June, 2001

Study of the Impact of the Statewide Systemic Initiatives Program

Norman L. Webb, Janet Kane, Darwin Kaufman, and Jung-Ho Yang

Technical Report to the National Science Foundation on the Use of State NAEP Data to Assess the Impact of the Statewide Systemic Initiatives

Wisconsin Center for Education Research
School of Education University of Wisconsin–Madison

Study of the Impact of the Statewide Systemic Initiatives

TECHNICAL REPORT

Norman L. Webb, Janet Kane, Darwin Kaufman, and Jung-Ho Yang

June, 2001

Funded by the National Science Foundation

Table of Contents

List of Tables vii

List of Figures xvii

Executive Summary xxv

Chapter 1 Introduction 1

Chapter 2 Review of Recent Studies 11

Chapter 3 Methodological Issues 23

Chapter 4 A Description of the Demographics of SSI Compared to Non-SSI States 37

Chapter 5 SSI and Non-SSI Achievement Using State NAEP Data: Descriptive Analysis 65

Chapter 6 Comparing SSI and Non-SSI States Using Indicators of Mathematics Curricular Reform from the State NAEP Teacher Questionnaires 107

Chapter 7 SSI and Non-SSI Achievement Using State NAEP Data: Empirical Bayes and Bayesian Analyses 313

Chapter 8 Conclusions and Future Directions 339

References 345

List of Tables

Table 2.1
Correlates of Student Achievement

Table 4.1
States Participating in the State NAEP

Table 6.1
Number and Percentage of SSI and Non-SSI States Participating in State NAEP Each Year

Table 6.2
States in the Two- and Three-Point Trend Samples at Grade 8 and the Two-Point Trend Sample at Grade 4

Table 6.3
States Not Meeting the NCES Participation Rate Guidelines

Table 6.4
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96),}$ Grade 8

Table 6.5
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, 1996

Table 6.6
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 8, 1996

Table 6.7
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96)}$, Grade 4

Table 6.8
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 4, 1996

Table 6.9
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 4, 1996

Table 6.10
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(92)}$, Grade 8

Table 6.11
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, 1992

Table 6.12
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 8, 1992

Table 6.13
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(92)}$, Grade 4

Table 6.14
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 4, 1992

Table 6.15
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 4, 1992

Table 6.16
Mean and Standard Deviation for SSI and Non-SSI States on $I_{RC(90)}$, Grade 8

Table 6.17
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, 1990

Table 6.18
Intercorrelations of State Means for Emphasis Given to the Four Skill Areas, Grade 8, 1990

Table 6.19
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 8

Table 6.20
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 8, for the Subsample that Followed the NCES Participation Rate Guidelines

Table 6.21
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 4

Table 6.22
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 4, for the Subsample that Followed the NCES Participation Rate Guidelines

Table 6.23
Predicting $I_{RC(96)}$ from $I_{RC(90)}$ and SSI Status, Grade 8

Table 6.24
Mean and Standard Deviation of $I_{MD(96)}$ in SSI and Non-SSI States, Grade 8, 1996

Table 6.25
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8, 1996

Table 6.26
Intercorrelations of State Means on the Nine Mathematical Discourse Items, Grade 8, 1996

Table 6.27
Mean and Standard Deviation of $I_{MD(96)}$ in SSI and Non-SSI States, Grade 4, 1996

Table 6.28
Mean and Standard Deviation of Individual Mathematics Discourse Items in SSI and Non-SSI States, Grade 4, 1996

Table 6.29
Intercorrelations of State Means on the Nine Mathematical Discourse Items, Grade 4, 1996

Table 6.30
Mean and Standard Deviation of $I_{MD4(92)}$ and $I_{MD(92)}$ in SSI and Non-SSI States, Grade 8

Table 6.31
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8, 1992

Table 6.32
Intercorrelations of State Means on the Seven Mathematical Discourse Items, Grade 8, 1992

Table 6.33
Mean and Standard Deviation of $I_{MD4(92)}$ and $I_{MD(92)}$ in SSI and Non-SSI States, Grade 4

Table 6.34
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and on-SSI states, Grade 4, 1992

Table 6.35
Intercorrelations of State Means on the Seven Mathematical Discourse Items, Grade 4, 1992

Table 6.36
Mean and Standard Deviation of $I_{MD(90)}$ and the Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8

Table 6.37
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States in 1992 and 1996, Grade 8

Table 6.38
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States that Met the NCES Participation Rate Guidelines in 1992 and 1996, Grade 8

Table 6.39
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status, Grade 8

Table 6.40
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

Table 6.41
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States in 1992 and 1996, Grade 4

Table 6.42
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States that Met the NCES Participation Rate Guidelines in 1992 and 1996, Grade 4

Table 6.43
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status, Grade 4

Table 6.44
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 4

Table 6.45
Predicting $I_{MD(96)}$ from $I_{MD(90)}$ and SSI Status, Grade 8

Table 6.46
Predicting $I_{MD(96)}$ from $I_{MD(90)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

Table 6.47
Mean and Standard Deviation of SSI and Non-SSI States on $I_{S(96)}$

Table 6.48
Percentage of Students in Each Category of Teachers' Knowledge of the NCTM Standards, Grade 8, 1996

Table 6.49
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96)}$

Table 6.50
Percentage of Students in Each Category of Teachers' Knowledge of the NCTM Standards, Grade 4, 1996

Table 6.51
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(96)}$, Grade 8

Table 6.52
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(96)}$, Grade 4

Table 6.53
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(92)}$, Grade 8

Table 6.54
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(92)}$, Grade 4

Table 6.55
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(90)}$, Grade 8

Table 6.56
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(96)}$, Grade 8

Table 6.57
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 8, 1996

Table 6.58
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(96)}$, Grade 4

Table 6.59
Seven Reform-Related Topics and the Percentage of Students Whose Teachers Had Studied the Topic, Grade 4, 1996

Table 6.60
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(92)}$, Grade 8

Table 6.61
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 8, 1992

Table 6.62
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(92)}$, Grade 4

Table 6.63
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 4, 1992

Table 6.64
Percentage of Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds, Grade 8, 1990

Table 6.65
Percent of Grade 8 Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds in the Three-Point Trend Sample

Table 6.66
Percent of Grade 8 Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds in the Three-Point Trend Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6.67
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 8

Table 6.68
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1996

Table 6.69
Intercorrelations of State Means on the Five Calculator Use Items, Grade 8, 1996

Table 6.70
Use of Calculators on State Assessments and State's SSI Status, for States Included in State NAEP in 1990, 1992, or 1996

Table 6.71
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

Table 6.72
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 4

Table 6.73
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 4, 1996

Table 6.74
Intercorrelations of State Means on the Five Calculator Use Items, Grade 4, 1996

Table 6.75
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 4 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

Table 6.76
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 8

Table 6.77
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1992

Table 6.78
Intercorrelations of State Means on the Seven Calculator Use Items, Grade 8, 1996

Table 6.79
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

Table 6.80
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 4

Table 6.81
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 4, 1992

Table 6.82
Intercorrelations of State Means on the Five Calculator Use Items, Grade 4, 1992

Table 6.83
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 4 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

Table 6.84
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(90)}$, Grade 8

Table 6.85
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1990

Table 6.86
Intercorrelations of State Means on the Four Calculator Use Items, Grade 8, 1990

Table 6.87
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(90)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

Table 6.88
Predicting $I_{C(96)}$ from $I_{C(92)}$ and SSI Status, Grade 8

Table 6.89
Predicting $I_{C(96)}$ from $I_{C(92)}$, Calculator Use Policy, and SSI Status, Grade 8

Table 6.90
Predicting $I_{C(96)}$ from $I_{C(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

Table 6.91
Predicting $I_{C(96)}$ from $I_{C(92)}$, Calculator Use Policy, and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8.

Table 6.92
Mean and Standard Deviation of I_C in SSI and Non-SSI States in 1992 and 1996, Grade 4

Table 6.93
Mean and Standard Deviation of I_C in SSI and Non-SSI States in 1992 and 1996, Grade 4

Table 6.94
Predicting $I_{C(96)}$ from $I_{C(90)}$ and SSI Status, Grade 8

Table 6.95
Predicting $I_{C(96)}$ from $I_{C(90)}$, Calculator Use Policy, and SSI Status, Grade 8

Table 6.96
Predicting $I_{C(96)}$ from $I_{C(90)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

Table 6.97
Predicting $I_{C(96)}$ from $I_{C(90)}$, Calculator Use Policy, and SSI Status, Grade 8

Table 6.98
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8 and Grade 4, 1996

Table 6.99
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8 and Grade 4, 1992

Table 6.100
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8, 1990

Table 6.101
Classification Results for Descriptive Discriminant Function, Grade 8, 1996

Table 6.102
Classification Results for Descriptive Discriminant Function, Grade 4, 1996

Table 6.103
Correlations Between the Individual Indicators and the Mean NAEP Mathematics Composite in 1996, 1992, and 1990

Table 7.1
Longitudinal Analysis of Grade 8 Data over 1990, 1992, and 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

Table 7.2
Longitudinal Analysis of Grade 4 Data in 1992 and 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

Table 7.3
Longitudinal Analysis of Cohort Data for Grade 4 in 1992 and Grade 8 in 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

Table 7.4
Cross-Sectional Analysis of Grade 8 Data: Empirical Bayes Estimates After Considering Jackknife Standard Errors

Table 7.5
Cross-Sectional Analysis of Grade 4 Data: Empirical Bayes Estimates After Considering Jackknife Standard Errors

List of Figures

Figure 2.1. TAAS Mathematics Domains and Objectives.

Figure 3.1. A model of systemic reform

Figure 3.2. Components and states used for Clune rating of SSI breadth and depth.

Figure 4.1. Percentage distribution of grades 8 and 4 students, by gender and SSI status: 1996 Group (22 SSI and 18 non-SSI states).

Figure 4.2. Percentage distribution of grades 8 and 4 students, by race and SSI status: 1996 Group (22 SSI and 18 non-SSI states).

Figure 4.3. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: 1996 Group (22 SSI and 18 non-SSI states).

Figure 4.4. Percentage distribution of grades 8 and 4 students on the basis of home environment, by SSI status: 1996 Group (22 SSI and 18 non-SSI states).

Figure 4.5. Percentage distribution of grades 8 and 4 students, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.5a. Percentage distribution in gender of grade 8 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.5b. Percentage distribution in gender of grade 4 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.6. Percentage distribution of grades 8 and 4 students, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.6a. Percentage distribution by race of Grade 8 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.6b. Percentage distribution by race of grade 4 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.7. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.8. Percentage distribution of grades 8 and 4 students, by home environment criteria by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 4.9. Percentage distribution of grades 8 and 4 students, by gender and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.9a. Percentage distribution by gender of grade 8 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.9b. Percentage distribution by gender of grade 4 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.10. Percentage distribution of grades 8 and 4 students, by race and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.10a. Percentage distribution by race of grade 8 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.10b. Percentage distribution by race of grade 4 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.11. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 4.12. Percentage distribution of grades 8 and 4 students, by home environment criteria by SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5.1. Trends in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.2. Trends in average scale scores, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.3. Trends in average scale scores, by race and SSI status: Trend Group 90-96 grade 8 and 92-96 grade 4 (17 SSI and 11 non-SSI states*).

Figure 5.4. Trends in average scale scores on content strands, by SSI status: Trend Group 90-96 grade 8 and 92-96 grade 4 (17 SSI and 11 non-SSI states).

Figure 5.5. Trends in average scale scores on content strands, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.6. Trends in average scale scores on content strands, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.7. Gender differences (males minus females) in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.8. Differences in average scale scores between White and Black students, by SSI Status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.9. Differences in average scale scores of White and Hispanic students, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.10. Cohort growth in average scale scores from 1992 to 1996, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.11. Cohort growth in average scale scores, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.12. Cohort growth in average scale scores, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.13. Cohort growth in average scale scores on content strands, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.14. Cohort growth in average scale scores on content strands, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI).

Figure 5.15. Cohort growth in average scale scores on content strands, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.16. Gender differences in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

Figure 5.17. Differences in average scale scores between White and Black students from 1992 to 1996, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 5.18. Differences in average scale scores between White and Hispanic students, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

Figure 6.1. State means on $I_{RC(96)}$ ordered from highest to lowest, grade 8.

Figure 6.2. State means on $I_{RC(96)}$ ordered from highest to lowest, grade 4.

Figure 6.3. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the two-point trend sample, grade 8.

Figure 6.4. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed NAEP's participation rate guidelines, grade 8.

Figure 6.5. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the two-point trend sample, grade 4.

Figure 6.6. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed NAEP's participation rate guidelines, grade 4.

Figure 6.7. Scatterplot of $I_{RC(90)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the three-point trend sample, grade 8.

Figure 6.8. Scatterplot of $I_{RC(90)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed the NCES participation rate guidelines, grade 8.

Figure 6.9. State means on $I_{MD(96)}$ ordered from highest to lowest, grade 8.

Figure 6.10. State means on $I_{MD(96)}$ ordered from highest to lowest, grade 4.

Figure 6.11. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, grade 8.

Figure 6.12. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, grade 8, for the subsample that followed the NCES participation rate guidelines.

Figure 6.13. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for SSI and non-SSI states, grade 8.

Figure 6.14. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.

Figure 6.15. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, grade 4.

Figure 6.16. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, grade 4, for the subsample that followed the NCES participation rate guidelines.

Figure 6.17. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for SSI and non-SSI states, grade 4.

Figure 6.18. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 4.

Figure 6.19. Relationship between $I_{MD(90)}$ and $I_{MD(96)}$ for SSI and non-SSI states.

Figure 6.20. Relationship between $I_{MD(90)}$ and $I_{MD(96)}$ for SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.

Figure 6.21. State means on $I_{S(96)}$ ordered from highest to lowest, grade 8.

Figure 6.22. State means on $I_{S(96)}$ ordered from highest to lowest, grade 4.

Figure 6.23. State means on $I_{PD(96)}$ ordered from highest to lowest, grade 8.

Figure 6.24. State means on $I_{PD(96)}$ ordered from highest to lowest, grade 4.

Figure 6.25. Mean for last year's staff development in 1992 and 1996, grade 8.

Figure 6.26. Mean for last year's staff development in 1992 and in 1996 for states following the NCES participation rate guidelines, grade 8.

Figure 6.27. Mean for last year's staff development in 1992 and in 1996, Grade 4.

Figure 6.28. Mean for last year's staff development in 1992 and 1996, for states following the NCES participation rate guidelines, grade 4.

Figure 6.29. Mean for last year's staff development in 1990, 1992, and 1996, grade 8.

Figure 6.30. Mean for last year's staff development in 1990, 1992 and 1996, for the SSI and non-SSI states following NCES participation rate guidelines, grade 8.

Figure 6.31. State means on $I_{RT(96)}$ ordered from highest to lowest, grade 8.

Figure 6.32. State means on $I_{RT(96)}$ ordered from highest to lowest, grade 4.

Figure 6.33. Mean of I_{RT} for SSI and non-SSI states at grade 8 in 1992 and 1996.

Figure 6.34. Mean of I_{RT} at grade 8 in 1992 and 1996, for SSI and non-SSI states following the NCES participation rate guidelines.

Figure 6.35. Mean of I_{RT} for SSI and non-SSI states at grade 4 in 1992 and 1996.

Figure 6.36. Mean of I_{RT} at grade 4 in 1992 and 1996, for SSI and non-SSI states following the NCES participation rate guidelines.

Figure 6.37. Proportion of students whose teacher had studied teaching students from different cultural backgrounds, grade 8.

Figure 6.38. Proportion of students whose teacher had studied teaching students from different cultural backgrounds in the subsample that followed the NCES participation rate guidelines.

Figure 6.39. State means on $I_{C(96)}$ ordered from highest to lowest, grade 8.

Figure 6.40. State means on $I_{C(96)}$ ordered from highest to lowest, grade 4.

Figure 6.41. Relationship between $I_{C(96)}$ and $I_{C(92)}$ for SSI and non-SSI states, grade 8.

Figure 6.42. Relationship between $I_{C(96)}$ and $I_{C(92)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.

Figure 6.43. Change in the mean of I_C for SSI and non-SSI states from 1992 to 1996, grade 4.

Figure 6.44. Change in the mean of I_C for the subsample of SSI and non-SSI states that followed the NCES participation rate guidelines, grade 4, 1992 to 1996.

Figure 6.45. Relationship between $I_{C(96)}$ and $I_{C(90)}$ for SSI and non-SSI states, grade 8.

Figure 6.46. Relationship between $I_{C(96)}$ and $I_{C(90)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.

Figure 6.47. Relationships among the indicators within a model of systemic reform.

Figure 6.48. Indicators of mathematics curricular reform and their relationship to student achievement.

Figure 7.1. Posterior distribution of the variance: Unconditional model.

Figure 7.2. Posterior distribution of the variance: Conditional model.

Figure 7.3. Posterior distributions of average scale scores of grade 8 in 1990.

Figure 7.4. Posterior distributions of growth rates of grade 8 from 1990 to 1996.

Figure 7.5. Posterior distribution of the variance: Unconditional model.

Figure 7.6. Posterior distribution of the variance: Conditional model.

Figure 7.7. Posterior distributions of average scale scores of grade 4 in 1992.

Figure 7.8. Posterior distributions of growth rates of grade 4 from 1992 to 1996.

Figure 7.9. Posterior distribution of the variance: Unconditional model.

Figure 7.10. Posterior distribution of the variance: Unconditional model.

Figure 7.11. Posterior distributions of average scale scores of cohort—grade 4 in 1992 and grade 8 in 1996.

Figure 7.12. Posterior distributions of growth rates of cohort—grade 4 in 1992 and grade 8 in 1996.

Project Staff

WCER Research Staff
Norman L. Webb, Co-Principal Investigator
Allan Cohen, Consultant
Janet Kane, Research Scientist
Michael Kane, Consultant
Tae-Hoon Kang, Project Assistant
Darwin Kaufman, Research Scientist
Jong-Baeg Kim, Project Assistant
Linda Wilson, Research Scientist
Jung-Ho Yang, Research Scientist
Margaret Powell, Editor
Lynn Lunde, Secretary
Mary Tedeschi, Secretary

Horizon Research Staff
Iris R. Weiss, Co-Principal Investigator
Rebecca Dotterer, Researcher
Dan Heck, Researcher
Diana Montgomery, Researcher
Susan Hudson, Secretary

Consultant
Jonathan Supovitz, University of Pennsylvania

NSF Program Officer
Bernice Anderson

Advisory Board
Dennis Bartels, Exploratorium, San Francisco
Audrey B. Champagne, SUNY, Albany
Jere Confrey, University of Texas
Andrew C. Porter, University of Wisconsin
Edward A. Silver, University of Michigan

Acknowledgements

All of the authors contributed to the total document, but major authors of specific chapters were as follows:

- Chapter 1 Introduction, *Norman L. Webb*
- Chapter 2 Review of Recent Studies, *Darwin Kaufman*
- Chapter 3 Methodological Issues, *Darwin Kaufman*
- Chapter 4 A Description of the Demographics of SSI Compared to Non-SSI States, *Jung-Ho Yang* and *Janet Kane*
- Chapter 5 SSI and Non-SSI Achievement Using the State NAEP: Descriptive Analysis, *Jung-Ho Yang* and *Norman L. Webb*
- Chapter 6 Comparing SSI and Non-SSI States Using Indicators of Mathematics Curricular Reform from the State NAEP Teacher Questionnaire, *Janet Kane*
- Chapter 7 SSI and Non-SSI Achievement Using State NAEP Data: Empirical Bayes and Bayesian Analysis, *Jung-Ho Yang*
- Chapter 8 Conclusions and Future Directions, *Norman L. Webb*

Executive Summary

By the time the State NAEP had been administered in 1996, the Statewide Systemic Initiatives Program of the National Science Foundation had funded 25 states and Puerto Rico for at least three years, the minimum period of implementation time necessary to detect an impact of the systemic initiatives. The purpose of this study has been to ascertain the impact of the Statewide Systemic Initiatives (SSIs) on student performance and classroom practices as measured by the State NAEP. Grade 8 mathematics data from the 1990, 1992, and 1996 State NAEP and grade 4 mathematics data from the 1992 and 1996 NAEP were analyzed, along with teacher questionnaire information, to determine the effect of the SSI program on improving mathematics learning and instruction in the participating states. A number of issues in using the existing NAEP data to analyze SSI impact arose, including variation in state participation in the NAEP for the three testing times, application of different NAEP weighting procedures, lack of consistency in the questions asked teachers about classroom practices, and failure of some states to reach the 90 percent participation rate of sample schools. Because of these issues, data were analyzed in a number of ways in order to produce findings and to judge the stability of results.

SSI states had a higher percentage of minority students than non-SSI states, but were essentially the same as the non-SSI states on socio-economic variables. Both SSI and non-SSI states improved in mathematics achievement in grades 4 and 8 from 1990 to 1996. In 1990 and 1992, the SSI states' mean achievement on the mathematics composite scale was below the mean achievement of the non-SSI states by about six scale points. By 1996, the SSI states had improved at a slightly faster rate than non-SSI states, reducing the gap by one scale point. Non-SSI states were more successful in achieving gender equity. The difference in composite scale scores between males and females was eliminated between 1992 and 1996 by non-SSI states, but remained the same in SSI states. Both SSI and non-SSI states maintained the achievement gap between Black and White students over the three testing times. However, there was some evidence that the achievement gap between Black and White students in SSI states declined on specific subtopic scales, including geometry and algebra and functions. The gap increased in non-SSI states on all five subtopic scales for both grades 4 and 8. Of particular note, on the algebra and function scale Black students in SSI states gained more between grade 4 and grade 8 than did White students in SSI states. Hispanic students in SSI states gained more than Hispanic students in non-SSI states over the four years from 1992 to 1996 on five of the six scales. Non-SSI students only gained more on the measurement scale.

Six indicators of reform practices were analyzed, including time spent in professional development during the last year, the number of reform-related topics studied by teachers, teachers' knowledge of the NCTM *Standards*, relative emphasis on reasoning and communication, students' opportunities for mathematical discourse, and students' use of calculators. In 1996, these reform indicators discriminated between SSI and non-SSI states, suggesting a higher prevalence of reform practices in SSI states than in non-SSI states. At grade 8 SSI states had a significant increase in opportunities for mathematical discourse over non-SSI states, but this indicator was not correlated with student achievement. The large gap between state differences within groups and the interaction among reform indicators requires further analyses to decipher true differences.

CHAPTER 1

INTRODUCTION

Tracking the impact of any large-scale educational reform confronts the researcher with a number of methodological and conceptual issues. In this first technical report on the Study of the Impact of Statewide Systemic Initiatives, our purpose is to describe in some detail the use of data from the State National Assessment of Educational Progress (State NAEP) to detect effects that we believe can be attributed to a state's participation in the National Science Foundation's Statewide Systemic Initiatives (SSI) Program, which was initiated in 1991. Although we have sought definitive findings on the impact of the SSI Program, we believe an equally important outcome of the study is the development of viable procedures for analyzing NAEP data and for using NAEP data to study large-scale educational reform. NAEP has surveyed the achievement of students at ages 9, 13, and 17 since 1969 and at grades 4, 8, and 12 since the 1980s, but NAEP results reported by states have only been available in mathematics since 1990. With the advent of the State NAEP, an important database is now accessible for contrasting differences among states in student achievement and educational practices as reported by teachers, students, and administrators on questionnaires administered by NAEP. In this study, we have mined the data from the State NAEP tests administered in grade 8 mathematics in 1990, 1992, and 1996 and in grade 4 mathematics in 1992 and 1996.

In this initial technical report, we describe our general approach to use of the State NAEP data to study the impact of the SSI Program, the complexities we encountered, how we dealt with these complexities, and our initial findings—obtained in contrasting the group of 25 states that received NSF funding with the 25 non-SSI states. The Commonwealth of Puerto Rico also received NSF funding for an SSI, which completed the set of 26 jurisdictions that constituted the SSI program. However, Puerto Rico did not participate in any year of the State NAEP; in this report, we refer, therefore, to the 25 SSI states participating in NAEP testing. After a description in Chapter 2 of the general design of the study and a review of other recent studies, we discuss our methodology in Chapter 3, where we highlight some of the technical issues we have faced and the decisions made to overcome these issues. Most of the findings that are reported in this technical report are descriptive. Chapters 4, 5, and 6 contrast the SSI and non-SSI states that participated in the State NAEP on demographics, student achievement, and classroom practices. In Chapter 7, we report our progress in doing a longitudinal analysis of the SSI and non-SSI states using an empirical Bayes and Bayesian analysis.

The National Science Foundation's Statewide Systemic Initiatives Program

Over ten years ago, in 1989, the National Council of Teachers of Mathematics published the first set of K-12 national content standards. Since then, nearly all of the states have developed content standards and assessments for mathematics, as well as the other content areas. The advancement of systemic reform has coincided with this massive effort on the part of states and districts to describe and assess more clearly what students should be able to know and to do in a multiplicity of content areas. Coinciding with and closely linked to standards-based reforms, systemic reform has evolved from the theory developed by Smith and O'Day in 1991 into practice as a change strategy for surmounting the difficult problem of enabling all students to meet challenging content standards.

Since 1991, the National Science Foundation (NSF) has funded Statewide Systemic Initiatives (SSIs) in 25 states and Puerto Rico, Urban Systemic Initiatives (USIs) in more than 40 cities, and Rural Systemic Initiatives (RSIs) in more than six regions. The Systemic Initiatives (SIs) represent a commitment of over $600 million by NSF. During the first five years of NSF funding, 1991 through 1996, SSI programs leveraged more than $500 million in additional funds. This was more than twice the amount the NSF invested in statewide programs over the same period (Zucker, Shields, Adelman, Corcoran, & Goertz, 1998).

Simply stated, systemic reform

> is a process that extends over a long period of time and that has to engage a number of people in system improvement through changing multiple system components and their interconnections concurrently.

Systemic reform in education does not imply uniform practice or the prevention of innovation. It does not imply only one strategy for change. Nor does it imply that there has to be a strong centralized system rather than a more locally controlled system. It does imply that an education system needs to add greater stability, improve alignment, remove barriers or other countervailing forces to change, create stronger links among components, and work with all teachers so that all students will have the chance to obtain knowledge of important science and mathematics.

In 1990, NSF instituted a new Directorate for Education and Human Resources (I) to promote the health and vitality of science and mathematics education in the country. In order to have a national impact, it adopted a systemic approach that would address entire systems of mathematics and science education, rather than isolated components such as curriculum, professional development, or pedagogy. In its strategy for large-scale change, NSF prominently advocated that state systemic initiatives adhere to high, explicit local and national standards for teaching and learning, such as the newly released National Council for Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics* (1989). NSF encouraged states to seek statewide change in pedagogy, including “hands-on” and “inquiry-based” education, that would relieve students of the unproductive burden of rote learning (Westat*McKenzie Consortium, 1998). Finally, the agency strongly advocated that the newly implemented methods be monitored by student achievement assessments designed to measure students’ learning of challenging content.

Beginning in 1991, NSF awarded cooperative agreements to states that proposed initiatives directed toward achieving NSF’s vision of reform. NSF gave each successful state up to $10 million over five years. It recognized that this level of funding was very small compared to states’ education budgets, but the agency expected these funds to be used as a catalyst that would garner other resources needed to mount a reform that would bring large-scale change to student learning throughout the state. A total of 26 grants were awarded in three cohorts:

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

To be considered for an SSI grant, states were required to articulate a vision for science and mathematics education and to indicate how the state intended to develop and manage the project, what partnerships would be created to further the reform, and how the planning process and progress would be evaluated.

Systemic reform was new for all involved in the program, including the National Science Foundation. Over time, the NSF administrators of the SSI program grew in their understanding and conceptualization of systemic change. As the SSI program evolved, essential components, or "drivers," of systemic reform were identified: Four process drivers and two outcome drivers became a dominant means for focusing the vision of systemic reform for the statewide systemic initiatives, for the urban systemic initiatives, and for the rural systemic initiatives.

Six Drivers of Educational System Reform

PROCESS DRIVERS

1. **Standards-Based Curricula**
Implementation of comprehensive, standards-based curricula—as represented in instructional practice and student assessment—in every classroom and laboratory, as well as other learning experiences provided through the system and its partners.

2. **Coherent Policies**
Development of a coherent, consistent set of policies that supports: provision of high quality mathematics and science education for each student; excellent preparation, continuing education, and support for each mathematics and science teacher (including all elementary teachers); and administrative support for all persons committed to dramatically improving achievement among all students served by the system.

3. **Convergent Resources**
Convergence and usage of all resources that are designed for or that reasonably could be used to support science and mathematics education—fiscal, intellectual, material, curricular, and extra-curricular—into a focused and unitary program to constantly upgrade, renew, and improve the educational program in mathematics and science for all students.

4. Broad-Based Support
Broad-based support from parents, policymakers, institutions of higher education, business and industry, foundations, and other segments of the community for the goals and collective value of the program, based on rich presentations of the ideas behind the program, the evidence gathered on its successes and its failures, and critical discussions of its efforts.

OUTCOME DRIVERS

5. **Significantly Higher Student Achievement**
Accumulation of a broad and deep array of evidence that the program is enhancing student achievement, using a set of indices that might include achievement test scores, higher level courses passed, college admission rates, college majors, advanced placement tests taken, portfolio assessment, and ratings from summer employers that demonstrate that students are generally achieving at a significantly higher level in science and mathematics than their predecessors.

6. **Improved Achievement of All Students**
Improvement in the achievement of all students, including those historically underserved.

The six "Drivers of Systemic Reform" evolved out of the National Science Foundation's systemic initiatives program over a number of years (National Science Foundation, 1997). Their significance first became evident toward the end of the five-year funding period for the first cohort of recipients of Statewide Systemic iinitiatives grants, which were awarded in 1991. The Drivers were developed by NSF staff to better guide the management of the program and the gathering of data from each of the SSI states, which, at the programs' peak, included 26 jurisdictions. The Drivers were used to identify core data elements that SSIs were required to supply to NSF in partial fulfillment of the terms of their grants. Later, NSF's Division of Research, Evaluation, and Communication incorporated the Drivers into their requests for proposals for funding evaluation studies of the systemic initiatives. Evaluators who submitted proposals were asked to attend to the "NSF framework of four process and two outcome drivers," or similar frameworks, as they assessed the impact of classroom change on student achievement within SSIs and institutionalization, scale-up, and continuous improvement by SSIs (Dear Colleague Letter, May 18, 1998, http:/www.I.nsf.gov/I/rec/pubs/ssi-impact.htm, p. 3). Examples of what a prospective grantee could propose are addressed by questions such as:

- What research studies and literature support the use of the hypothesized process drivers?
- What would be a credible assessment system for monitoring progress by a driver to facilitate the implementation of high-quality mathematics and science education for all students?
- What is the relation of process drivers to learning infrastructure indicators and student achievement drivers and indicators?

More traditional reforms focus on a single component or unit and on incremental change, whereas systemic reform considers all of the components, their interactions with each other, and their alignment in attaining common goals. In theory, school-based reform, curriculum reform, and other singularly focused reform initiatives are insufficient to sustain an effort to attain significant improvement in student learning without attending to other system components. Those successes that can be achieved through school-based reform are deterred or inhibited by shifts in policy through state and district mandates or a diminishing teaching force of knowledgeable and well-trained teachers. Standards-based reform is important to a systemic reform, but does not imply that the reform is directed toward systemic change. Other components within the system, such as professional development, accountability, teacher preparation, and resource allocation, need to be addressed to achieve standards-based systemic reform. A state or district education system will make progress towards systemic reform when policies, administration, teaching, and curriculum are working in concert with each other in an effort directed toward promoting improved learning of challenging content by all students.

NSF closely monitored the progress of its highly visible SSI program, expending hundreds of millions of dollars. Each SSI was visited, annual principal investigators' meetings were held, SRI International was engaged as an external evaluator (Zucker et al., 1998), and Abt Associates Inc. was engaged to monitor each site by conducting site visits. As a result of an accumulation of information and data, it was evident that some of the 26 SSIs were not fulfilling the full intent of the program. NSF withdrew its funding from four sites, one (Rhode Island) after only one year. These states—Rhode Island, North Carolina, Florida, and Virginia—therefore received funding for less than five years.

States varied in the strategies they adopted to attain systemic reform. Nearly all states claimed to have mathematics and science as a major focus. Eleven focused on grades K through 16. Another six focused on grades K through 12. The other states concentrated their initiatives on the middle or primary grades. Only the Montana SSI addressed primarily high school. Eighty percent of the SSIs had 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 the SRI International 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), reforming higher education and the preparation of teachers (13 SSIs), and mobilizing public and professional opinion (14 SSIs).

Study of the SSI Program Using NAEP Data

After the first phase of funding, a large question remained about the actual impact of the SSIs on student learning. A number of reports have been published on the SSI program (Laguarda, 1998; Shields, Marsh, & Adelman, 1998; Corcoran, Shields, & Zucker, 1998; Shields, Corcoran, & Zucker, 1994; Zucker & Shields, 1995; Zucker, Shields, Adelman, & Powell, 1995; Barley & Jenness, 1995; Horizon Research, Inc., 1995; Inverness Research Associates, 1995). Additionally, burgeoning research and literature on the evaluation of systemic education reform has produced an emerging analytical foundation in the field. Much, however, remains to be learned about the practice and theory of systemic evaluation in education. Nearly all of the evaluations primarily focused on generating formative and descriptive information.

Very few addressed the question of student impact. In fact, some even questioned whether five years was enough time for any state to mount an effort that would be large enough to have an impact on student learning (St. John, 1999). At best, a state could engage in capacity building, but the sequence of changes envisioned from reform efforts to teacher knowledge to classroom practices to student learning would not have sufficient time to develop on a scale adequate to influence achievement levels on state assessments. In addition, states were engaged in other reform efforts besides those funded by NSF, including accountability, increasing graduation requirements, grade-to-grade promotion, and state curriculum standards and assessments.

