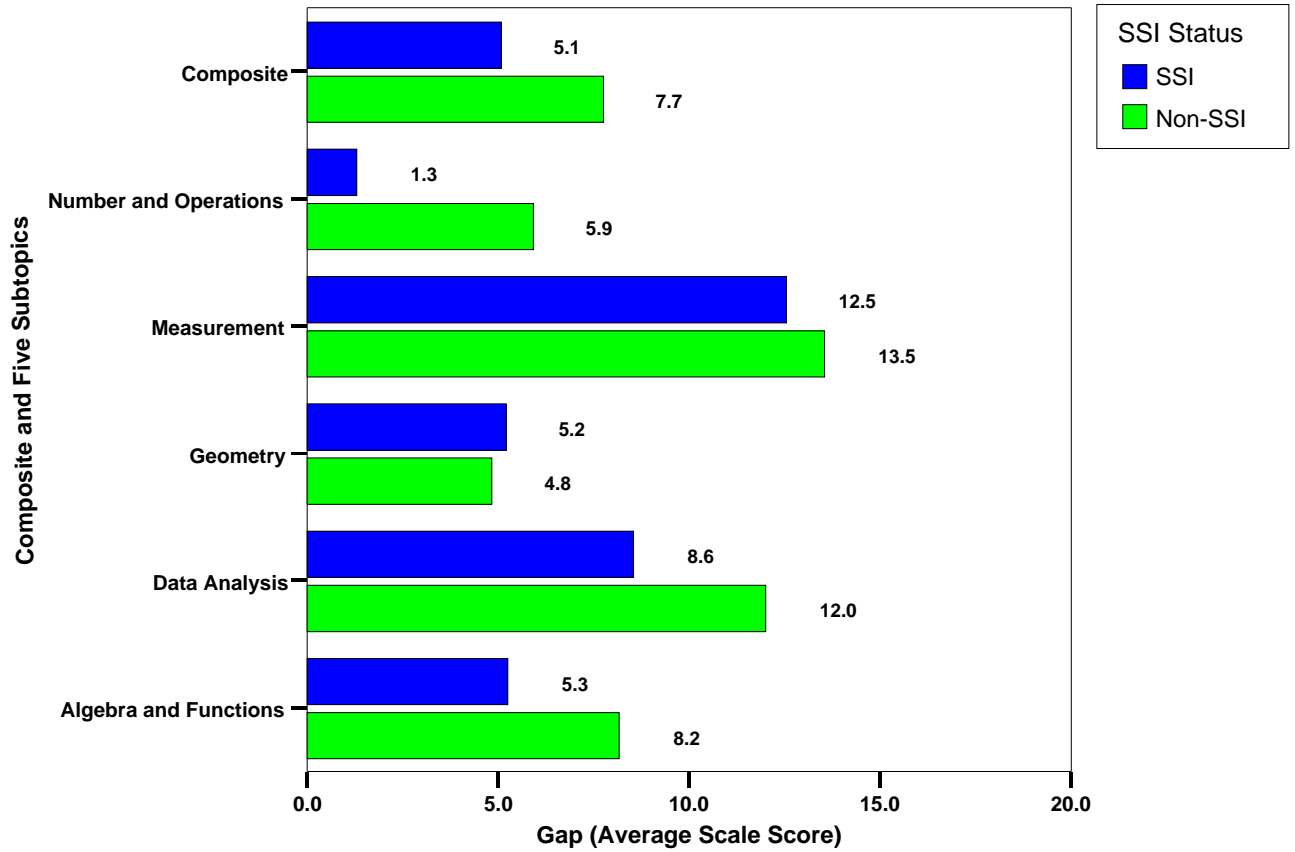
*Ethnicity*

Figure A-3. Differences in average scale scores between White and Black students from 1996 to 2000, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).



\* Due to the insufficient sample size of these subgroups, results are based on 12 SSI states and 8 non-SSI states.

Figure A-4. Differences in average scale scores between White and Hispanic students from 1996 to 2000, by SSI status: Trend Group 92-00 (14 SSI and 13 non-SSI states\*).



\* Due to the insufficient sample size of these subgroups, results are based on 11 SSI states and 12 non-SSI states.