This research seeks to study the possible impact of the SSI program on student achievement and to glean lessons that can be learned about designing, implementing, evaluating, and supporting statewide systemic reform. The National Assessment of Education Progress (NAEP), for the first time in 1990, administered an assessment in mathematics (grade 8) that produced state results. In 1992 and 1996, NAEP produced state results in mathematics for grades 4 and 8.

The Study of the Impact of Statewide Systemic Initiatives Project has engaged in analyzing data from the State NAEP to determine what the impact of the statewide systemic initiatives on student learning has been and to identify other significant outcomes of the SSI program. We have focused our analysis on grade 8 mathematics data for the years 1990, 1992, and 1996 and grade 4 data for 1992 and 1996. Grade 4 state-by-state data were not collected by NAEP in 1990. Our approach is to develop a profile for each state, SSI states and non-SSI states, using the NAEP database, including data on student achievement, demographics, and instructional practices. These profiles are then analyzed, using several analytical methods, to draw inferences about the impact of the SSIs. Currently, we are in the middle of this process.

The NAEP database is complex because of the sampling procedures employed in collecting the data, the weighting procedures, and how the data are structured to compute an estimate of the error in the findings reported. Although we are in the midst of our analyses, at this time we do have some preliminary findings that we can share. These observations and findings are from our analysis of summary data as reported in the NAEP 1996 Compendium and our analysis of student data from the NAEP database for the three test years.

This report describes the technical issues raised from the beginning in a series of studies designed to investigate very gross effects that can be associated with the Statewide Systemic Initiatives over a limited period of time relatively early in the evolution of these reforms. In the view of some analysts, there are a number of reasons for not using NAEP data at this stage to study the impact of SSIs. The most frequently cited are: 1) the NAEP instruments are not sensitive to the changes in student learning advanced by most of the SSIs (poor alignment); and, 2) the timing of the most recent NAEP in mathematics, 1996, is too early for any SSI to reach scale in order for changes in achievement related to SSIs to be detected (inadequate time frame). Any study of SSIs using NAEP data needs to address these issues directly. One purpose of this report is to raise these and other technical issues that we faced and to indicate how we addressed these issues.

The State NAEP data 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. The analysis of these data provides a good entry point into study of the impact of SSIs, if only to provide baseline information and to help identify SSI states worthy of more detailed analyses. However, enough questions exist about the 1990 through 1996 NAEP data and the relationship to SSI states to warrant some careful analyses. The central research question addressed by this study is:

> 1A. What differences were there on mathematics achievement and student participation variables (course completion) as measured by NAEP between SSI states and non-SSI states over the period 1990-96?

In this study, we have mined the NAEP data and other existing studies that use the NAEP mathematics data for the three years 1990, 1992, and 1996. NAEP mathematics data for 1990 and 1996 are available for 30 states. NAEP mathematics data for 1992 and 1996 are available for 35 states, 20 SSI states and 15 non-SSI states (Phelps, Cullen, Easton, & Best, 1997). Of the 14 states that had a significant increase in scores on grade 8 mathematics average proficiency scores between 1990 and 1996 and were above the national average, seven had an SSI. Of the 11 states that met these criteria for increase in scores between 1992 and 1996, eight were SSI states. In a value-added analysis of NAEP data by state between 1992 and 1996, Barton and Coley (1998) concluded, “Most states are not significantly different from each other in terms of cohort growth from the fourth to the eighth grade” (p. 11). However, the two states that had the highest growth were SSI states (Nebraska and Michigan). These very global results warrant closer scrutiny.

It is most reasonable to expect to see improvement in mathematics achievement at the middle grades. All 26 SSIs targeted specific grade levels in mathematics (Zucker, Shields, Adelman, & Powell, 1995); all but Montana targeted change in the middle grades, 19 targeted elementary grades, and 12 targeted high school. In 1994-95, 11% of all mathematics and science teachers in the SSI states were directly participating in the SSIs (Shields, Marsh, & Adelman, 1998, p. 7); nearly 20% of the middle grades mathematics and science teachers participated in some way that year. These figures are not cumulative since the initial funding of the SSIs, but represent a lower boundary for the number of teachers reached up to that school year.

Design of Study

In this report, we explain in some depth the technical issues we have faced and resolved in undertaking a comparison between SSI states as one group and non-SSI states as a contrast group. One in a series, this study is designed to examine differences and similarities between the two groups of states. In following studies, we will provide data on individual SSI states in state profiles. The within-SSI program differences among states are large. The state profiles will be used to describe performance and process in the individual states so that cluster, or individual states, can be studied in more detail. In another study, we will analyze the state assessment data we have acquired from three states. We are looking at state assessment data to both replicate the State NAEP findings from 1990 through 1996 and to project the trajectory of achievement in subsequent years. In our final study, we will analyze individual items and item types to study the pattern of performance on specific topics, NAEP mathematical abilities (conceptual understanding, procedural knowledge, and problem solving), and item formats.

This research is designed to study the impact the Statewide Systemic Initiatives (SSI) have had using the National Assessment of Educational Progress and state assessments as criteria. Beginning in 1990, NAEP provided uniform data for a number of states. Mathematics achievement data along with teacher, student, and school policy information is available for grade 8 in 1990, 1992, and 1996 and grades 4 in 1992 and 1996. This time span is fortuitous for evaluating the impact of SSI during its implementation. 1990 and, for some states, 1992 data serve as a baseline, providing information about the status of mathematics achievement and related practices just prior to the beginning of SSI. Information available from the 1996 NAEP project allows for an initial study of the impact of SSI on both students' mathematics performance and associated policies and procedures.

The overall strategy for determining SSI impact involves comparison of SSI and non-SSI states on a variety of variables. Three types of comparisons were made: status at all three years; three-point trend analysis using data points at 1990, 1992 and 1996; and two-point trends using data from 1992 and 1996. Because of the voluntary nature of the State NAEP, each of the comparisons uses a different sample of states. In addition to examining differences in means for SSI and non-SSI states, where sample sizes permit, achievement levels of minority and majority groups are compared to assess the extent of the gap-closing purpose of the National Science Foundation in creating the SSI program.

Three types of variables are available from the State NAEP: 1) cognitive achievement as measured by several types of mathematics questions; 2) demographic analysis; and, 3) policy and practice indicators that are based on teacher, student, or principal questionnaires. Mathematics achievement and six teacher questionnaire-based items and scales serve as the dependent variables for the studies. The questionnaire-related variables are: relative emphasis on reasoning and communication; opportunities for mathematical discourse; reform topics studies; the NCTM *Standards*; last year's professional development; and, calculator use.

In order to understand the characteristics of the two groups of states, descriptive studies based on the demographics variables collected by NAEP were conducted. Because of the differences in the composition of the state groups as a function of participation in the State NAEP in the three assessment years and its impact on states available for various trend studies, the demographics of several state groupings were studied separately. In addition, descriptive studies of trends in average scale scores over 1990, 1992, and 1996 and cohort growth in average scale scores from grade 4 (1992) to grade 8 (1996) were done for the total group, as well as gender and ethnic breakdowns for composite scores, subtopic scores, and gaps between the different groups.

Multiple linear regression models are used to evaluate the differences between SSI and non-SSI states with respect to the teacher questionnaire-based indicators of curricular reform. Hierarchical Linear Models are employed to compare SSI and non-SSI states with respect to status and trends associated with State NAEP mathematics achievement data.

A second phase of the impact study will use data obtained from state assessments in Texas, Maine, and Massachusetts. The results of these local tests will be evaluated to determine the degree to which they are consistent with NAEP status and trends. In addition, because the data for each of the three states is available through 1999, it will be possible to extend the study of achievement trends beyond what is possible with existing NAEP data. Further benefits from the study of state assessment data come from their closer alignment with the states' standards and more detailed information about mathematics subtopics. Since two of the three states provide student level information about individual items, greater flexibility in the study of the impact of reform on various item content and type (e.g., multiple-choice vs. open-ended) configurations is possible.

CHAPTER 2

REVIEW OF RECENT STUDIES

Correlates of Student Achievement

Several recent studies (Grissmer et al., 2000; Raudenbush et al., 1998; Raudenbush et al., 1999; and, Wenglinski, 2000) have examined the impact of a variety of factors on achievement as measured by the National Assessment of Educational Progress (NAEP). Klein et al. (2000b) examined a similar question using local tests as the measure of mathematics achievement. Table 2.1 lists the variables used in these studies within the following categories: student, family, and home characteristics; educational resources and teacher characteristics; schooling characteristics; and, classroom practices. There is considerable overlap among the factors investigated by the several projects. The findings reviewed below reveal not only direct relationships between independent variables and achievement but also complex interrelationships among the independent variables. For example, in some cases, student, family, and home characteristics correlate with educational resources and teacher characteristics in a way that reflects the fact that more advantaged students tend to have access to better education (Raudenbush et al., 1999).

Among student, family, and home characteristics, the strongest associations with achievement are typically found for parental educational levels, family income, and race/ethnicity (Grissmer et al., 2000). Raudenbush et al. (1999) found significant associations with NAEP mathematics achievement for all of the variables in this category. Furthermore, these authors found that home and family characteristics tend to be reflected indirectly in achievement through their impact on school resources, as well as having a direct effect on learning. Poverty appears to impact achievement at both the individual and school level. Regardless of their proportion, poor students in schools tend to show lower achievement. Raudenbush (1998) observed an impact of poverty on school quality in that schools with large numbers of students living below the poverty line have fewer resources than schools that have a smaller number of students from homes with poverty-level incomes. Similarly, he found that median family income of a school population, even with family income level controlled, still accounts for a significant portion of the variation in NAEP mathematics achievement. Bolstering Phillips' (2000) observation that the effect of racial and economic factors on achievement are distinct, Raudenbush et al. (1999) found that the percentage of minority students in a school is negatively related to mathematics achievement, in addition to the variation accounted for by family income level

Focusing on educational resources and teacher characteristics and controlling for student, family, and home characteristics, Grissmer et al. (2000) showed that higher per-pupil expenditures, lower pupil-teacher ratios at early grade levels, higher reported adequacy of teacher-reported resources, and lower teacher turnover were positively related to student achievement as indicated by an aggregate of NAEP mathematics and reading scores. These variables, together with higher levels of participation in public pre-kindergarten, reportedly

Table 2.1

Correlates of Student Achievement

	Grissmer	Raudenbush	Klein	Wenglinski
Student, family, home characteristics	• Race • Parent education • Family income • Family composition • Mobility • Working mothers • Teen Births	• Gender • Race • Parent education • TV watching • Mobility • Regular newspapers • Books in home • Regular magazines • Family income • Family structure	• Race • Gender • Free/reduced-cost lunch • Language • Special Ed • Previous years performance	• Parents education • Newspapers and magazines in the home • Number of books in the home
Educational resources and teacher characteristics	• Pupil-teacher ratio • Per-pupil expenditure • Teacher salary • Teacher education • Teacher experience • Teacher mobility • Adequacy of teacher resources • Community type	• Teacher experience • Teacher education • Teacher major • Instructional expenditures per student • Percent minority • Location • Median family income	• Teacher Degree • Teacher course work • Course work • Teacher gender • Teacher ethnicity • Teacher experience	• Teacher experience • Teacher education • Teacher major • Professional development o Different populations o Ongoing assessment o Higher-order thinking o Interdisciplinary teaching o Classroom management o Cooperative learning
Schooling characteristics	• % Pre-kindergarten	• Taking algebra • Taking pre-algebra • Eighth grade algebra • Availability of computers • School climate • Reasoning		
Classroom practices			• Time on mathematics • Approach to introducing topics • Typical instructional practices • Typical student activities • Types of written assignments • Use of written work • Methods of assessing	• Working in groups • Using written materials • Writing about mathematics • Hands-on learning • Point-in time assessments • On-going assessments • Talk about mathematics • Address routine problems • Address algebra • Address unique problems • Address geometry • Assign homework

accounted for half of the variation in aggregate NAEP achievement not attributable to student, family, and home factors. Teacher salaries, teacher educational levels, and increased experience over the previous three years did not account for a significant amount of achievement variation. Wenglinski (2000) also examined teacher characteristics, emphasizing professional development, and found that teachers' majors/minors in mathematics and professional development that focused on working with different student populations and teaching higher-order thinking skills were, when student, family and, home characteristics were controlled for, positively related to mathematics achievement on the NAEP.

School-level variables are of particular interest because they are subject to policy decisions at the local level. Grissmer et al. (2000) found that the percentage of children participating in public pre-kindergarten was positively related to NAEP achievement. A second school-level policy variable that is related to higher NAEP mathematics achievement is the availability of high school algebra for grade 8 students. Raudenbush et al. (1999) reported an effect of 1.0 *SD* for taking algebra in the eighth grade. Those taking pre-algebra outperformed those taking eighth grade mathematics or other non-algebra mathematics by .4 *SD*.

In their study on the impact of classroom practices in science and mathematics of the National Science Foundation's Statewide Systemic Initiatives, Shields, Marsh, and Adelman (1998) identified a number of strategies that are believed to be effective for improving student achievement. They listed: greater emphasis on understanding mathematics concepts; application of knowledge to everyday situations; integration of concepts across subjects; the engagement of students in their own learning; sensitivity to individual students' learning styles; increased use of technology; use of new forms of assessment for instructional planning; more emphasis on data gathering and analysis, statistics, geometry and visualization; discovery learning; and, a constructivist approach. NAEP's teacher questionnaire items have served as the basis for several studies of the impact of classroom practices on mathematics achievement (Grissmer et al., 2000; Raudenbush et al., 1998; and, Wenglinski, 2000). Klein et al. (2000a) used a longer survey of reform practices; his study found small, positive but "rarely significant" relationships between teaching practices and student mathematics performance on open-ended items. Three of six sites showed significance on open-ended items. In order to provide a sense of effect size, the researchers noted that for the largest observed positive relationship, "Our model suggests that with a teacher at this site using all of the reform practices monthly, the average student was predicted to score at about the 48^{th} percentile on the test, while for a teacher using all of the reform practices weekly, we would predict that a similar student would score at about the 54^{th} percentile" (p. 27). Three of the sites showed negative, or insignificant relationships between traditional classroom practices and student mathematics achievement as measured by open-ended items. It was observed that the direction of the reported relationships were what would be expected and that the modest correlations were not unexpected due to students' brief period of exposure to reformed classroom practices. These researchers noted the insensitivity of the test instruments due to lack of alignment with curriculum and instruction and lower than desirable reliability coefficients.

Specific classroom practices can be related to NAEP mathematics achievement. Wenglinski (2000) found that students who were exposed (as indicated by their teachers' answers on their questionnaire) to weekly hands-on learning and "a lot" of teaching of higher-order thinking skills were 39% of a grade level ahead of their peers. Students exposed to on-going assessment on a frequent basis were 46% of a grade level behind those of their peers who encountered such practices on a less frequent basis. Further reinforcing the value of teaching higher-order thinking, Raudenbush (2000) also found that students whose teachers indicated an emphasis on teaching mathematical reasoning skills had higher NAEP mathematics scores than students whose teachers emphasized these skills less.

Grissmer et al. (2000), controlling for a number of student, family, and home characteristics, attributed "differences in scores by state for students from similar families" (p. 47), in part, to pupil-teacher ratios, pre-kindergarten participation, teacher mobility, and adequacy of teacher resources. This suggests that policy decisions can affect student performance at the state level.

Beginning in 1990, the National Assessment of Educational Progress has offered a voluntary testing program that is designed to allow comparisons among student achievement levels of participating states (Allen et al., 1997). Observing the strong relationship between demographic factors and educational achievement, Raudenbush et al. (1999) addressed the questions about the appropriate use of state achievement scores when making comparisons of states' educational accomplishments. Noting the positive relationship of student social and economic background factors with both achievement scores and effective school practices, these researchers asserted that fair, statistically unbiased comparisons require a model that includes "social composition, school policy, and practice" (p. 434). As a result of applying a two-stage approach using a hierarchical linear model for within-state analysis and a Bayesian synthesis for the second-stage, between-state analysis, the authors concluded that "Most of the state-to-state heterogeneity seems to be explainable on the basis of covariates defined on students, teachers and schools (p. 431)" (see Table 2.1 for a list of the variables used in this study). Based on the results of their analysis, which virtually eliminated most between-state differences by controlling correlates, these researchers believe that rather than compare states on mean achievement scores, the more meaningful comparisons should be based on measures of school policy and practice.

While Raudenbush and his colleagues (1999) focused their study on 1996 NAEP mathematics results, Grissmer (2000) aggregated NAEP data across all available state results for reading and mathematics from the 1990, 1992, and 1996 NAEP tests. To explain differences in state achievement, his project studied the impact of educational resource factors that were influenced by policy.

Mathematics Content

The National Assessment of Educational Progress (NAEP) has been conducted since 1969 in some form (http://nces.ed.gov.nationsreportcard/site/whatis03/). Over time, three distinct NAEP projects have evolved: the Main NAEP, the long-term Trend NAEP, and the State NAEP. The Main NAEP periodically assesses students' achievement in reading, mathematics, science, writing, U.S. history, civics, geography, the arts, and other subjects at grades 4, 8, and 12. The State NAEP has measured writing, reading, mathematics, and science at grades 4 and 8. Student samples for this program are drawn to permit inferences about the achievement levels for each participating state. The content of both the Main and State NAEP programs follow curriculum frameworks, developed by the National Assessment Governing Board (The College Board, 1996), which adapt to changes in the nation's curricula. Since 1989, the mathematics tests have followed the recommendations of the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) (often referred to as the NCTM *Standards*). Test-item types for the Main and State NAEP assessments that are consistent with the current state-of-the-art in achievement testing also have evolved.

In contrast, both the student sampling frame and the content of the long-term Trend NAEP has remained essentially unchanged. The Trend program, which began 30 years ago and was intended to monitor general trends in achievement, was NAEP's original program. Rather than focusing on grade levels, the Trend assessment targets students at ages 9, 13, and 17 in mathematics, reading, and science. Unlike the Main and State NAEPs, the content of which evolves to match changes in curriculum and instructional practice, the content blueprints of the Trend tests have not changed.

Since 1990, the frameworks for Main and State NAEP assessments in mathematics have covered five content areas and three mathematical abilities. The content areas are: number sense, property, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and, algebra and functions. The mathematical abilities measured are conceptual understanding, procedural knowledge, and problem solving. Three types of items were employed: multiple-choice, open-ended, and extended open-ended, first used in 1992 (The College Board, 1996).

Wilson and Blank (1999) analyzed results by item type and concluded that students performed most poorly on those items that required the most student written work. They observed that the open-ended items that assess higher-order thinking and require skill in communicating about mathematics were the most difficult.

Comparing changes over time for mathematics for the Trend and Main NAEPs, Loveless and Diperna (2000) observed that results on the Main NAEP between 1990 and 1996 have not evinced the gains shown on the Trend NAEP. Although the Trend NAEP continues the original NAEP practice of testing 9- and 13-year-olds, while the State and Main programs assess at grade 4 and 8 levels, these authors assume that the variation in samples is small enough so as not to invalidate meaningful comparisons between the programs. For example, at grade 8, the gain on the Main assessment was 9 points, while, for 13-year-olds, there was no gain on the Trend measure. The authors' comparison of the content of the two assessments revealed an increase in the proportion of geometry items on the State NAEP. They also noted the introduction of calculator use and the provision of manipulatives for the Main NAEP in 1990. In order to further understand the changes in mathematics achievement, Loveless and Diperna charted the gains and losses in correct response rates for various clusters of items (e.g., geometry, problem solving, data analysis, addition of whole numbers, and fractions) on the two versions. Across all three age levels on the Trend NAEP, they found increases in performance on what they described as "NCTM-like topics" (p. 18), such as geometry, problem solving, and data analysis, and decreases in performance on arithmetic items, such as addition and subtraction of whole numbers and fractions.

Wilson and Blank (1999) observed that the open-ended items that were most difficult for students—items that assess the higher-order thinking and mathematical communication skills—are central to the reforms recommended by the NCTM *Standards*. In order to improve achievement in these areas, the authors suggest that "students in mathematics classes need more opportunities to work non-routine problems, to use higher-order thinking skills, and to communicate their mathematical ideas" (p. 19). While noting that the Trend NAEP showed that students improved on content, such as geometry, which is given more emphasis in current reforms, Loveless and Diperna (2000) concluded that performance in geometry "remained abysmal" (p. 19). Thus, at least these researchers evaluating NAEP trends have agreed that there is a considerable need for improved student learning of the content and skills deemed critical by the NCTM *Standards*. Furthermore, observing the slippage in performance on arithmetic items on the Trend NAEP, Loveless and Diperna (2000) warned that efforts to improve learning of reform skills should not come at the expense of the basics of computation with whole numbers, decimals, and fractions.

Trends

Typically, two approaches are employed for tracking change in educational achievement (Barton & Coley, 1998). Cross-sectional studies monitor performance at established ages or grade levels. This method yields information about the differences in the amount of learning achieved at different ages or grade levels. In cross-sectional studies, differences between groups are due to at least two factors: 1) age or grade level, and 2) differences between groups of students. A second approach, referred to as cohort analysis, follows the performance of a defined group of students as they mature. Cohort studies describe how much a specific group of students learn within a fixed period of their schooling. NAEP produces data on representative samples within grades and ages; however, since different students are sampled, the equivalence of these samples is open to question. Thus, changes in performance may be due to learning or differences in the groups sampled at two grades or age levels. State census testing with student IDs allows researchers to follow the same subjects over time, thus assuring the testing of true cohorts at different grade levels.

The picture of change in mathematics learning over time, as measured by NAEP, also differs depending on whether cohort or cross-sectional methods are used. Using data from the Trend NAEP to compare mathematics achievement of the student cohort that grew from age 9 to 13 in the years 1978 to 1982 to that of the cohort spanning the same age range during the period from 1992 to 1996, Barton and Coley (1998) found no significant difference in the amount of mathematics learned during those four-year intervals. However, if the question about change is framed differently, comparing a cross-section of 9-year-olds' performance in 1978 to the performance of 9-year-olds in 1996, mathematics achievement increases. The same is true for age 13, indicating that students ages 9 and 13 in 1996 demonstrated that they knew more mathematics than had their counterparts 18 years earlier. However, the gain in achievement between age 9 and age 13 remained constant. Using State NAEP data, Grissmer et al. (2000) identified significant gains in mathematics achievement between 1990 and 1996 for the states taken as a whole. Grade 8 scores improved more than grade 4 scores. States differed in the amount of gain, most being significant, with mean improvements per year ranging from zero to two percentile points.

Depending on how change is defined, one could draw conflicting conclusions about trends between 1978 and 1996. Barton and Coley (1998) suggest that cohort studies, because they focus on similar groups of students (either by following identical students or sampling from the same population) and have a built-in control for certain demographic and family factors, may be a better measure of educational effectiveness than cross-sectional studies. However, these authors do not take the position that one approach to studying educational achievement trends is generally preferable; rather, they contend that both types of information should be considered. Since 1984, the State and Main NAEP assessments have been spaced to allow sampling of the same cohort of students to be tested first in grade 4 and subsequently in grades 8 and 12. The NAEP state-level assessment that began in 1990 allows tracking of cohorts from grade 4 in 1992 to grade 8 in 1996 (Allen et al., 1997).

Comparative judgments about states' educational quality are often made by ranking mean proficiency scores at a given grade level. For the State NAEP, grades 4 and 8 are used. Barton and Coley (1998) pointed out that comparing the cohort gains of states often shows a different picture and may be the preferred way for judging the effectiveness of schooling. These researchers used as examples the results of Arkansas and Maine. In 1992, Maine fourth graders led the nation with an average scale score of 232 on the NAEP mathematics test, while Arkansas was at the bottom with a score of 210. However, between 1992 and 1996, the gain in each state between grade 4 and grade 8 was the same, 52 points. Based on this comparison of students in the two states, one can conclude that the effectiveness of mathematics education in those states between grades 4 and 8 was equivalent.

For over thirty years, closing the educational achievement gaps between advantaged and disadvantaged children has been a primary focus of state and federal policy (US Department of Health, Education, and Welfare, 1972; Odden, 1991). Phillips (2000) shows that if parents' education and income, percent of students in a school receiving free or reduced-cost lunch, and the locale of a school are controlled for, a Black/White gap in achievement remains. Based on a meta-analysis of a number of cross-sectional studies conducted between 1965 and 1996, Phillips (2000) reported an effect size of the Black/White gap of about .8 *SD* averaged over 12 grades and that, when controlling for historical trends, it increases roughly .18 *SD* between grades 1 and 12.

Phillips (2000) also reports the implications of combining the results of two cohort studies—the National Education Longitudinal Survey (NELS) and Prospects (Phillips, Crouse, & Ralph, 1998)—of the trajectory of the Black/White gap across the grades. For mathematics, Blacks trailed Whites by less than .1 *SD* through grade 6. Between the end of grade 6 and grade 9, the educational achievement gap appears to widen by roughly .1 *SD* per grade and then levels off. While these data appear to be among the best available for understanding the Black/White gap over an extended period, the author cautions that the results of her analysis is "very imprecise" (p. 108). Comparison of the grade 4 and 8 NAEP results for 1992 and 1996 did not detect such a widening of the Black/White gap. Rather, these results show that the gains for Whites and Blacks are roughly the same, indicating no significant change (Barton & Coley, 1998) in the size of the gap during these grade intervals (Shaughnessy et al., 1997).

Comparing State Assessments and NAEP Results

Comparability of NAEP to other assessments is difficult due to the differences in content coverage, item format, test-administration procedures, intended use, and the consequences associated with the use of the results (Feuer et al., 1999). Linn (2000) observed that comparability of state assessments with NAEP scores is sometimes compromised when purposes differ. Evidence suggests that variation in item format may result in different estimates of student knowledge and skills across content topics and content dimensions (Linn et al., 1991; Kenney & Silver, 1998).

The Main and State NAEP tests are designed to reflect "many of the state's curricular emphases and objectives in addition to what various scholars, practitioners, and interested citizens believed should be included in the curriculum." Its purpose "is to provide information about the progress and achievement of students in general" (Allen et al., 1997, p. 20). NAEP Main and State mathematics tests attempt to represent a broad nationwide consensus regarding what is deemed important content for curricula and the learning that should result, both with respect to topics—e.g., geometry, measurement, algebra—and cognitive dimensions—e.g., problem solving, reasoning, or recall of facts (Kenney & Silver, 1998). The NAEP multi-matrix sampling design allows for the administration of over 160 items in various formats with nearly an equal distribution of multiple-choice and extended-response items (Allen et al., 1997). Finally, it is important to note that many of the NAEP administration procedures are different from other large-scale assessments: e.g., testing for any one student usually lasts only an hour; test administrators are well trained and often monitored; and, members of testing groups are randomly selected from school grade-level populations rather than being intact classes (Allen et al., 1997). Furthermore, NAEP participants often lack extrinsic motivation to perform well (Feuer et al., 1999).

To the extent that state assessment procedures differ from those of NAEP, comparability of results is likely to be compromised. Some differences are pervasive. Klein et al. (2000b) noted that, because of multi-matrix item sampling, content coverage of NAEP tests is much broader than is feasible for typical state assessments, which are usually designed to report comparable results for each student at a grade level. Kenney and Silver (1998) reported that even where the state and NAEP test frameworks match well, subtle differences between a state's curriculum and instruction and those targeted by NAEP may compromise the comparability of results from the two assessments. Wilson and Blank's (1999) observation that students scored poorly on items demanding high production (open-ended and extended open-ended items) is another reason why NAEP results may not be comparable with those from state assessments, particularly if the different assessments vary in the proportion of item formats used. It is suggested (Kane, personal communication, February, 2001) that the observed discrepancy between results on open-ended and multiple-choice items may be attributable to lack of motivation associated with NAEP tests.

Comparability of trends may also be affected by the differing characteristics of assessment and accountability programs. Linn (2000) concluded that the stakes and regularity of testing are important considerations when comparing achievement levels from year to year. He observed that where the same test is administered annually within a district or state, achievement levels increase. When new tests, covering similar content are introduced, mean achievement levels drop. Thus, within a particular state or district context, familiarity with a test over time can result in improved performance. Linn (2000) suggests that that this familiarity effect actually represents an upwards bias in the estimates of student learning.

Where promotion decisions depend on test results, cohort achievement gain indicators may be biased upwards by retention (Haney, 2000). This occurs because the anticipated lower scores (of retained students) are removed from populations at grades subsequent to grades where promotion decisions are made. Cross-sectional comparisons may be biased downward at grades in which students are held back and biased upwards in subsequent years (Klein et al., 2000b). Again, this occurs because of changes in grade-level population characteristics that increase the proportion of low performers in decision years and decrease the proportion of these students in subsequent years. Because NAEP testing involves random sampling at all levels and comes every two or four years, the familiarity effect is unlikely to be caused by NAEP. However, both NAEP and state gain indicators should reflect the bias that may occur because of retaining students. One must be wary of gains on both types of assessment in such high-stakes environments.

NAEP scores have been proposed as a criterion for judging the validity of state assessments (Klein et al., 2000b). As Cronbach observed over 30 years ago, “one validates, not a test, but an interpretation of data arising from a specified procedure” (Cronbach in Thorndike, 1971). Thus, questions of a test’s validity depend on the interpretation or use of the results stemming from the test. Such use of results depends on the purpose for which the test is designed. It seems that NAEP’s results could prove useful as a criterion for judging the validity of a state’s assessment to the extent that it shares the same design, purpose, and content as NAEP.

Texas presents a case in point regarding the use of the NAEP in judging the validity of its state assessment, the Texas Assessment of Academic Skills (TAAS). The content strands for the State NAEP are: number sense, property, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions. The NAEP item types include multiple-choice, open-ended, and extended open-ended (The College Board, 1996); TAAS uses only multiple-choice items. The grade 6 TAAS from 1994 to 1999 sampled 13 objectives that are grouped into three domains, as indicated in Figure 2.1. The eighth grade TAAS had 60 items, while the 1996 State NAEP used 162 items. Nine of the TAAS objectives seem to fit within the single NAEP strand of number sense, property, and operations. Clearly, the variety of item types used for the NAEP would suggest that it has greater breadth than TAAS. While there are TAAS objectives that relate to each of the NAEP strands, the weighting of topics is much different.

Figure 2.1. TAAS Mathematics Domains and Objectives.

Domain: Concepts

1. The student will demonstrate an understanding of number concepts.
2. The student will demonstrate an understanding of mathematical relations, functions, and other algebraic concepts.
3. The student will demonstrate an understanding of geometric properties and relationships.
4. The student will demonstrate an understanding of measurement concepts using metric and customary units.
5. The student will demonstrate an understanding of probability and statistics.

Domain: Operations

6. The student will use the operation of addition to solve problems.
7. The student will use the operation of subtraction to solve problems.
8. The student will use the operation of multiplication to solve problems.
9. The student will use the operation of division to solve problems.

Domain: Problem Solving

10. The student will estimate solutions to a problem situation.
11. The student will determine solution strategies and will analyze or solve problems.
12. The student will express or solve problems using mathematical representation.
13. The student will evaluate the reasonableness of a solution to a problem situation. (Texas Education Agency, 1999)

Mehrens (2000) points out that the TAAS is well designed to measure the Texas Essential Elements, which comprise the state's written curriculum developed by Texas educators. If one believes that curriculum and test domains are the purview of states and local school districts, then the content validity of instruments intended to measure the domains should be judged with respect to the specification made by those jurisdictions. Based on his observation that every objective the TAAS purports to measure is tested every year and is clear documentation of sound procedures for matching test items to the Texas Essential Skills (Texas Educational Agency, 1999), Mehrens argues that the TAAS instruments meet the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council of Measurement in Education, 1999) for content validity.

Believing that the State NAEP mathematics assessment should be considered the "gold standard" for mathematics content and thus a proper "benchmark" to use in evaluating the validity of TAAS, and because the effect size of annual cross-sectional gains on NAEP are considerably smaller than those of TAAS, Klein et al. (2000b) claim that TAAS is an inflated indicator of mathematics achievement for Texas students. Consistent with Cronbach's definition, the validity of TAAS depends on the interpretations made from its results. Based on the Mehrens study, claims of validity for TAAS mathematics tests seem justified as long as interpretations of TAAS results are clearly limited to the Texas Essential Elements and the specific objectives of the TAAS. However, more general claims about a construct as broad as "mathematics achievement" based on interpretations of TAAS may not be considered valid because of the narrowness of the content of TAAS vis-à-vis the widely accepted NCTM *Standards*. In terms of gains, the TAAS would seem to be a valid indicator for the Essential Elements and TAAS objectives, but not of mathematics achievement in general. Klein et al. (2000b) suggest that TAAS is not valid because teachers are in effect teaching to the test. However, it is clear, because test forms are changed each year and are secure, that such teaching can only be to the well-defined content domain of the instruments and not to specific items. As long as the inferences intended for the results are limited to this well-defined content, there should be no question of the test's validity for judging gain. However, interpretations of TAAS longitudinal data that go beyond the specific content of the tests would be questionable. It is reasonable and valid to attribute the steady improved performance on the TAAS to better teaching and learning of the Texas Essential Elements. Claims of more general improvement in mathematics by Texas, while plausible, should not be based on TAAS results.

CHAPTER 3

METHODOLOGICAL ISSUES

The State NAEP database is complex and idiosyncratic. Its use for the Study of the Impact of the Statewide Systemic Initiatives in mathematics raised several methodological issues. In the sections that follow, methodological questions that arose during this research project and the solutions we applied are reviewed.

NAEP Sampling and Weighting

Weighting

The State NAEP employed complex, sophisticated student and item sampling in order to maximize the content coverage of the assessment while, at the same time, minimizing the number of students involved and the time they needed to spend taking the tests. Approximately 100 schools were sampled in each participating state. Thirty students were randomly chosen in each school, resulting in sample sizes of approximately 3,000 students in each state. The sampling plan was designed to provide estimates for the public school population of an entire state, as well as of certain selected subpopulations. Stratification was by urbanization, percentage of Black and Hispanic students enrolled, and median household income within the ZIP code area of the school (Allen et al., 1997; Mullis et al., 1993). Even though the sampling was stratified by school and district, it yielded valid representation only at the state level.

In 1992 and 1996, State NAEP databases provided weights to allow for equal or proportional weighting of states. While the choice of weighting procedure does not affect state means, the proportional approach weights states by their populations, so larger states contribute more to population estimates. Use of the equal weighting approach has the effect of treating each state as though it has the same-sized population. Since the SSI focused on the state level, for this study we have, where available, used equal weighting in order to consider each SSI state as an equal and independent replication of SSI reform efforts. The 1990 State NAEP database includes only proportional weights; using techniques similar to those used by NAEP for 1992 and 1996, we computed equal weights for the 1990 analysis.

State means were computed using the equal weights provided by the State NAEP database. In this approach, the state is considered the unit of analysis, with each state's NAEP results reflecting either its SSI or non-SSI status. While states are the unit of analysis, students are the sampling unit within the NAEP design. Weights are applied to individual student data that are then aggregated to determine weighted state values.

In addition to measuring students' mathematics achievement, teachers whose students were sampled by the State NAEP were asked to complete a questionnaire on their background, training, and instructional practices (Allen et al., 1997). These teacher responses are merged with the achievement item responses of each of the sampled students to which they were teaching mathematics at the time of assessment, making one record in the data file.

Results of the teacher questionnaire are reported in terms of the percentage of students with teachers choosing each questionnaire response. The same student weights are applied to the cognitive and teacher questionnaire items.

Population Subgroups

Analysis of student performance by certain demographic categories is a desirable feature of the Study of the Impact of Statewide Systemic Initiatives. However, in a few states, the numbers of students in some racial/ ethnic categories are too small to allow accurate estimates of population values for these subgroups. Because of this, we reduced the number of states in the analyses that compared the performance of students by racial/ethnic groups to include only those states with a sufficient student ethnic population to make valid inferences.

Participation Guidelines

Because district and school cooperation with the State NAEP is typically voluntary, in order to ensure the integrity of state samples, the National Center for Education Statistics (NCES) established participation rate standards for schools and students. These standards seek to reduce bias due to school non-response, as well as that due to inadequate strata-specific representation of a population with respect to students with disability, limited English proficiency, types of assessment session (monitored or unmonitored), school level of urbanization, minority enrollment, and median household income of the area in which the school is located. State NAEP results are reported for some states that did not fully meet these NCES standards. In the portion of this study that focuses on the impact of an SSI on teacher practices, all analyses were done twice—once for all states and a second time for only those states that met the NCES participation rate standards. Results of the two analyses were then compared to consider whether the findings of the larger group of states were affected by results from states that did not meet the standards. Appendix A contains a description of the participation rate standards.

Groupings of States

In order to evaluate the relation of the Statewide Systemic Initiatives to teacher practices and characteristics and to student achievement, the mean of SSI states was compared to the mean of non-SSI states. We did cross-sectional comparisons to examine differences at given years and longitudinal comparisons to look at trends over two or three testing points. All states included in the State NAEP database in a given year were included in the comparison for that year. Data from states that participated in the State NAEP in consecutive years was used for the longitudinal comparisons.

While the jurisdictions of Guam, Puerto Rico, the Virgin Islands, Washington, DC, and Department of Defense Schools were included in the State NAEP, they were not included in this study because our focus was on states rather than on jurisdictions generally.

Because of self-selection, a different set of states participated in the State NAEP in each testing year (1990, 1992, and 1996), resulting in different groupings of both SSI and non-SSI states for each of the three years. Three sets of comparison groups were used for the study. Separate comparison groups were employed to study yearly status, and two-point and three-point trends. The number of states available for various analyses declined as the number of points increased. Because the State NAEP was not administered at grade 4 in 1990, three-point analysis was limited to grade 8. While the self-selection of states for NAEP participation may affect the results of this study of SSI impact, it is not possible to isolate such an effect. The numbers, percentages, and names of states included in the three types of comparisons are provided in Appendix B of this chapter.

There is no basis for believing that SSI and non-SSI groups were equivalent on any of the variables of interest prior to implementation of the SSI. In fact, SSI states, as a whole, had a higher proportion of minority students than did the non-SSI states in 1990, prior to NSF's startup of the program.

Indicator Development

The teacher questionnaire responses associated with each sampled student in the State NAEP database provide information about teacher background, training, and instructional practices. One aspect of this study was to discern the impact of SSI initiatives on teacher variables captured in the questionnaire data. Clune's (1998) theory of systemic reform was used as a classification scheme to identify those items that might reflect goals of reform. Based on classification within this framework, sets of items were identified as potential elements of reform variables. Combining items into scales has several benefits:

- sets of items delineate theoretical constructs;
- random error is reduced and true score variability increased;
- reporting is parsimonious; and,
- parametric analysis may be used.

Changes in teacher questionnaires in each year of the State NAEP have complicated longitudinal comparisons of reform variable values. Few of the questionnaire items remained exactly the same in 1990, 1992, and 1996. Wording and the number and descriptions of response options were changed. In addition, a number of items were added. Longitudinal comparisons have been most affected by this lack of consistency.

Longitudinal Analysis

Student Achievement

To study relationships between SSI status and gains in mathematical achievement, we used two kinds of statistical analyses. State NAEP data have a hierarchical structure: students are nested within schools, and schools are nested within states. Hierarchical linear modeling is often used with such data (Bryk & Raudenbush, 1992; Goldstein, 1995; Snijders & Bosker, 1999). However, the State NAEP data have special characteristics that might make other methods more appropriate. Three unique features of the State NAEP data are: the small number of state-level cases; states' voluntary participation in tests, violating the assumption of randomness; and, the heterogeneous variance structure of each state, instead of the homogeneity assumed by the model (Raudenbush et al., 1999).

Aware of these limitations, we used two different methods for the longitudinal analyses of mathematics achievement: Descriptive Trend Analysis (Barton & Coley, 1998; Grissmer et al., 2000) and Empirical Bayes and Bayesian Analysis (Raudenbush et al., 1999).

Due to the "comparable metrics" of State NAEP scores across data collection years and across grades (Allen et al., 1997), we can study the changes in achievement scores of each grade across years and cohort growth from grade 4 to grade 8.

With descriptive trend analyses, we compared the average performance of SSI states and non-SSI states across gender and race over the assessment years to identify the longitudinal growth of student achievement on the NAEP mathematics composite achievement scale as well as in the five content strands (i.e., number and operations, measurement, geometry, data analysis, and algebra and functions).

Empirical Bayes and Bayesian analysis is similar to a meta-analysis, as described in Bryk and Raudenbush (1992, Chapter 7). This method incorporates jackknife standard errors (as described below) to create confidence intervals around each state mean for grades 4 and 8 in each test year. Then the individual state estimates are combined to estimate an overall mean for the SSI and non-SSI states. From a longitudinal perspective, this method obtains estimates of the average state mean in 1990, along with the state growth rate per year from 1990 to 1996.

Reform Indicators

In addition to examining the effect of an SSI on student achievement, we also used teacher questionnaire data to evaluate the impact of SSI on instructional practices associated with reform. For this technical report, comparisons between SSI and non-SSI states on the reform indicators are limited analyses of the state means for the SSI and non-SSI states. Within state variability on the reform indicators is not addressed as part of this technical report.

Some indicators have a common scale across two or three years. For these, repeated measures analysis of variance (ANOVA) were used to examine the main effects of SSI status and time as well as the interaction of time and SSI status.

Other indicators had different scales in different years. For these, linear regression was used to determine whether a state's SSI status was related to the reform indicator in 1996. A two-step regression model was used. At Step 1, the value of the 1996 indicator was predicted from prior values (either 1990 or 1992). At Step 2, SSI status was added as a predictor. The purpose of this two-step approach is to examine the contribution of SSI to changes in reform-related activities, beyond the general trend across all states. The F test for SSI status in Step 2 was used to evaluate whether SSI status contributed to the 1996 indicator.

To provide another perspective for interpreting the relationship between the indicator variables and SSI status, a descriptive discriminant function analysis and canonical correlations were calculated. These methods provided information about the utility of the indicator variables as a basis for classifying states as SSI or non-SSI. Classification based on the obtained discriminant function was compared with the actual SSI status of states. The canonical correlation between the indicators and SSI status provided a familiar index of the relationship of the indicator variables taken together with SSI status.

Interpreting Differences

Standard Errors of Estimation

Analysis of State NAEP data for this project required calculating means and standard errors on a variety of measures for state populations at two grade levels and in three different years, as well as for population subgroups within states. Because State NAEP data collection procedures use complicated sampling plans, routine procedures for calculating means and standard errors are inappropriate. In particular, the State NAEP employed stratified samples in each state and did not sample at the same rate in different strata. The NAEP database includes replicate weights to use with a re-sampling technique to estimate the standard error of the means. For this report, we used the jackknife procedure (Bradley, 1982) to estimate the standard errors in the individual state means for the within-state analyses. This re-sampling method involves five steps:

1. An estimate of the mean is obtained for the complete set of data in the state, using the base weights for the state.
2. Multiple subsets of the data are sampled to generate multiple estimates of the mean. For the State NAEP, the re-sampling is achieved through the use of sets of replicate weights, which effectively select subsets of the data.
3. Each set of replicate weights is then used to generate an estimate of the mean. These separate estimates are sometimes called pseudo-estimates.
4. The mean of the pseudo-estimates is the jackknife estimate of the mean.
5. Using the pseudo-estimates as data, the jackknife procedure generates an estimate of the standard estimate of the mean for each state.

The result of the jackknife procedure is an unbiased estimate of the mean and the standard error, based on the sampling and weighting procedures developed for the State NAEP data. The standard errors were used to compute confidence intervals around means, indicating the extent to which the mean would vary over independent samples from the state.

For analyses based on state means (e.g., estimating the mean effect across the SSI states), the usual formulas were used to estimate the standard error. This approach weights states equally, using the state mean as the basic unit of analysis and limiting conclusions to those about state means.

Conceptual Analysis

In this research program, we examine a large number of variables over several subgroups within the State NAEP database and from some selected state databases. A number of the comparisons are based on the NAEP achievement test results, but most also involve demographic variables (e.g., gender, race) and reform indicators (e.g., emphasis on reasoning and communication).

In this kind of correlational and retrospective analysis, some statistically significant findings represent substantive differences while others are due to chance. Chance findings are especially problematic in a very large and complex database, like that of the State NAEP results, in which there are many variables and many observed differences are statistically significant because of the large sample sizes.

One way to minimize the misinterpretation of chance findings is to examine the data within an explanatory framework. In this study we have used a model of systemic reform to identify comparisons of interest. In addition, we examined each specific question under consideration from several perspectives in order to build a set of results that, taken together, could support a general conclusion. In the absence of experimental controls, this kind of multiple testing of the network of relationships can be an effective way to use data in building confidence in causal hypotheses. However, firm conclusions about causality depend on future research specifically designed to test the hypotheses.

Effect Size

Research on factors that influence student achievement is moving away from statistical significance testing to estimates of effect size (Cohen, 1969). Rather than simply evaluating the null hypothesis, differences can be evaluated in relation to the standard deviations of the measures. Klein et al. (2000b) consider effect sizes of trend data in the Texas Assessment of Academic Skills state assessment results between .31 and .49 *SD* units to be indicative of "very large improvement" (p. 6). Phillips's (2000) summary of a meta-analysis of studies of Black/White achievement differences showed average effect sizes across twelve grades of roughly .8 *SD* units. At this point in our study of SSI impact, we are exploring the feasibility of developing conventions for the interpretation of effect sizes.

Attribution

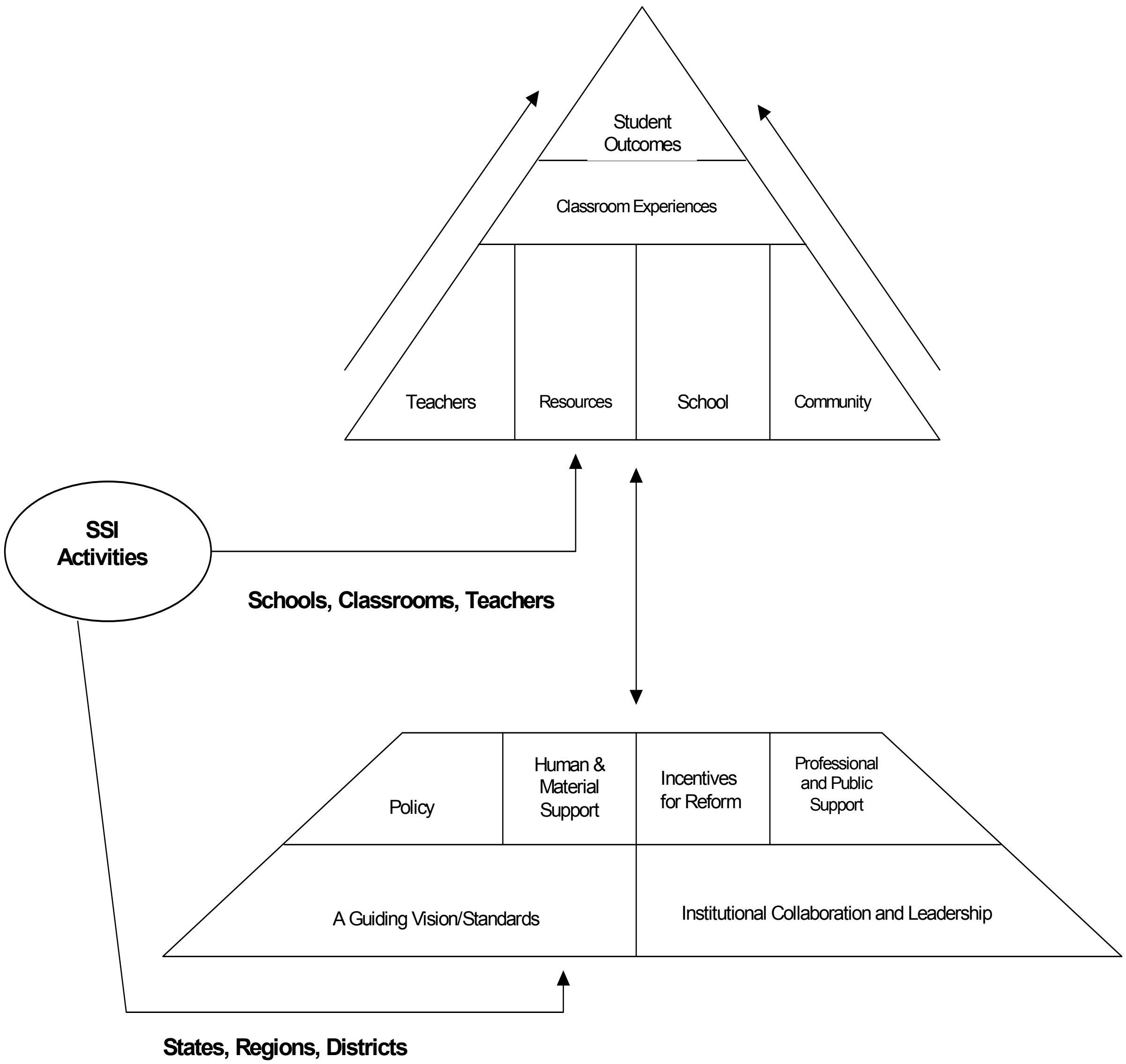
An evaluation team at Stanford Research Institute (SRI) developed a conceptual model (Figure 3.1) of systemic reform (Zucker et al., 1998) that assigns SSI activities to "two related but distinct channels" (p. 3). The two sets of activities can be distinguished in terms of their distance from the classroom. One aims to build the state, region, and district infrastructure necessary to support and sustain reform. An example of activities at this level is adoption and diffusion of state policy building on the NCTM *Standards* (1989). The second set of activities aims more directly at student outcomes by focusing on initiatives that improve both teaching and the quality of students' learning experiences. An example of this type of activity is professional development designed to improve teacher knowledge and skills for implementing reform practices in their classroom methods, such as having students write about mathematics or work on team problem solving.

The goal of this project is discover the utility of the State NAEP database and the results from several state assessments for detecting the impact of the SSIs. The model depicted in Figure 3.1 suggests that, in addition to looking for evidence of impact in student *achievement results*, where possible evidence of *policy development and implementation* as well as specific *processes* aimed at improving student learning of mathematics should be considered.

Data presented in Chapter 4 show that states differ on both demographic and educational variables. Not only are there differences apparent among states but, also, between regions and, perhaps most importantly, among the sets of SSI and non-SSI states. For example, prior to the advent of the SSI, states that were later awarded grants differed from non-SSI states both on levels of student mathematics achievement and on racial composition. At the beginning of the SSI program, SSI states, as a group, evidenced lower mean mathematics achievement levels than did non-SSI states. Overall, SSI states had higher proportions of minorities and students from economically disadvantaged families than did non-SSI states.

While these differences are perhaps a function of a deliberate effort by NSF to focus resources in the states with greatest need, the lack of initial comparability of SSI and non-SSI

Figure 3.1. A model of systemic reform (Zucker et al., 1998).



states complicates efforts to detect evidence of SSI impact. At first glance, one might consider use of traditional regression procedures that control for pre-existing differences. However, the lack of independence of the variables that reflect group differences from aims of SSI makes these approaches problematic, since removal of the effects of variance due to pre-existing conditions will neutralize variance effected by an SSI. Based on the work of Raudenbush et al. (1999), we are experimenting with Bayesian linear modeling approaches as a means of isolating SSI impact.

Because of limitations associated with pre-existing differences within the groupings of SSI and non-SSI states, an alternative approach for studying SSI impact was needed. We sought to construct a plausible argument that using NAEP data and other existing information to
relate improved student achievement to SSI activities. A new focus on trends in individual states seemed to hold promise. Profiles of SSI states, including demographic and prior achievement data along with detailed information about the focus of SSI activities and theory of action vis-à-vis impact, should serve as a solid basis for establishing an expected locus for impact evidence.

Clune (1998) rated 13 SSI states on the breadth and depth of a number of activity components. His group will eventually have ratings on all SSI states. Figure 3.2 lists the states and components. These ratings, along with additional information on a state's SSI target population, saturation, form, systemicness, and conceptualization of mathematics, (obtained from SSI states as well as various research and evaluation reports) will serve as the basis for the classification of state SSI activities.

One way in which NAEP achievement data can be broken down is by the content strands of the NAEP mathematics framework. In addition, results can be classified by item types, such as short- or extended-constructed response. Results based on such classifications, when compared with a state's emphasis on specific mathematics topics, will be particularly useful when analyzing achievement trends of SSI states. Information about item type and content, which is available from some state assessment databases, will allow similar analysis.

The NAEP sampling, coupled with the uneven saturation of SSI implementation activities within states, presents another complication for detecting the impact of SSI activities and investments. Even when quality information about the nature of SSI activities within states exists, there is no way of knowing the extent to which SSI-targeted schools are represented in the NAEP sample for a given year because most schools are not involved in the NAEP sample and SSI participation is usually not random. This is another reason why the state assessment data can prove useful. With state assessment information available, there is a certainty that SSI-targeted schools are included in the database. Furthermore, where there is specific information about the SSI targeting of schools or districts, the achievement results of those entities can be compared with others to detect impact on student achievement of mathematics outcomes.

Attribution is further hindered by the correlational nature of the statistical methodology available for this study. While we are unable to presume causality for relationships because of the constraints on this study, the possibility that certain relationships exist is important because it suggests areas for further, more refined research. In a similar vein, in order to ensure that all promising relationships are documented, *p* values are usually reported as descriptive information rather than for hypothesis-testing purposes.

In summary, there is great value in accurately describing the group of SSI states in contrast with the group of non-SSI states on common measures. Even if no significant differences are found between the SSI states and the non-SSI states, the accurate description of data can be used to lay the groundwork for more detailed studies and to inform the discussions about other initiatives supporting statewide reform. The design of the State NAEP does allow the investigation of differences among the SSI and non-SSI states beginning with the 1990 measures, prior to the implementation of the SSI program, and at two subsequent points, 1992 and 1996, during the implementation of the SSI program. However, because not all states participated in all three testing times and due to the existence of other co-variables, building analytic models using techniques such as Bayesian linear modeling can uncover some relationships, but cannot capture the impact across a large number of the SSI states. Therefore, in the work we report in the future we have turned to other approaches.

One approach is to develop a data profile on each SSI state, including the emphasis each SSI gives to the five mathematics content topics tested in NAEP. These patterns of emphasis will be compared to the pattern of growth in achievement on the same topics. Another approach is to compare findings from the State NAEP with state assessment data. If the two datasets produce comparable findings, this may confirm improved student achievement. Then the state assessment data, collected from a larger sample than the State NAEP, can be used to relate SSI activities to student achievement by identifying the student performance on schools and districts most active in the SSI. Establishing consistent patterns among SSI activities, state assessment results, and State NAEP results in two or three states will increase the confidence that NAEP data can be used to detect effects related to SSIs. Finally, we will draw upon findings of other studies and our own study of process indicators to link SSI activities to teacher and student report data of classroom activities and to student achievement in the attempt to establish a likely chain of evidence.

The limited data that are available place constraints on fully attributing student increases in learning to SSI activities. However, using these multiple approaches will allow us to make an informed decision about the likelihood of such a relationship.

Figure 3.2. Components and states used for Clune's rating of SSI breadth and depth.

Components	***States***																									
	AR		**CA**		**CT**		**DE**		**GA**		**KY**		**LA**		**ME**		**MI**		**MT**		**NY**		**TX**		**VT**	
	B[1]	***D***[2]	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***	***B***	***D***
Pre-SSI Policy																										
Standards/Frameworks																										
Professional Development																										
Assessment																										
Accountability																										
Other Inst. Guidance Policies																										
Pre-SSI Infrastructure																										
SSI Leadership and Change Strategy																										
Educational Change Strategy																										
Org./System Change Strategy																										
Resources																										
SSI Policy																										
Standards/Frameworks																										
Professional Development																										
Assessment																										
Accountability																										
Other Inst. Guidance Policies																										
SSI infrastructure																										
SSI standards-Based Ins. Reform																										
Individual Capacity Building																										
Org. Capacity Building																										
Classroom Practice																										

[1] Breadth
[2] Depth

LIST OF APPENDICES

Appendix 3A NCES Participation Rate Standards

Appendix 3B Comparison Groups

Table 3B.1a
Number and Percentage of SSI and Non-SSI States Included in Various Comparison Groups

Table 3B.1b
States Tested in Various Years and Included in the Trend Comparison Groups

APPENDICES

Appendix 3A NCES Participation Rate Standards

- A jurisdiction will receive a notation if its weighted participation rate for the initial sample of schools was below 85% AND the weighted school participation rate after substitution was below 90%. (Appendix A in Shaughnessy, Nelson, & Norris, 1997, p. 282)

- A jurisdiction that is not already receiving a notation for problematic overall school or student participation rates will receive a notation if the sampled students within participating schools included a class of students with similar characteristics that had a weighted student response rate of below 80%, and from which the non-responding students together accounted for more than five percent of the jurisdiction's weighted assessable student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the students, whether or not the student was classified as a student with a disability (SD) or of limited English proficiency (LEP), and the type of assessment session (monitored or unmonitored). In addition, for public schools, classes of schools were determined by school level of urbanization, minority enrollment, and median household income of the area in which the school is located. (Appendix A in Shaughnessy, Nelson, & Norris, 1997, p. 283)

Appendix 3B Comparison Groups

Table 3B.1a
Number and Percentage of SSI and Non-SSI States Included in Various Comparison Groups.

	SSI States (n = 25)		Non-SSI States (n = 25)	
	Number	Percent	Number	Percent
Yearly				
Grade 8 1990	20	80	17	68
1992	22	88	19	76
1996	22	88	18	72
Grade 4 1992	22	88	19	76
1996	23	92	20	80
2- Point Trend (1992 –1996)				
Grade 8	20	80	15	60
Grade 4	21	84	16	64
3- Point Trend (1990, 1992, 1996)				
Grade 8	17	68	11	44

Table 3B.1b

States Tested in Various Years and Included in the Trend Comparison Groups

	Grade 4			Grade 8				
	Yearly		Trend	Yearly			Trend	
	1992	1996	2	1990	1992	1996	2	3
SSI								
Arkansas	X	X	X	X	X	X	X	X
California	X	X	X	X	X	X	X	X
Colorado	X	X	X	X	X	X	X	X
Connecticut	X	X	X	X	X	X	X	X
Delaware	X	X	X	X	X	X	X	X
Florida **	X	X	X	X	X	X	X	X
Georgia	X	X	X	X	X	X	X	X
Kentucky	X	X	X	X	X	X	X	X
Louisiana	X	X	X	X	X	X	X	X
Maine	X	X	X		X	X	X	
Massachusetts	X	X	X		X	X	X	
Michigan	X	X	X	X	X	X	X	X
Montana		X		X		X		
Nebraska	X	X	X	X	X	X	X	X
New Mexico	X	X	X	X	X	X	X	X
New York	X	X	X	X	X	X	X	X
New Jersey	X	X	X	X	X			
N. Carolina **	X	X	X	X	X	X	X	X
Ohio	X			X	X			
Rhode Is **	X	X	X	X	X	X	X	X
South Dakota								
S. Carolina	X	X	X		X	X	X	
Texas	X	X	X	X	X	X	X	X
Vermont		X				X		
Virginia **	X	X	X	X	X	X	X	X
Number	**22**	**23**	**21**	**20**	**22**	**22**	**20**	**17**

Non-SSI								
Alabama	X	X	X	X	X	X	X	X
Alaska		X				X		
Arizona	X	X	X	X	X	X	X	X
Hawaii	X	X	X	X	X	X	X	X
Indiana	X	X	X	X	X	X	X	X
Iowa	X	X	X	X	X	X	X	X
Idaho	X			X	X			
Illinois				X				
Kansas				X				
Maryland	X	X	X	X	X	X	X	X
Minnesota	X	X	X	X	X	X	X	X
Mississippi	X	X	X		X	X	X	
Missouri	X	X	X		X	X	X	
Nevada		X						
NewHampshire	X			X	X			
North Dakota	X	X	X	X	X	X	X	X
Oklahoma	X			X	X			
Oregon		X		X		X		
Pennsylvania	X	X	X	X	X			
Tennessee	X	X	X		X	X	X	
Utah	X	X	X		X	X	X	
Washington		X				X		
W. Virginia	X	X	X	X	X	X	X	X
Wisconsin	X	X	X	X	X	X	X	X
Wyoming	X	X	X	X	X	X	X	X
Number	**19**	**20**	**16**	**17**	**19**	**18**	**15**	**11**

** SSI states with less than five years SSI funding

CHAPTER 4

A DESCRIPTION OF THE DEMOGRAPHICS OF SSI COMPARED TO NON-SSI STATES

Introduction

When comparing two groups, it is important to understand the salient characteristics of the groups being compared. This chapter will contrast the SSI states with non-SSI states on the basis of demographic variables utilized in the State NAEP. The main variables are gender, ethnicity, parents' education, and home environment composite. These variables will be presented for different groupings of states by participation in SSI, by year of participation in NAEP, and by analytic group.

In grade 8, 22 SSI states and 18 non-SSI states participated in the State NAEP (Table 4.1). In grade 4, 23 SSI states and 19 non-SSI states participated. Two additional states participated at grade 4, one an SSI state, New Jersey, and one a non-SSI state, Pennsylvania.

Different analytic groupings of states are used for the different analyses. Not all of the states participated in all three years of the State NAEP—1990, 1992, and 1996. In order to do trend analyses over these three years, we have used only those states that participated in all three State NAEP tests (Table 4.2). This, Trend Group 90-96, includes 17 SSI states and 11 non-SSI states. When we use only the 1992 and 1996 data, we increase the number of states participating in the State NAEP to 20 SSI states and 15 non-SSI states. These 34 states will constitute Trend Group 92-96. The largest number of states participated in the State NAEP in 1996—22 SSI states and 18 non-SSI states. This group of states is referred to as the 1996 Group.

We begin this chapter by reporting the demographic variables for the 1996 Group. This is the best representation we have for contrasting the SSI states and non-SSI states using NAEP data. By 1996 most states had been in the SSI program three or more years. We then report the data for Trend Group 90-96, followed by data for Trend Group 92-96. The latter group best reflects any changes in demographics over the first years of the SSI program because 1992 coincides with the beginning of the Statewide Systemic Initiatives program. We compare the 1996 data for each of the trend groups with the 1996 Group to disclose how the demographics of the two groups change when states that did not participate in all three State NAEP assessments are excluded from the analysis (See Appendix A).

Table 4.1
States Participating in the State NAEP

	Grade 8						Grade 4			
	90		92		96		92		96	
	SSI	Non-SSI	SSI	Non-SSI	SSI	Non-SSI	SSI	Non-SSI	SSI	Non-SSI
Complete Sample (17 SSI, 11 Non-SSI)	Arkansas	Alabama	Arkansas	Alabama	Arkansas	Alabama	Arkansas	Alabama	Arkansas	Alabama
	California	Arizona	California	Arizona	California	Arizona	California	Arizona	California	Arizona
	Colorado	Hawaii	Colorado	Hawaii	Colorado	Hawaii	Colorado	Hawaii	Colorado	Hawaii
	Connecticut	Indiana	Connecticut	Indiana	Connecticut	Indiana	Connecticut	Indiana	Connecticut	Indiana
	Delaware	Iowa	Delaware	Iowa	Delaware	Iowa	Delaware	Iowa	Delaware	Iowa
	Florida	Maryland	Florida	Maryland	Florida	Maryland	Florida	Maryland	Florida	Maryland
	Georgia	Minnesota	Georgia	Minnesota	Georgia	Minnesota	Georgia	Minnesota	Georgia	Minnesota
	Kentucky	North Dakota	Kentucky	North Dakota	Kentucky	North Dakota	Kentucky	North Dakota	Kentucky	North Dakota
	Louisiana	West Virginia	Louisiana	West Virginia	Louisiana	West Virginia	Louisiana	West Virginia	Louisiana	West Virginia
	Michigan	Wisconsin	Michigan	Wisconsin	Michigan	Wisconsin	Michigan	Wisconsin	Michigan	Wisconsin
	Nebraska	Wyoming	Nebraska	Wyoming	Nebraska	Wyoming	Nebraska	Wyoming	Nebraska	Wyoming
	New Mexico		New Mexico		New Mexico		New Mexico		New Mexico	
	New York		New York		New York		New York		New York	
	North Carolina		North Carolina		North Carolina		North Carolina		North Carolina	
	Rhode Island		Rhode Island		Rhode Island		Rhode Island		Rhode Island	
	Texas		Texas		Texas		Texas		Texas	
	Virginia		Virginia		Virginia		Virginia		Virginia	
Inconsistent Participation						Alaska				Alaska
		Idaho	Maine	Idaho	Maine		Maine	Idaho	Maine	
		Illinois	Massachusetts		Massachusetts		Massachusetts		Massachusetts	
	Montana			Mississippi	Montana	Mississippi		Mississippi	Montana	Mississippi
	New Jersey		New Jersey	Missouri		Missouri	New Jersey	Missouri	New Jersey	Missouri
	Ohio	New Hampshire	Ohio	New Hampshire			Ohio	New		
		Oklahoma	South Carolina	Oklahoma	South Carolina		South Carolina	Oklahoma	South Carolina	
		Oregon			Vermont	Oregon			Vermont	Oregon
		Pennsylvania		Pennsylvania				Pennsylvania		Pennsylvania
				Tennessee		Tennessee		Tennessee		Tennessee
				Utah		Utah		Utah		Utah
						Washington				Washington
Excluded Sample		District of Columbia		District of Columbia		District of Columbia		District of Columbia		District of Columbia
		Guam		Guam		Guam		Guam		Guam
		Virgin Islands		Virgin Islands				Virgin Islands		

Table 4.2
State NAEP Analytic Sample

1996 Group		**Trend Group 90-96**		**Trend Group 92-96 or Cohort Group**	
SSI	Non-SSI	SSI	Non-SSI	SSI	Non-SSI
N = 22	N = 18	N = 17	N = 11	N = 20	N = 15
Arkansas	Alabama	Arkansas	Alabama	Arkansas	Alabama
California	*Alaska*	California	Arizona	California	Arizona
Colorado	Arizona	Colorado	Hawaii	Colorado	Hawaii
Connecticut	Hawaii	Connecticut	Indiana	Connecticut	Indiana
Delaware	Indiana	Delaware	Iowa	Delaware	Iowa
Florida	Iowa	Florida	Maryland	Florida	Maryland
Georgia	Maryland	Georgia	Minnesota	Georgia	Minnesota
Kentucky	Minnesota	Kentucky	North Dakota	Kentucky	*Mississippi*
Louisiana	*Mississippi*	Louisiana	West Virginia	Louisiana	*Missouri*
Maine	*Missouri*	Michigan	Wisconsin	*Maine*	North Dakota
Massachusetts	North Dakota	Nebraska	Wyoming	*Massachusetts*	*Tennessee*
Michigan	*Oregon*	New Mexico		Michigan	*Utah*
Montana	*Tennessee*	New York		Nebraska	West Virginia
Nebraska	*Utah*	North Carolina		New Mexico	Wisconsin
New Mexico	*Washington*	Rhode Island		New York	Wyoming
New York	West Virginia	Texas		North Carolina	
North Carolina	Wisconsin	Virginia		Rhode Island	
Rhode Island	Wyoming			*South Carolina*	
South Carolina				Texas	
Texas				Virginia	
Vermont					
Virginia					

1996 Group (22 SSI and 18 Non-SSI States)

As expected, the number of students in both SSI and non-SSI states in 1996 was nearly evenly divided between males and females (Figure 4.1). At both grades 4 and 8, the SSI states had, as a group in 1996, a slightly lower percentage of White, Asian, and American Indian students than did the non-SSI states (Figure 4.2). The 22 SSI states had more Black students in both grades 4 and 8, 4% and 5% respectively, and more Hispanic students in both grades 4 and 8, 3% and 5% respectively. These data corroborate that SSI states tended to have a higher percentage of minority students than did non-SSI states.

Figure 4.1. Percentage distribution of grades 8 and 4 students, by gender and SSI status: 1996 Group (22 SSI and 18 non-SSI states).

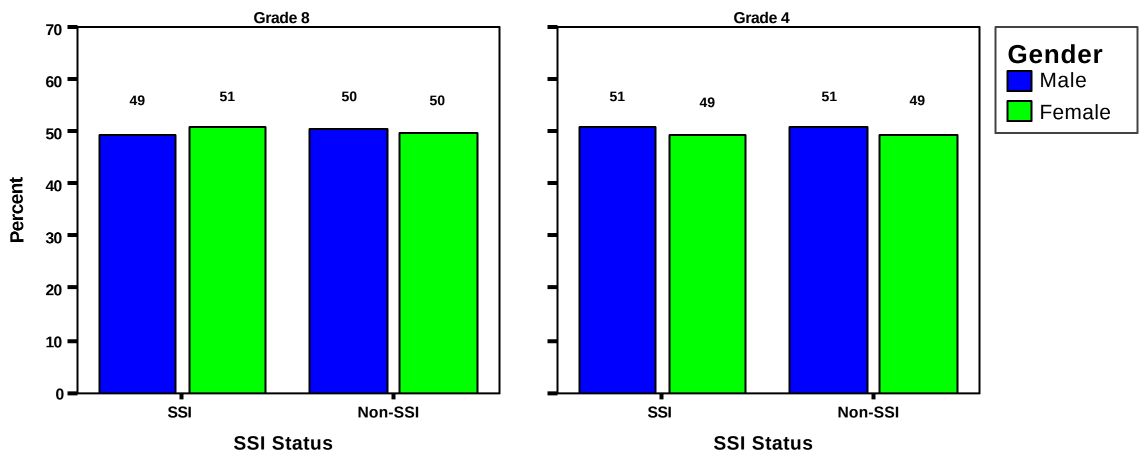
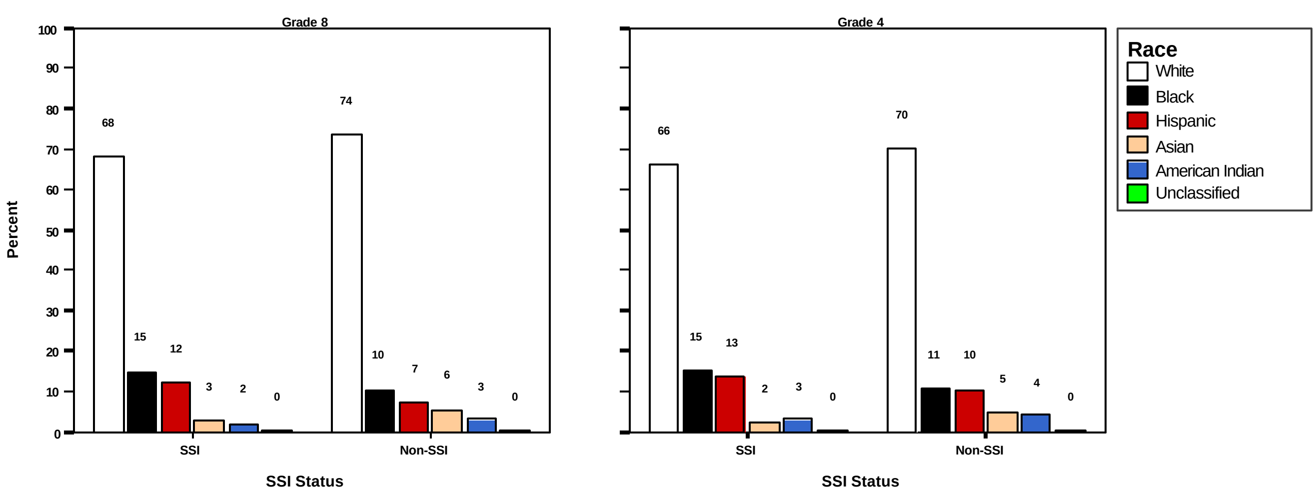
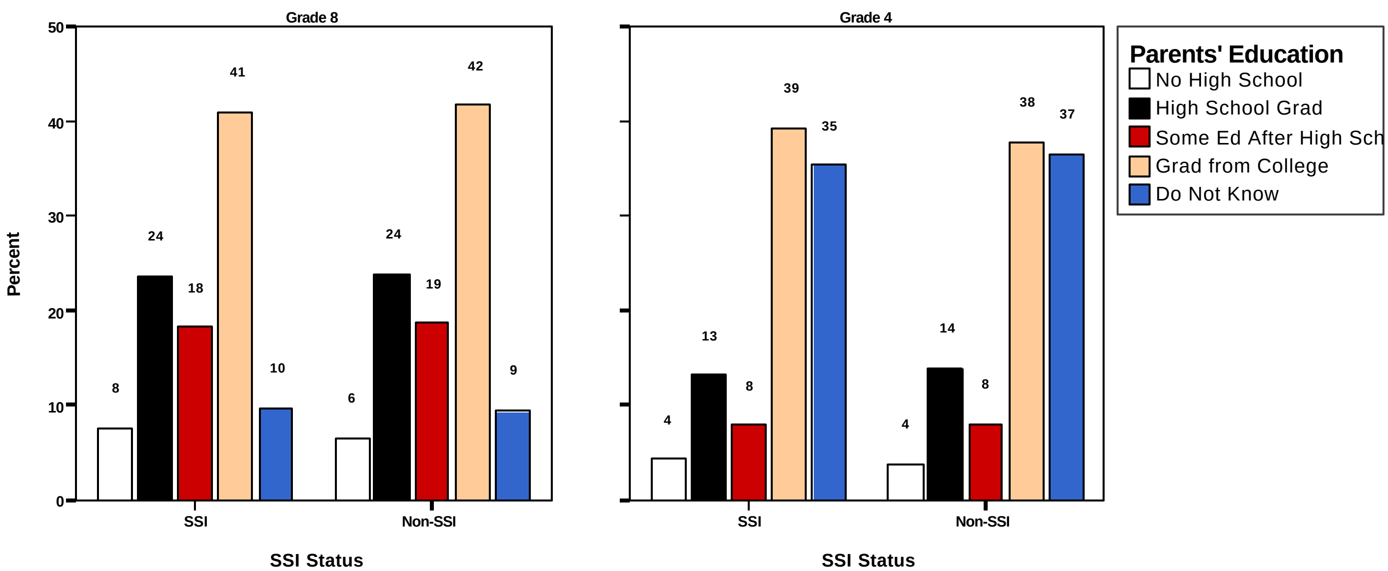


Figure 4.2. Percentage distribution of grades 8 and 4 students, by race and SSI status: 1996 Group (22 SSI and 18 non-SSI states).



The State NAEP did not collect information on the socioeconomic status (SES) of students, but did include questions that can be used to infer students' SES. Two variables are used to indicate the economic status of the students—parents' education and home environment. The NAEP student questionnaire asked students to report the education level of their fathers and their mothers—that is, the highest educational level attained by the father and the mother. Students' reports of parent's education levels were very similar in the SSI and non-SSI states in 1996 (Figure 4.3). SSI states and non-SSI states varied at most by 2% in any one category. If information for one parent was missing, the education level of the other parent was used. For both grade 4 and grade 8, roughly 40% of the students reported that at least one parent graduated from college. In fourth grade, about a third of the students responded that they did not know their parents' educational level.

Figure 4.3. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: 1996 Group (22 SSI and 18 non-SSI states).

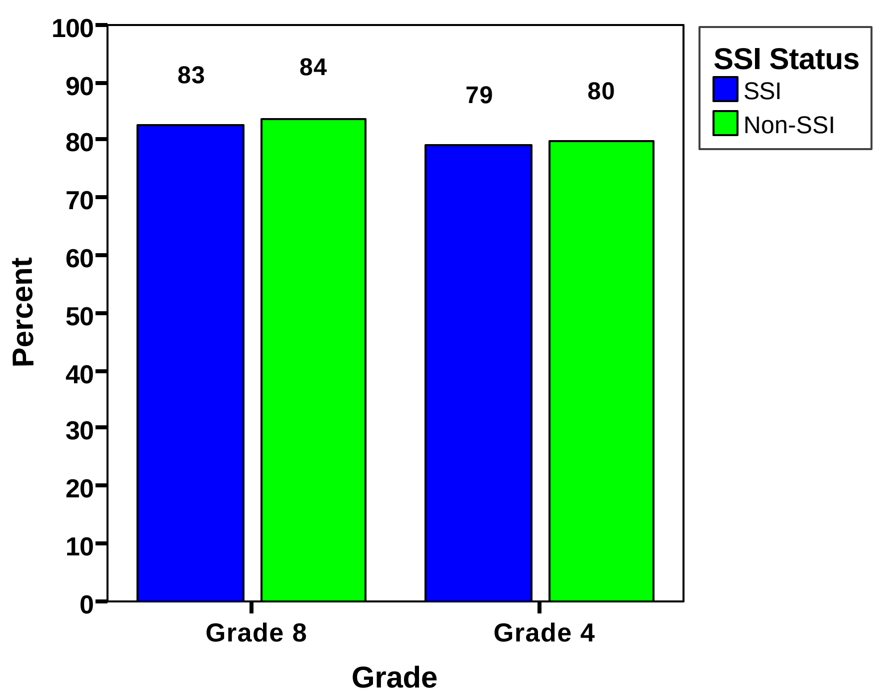


Students in the SSI states varied little from those in the non-SSI states in home environment (Figure 4.4). The home environment measure combines student responses to four family background items from the NAEP student questionnaire:

> Does your family get a newspaper regularly?
> Is there an encyclopedia in your home?
> Are there more than 25 books in your home?
> Does your family get any magazines regularly?
>
> *Response Options*
>
> Yes
> No
> I don't know

In both SSI and non-SSI states, generally less than 20% of the students responded "No" to any of these items. On average, just one percent more of the students in the non-SSI states had these advantages, compared to students in the SSI states.

Figure 4.4. Percentage distribution of grades 8 and 4 students on the basis of home environment, by SSI status: 1996 Group (22 SSI and 18 non-SSI states).



Trend Group 90-96 (17 SSI and 11 Non-SSI States)

The proportion of males and females in both SSI and non-SSI states remained nearly the same for the three testing times—1990, 1992, and 1996 (Figure 4.5). However, the constancy in the overall mean for each set of states obscures differences among them. Some states fluctuated over the three testing times in the percentages of male and female students (Figure 4.5). At grade 8, a higher percentage of SSI states had more females than males participate in the NAEP assessments at the three testing times than non-SSI states. The reasons for this are not apparent, but the pattern of the proportion of females relative to the proportion of males clearly indicates differences between the sets of states (Figure 4.5a). At grade 4, the proportion of females and males participating in NAEP is more comparable between the SSI and non-SSI states (Figure 4.5b).

Figure 4.5. Percentage distribution of grades 8 and 4 students, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

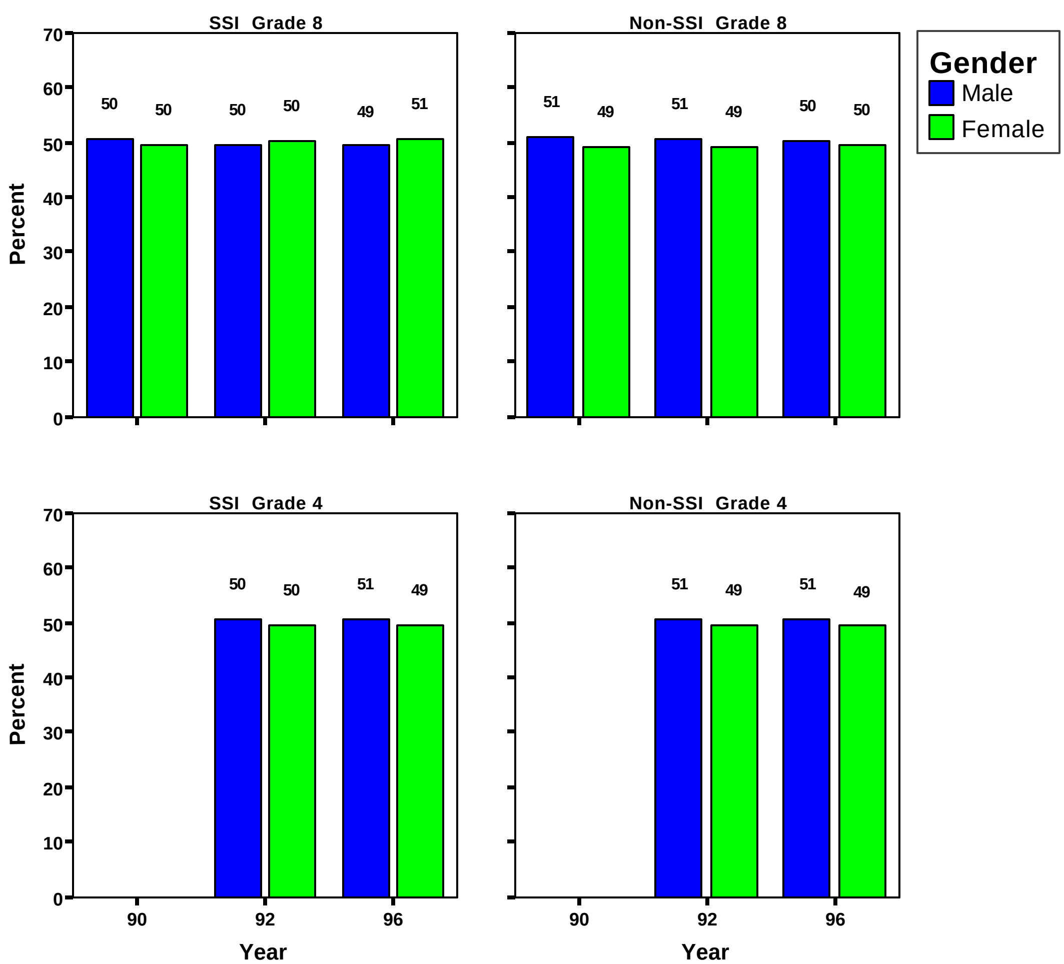
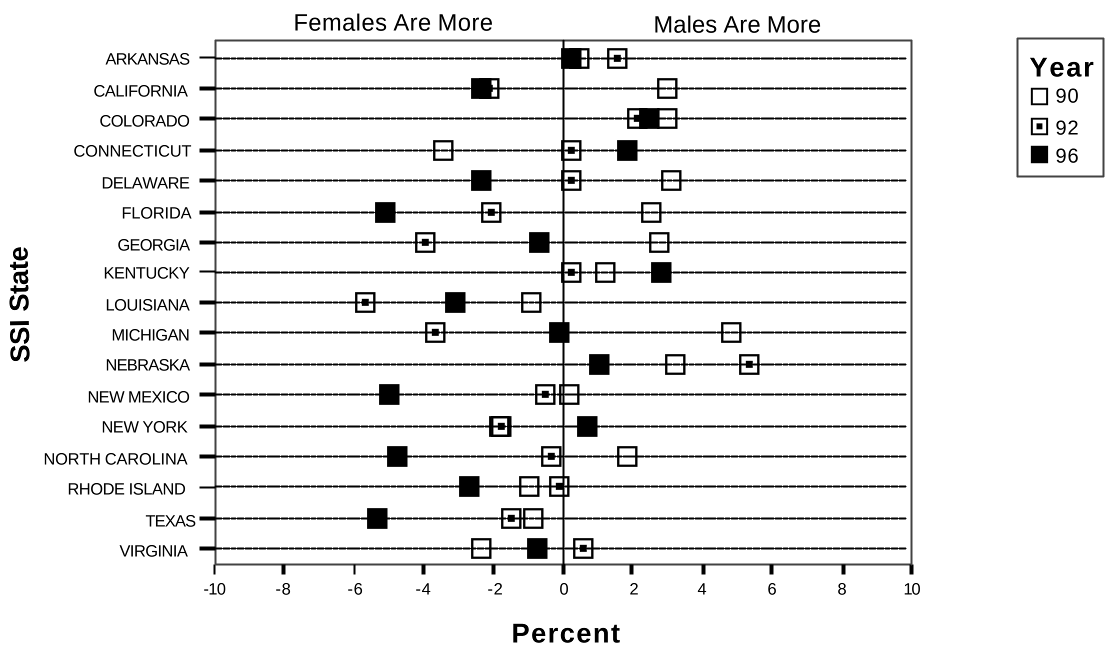


Figure 4.5a. Percentage distribution in gender of grade 8 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).



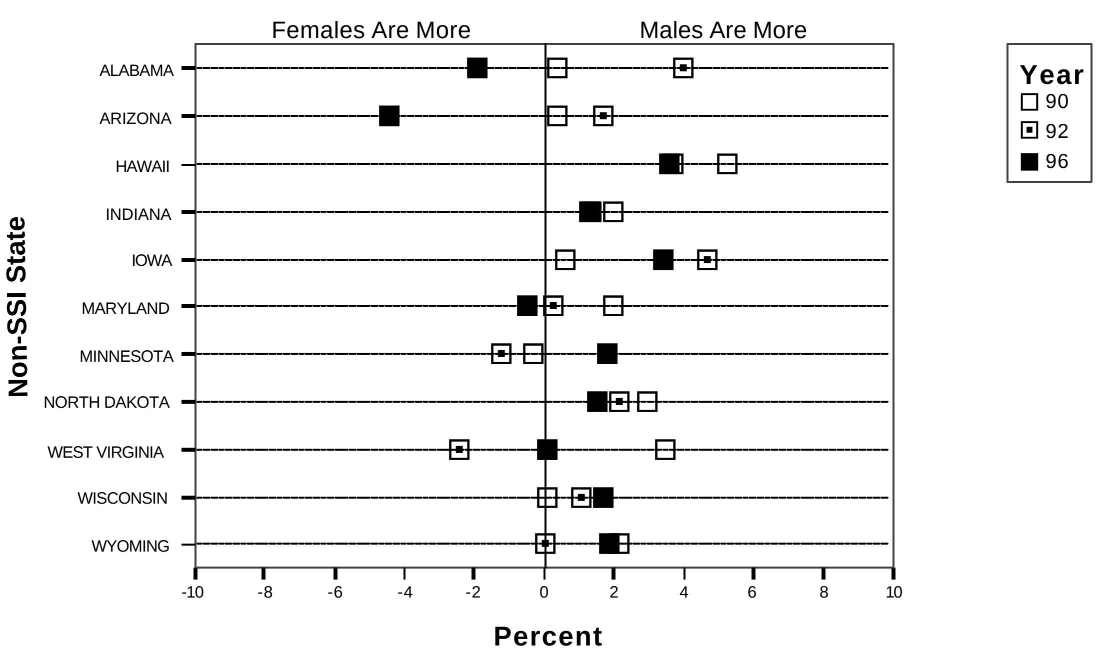
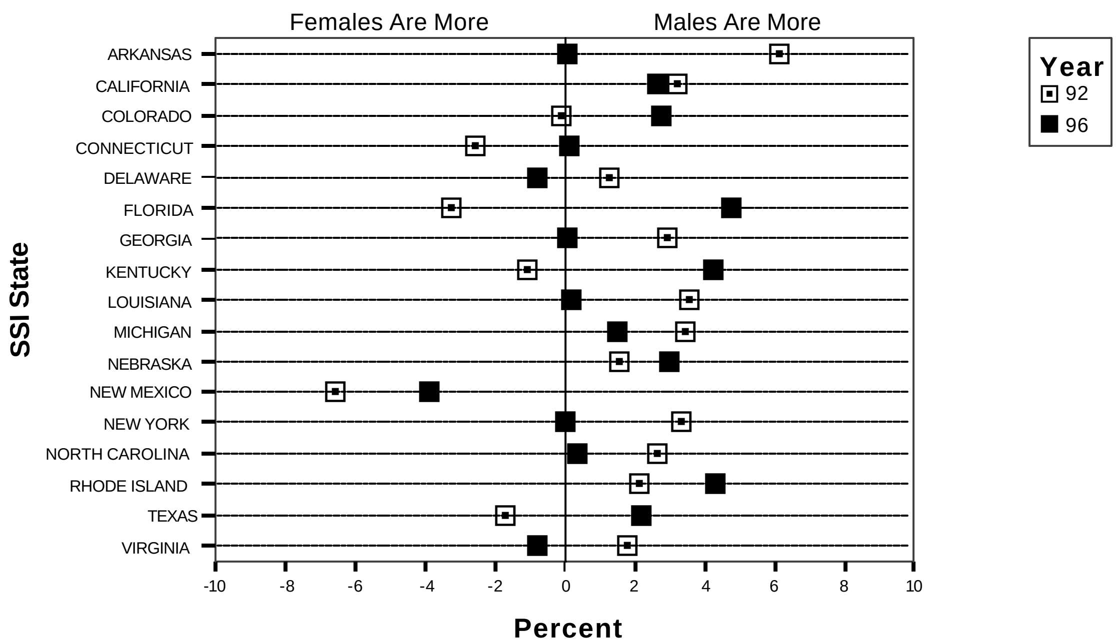
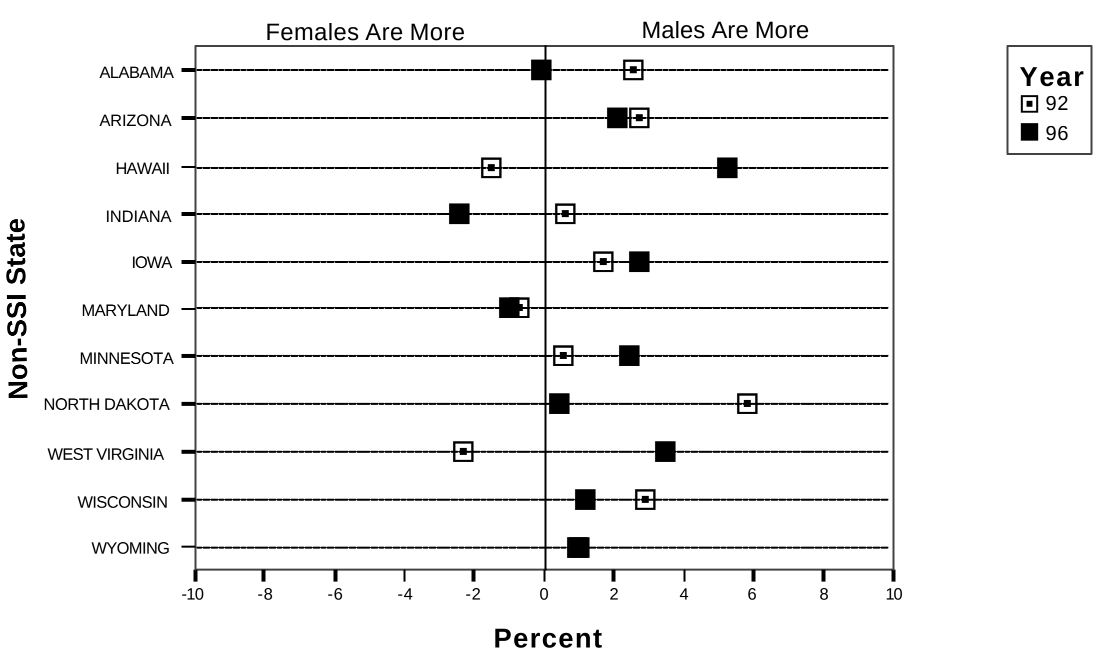


Figure 4.5b. Percentage distribution in gender of grade 4 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).





As for the total 1996 sample, the 17 SSI states in Trend Group 90-96 had fewer White and Asian students and a relatively greater number of Black and Hispanic students than the 11 non-SSI states in that group (Figure 4.6). In 1996, the 17 SSI states in Trend Group 90-96 did have a lower percentage of White students (4% less) and a higher percentage of Black and Hispanic (a total of 3% more) students than did the larger group of the 22 SSI states in the 1996 Group. This indicates that the smaller number of SSI states that constitute Trend Group 90-96 did not have the same racial composition as the SSI states that participated in NAEP. Reducing the sample size from 18 non-SSI states in the 1996 Group to 11 non-SSI states in Trend Group 90-96 resulted in less variation in racial distribution than for the SSI states. The percentage by racial group only varied by one or two percentage points.

Both SSI states and non-SSI states had a small decline in the percentage of White students from 1990 to 1996, 2% for SSI states and 1% for non-SSI states. The variation among states in the difference between White students and non-White students was similar for SSI states and non-SSI states (Figures 4.6a and 4.6b). In three SSI states, non-White students tested by NAEP outnumbered the White students—California, New Mexico, and Texas. Hawaii was the only non-SSI in which this was true. In seven of the 11 non-SSI states, over 80% of the students were White, while only two of the SSI states had over 80% White students.

Figure 4.6. Percentage distribution of grades 8 and 4 students, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

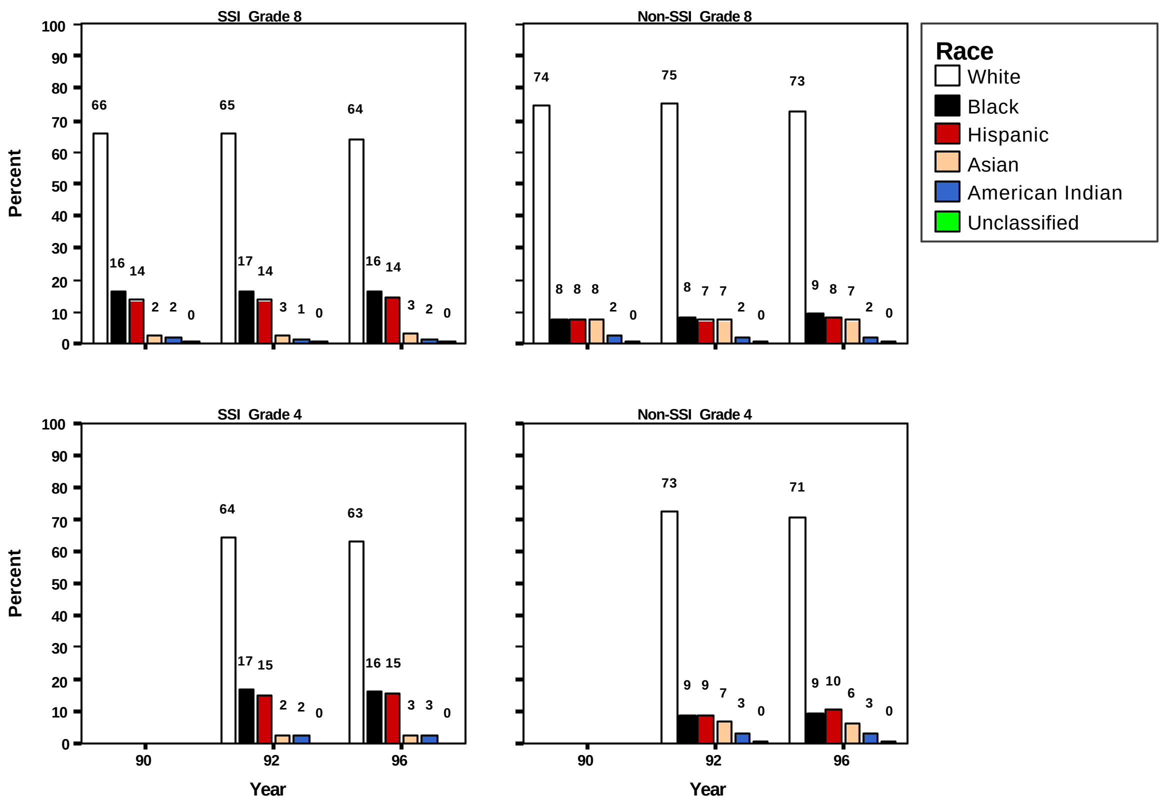
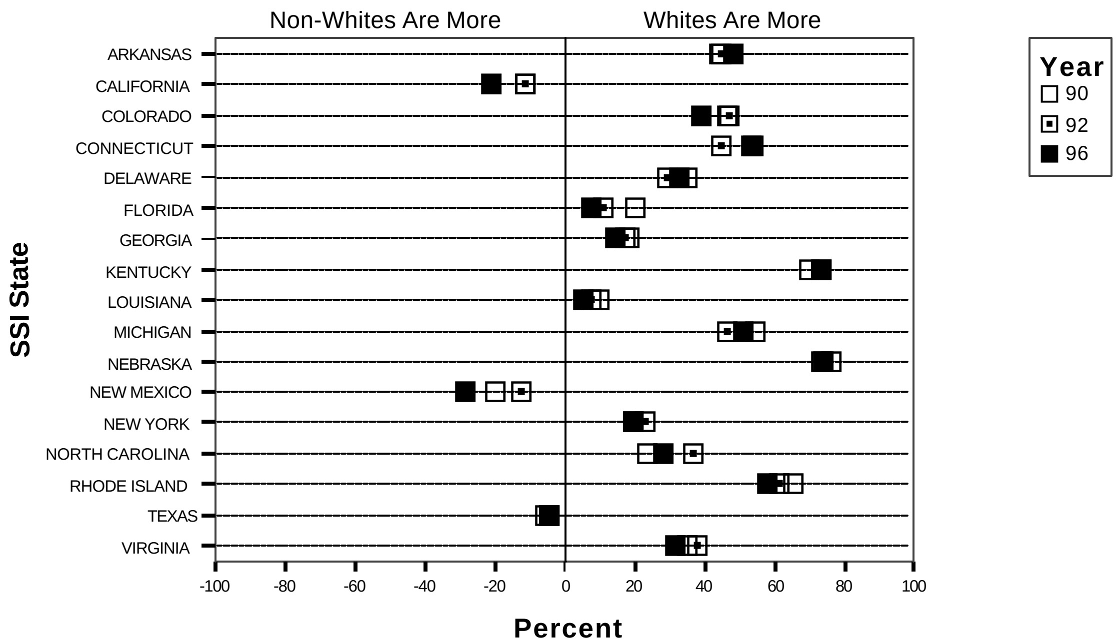


Figure 4.6a. Percentage distribution by race of Grade 8 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).



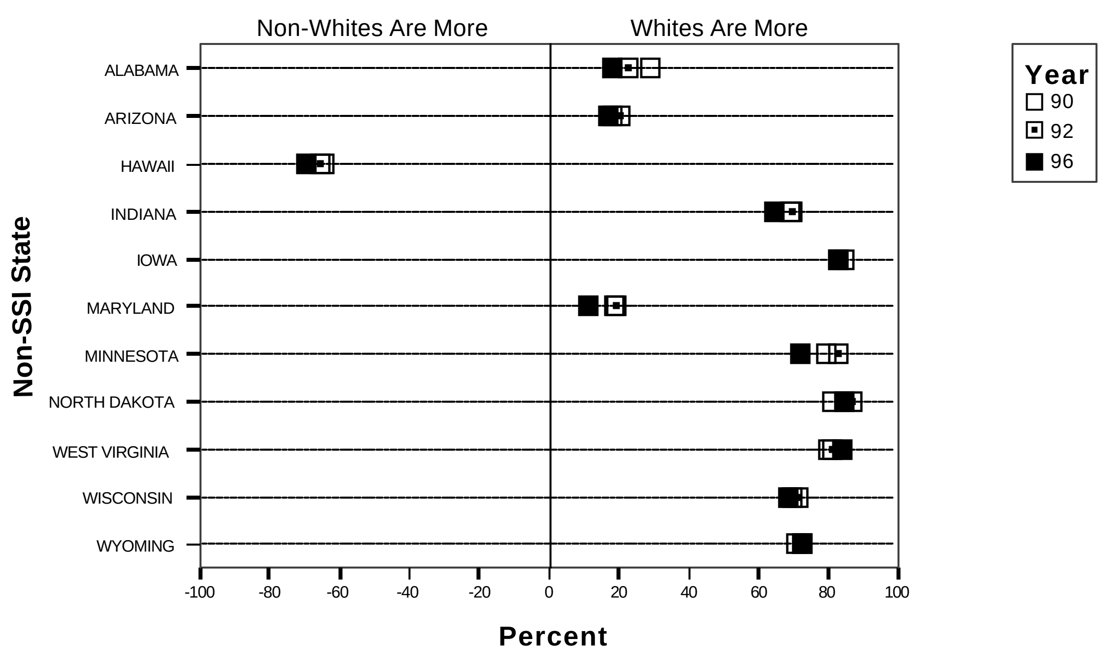
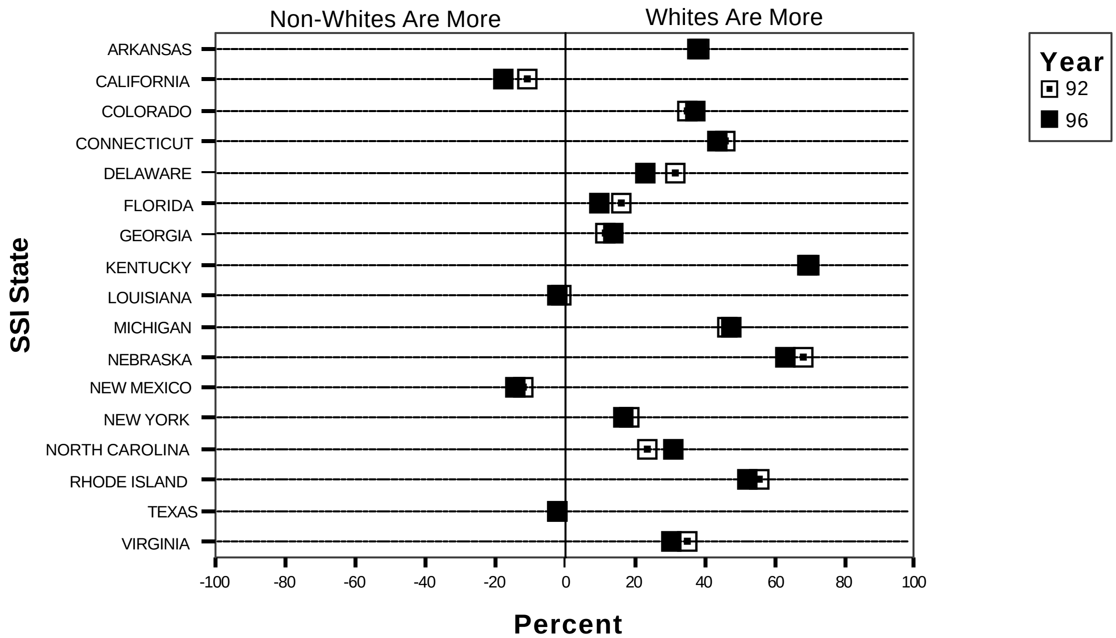
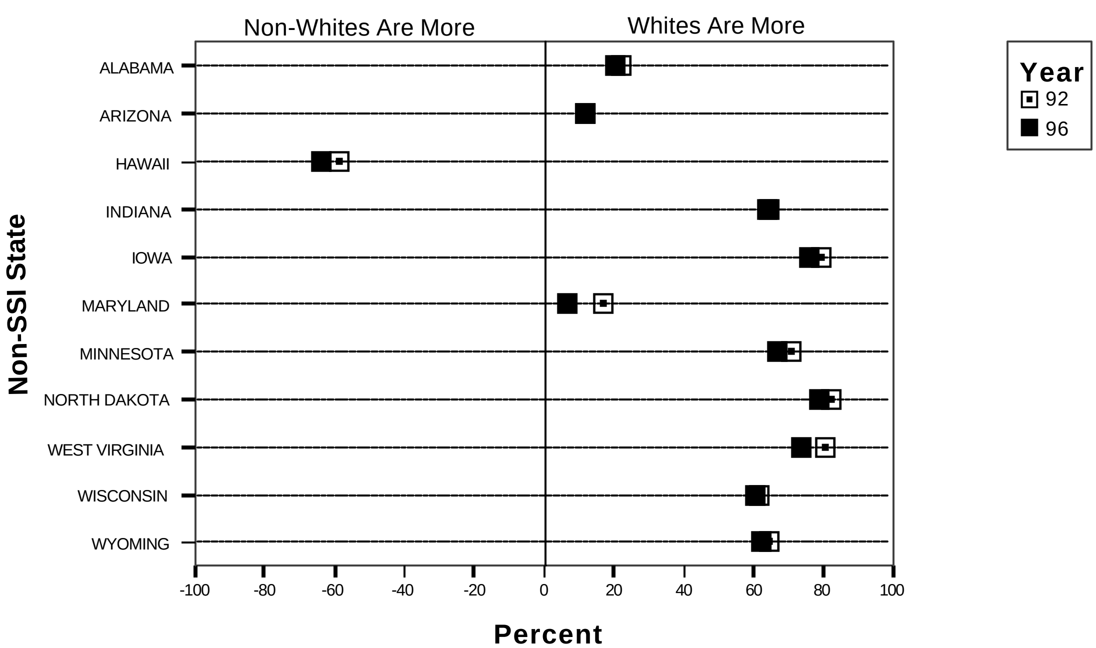


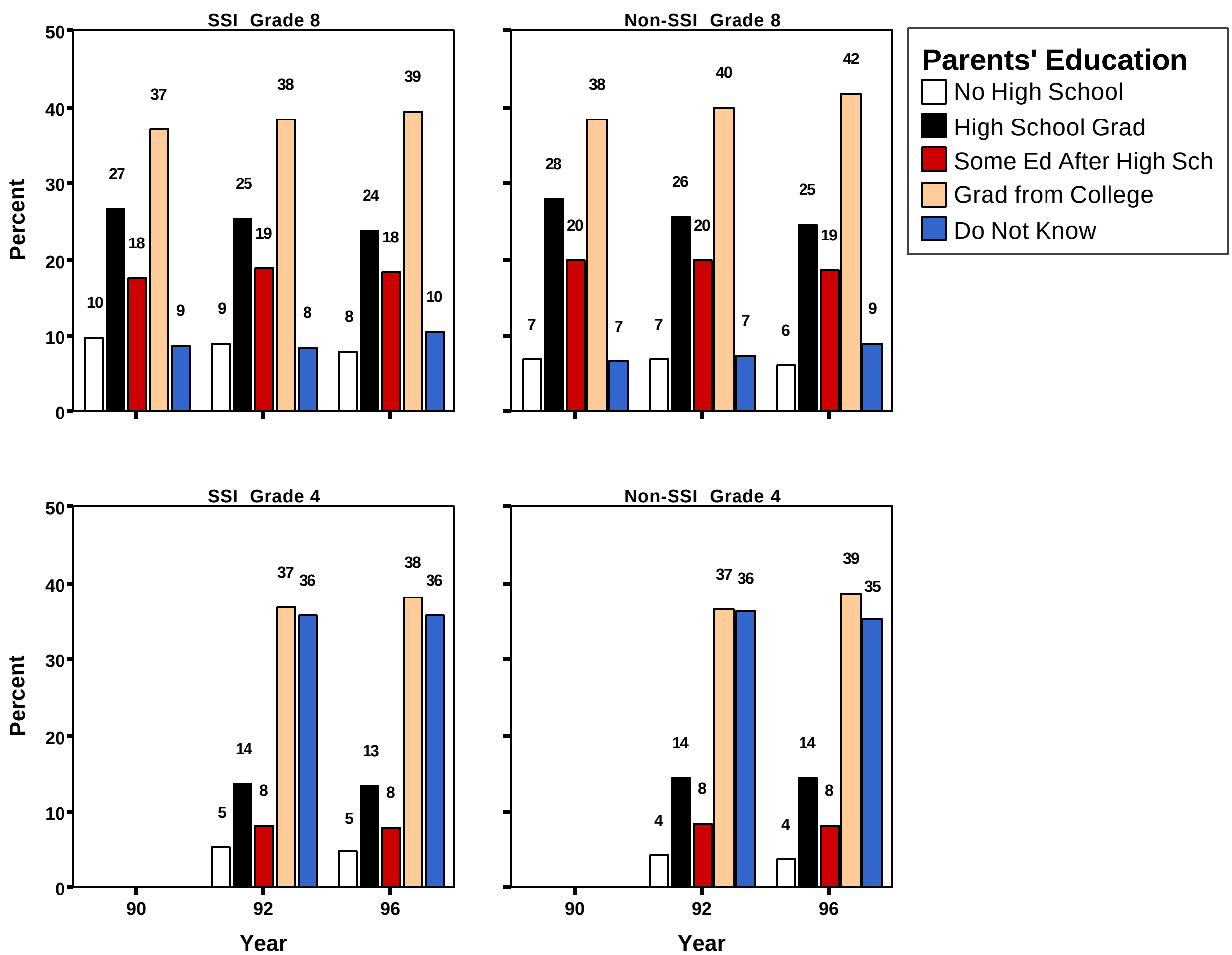
Figure 4.6b. Percentage distribution by race of grade 4 students, by SSI status and state: Trend Group 90-96 (17 SSI and 11 non-SSI states).





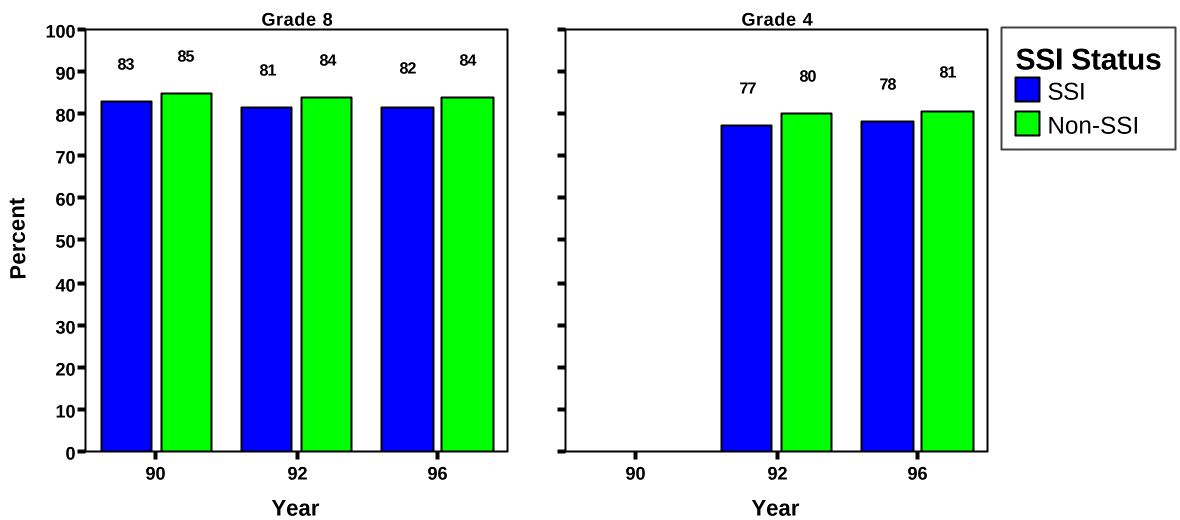
The education of parents of grade 8 students increased slightly from 1990 to 1996 (Figure 4.7). For each year, the percentage of parents who graduated from college was slightly higher in the non-SSI states compared to the SSI states, from 1% in 1990 to 3% in 1996. For grade 4, parents' education in SSI and non-SSI states was almost identical, with very little change from 1992 to 1996. Trend Group 90-96 varied by two percentage points or less from the 1996 Group on parents' education, indicating that very little reduction in parents' education is found as a result of using the smaller group of states to do the trend analysis.

Figure 4.7. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



In considering a home environment that included books, encyclopedias, magazines, and newspapers, a slightly higher percentage of students in non-SSI states had these advantages at both grade 8 and grade 4 levels and for all years of the State NAEP (Figure 4.8). The home environment reported by students in the 1990-96 Trend Group varied little, 1% or less, from that reported by the students in the 1996 Group. This, along with the lack of variation in parents' education, implies that the smaller Trend Group 90-96 is very similar to the larger group of states with respect to students' SES.

Figure 4.8. Percentage distribution of grades 8 and 4 students, by home environment criteria by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Trend Group 92-96 (20 SSI and 15 Non-SSI States)

As with other samples, the percentage of males and females in the total SSI and non-SSI groups is either 50/50 or 51/49 for the two-year trend sample (Figure 4.9). Five SSI states had a relatively large number (more than 5% difference) of females in 1996 at grade 8: Florida, New Mexico, North Carolina, South Carolina, and Texas. None of the non-SSI states had as high a percentage of female students assessed at grade 8. The two non-SSI states with the largest percentage of females tested in 1996 were Arizona with nearly 5% more females tested than males and Mississippi with nearly 4% more females tested than males (Figure 4.9a). In 1996, no state with more males exceeded a difference of 5%. At grade 4 in 1996, only one SSI state had a relatively large percentage of females—New Mexico (Figure 4.9b). Hawaii, a non-SSI state, had a relatively large percentage of males. As would be expected, Trend Group 92-96, with two fewer states in each category, varied very little from the 1996 Group. Since Trend Group 92-96 is nearly the same in the distribution by gender as is the larger 1996 Group, the analysis of the smaller group is not skewed by gender.

Figure 4.9. Percentage distribution of grades 8 and 4 students, by gender and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

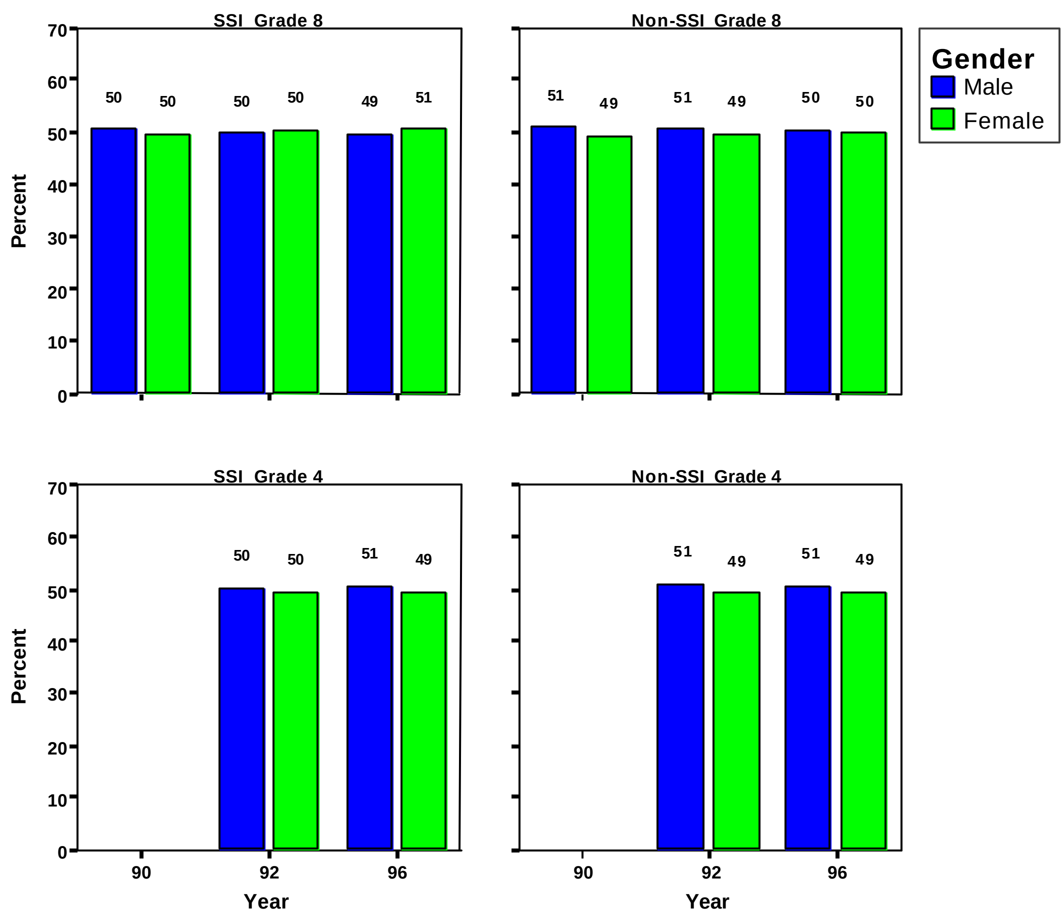
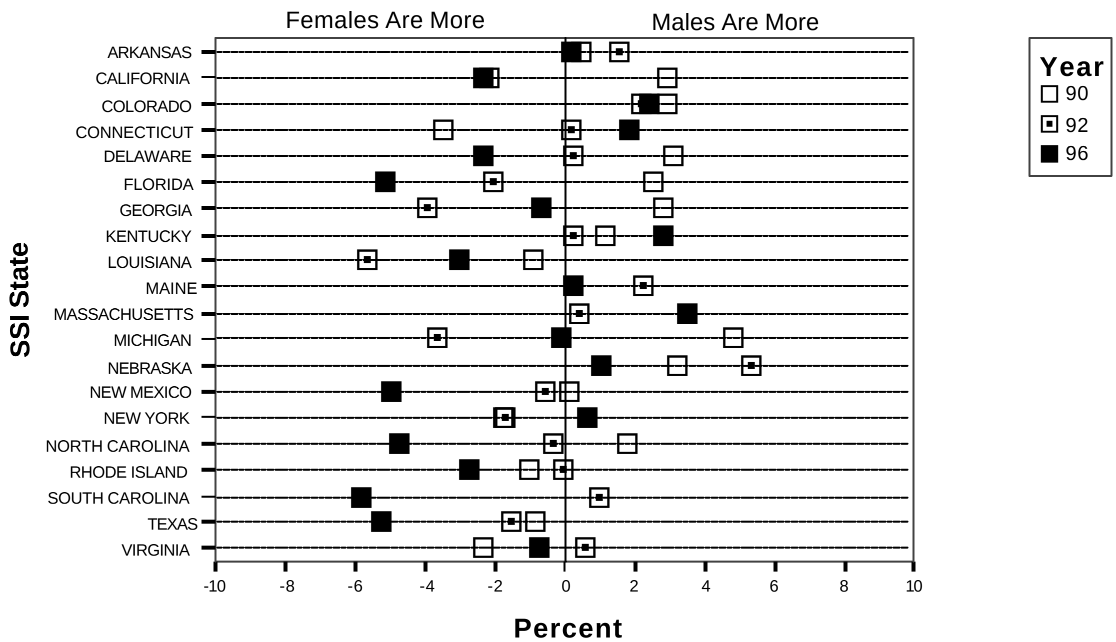


Figure 4.9a. Percentage distribution by gender of grade 8 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).



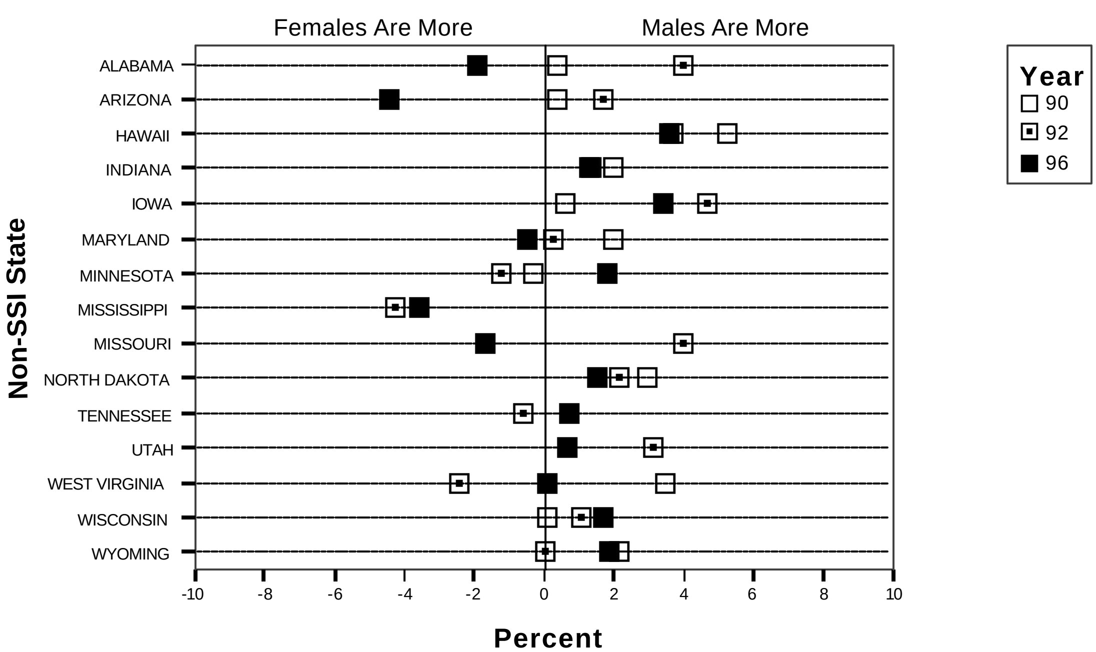
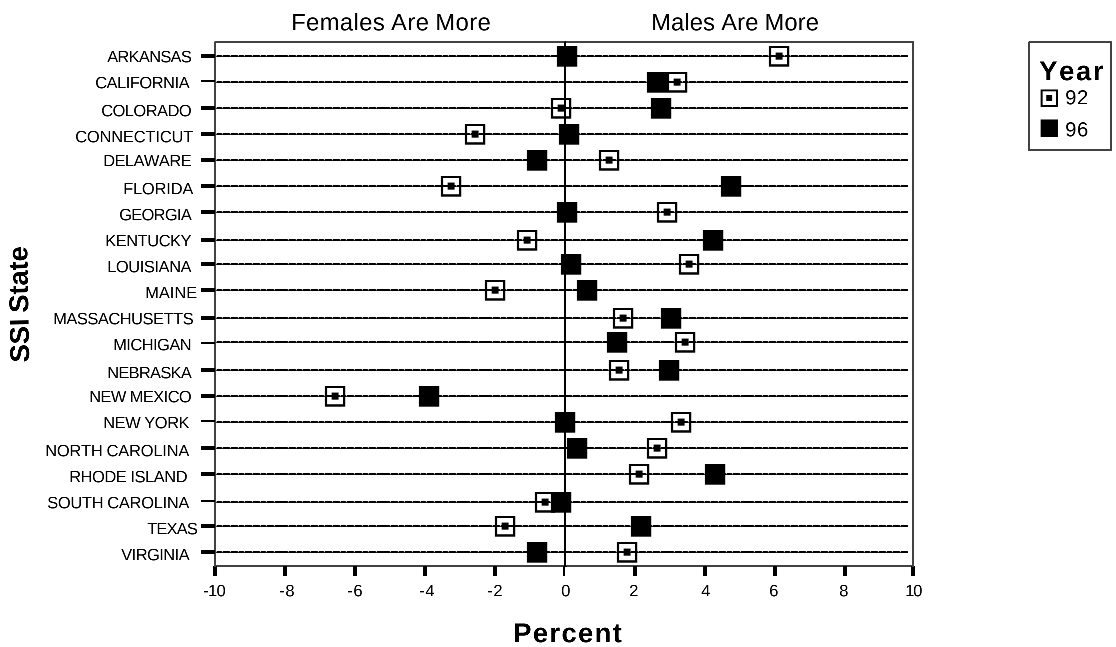
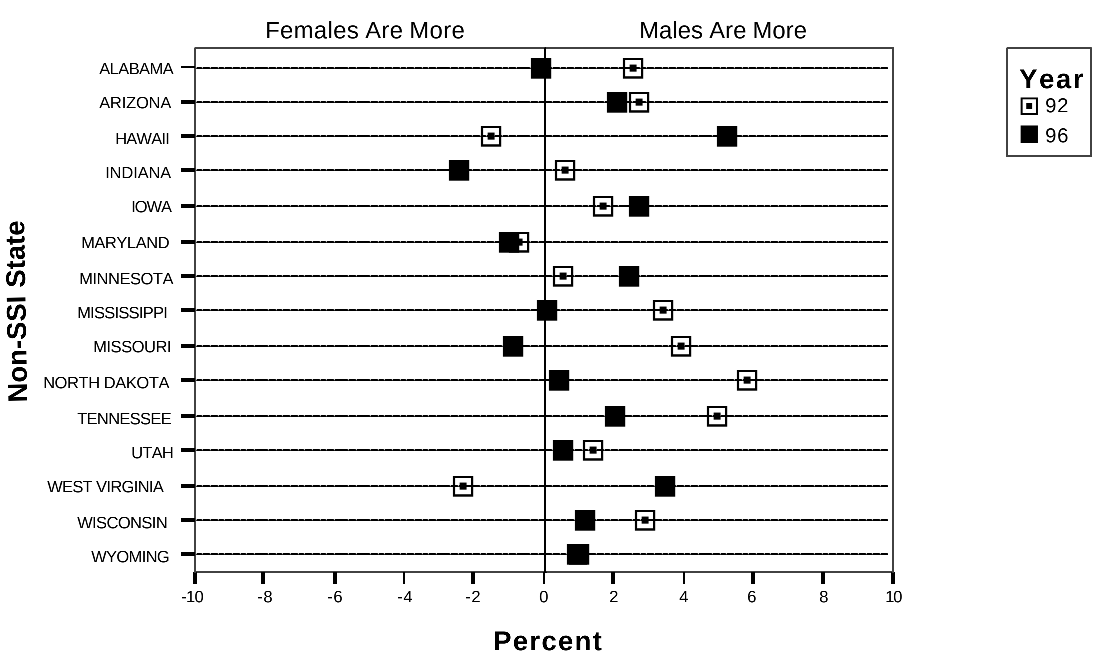


Figure 4.9b. Percentage distribution by gender of grade 4 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).





For Trend Group 92-96, the SSI states in both grade 8 and grade 4 had relatively fewer White and Asian students, and relatively more Black and Hispanic students than the non-SSI states as a group (Figure 4.10). In 1996, the 20 SSI states in grade 8 had 7% fewer White students than the 15 non-SSI states (Figure 4.10a) and in grade 4 the SSI states had 5% fewer White students (Figure 4.10b). The distribution by race of the tested population in Trend Group 92-96 varies very little from the distribution in the 1996 Group. No one category differs by more than 2% and most differ by 1% or are the same.

In nine of the 15 non-SSI states in 1996, at least 80% of the students were White. Only five of the 20 SSI states had such a large percentage of White students (Figure 4.10a). The grade 4 charts show a similar pattern, though with relatively fewer Whites in the SSI states (Figure 4.10b).

Figure 4.10. Percentage distribution of grades 8 and 4 students, by race and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).

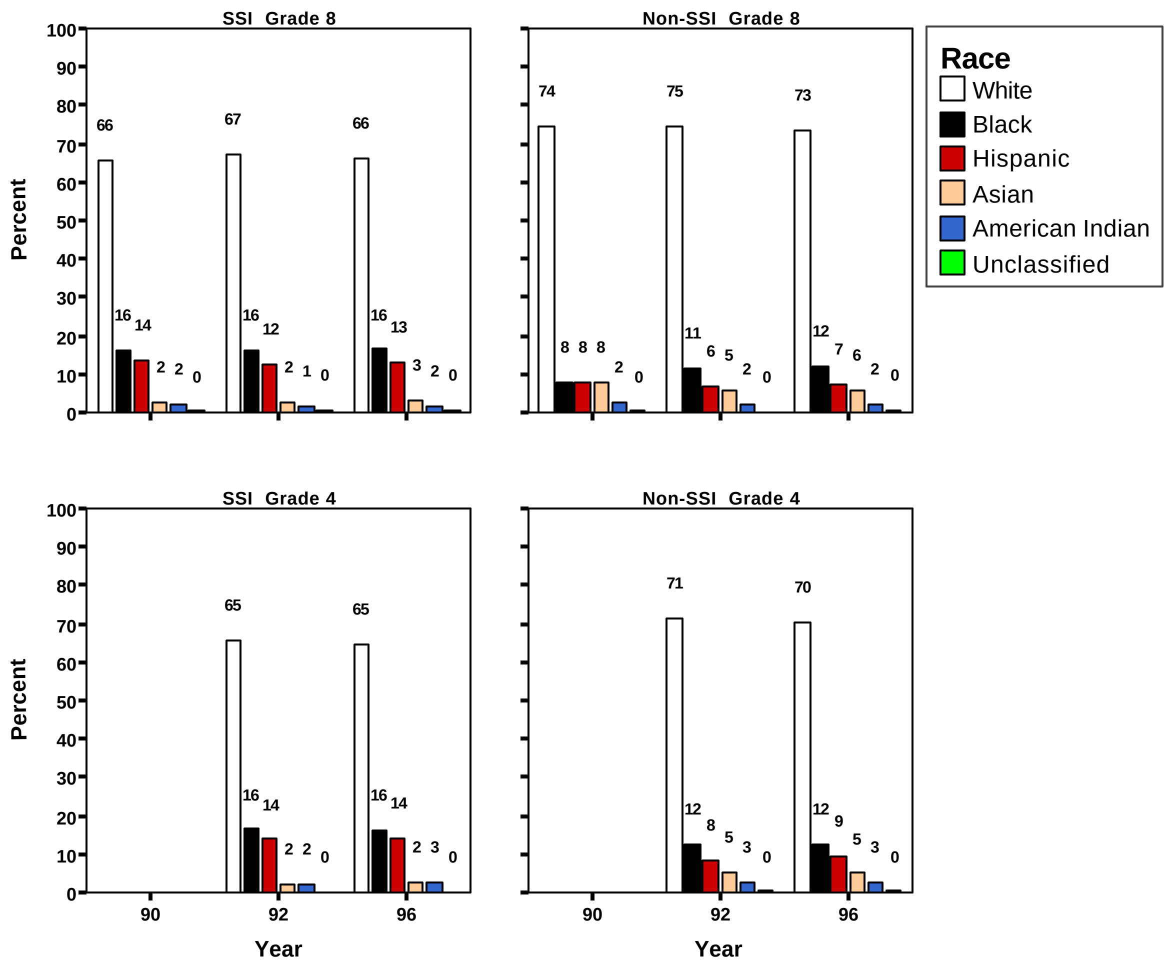
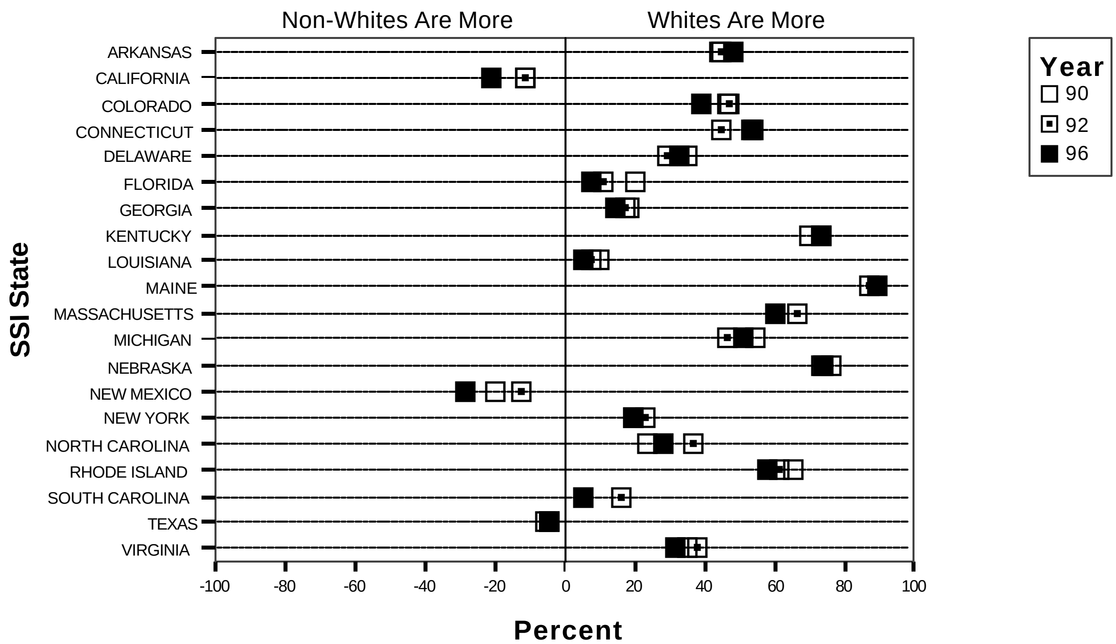


Figure 4.10a. Percentage distribution by race of grade 8 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).



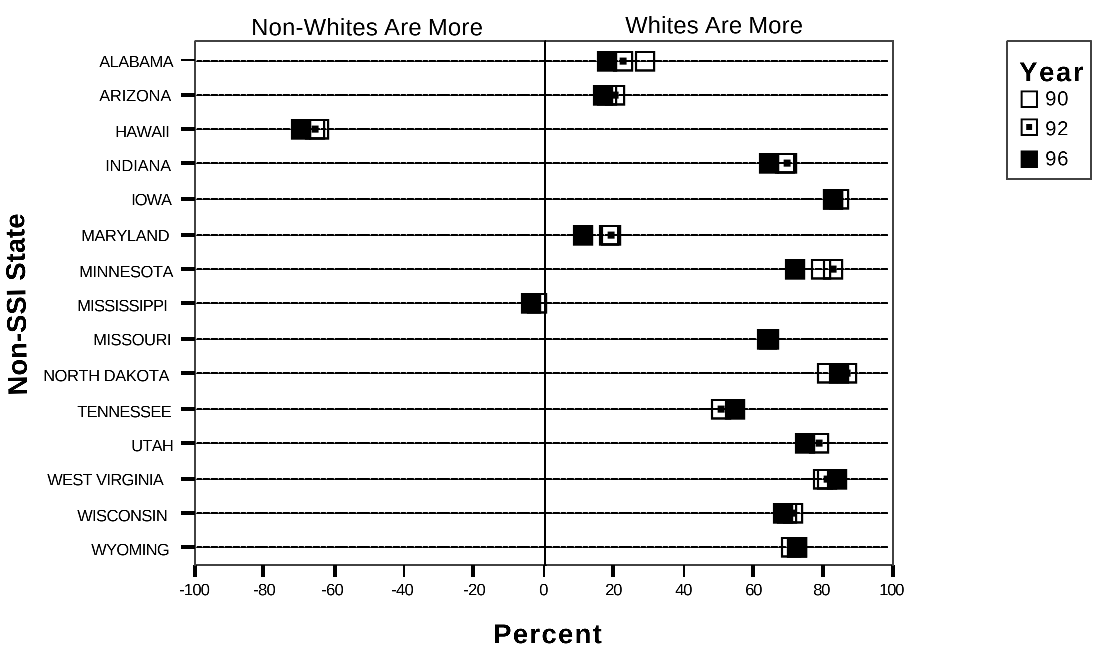
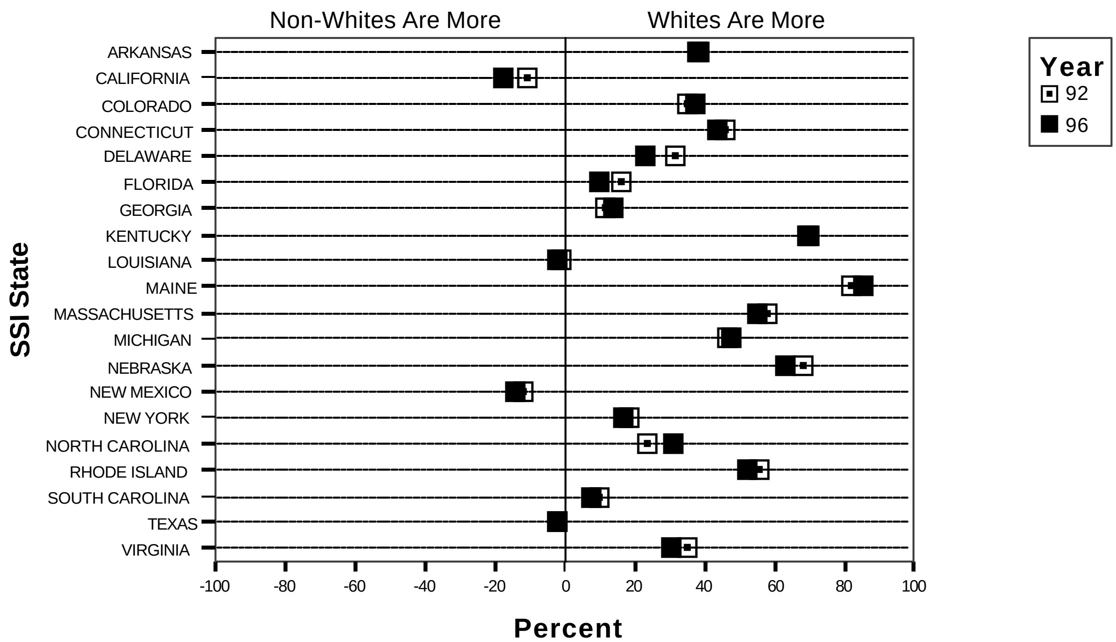
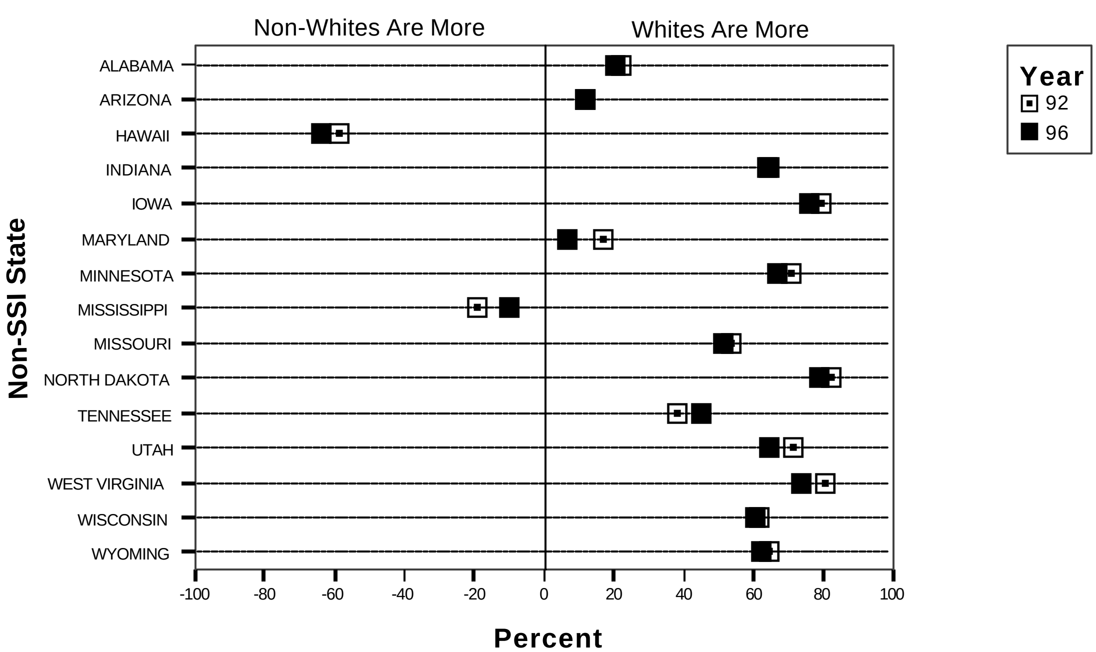


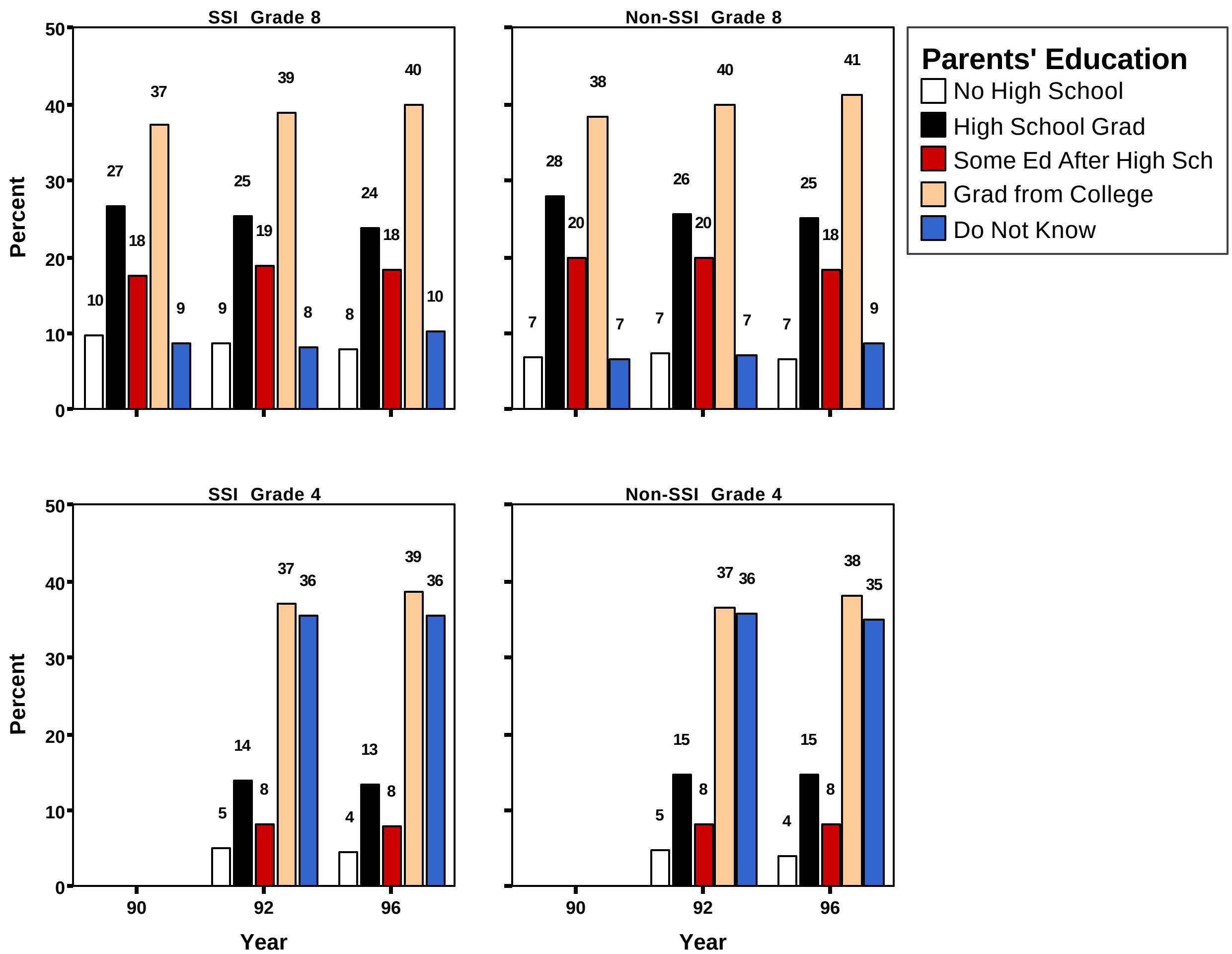
Figure 4.10b. Percentage distribution by race of grade 4 students, by SSI status and state: Trend Group 92-96 (20 SSI and 15 non-SSI states).





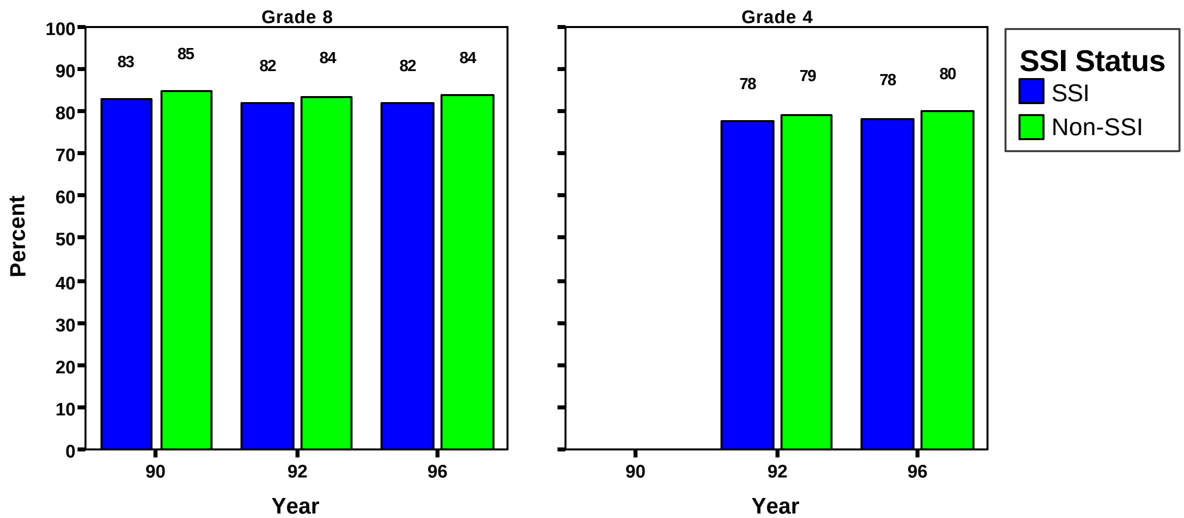
Parent education trends are comparable for SSI and non-SSI states, with the percentage of students with at least one parent who graduated from college just one percentage point higher in the non-SSI states at grade 8 (Figure 4.11) and one percentage lower in grade 4 in 1996. In both the SSI and the non-SSI groups at grade 8 in Trend Group 92-96 (Figure 4.11) compared to the 1996 Group (Figure 4.3), 1% fewer of the students reported a parent who had graduated from college. This difference is not large enough to indicate that Trend Group 92-96 is different from the 1996 Group. There were no differences between the two groups in grade 4. As in the other two groups, in grade 4 almost a third of the sample reported that they did not know either parent's education level.

Figure 4.11. Percentage distribution of grades 8 and 4 students, by parents' education and SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).



A slightly higher percentage of students, 2% in non-SSI states, reported a home environment that contained books, an encyclopedia, magazines, and newspapers, across all years and in both grade 4 and grade 8 (Figure 4.12). The SSI states in Trend Group 92-96 reported 1% fewer students with enriched home environments compared to the 1996 Group.

Figure 4.12. Percentage distribution of grades 8 and 4 students, by home environment criteria by SSI status: Trend Group 92-96 (20 SSI and 15 non-SSI states).



Summary of Demographics

Students tested by NAEP in the SSI states and the non-SSI states varied mainly in the proportion of White, Black, and Hispanic students in each group. SSI states had a higher proportion of Black and Hispanic students and a lower proportion of White students. The student population tested in the two groups of states was nearly evenly distributed by gender. However, when considering individual states, more states in the SSI group than the non-SSI group had a noticeably higher proportion (nearly six percent more) of females tested than males in grade 8. In 1996, in four of the 17 SSI states (23%) compared to one of 11 non-SSI states (9%), the population of females tested exceeded the proportion of males tested by 4%. In ten of the 17 SSI states, more females than males were tested compared to three of the 11 non-SSI states. The SSI states and non-SSI states did not vary on parents' education or home environment, two indices related to SES.

The two Trend Groups, 90-96 and 92-96, were both found to be, in general, comparable to the larger group of states that participated in the 1996 State NAEP. The change from the 1996 Group of 22 SSI states and 18 non-SSI states to Trend Group 90-96 of 17 SSI states and 11 non-SSI states did vary the racial distribution of the students tested. The smaller group had a lower proportion of White students in both the SSI and non-SSI states by 3% to 5%. Otherwise, the states participating in the two trend groups were very similar to the 1996 Group.

List of Appendices

Appendix A presents comparisons of the three analytic samples. The percentage of male and female students in each sample looks the same for the three analytic samples.

Figure 4A.1. Comparisons of three analytic samples by gender, by SSI status, and grade.

Figure 4A.2. Comparisons of three analytic samples by race, by SSI status, and grade.

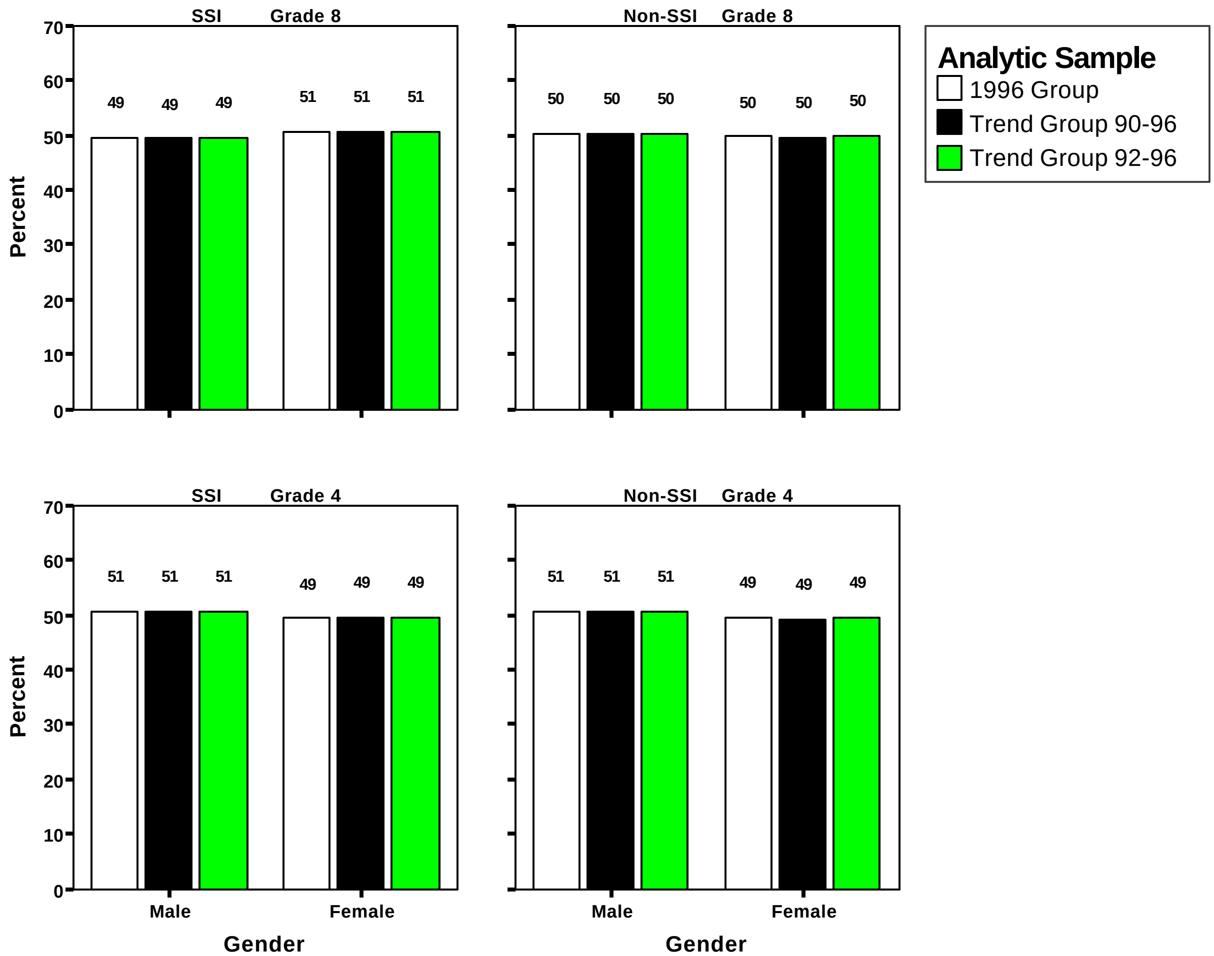
Figure 4A.3. Comparisons of three analytic samples by parents' education, by SSI status, and grade.

Figure 4A.4. Comparisons of three analytic samples on the basis of home environment criteria, by SSI status, and grade.

Appendix A

Appendix A presents comparisons of the three analytic samples. The percentage of male and female students in each sample looks the same for the three analytic samples.

Figure 4A.1. Comparisons of three analytic samples by gender, by SSI status, and grade.



In the SSI states, the 1996 sample has slightly more White students and slightly fewer Black and Hispanic students than the trend samples. In the non-SSI states, the percentages across samples seem fairly consistent.

Figure 4A.2. Comparisons of three analytic samples by race, by SSI status, and grade.

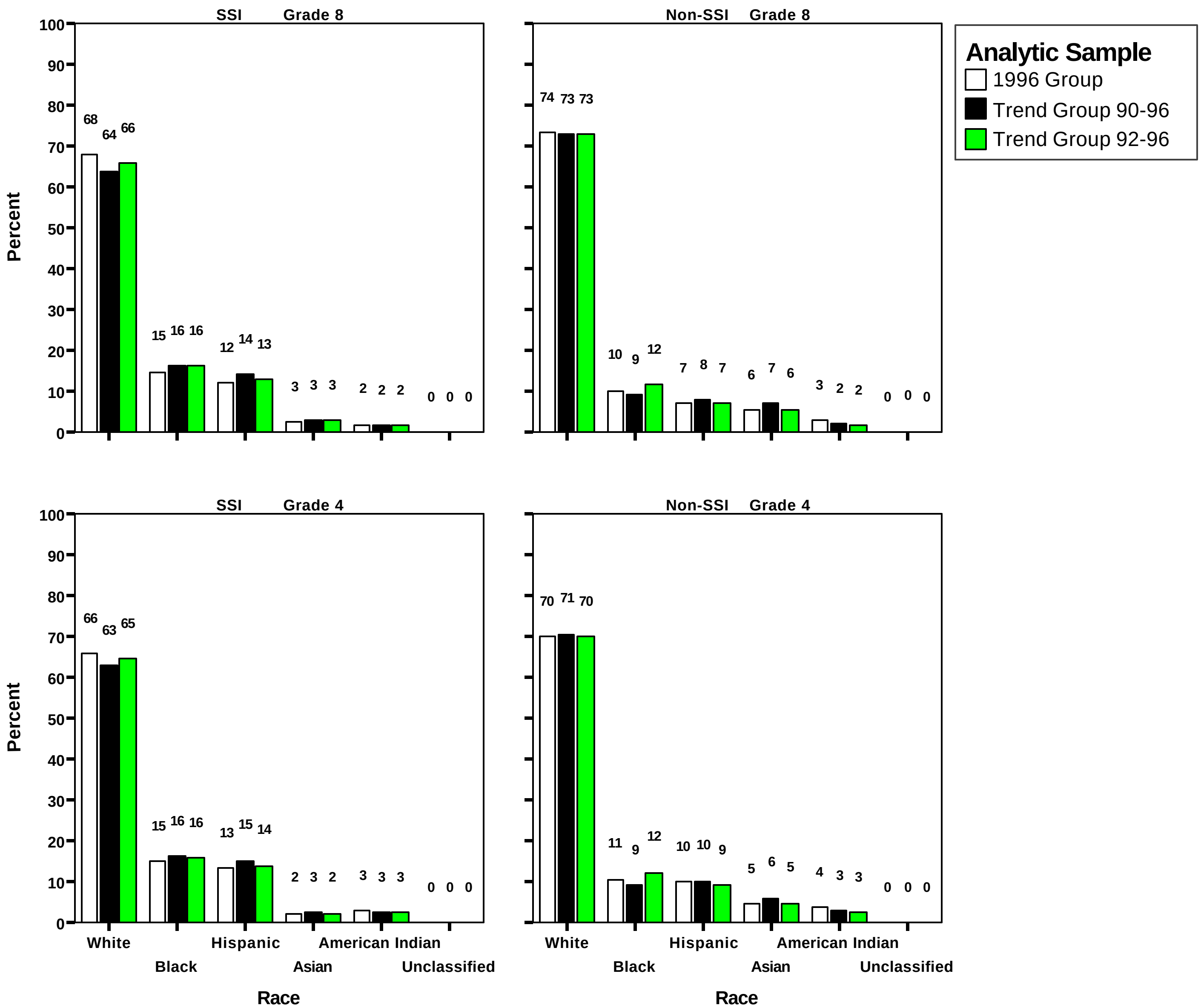
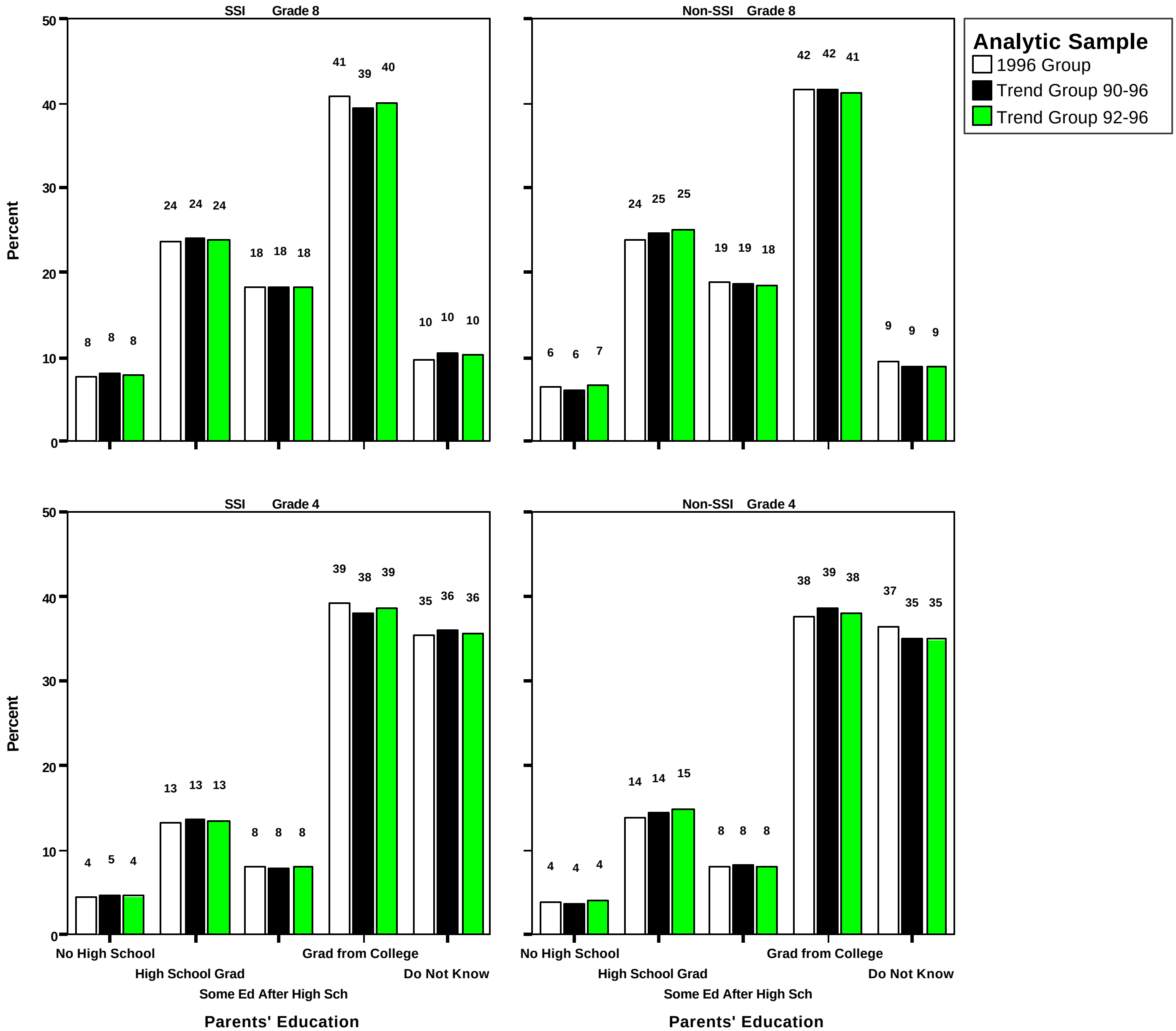
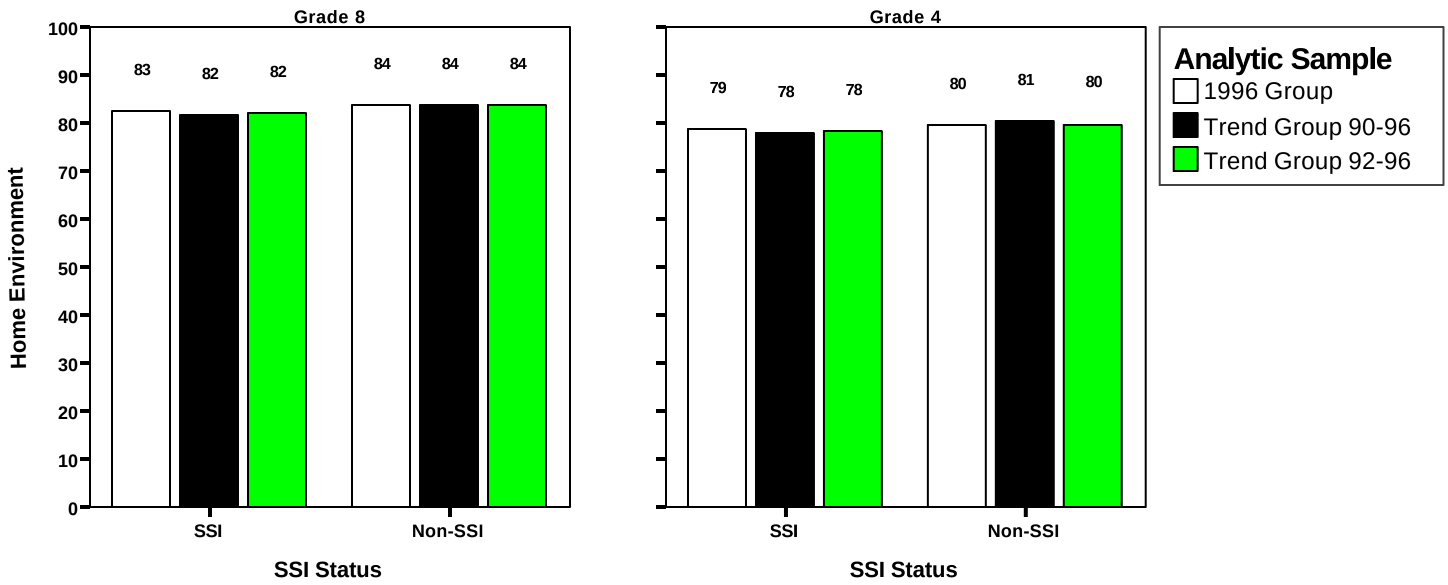


Figure 4A.3. Comparisons of three analytic samples by parents' education, by SSI status, and grade.



Parents' education and home environment seem consistent across the three samples.

Figure 4A.4. Comparisons of three analytic samples on the basis of home environment criteria, by SSI status, and grade.



CHAPTER 5

SSI AND NON-SSI ACHIEVEMENT USING STATE NAEP DATA: DESCRIPTIVE ANALYSIS

Introduction

Use of the State NAEP data allows us to track the change in academic performance in each state that voluntarily participated in the assessment. At present, State NAEP results are available for grade 8 students for three years—1990, 1992, and 1996—and for grade 4 students for two years—1992 and 1996. The achievement scales used in the State NAEP range from 0 to 500. The scales summarize the results for each of five mathematics content strands (i.e., number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions) and one overall composite score. Using IRT procedures, the scale scores from each of the State NAEP assessments are linked to each other to make them comparable across assessment years. Thus, these scores and procedures enable us to monitor the trends of student performance in each state over the years of 1990, 1992, and 1996 (Allen et al., 1997). In this chapter, we focus on identifying differences in mathematics scale scores between SSI and non-SSI states for grades 4 and 8. The results of the descriptive trend analysis are based on 28 states with data available over the three assessment years. Of the 28 states, 17 are SSI states and 11 are non-SSI states. Additional comparisons are given in Appendices A and B for this chapter, which present results for all participating states

This chapter consists of two main sections: one focuses on the results of the trends of grades 4 and 8 students, and the other on the cohort growth results from grade 4 (1992) to grade 8 (1996). Within each of these sections, results are presented for the total group, for males and females, and for Whites, Blacks, and Hispanics. In addition to the comparison of composites and of the five content strands, the gaps found between different gender and ethnic groups are also reported.

Trends in Average Scale Scores over 1990, 1992, and 1996

Composite Scores

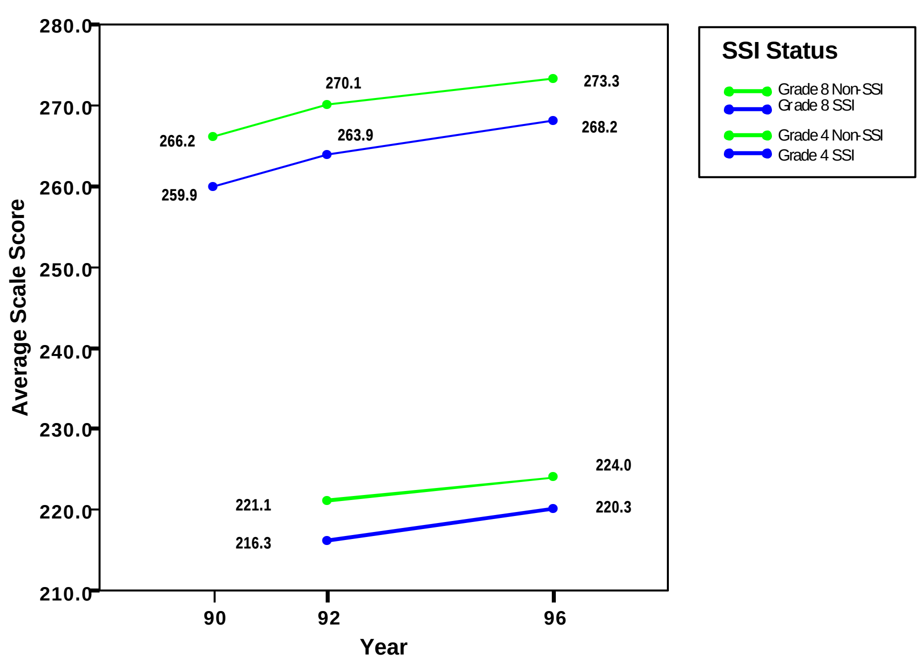
Total Group

In mathematics for both the SSI states and non-SSI states, students continually increased in average scale scores from 1990 to 1996 for grade 8 and from 1992 to 1996 for grade 4. Overall, the average performance across the 17 SSI states was lower than the average for the 11 non-SSI states across all three years. But, the initial gap between SSI states and non-SSI states narrowed slightly in 1996.

The average scale score for grade 8 mathematics from 1990 to 1996 showed an 8.3-point increase for the 17 SSI states and a 7.1-point increase for the 11 non-SSI states (Figure 5.1). The average increase by the SSI states was slightly higher than by non-SSI states, by 1.2 points. Prior to 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 1990, the 17 SSI states averaged 6 points less than the 11 non-SSI states. In 1992, the difference was still about 6 points, and in 1996 it was slightly less, around 5 points.

Similarly in grade 4, both the SSI states and non-SSI states made a slight gain in average scale scores from 1992 to 1996 (Figure 5.1). In 1992, the SSI average was 5 points lower than the non-SSI average. The performance gap was reduced by one point in 1996.

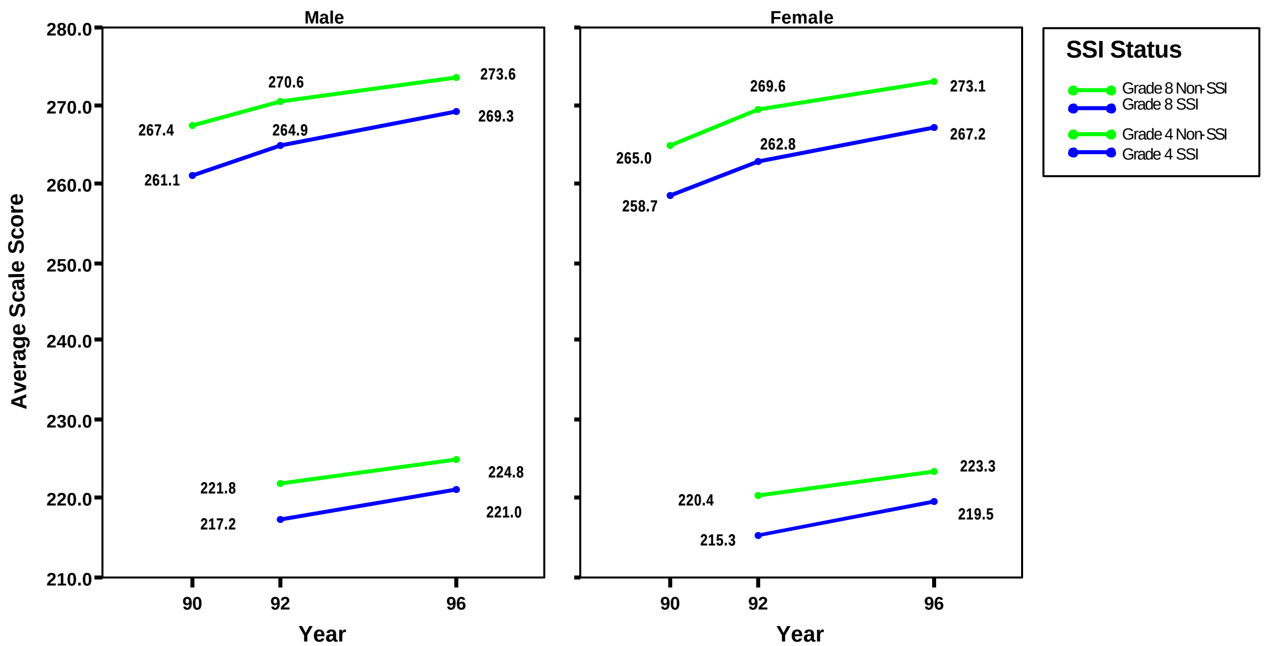
Figure 5.1. Trends in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Gender

The trend in average scale scores for male and female students shows similar patterns across grades and SSI status. Male and female students in both SSI and non-SSI states showed increases in average scale scores in all testing years, with male students scoring higher than female students. As shown in previous overall scale-score trends (Figure 5.2), for both male and female students the SSI states had lower average scale scores than non-SSI states. However, both male and female students in SSI states gained slightly more than those in non-SSI states. As a result, the gap between SSI states and non-SSI states dropped 1 to 2 points across grades. For example, in grade 8, male students in non-SSI states gained 6.2 points in average scale scores from 1990 to 1996, while over the same period male students in SSI states gained 8.2 points.

Figure 5.2. Trends in average scale scores, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Ethnicity

Overall, the average scale-scores of racial subgroups (Whites, Blacks, and Hispanics) show increases in mathematics performance across the assessment years. But substantial variety in mathematics performance among racial subgroups was evident at grades 4 and 8. In both grades, White students outperformed Black and Hispanic students, and Hispanic students scored higher than Black students (Figure 5.3).

From 1990 to 1996, White students in the 17 SSI and the 11 non-SSI states gained in their NAEP mathematics composite scores, reflecting the gains for the state as a whole. Grade 8 White students from both the 17 SSI states and the 11 non-SSI states gained in mathematics achievement from 1990 to 1992 and from 1992 to 1996. Grade 4 White students from both the 17 SSI states and the 11 non-SSI states gained in mathematics achievement from 1992 to 1996. In the SSI states, White students scored lower than those in non-SSI states in each of the three years. Over time, the increase in scores by White students in the SSI states was slightly greater than the increase in non-SSI states, so the difference between SSI and non-SSI states was smaller at both grades 8 and 4 in 1996.

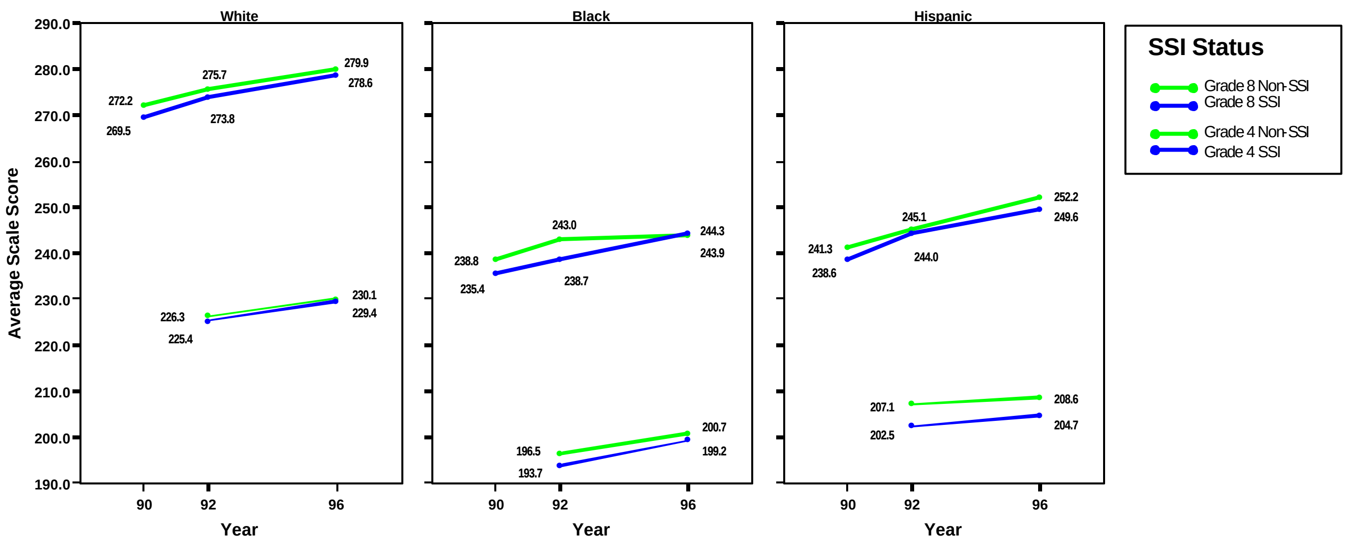
Average scale scores in both grades for Black students improved more for the SSI states than they did for the non-SSI states (16 SSI, 6 non-SSI)—the number of states with minority populations that were large enough to report data for all three years (i.e., a minimum sample size of 62 students per state was required to report the results for any subgroups) (Mullis et al., 1991).

In 1990, grade 8 Black students in SSI states had a mean mathematics score 3.4 points below Black students in the six non-SSI states. Six years later, Black students in the SSI states slightly outperformed those in the six non-SSI states by 0.4 points. From 1990 to 1992, the mean score for grade 8 Black students increased for both SSI states and non-SSI states. From 1992 to 1996, the mean score for grade 8 Black students increased considerably in the SSI states, compared to increases in non-SSI states, a 5.6 increase compared to a 0.9 increase.

In grade 4, the mean mathematics score of Black students in the SSI states increased by 5 points between 1992 and 1996, compared to the 4-point increase over this period by Black students in the six non-SSI states. In 1996, the mean score for grade 4 Black students in the SSI states was 1.5 points lower than the mean score for Black students in non-SSI states.

For Hispanic students, performance changes in mathematics were similar to those for White students except from 1992 to 1996 at grade 8. For grade 8 Hispanic students in SSI states, the gap between SSI states and non-SSI states was reduced by one point in 1992 from 3 points in 1990, but the gap in 1996 returned to the same level as it was in 1990 (Figure 5.3). The gap in average scale scores for grade 4 Hispanic students between SSI states and non-SSI states, however, remained stable at 4 points.

Figure 5.3. Trends in average scale scores, by race and SSI status: Trend Group 90-96 grade 8 and 92-96 grade 4 (17 SSI and 11 non-SSI states*).



* Due to the insufficient sample size of these subgroups, results are based on 16 SSI states and 6 non-SSI states for Blacks, and 15 SSI states and 10 non-SSI states for Hispanics.

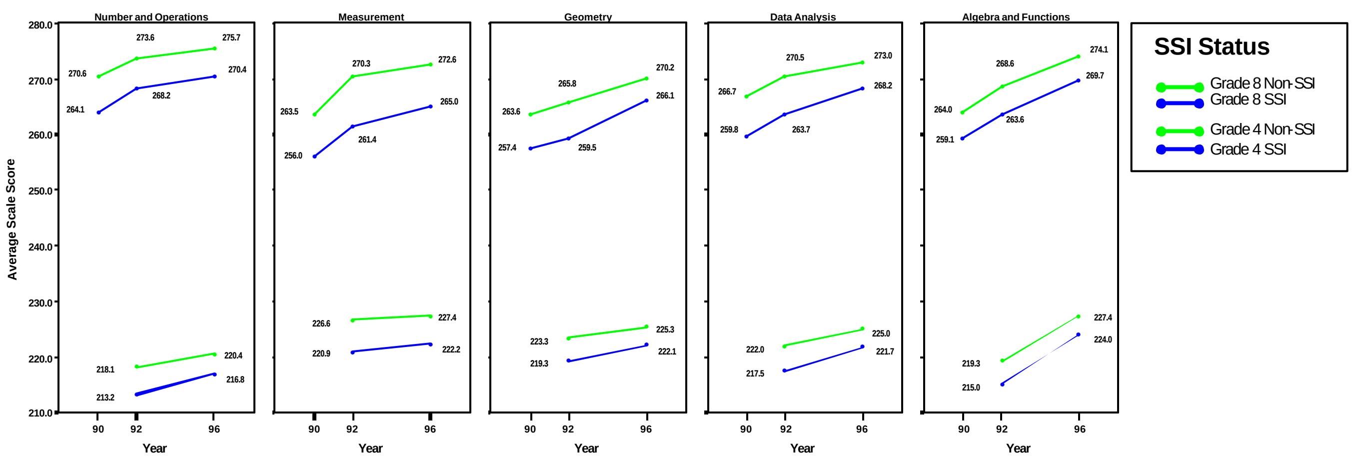
Subtopic Scores

Total Group

Very few differences existed 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 5.4).

On each of the five mathematics topics, SSI states had average scale scores below those of the non-SSI states for all testing times for both grade 4 and grade 8. However, both SSI states and non-SSI states had increased gain in mathematics performance on each of the five mathematical topics. In general, grade 8 students scored higher on number/operations followed by data analysis, algebra/functions, measurement, and geometry. In general, grade 4 students scored lower on number/operations than on the other four topics. The greatest gains at both grade 4 and grade 8 levels were in algebra/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 all five subscales.

Figure 5.4. Trends in average scale scores on content strands, by SSI status: Trend Group 90-96 grade 8 and 92-96 grade 4 (17 SSI and 11 non-SSI states).

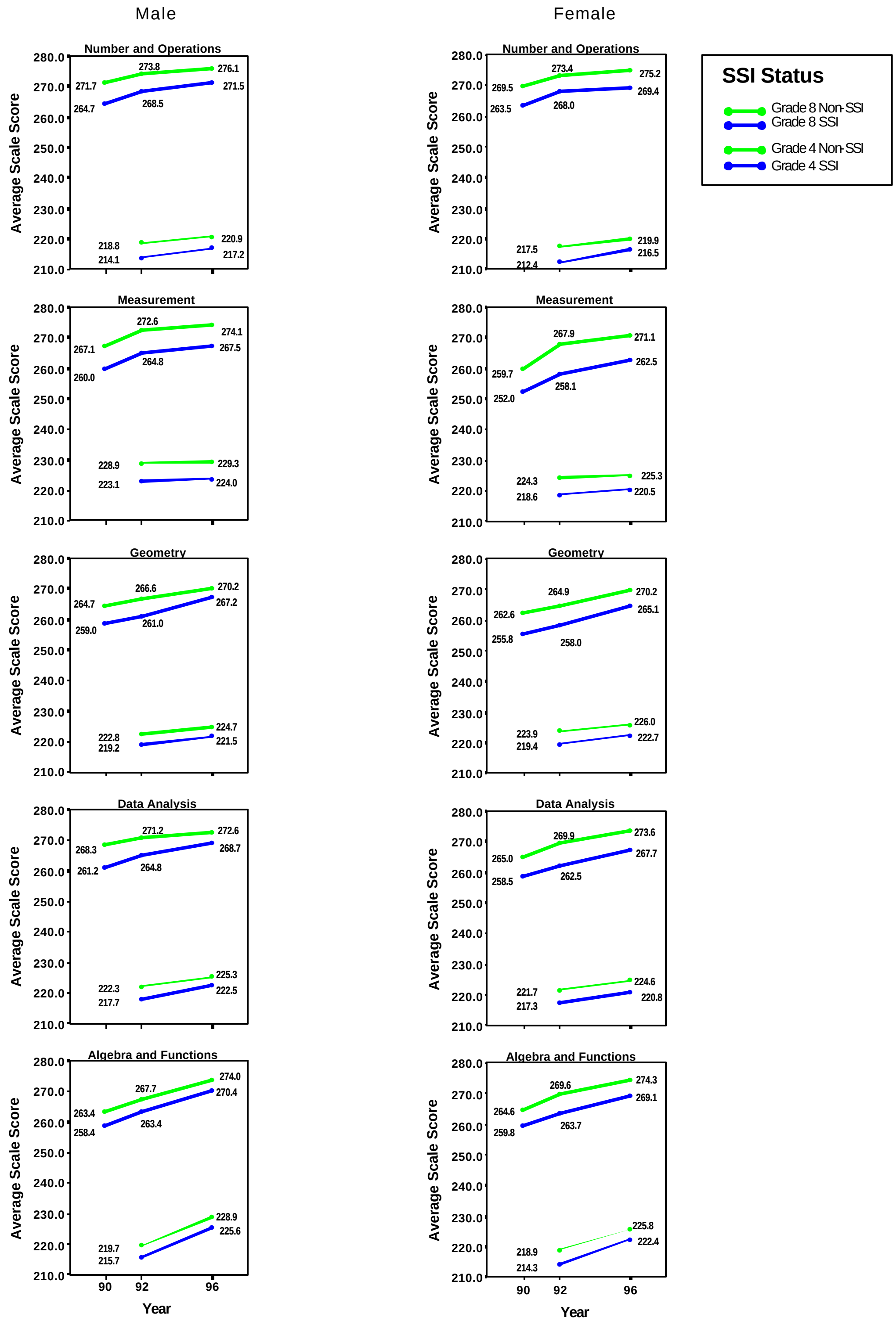


Gender

Gender trends for SSI states and non-SSI states were similar to those for all students (Figure 5.5). Substantial progress was made among male and female students in the five content strands over the three assessment years for each grade, while the differences between the two groups of SSI and non-SSI states remained. Male and female students in non-SSI states performed better than their counterparts in SSI states. There were gender differences across average content-strand scale scores, with males performing somewhat higher.

In grade 8, female students showed a larger increase in five content strands from 1990 to 1996 than male students. In 1996, grade 8 females in three of five content strands (measurement, geometry, and algebra and functions) gained at least 10 points above their 1990 score, regardless of SSI group status. The greatest gains on average scale scores in content strands were in algebra and functions at both grade levels. Male students in grade 8 gained 12 points in SSI states and 9.6 points in non-SSI states. The results for grade 4 were quite similar for both SSI states and non-SSI states (around 10 points). For female students in grades 4 and 8, SSI states and non-SSI states improved similarly, up to 10 points in the algebra and functions content strand. Despite the continuing gender gaps between SSI and non-SSI states in the five content strands, grade 8 male students in SSI states in 1996 reduced the gap by 3 points in geometry and by 4 points in data analysis from 5.7 and 7.1 points in 1990, respectively.

Figure 5.5. Trends in average scale scores on content strands, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Ethnicity

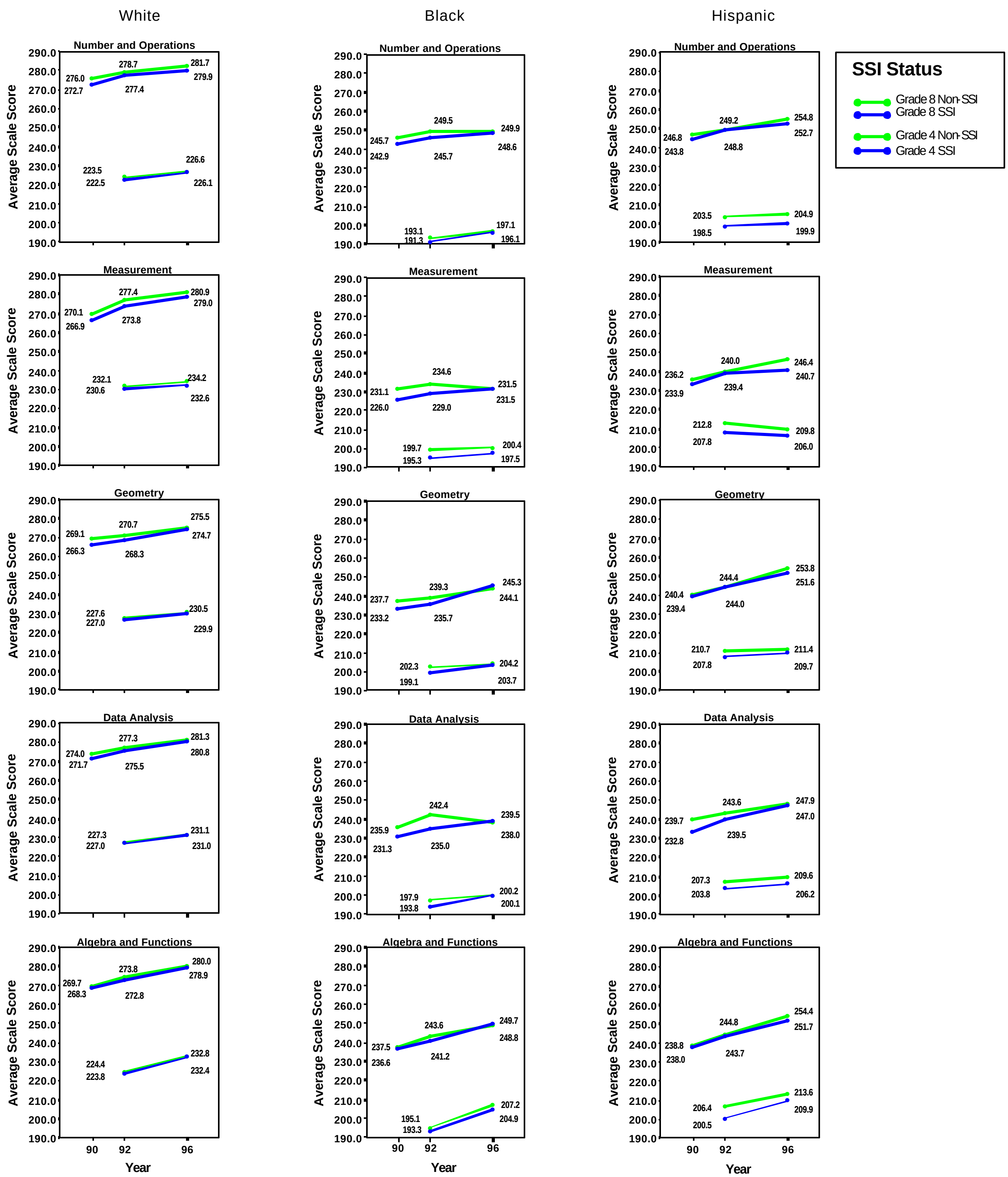
In the five mathematics content strands, the overall changes in performance among the three racial subgroups were small, but there were cumulative increases from the initial assessment in 1990 to 1996 in grade 4 and grade 8, except for a declining trend for Hispanic students in grade 4 in measurement. In 1996, average measurement scale scores for grade 4 Hispanic students in both SSI and non-SSI states decreased by 1.8 to 3 points. The score differences observed for racial subgroups in composite scale scores were also observed in the five content-strand scale scores. White students scored higher than Black and Hispanic students, while Hispanic students outperformed Black students across grades and assessment years (Figure 5.6).

There were varied patterns of average scale-score gains in five content strands for White, Black, and Hispanic students by SSI group status. Grade 8 White students in SSI states gained relatively more than those in non-SSI states from 1990 to 1996. The score differences between the two groups decreased across five content strands; for example, in geometry and data analysis, the gaps narrowed by less than one point in 1996, from 2 to 3 points in 1990.

For Black students in the SSI states, the gains are apparent over the assessment years. In all five content strands, score increases for grade 8 Black students were almost identical for both SSI and non-SSI states between 1990 and 1992. But, score increases in SSI states were substantially higher than in non-SSI states between 1992 and 1996. As a result, SSI states outperformed non-SSI states in 1996 in three of five content strands (e.g., geometry, data analysis, and algebra and functions). In measurement, scores in both SSI states and non-SSI states were the same in 1996; the gaps in the remaining two content strand scores narrowed by one point. In 1996, grade 4 gaps in average scale scores between SSI states and non-SSI states also decreased or remained stable.

The average scores for Hispanic students in SSI and non-SSI states generally showed gains in both grades, with some variations. In grade 8, SSI states reduced the gaps in average scale scores with non-SSI states by 0.4 points from 1990 and 1992, but the gaps widened in the later period over four of five content strand scores. Only scale scores in data analysis decreased in 1996. The score gaps for grade 4 students in both SSI and non-SSI states remained relatively stable.

Figure 5.6. Trends in average scale scores on content strands, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



* Due to the insufficient sample size of these subgroups, results are based on 16 SSI states and 6 non-SSI states for Blacks, and 15 SSI states and 10 non-SSI states for Hispanics.

Gaps Between Different Groups

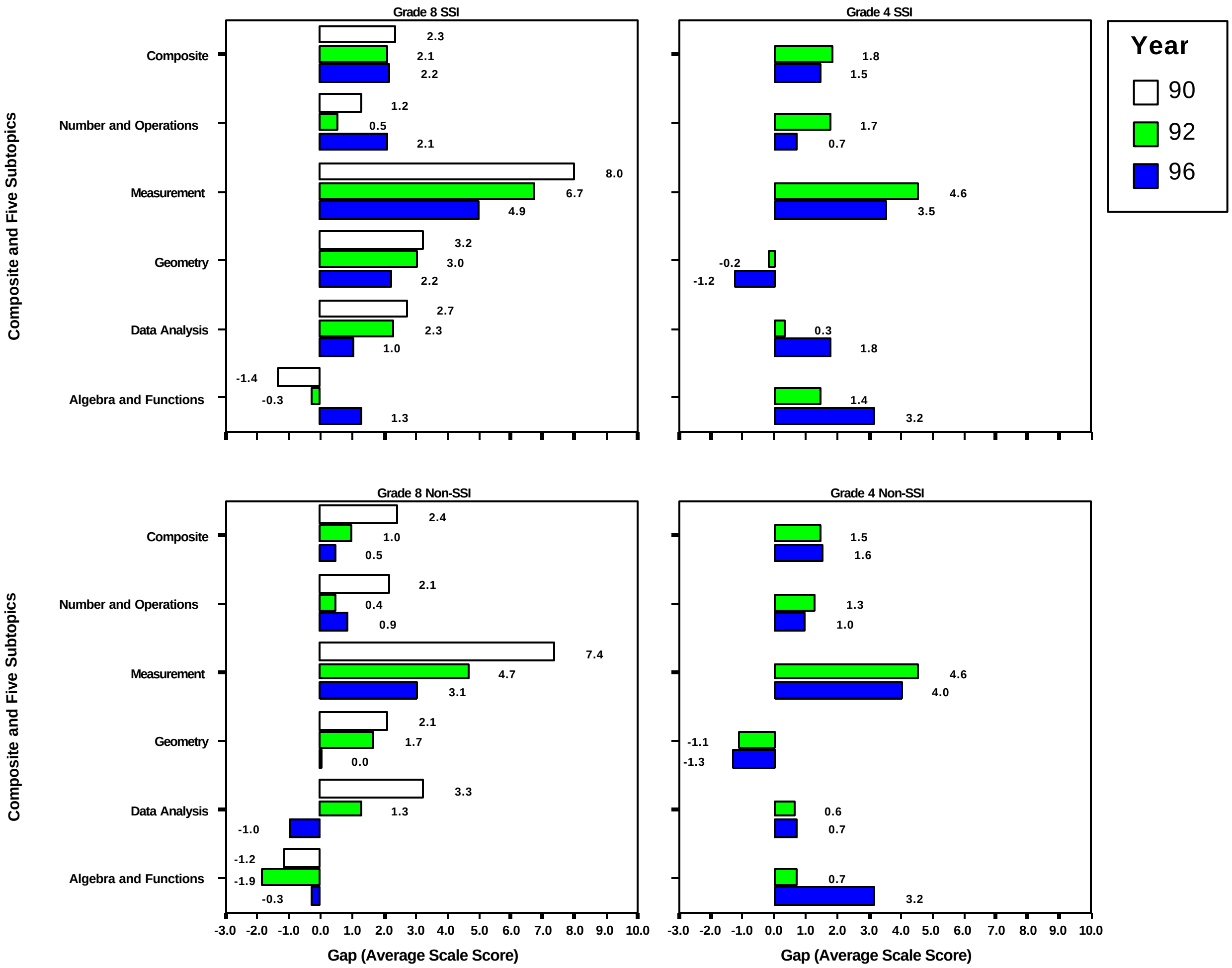
Gender

Overall, trends in the average scale score gaps between male and female students across years and grades show similar patterns: gender gaps are generally smaller in later assessment years than in early assessment years (Figure 5.7).

Grade 8 male students in non-SSI states scored an average of two points higher than females in their mean mathematics composite score in 1990 and one point higher in 1996. In SSI states, males consistently scored 2 points higher in 1990 to 1996. In grade 4, the gap between males and females for both SSI and non-SSI states remained nearly constant in 1992 and 1996, ranging between 1.5 to 1.8 points. The visible changes occurred in the measurement strand. In both grades in SSI states and in non-SSI states, male and female gaps in measurement were reduced from 1990 to 1996 for grade 8 (around 4 points) and from 1992 to 1996 for grade 4 (up to 1 point). More interestingly, the grade 8 gender gaps in algebra and functions for SSI states and in data analysis for non-SSI states were reversed in 1996: In SSI states male students averaged 1.3 points higher in algebra and functions compared to female students. In contrast to the previous two years in data analysis, female students in non-SSI states scored higher than their male counterparts in 1996.

But, there is little evidence from the NAEP data to indicate that SSIs had any effect in closing the achievement gap between male and female students. The mean mathematics composite score for female students and male students differed at most by 2 points at grades 4 and 8, at any of the testing times, and for both SSI and non-SSI states. The mean score for grades 4 and 8 for both male and female students increased over time. However, the 2-point difference in male scores in non-SSI states in grade 8 in 1990 had decreased by one point by 1996. The 2-point gap at grade 8 in SSI-states remained the same over the three testing times. In 1992, at grade 4 the mean mathematics composite score between male and female students differed by 1.8 and 1.5 points for both SSI and non-SSI states, respectively. This gap was lowered only slightly to 1.5 pointS in SSI states in 1996, but remained essentially the same in non-SSI states.

Figure 5.7. Gender differences (males minus females) in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

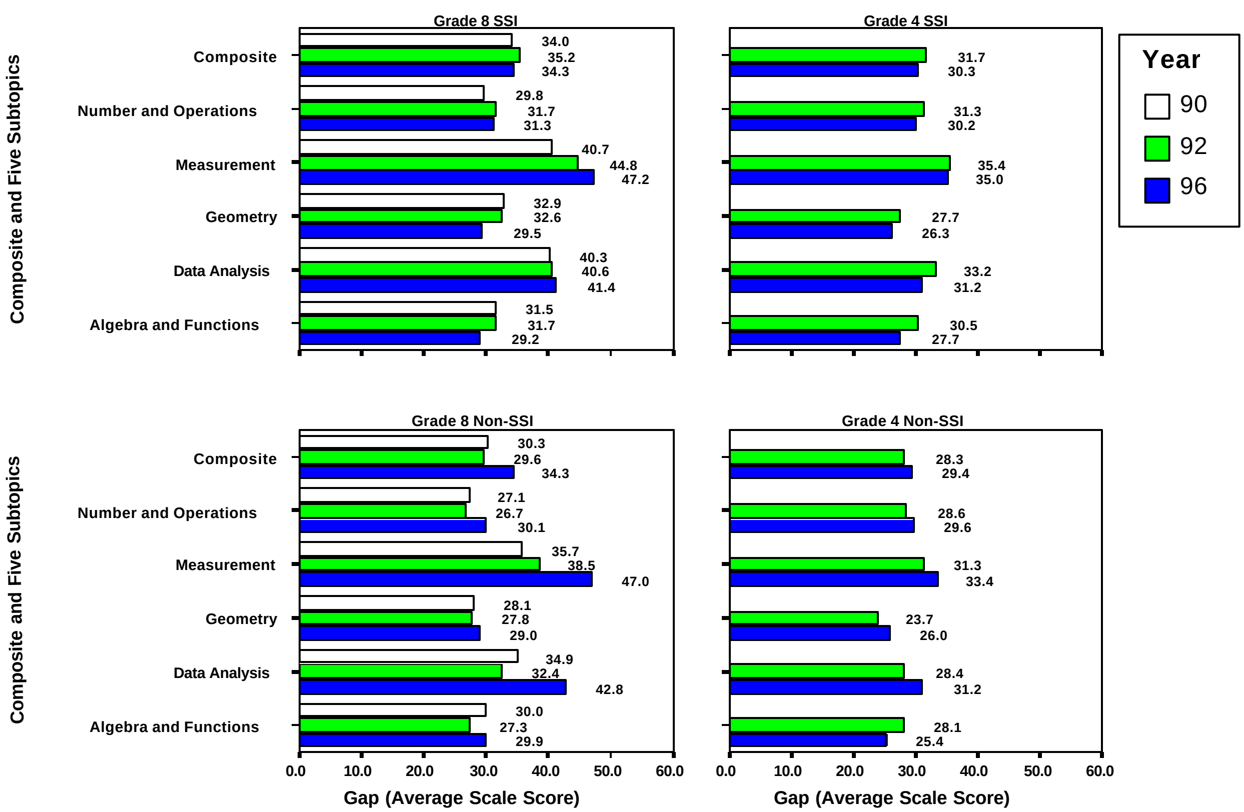


Ethnicity

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 5.8). 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 1992. The gaps in non-SSI states decreased in 1992 from 1990, but increased in 1996. For example, in data analysis, the gap decreased 2.5 points in 1992, but widened by 10.4 points in 1996, although SSI gaps remained stable. In both SSI and non-SSI states, the score gaps in measurement gradually increased from 1990 to 1996.

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 1996, but non-SSI states increased the gaps, with the exception of algebra and functions scale scores. Whereas the gap between White and Black students remained constant in grade 8 and declined slightly in grade 4 in SSI states, in non-SSI states the gap increased at both grades.

Figure 5.8. Differences in average scale scores between White and Black students, by SSI Status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).

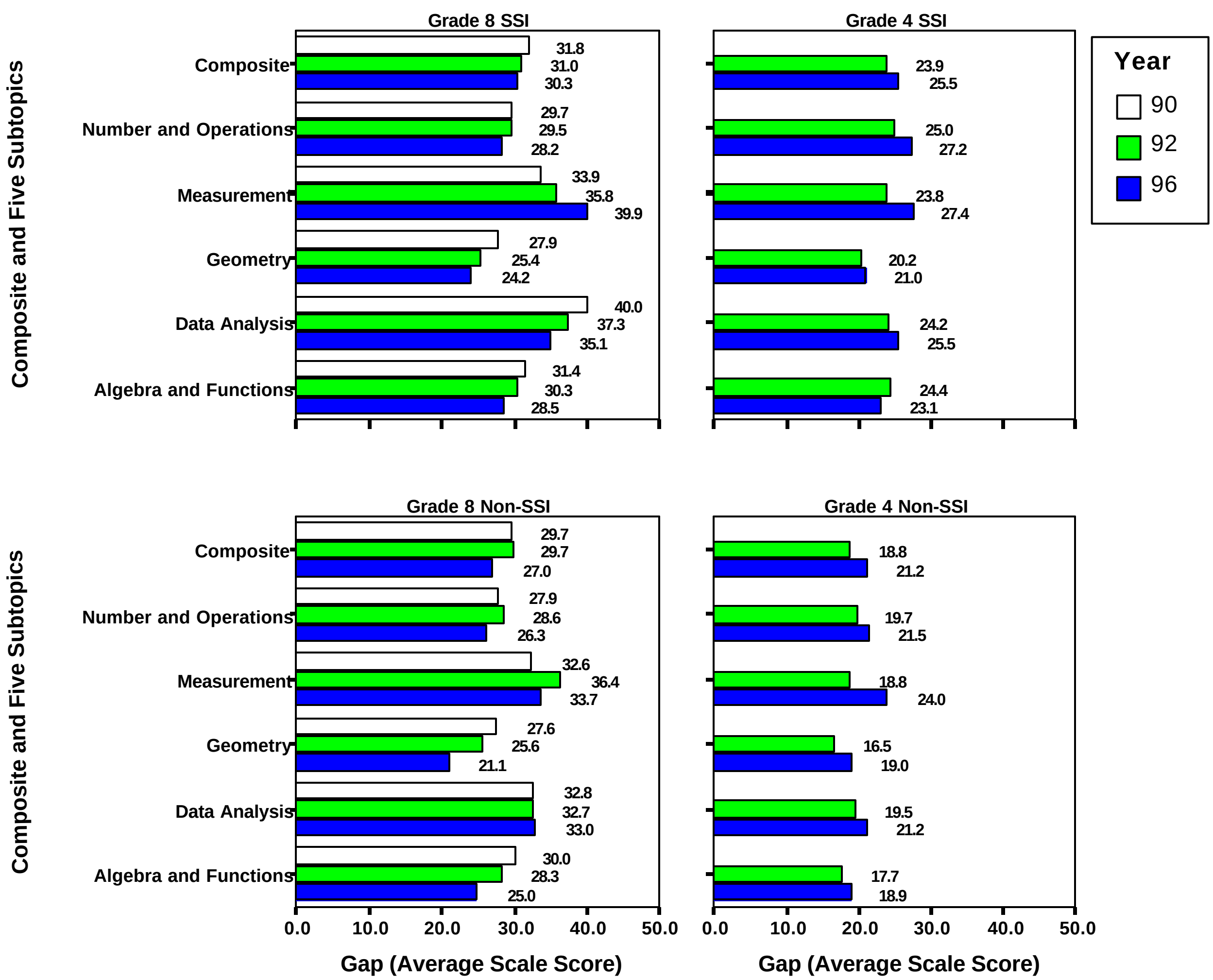


* Due to the insufficient sample size of these subgroups, results are based on 16 SSI states and 6 non-SSI states.

Regarding score differences of White and Hispanic students, there were different trends for students in grades 4 and 8 (Figure 5.9). In grade 8, the gaps in SSI states and non-SSI states narrowed between 1990 and 1996. However, the score gaps for grade 4 students widened from 1992 to 1996. Grade 8 on the measurement content strand showed a different pattern of score-gap change: The gap in SSI states increased across the assessment years, but the score gap in non-SSI states dropped in 1996 after a 4-point increase in 1992.

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

Figure 5.9. Differences in average scale scores of White and Hispanic students, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



* Due to the insufficient sample size of these subgroups, results are based on 15 SSI states and 10 non-SSI states.

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

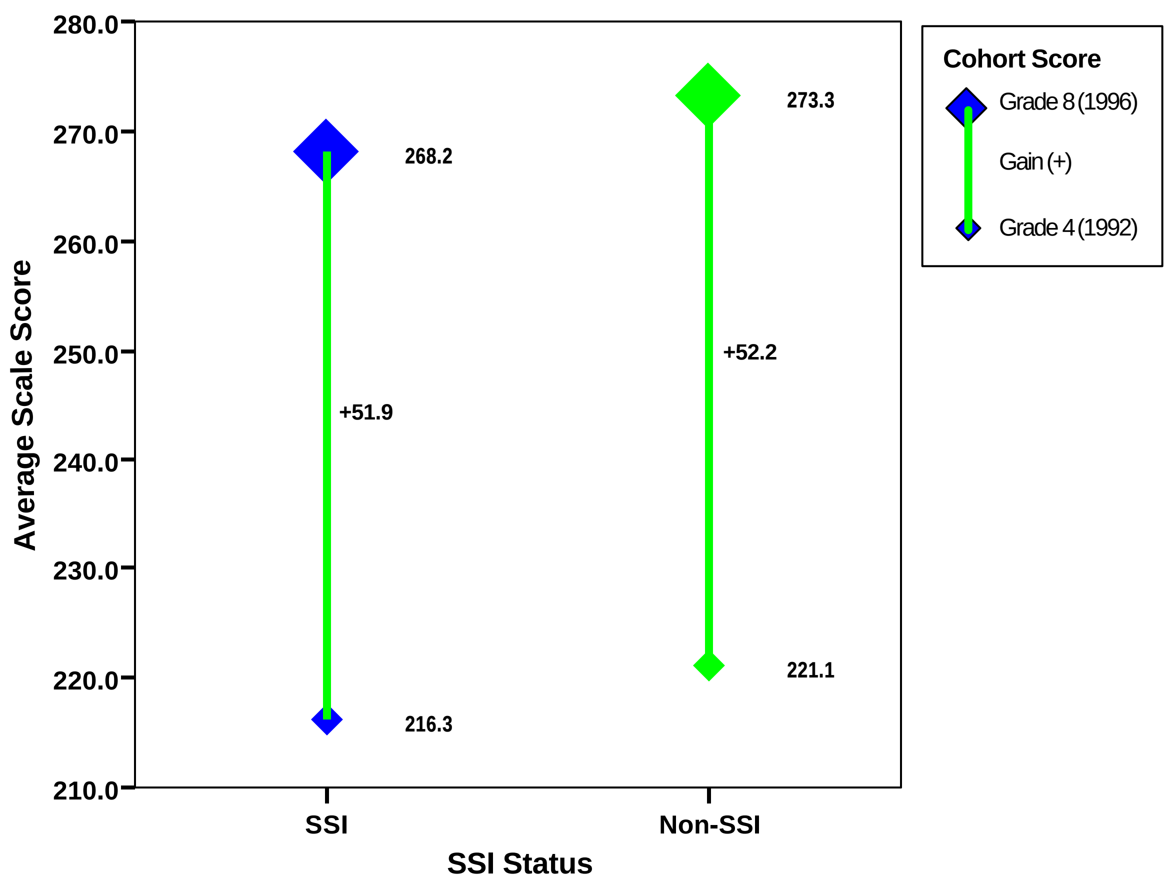
This performance comparison of cohort growth at two grade levels (grade 4 and grade 8) allows us to track achievement growth of the same group of students after four years.

Composite Scores

Total Group

Both SSI and non-SSI states had substantial cohort growth between grade 4 in 1992 and grade 8 in 1996 (Figure 5.10). Students in non-SSI states scored higher than those in SSI states in the two assessment years; for example, the grade 4 scale score in SSI states was 216.3, compared to 221.1 points in non-SSI states. After four years, grade 8 students in SSI states scored 268.2 points and their counterparts in non-SSI states scored 273.3. However, the cohort growth for SSI states and non-SSI states was nearly the same, 51.9 points and 52.2 points, respectively.

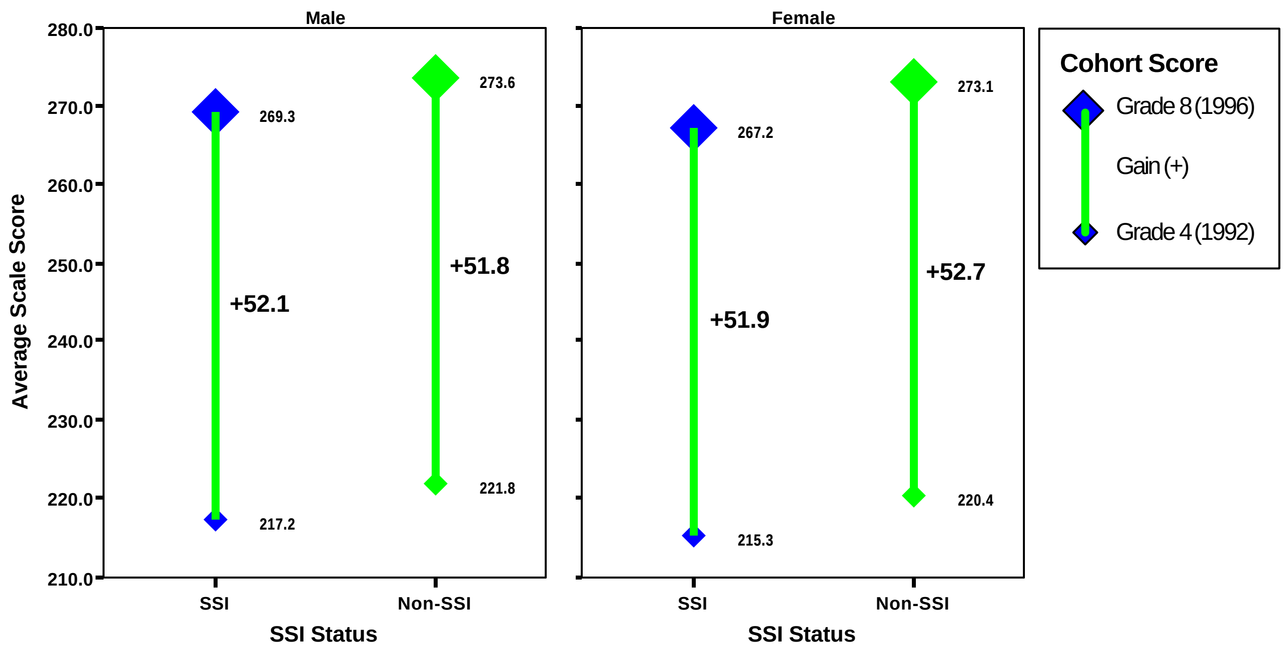
Figure 5.10. Cohort growth in average scale scores from 1992 to 1996, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Gender

Male and female cohorts in both SSI and non-SSI states showed performance improvement in average scale scores from 1992 to 1996. However, we can detect a different pattern of cohort growth between SSI and non-SSI states within gender groups (Figure 5.11). For the male cohort, SSI states gained slightly more, by 0.3 points, than non-SSI states, while the female cohort in non-SSI states outscored their counterparts in SSI states by 0.8 points.

Figure 5.11. Cohort growth in average scale scores, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).

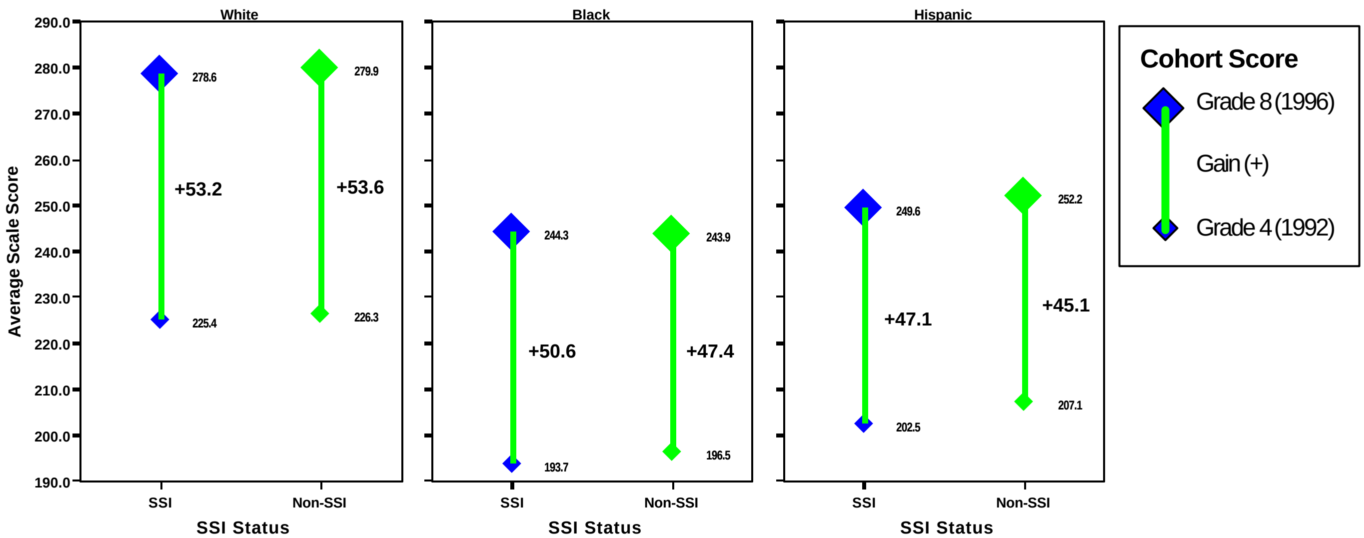


Ethnicity

The cohort growth of the three racial groups shows increases in average scale scores between the two assessment years, 1992 and 1996. The pattern of cohort growth varied among White, Black, and Hispanic students. White students made higher gains than Black and Hispanic students, while Black students made greater gains than Hispanic students (Figure 5.12).

The results for Black and Hispanic students are encouraging for SSI states. Cohort growth of Black and Hispanic students in SSI states was 3.2 points and 2 points respectively, compared to their counterparts in non-SSI states over the four-year timeframe.

Figure 5.12. Cohort growth in average scale scores, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



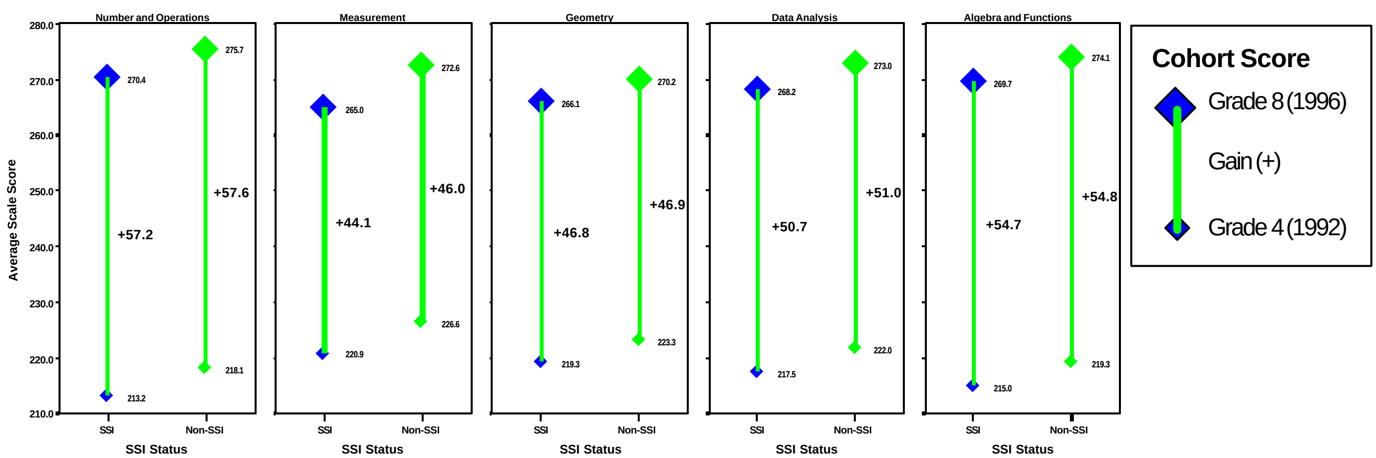
* Due to the insufficient sample size of the subgroups, results are based on 16 SSI states and 6 non-SSI states for Blacks, and 15 SSI states and 10 non-SSI states for Hispanics.

Subtopic Scores

Total Group

Both SSI and non-SSI states showed similar pictures of cohort growth in the five mathematics content strands over the four years. Students made gains from grade 4 in 1992 to grade 8 in 1996. In all five content strand scale scores, non-SSI states showed slightly higher cohort gains than students in SSI states; note that the non-SSI grade 4 students started at a slightly higher point than students in the SSI states. The results show some variations of cohort growth across five content strands (Figure 5.13). Cohort students in both SSI and non-SSI states were likely to make greater gains in number and operations (57 points), algebra and functions (55 points), data analysis (51 points) than in geometry (47 points) and measurement (up to 46 points).

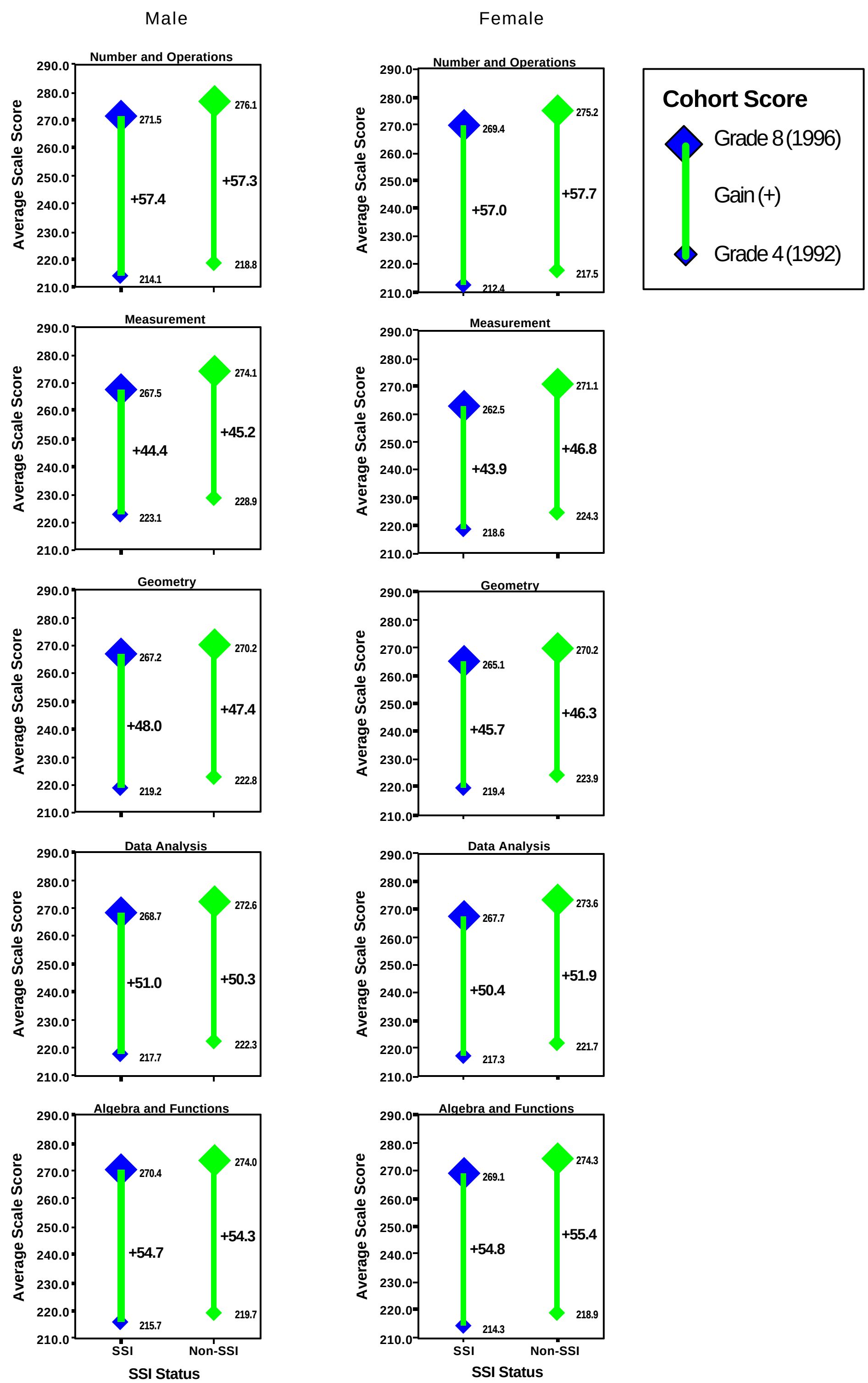
Figure 5.13. Cohort growth in average scale scores on content strands, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



Gender

We can see that the scores of male and female cohort students were quite similar in the five content strands (Figure 5.14). They showed cohort growth of between 44 points and 58 points. Most cohort gains were observed in number and operations and algebra and functions, and least growth in measurement for both male and female in both cohorts. But, there was a difference in cohort growth performance for male and female students by SSI status. Male students in SSI states gained more than those in non-SSI states in four of five content scale scores, while in all five content strands, female students in non-SSIs did better than their counterparts in SSI states.

Figure 5.14. Cohort growth in average scale scores on content strands, by gender and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI).



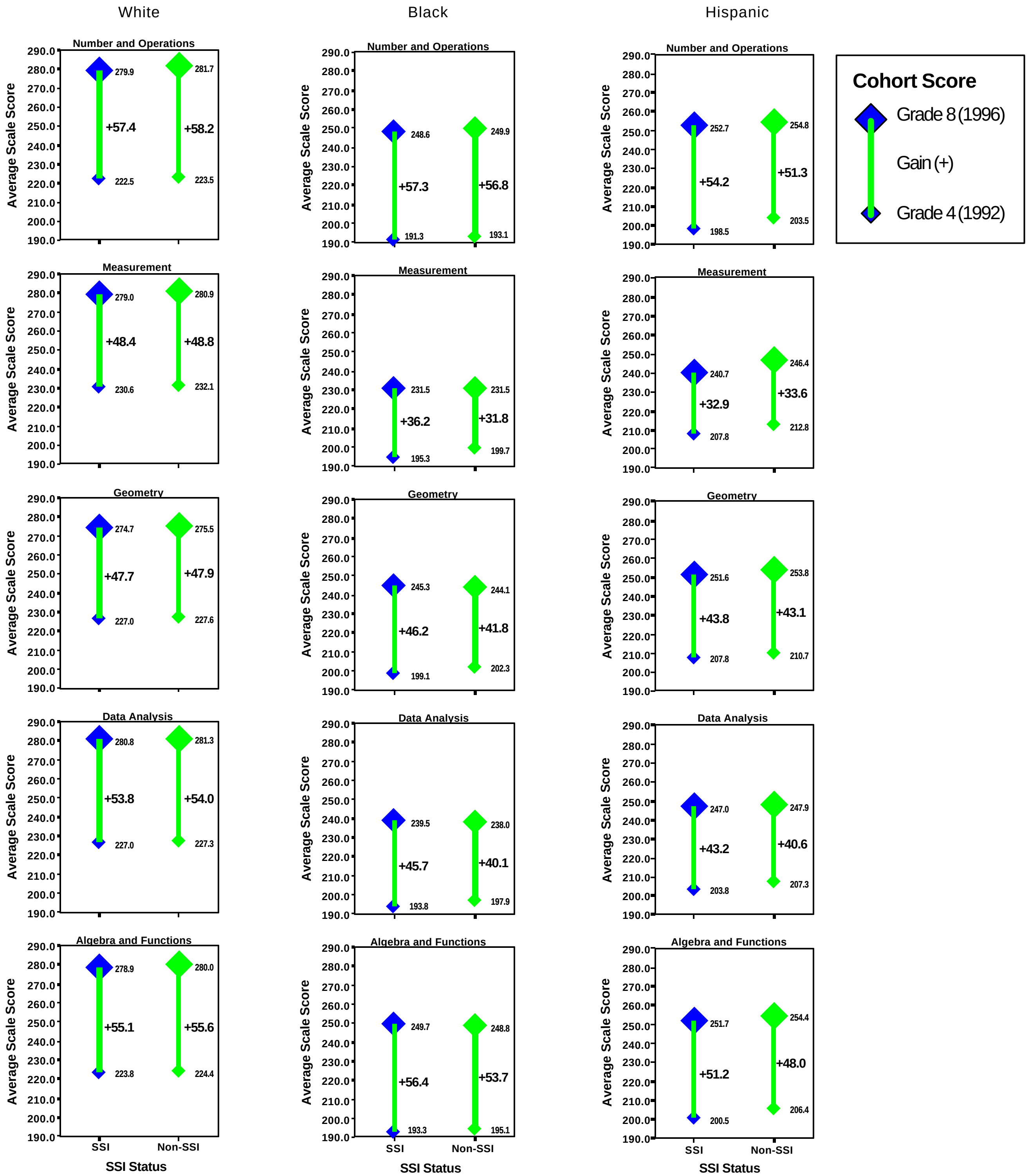
Ethnicity

The results for White, Black, and Hispanic students in cohort growth indicated differences among the three groups in five content strands (Figure 5.15). In general, White students outperformed Black and Hispanic students, and Black students gained more than Hispanic students. As observed in previous figures, Black and Hispanic students did better in number and operations, in algebra and functions, and in data analysis, than in measurement, and in geometry.

Black and Hispanic cohort students in SSI states outperformed their counterparts in non-SSI states in all five content strand scale scores. In particular, cohort growth of Black students in SSI states was substantially larger than the increases of their counterparts in non-SSI states. Thus, although grade 4 Black and Hispanic students in SSI states started below their counterparts in non-SSI states in 1992, four years later there were no gaps between SSI and non-SSI states; in fact, their gaps were reversed in three of five content strands.

Hispanic cohort students in SSI states also gained more than those in non-SSI states in four of the five content strands.

Figure 5.15. Cohort growth in average scale scores on content strands, by race and SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



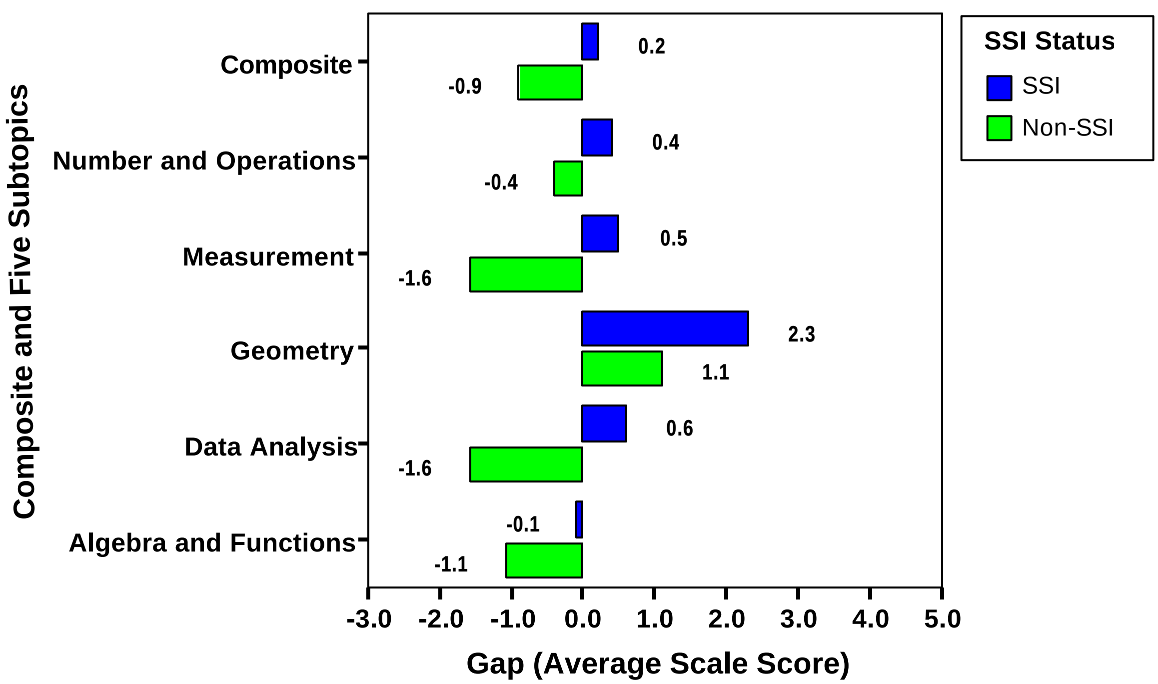
* Due to the insufficient sample size of the subgroups, results are based on 16 SSI states and 6 non-SSI states for Blacks, and 15 SSI states and 10 non-SSI states for Hispanics.

Gaps Between Different Groups

Gender

There was a distinct pattern for gender differences in cohort growth by SSI and non-SSI states (Figure 5.16). Across composite and content strands, male students generally scored higher in SSI states, but the pattern of cohort gender difference was reversed in non-SSI states. The only exceptions to this pattern were in geometry and in algebra and functions. Overall, the gender differences in cohort growth were smaller in SSI states than in non-SSI states. For example, cohort differences between male and female students in SSI states was 0.5 points in measurement, compared to –1.6 points in non-SSI states.

Figure 5.16. Gender differences in average scale scores, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states).



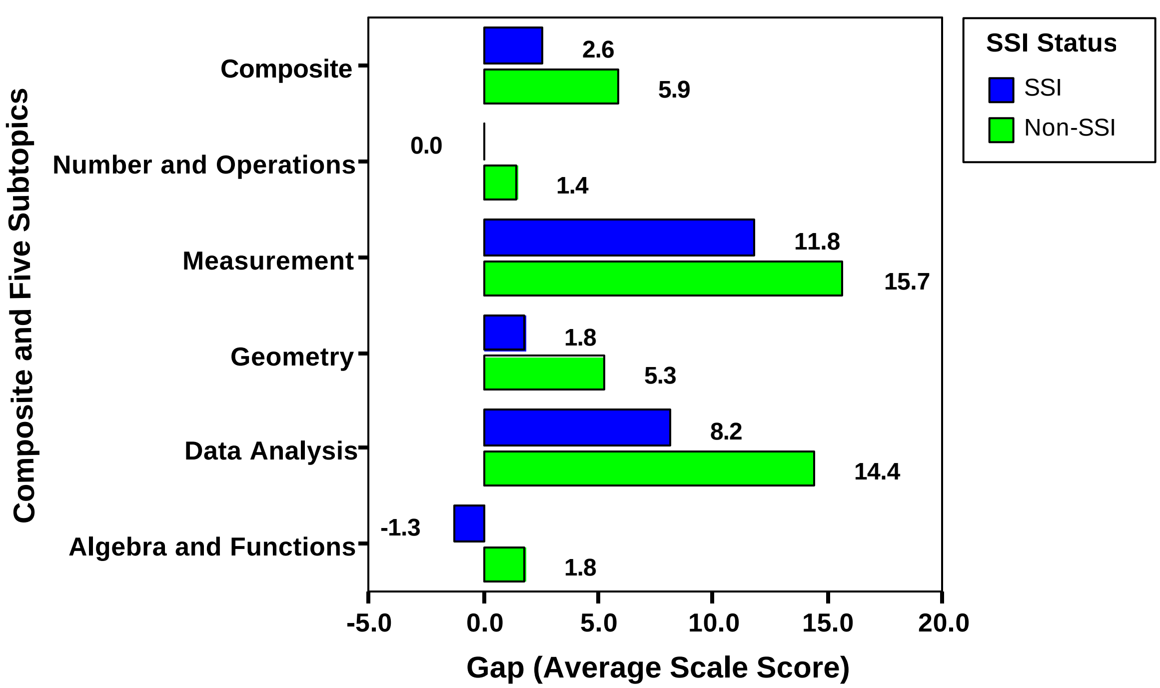
Ethnicity

The results for cohort growth differences between White and Black students were quite interesting. SSI states were successful in reducing the gap in cohort growth between White and Black students (Figure 5.17). 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 data analysis. Non-SSI states had a 14.4-point difference, while SSI states had an 8.2-point difference.

The most interesting picture in cohort growth differences was displayed in algebra and functions. In SSI states in 1996, Black students in the cohort of students who were in grade 4 in 1992 gained more in algebra and functions over four years than White students. As a result, the gap between White and Black students was reversed, with Black students gaining more than White students.

In non-SSI states, White students gained slightly more than Black students, 1.8 points. Although not statistically significant, Black students gaining more than White students is noteworthy considering all of the other comparisons show White students perform better than Black students.

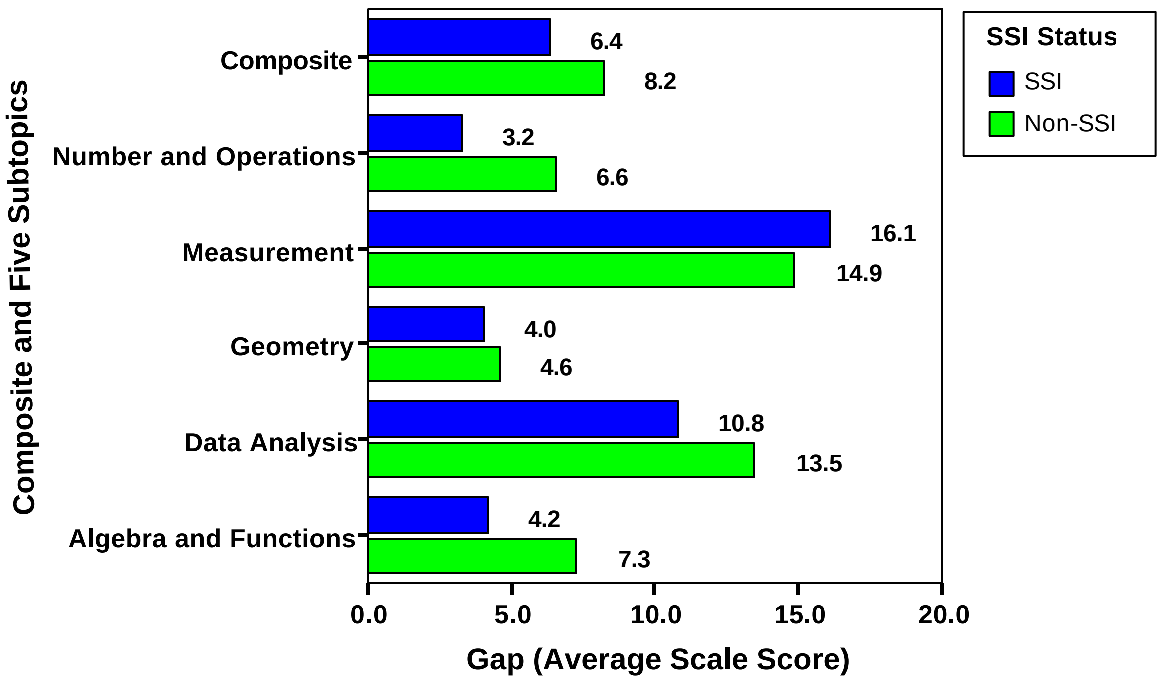
Figure 5.17. Differences in average scale scores between White and Black students from 1992 to 1996, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



* Due to the insufficient sample size of these subgroups, results are based on 16 SSI states and 6 non-SSI states.

SSI states also did better in reducing the gap between White and Hispanic cohort students than non-SSI states (Figure 5.18). In five of the six scale scores, the cohort growth differences narrowed more in SSI states than in non-SSI states: Composite (6.4 points for SSI, 8.2 points for non-SSI); number and operations (3.2 for SSI, 6.6 for non-SSI); geometry (4.0 for SSI, 4.6 for non-SSI); data analysis (10.8 for SSI, 13.5 for non-SSI); and algebra and functions (4.2 for SSI, 7.3 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 and in geometry, the cohort gaps between White and Hispanic students were smaller than in other content strands.

Figure 5.18. Differences in average scale scores between White and Hispanic students, by SSI status: Trend Group 90-96 (17 SSI and 11 non-SSI states*).



* Due to the insufficient sample size of these subgroups, results are based on 15 SSI states and 10 non-SSI states.

Summary and Conclusions

This chapter presents the results from the State NAEP mathematics assessments for grade 4 in 1992 and 1996, and grade 8 in 1990, 1992, and 1996, and for cohort students in SSI states and non-SSI states. We focused our analyses on the 28 states—17 SSI and 11 non-SSI—that participated in all three assessment years. The trend differences between SSI and non-SSI states in the composite score and in each of the five content strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions) were based on descriptive trend analyses that compared 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. Considerable improvements were also noted for students by gender and race/ethnicity. However, SSI and non-SSI states showed no clear patterns in the gaps between different subgroups across the assessment years.

Summaries of performance trends for different subgroups and gaps between males and females, as well as between Whites and Blacks and between Whites and Hispanics, are as follows:

Trends in average scale scores

- Both SSI and non-SSI states experienced an increase in the average composite scale scores from 1990 to 1996 at grade 8 and from 1992 to 1996 at grade 4.
- Male and female students in SSI states consistently scored lower than their counterparts in non-SSI states, at both grades 4 and 8, on the composite scale score and on each content strand. However, both male and female students in SSI states gained at a slightly higher rate over this period than those in non-SSI states on most scale scores. Males in SSI states had considerable higher gains on the geometry and data analysis scales than on the other scales.
- For both SSI and non-SSI states, the gaps between the performance on the different scales by male and female students were gradually reduced from 1990 to 1996. Some variation in gender differences for the composite score and on each of the five content strands was apparent.
- Substantial performance differences by ethnicity between SSI and non-SSI states existed on all six scale scores at both grades 4 and 8. Regardless of the SSI status, White students outperformed Black and Hispanic students on all scale scores at both grade levels. Hispanic students scored higher than Black students at both grade levels. Generally, White and Black students in SSI states gained substantially more than their counterparts in non-SSI states on most scale scores.
- The difference in performance between White and Black grade 8 students in SSI states decreased in geometry and algebra and functions, but increased very slightly on the composite score in number and operations, measurement, and data analysis. In contrast, the gap between White and Black grade 8 students in six non-SSI states increased substantially on all of the scales except in algebra and functions, where it remained constant. At grade 4 from 1992 to 1996, the gap between White and Black students in SSI

states decreased on all six scales, but in the six non-SSI states the gap increased on all scales except for algebra and functions.

- 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 1990 to 1996 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 nearly the same over the four years, 51.9 and 52.2 respectively.
- The cohort of male students in SSI states gained 0.3 points more between grade 4 and grade 8 than male students in non-SSI states, whereas female students in SSI states gained 0.8 points less than female students in non-SSI states.
- Comparing the four-year gain made by male students with the gain made by female students, male students in SSI states had a higher gain than female students on five of the six scales, with algebra and functions as the only exception. In contrast, male students in non-SSI states gained less than female students on five of the six scales, with the measurement scale as the only exception.
- Considering ethnicity, White students in the 1992 grade 4 cohort gained about the same in both SSI states and non-SSI states over four years, 53.2 and 53.6 respectively.
- Black and Hispanic students in both SSI states and non-SSI states gained less than White students over the four years between grade 4 and grade 8. This indicates that Black and Hispanic students continued to lose ground over these four years. However, Black and Hispanic students in SSI states gained more than Blacks and Hispanics did in non-SSI states. Thus, Black and Hispanic students did not lose as much ground in SSI states as in non-SSI states.
- 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 algebra and functions scale, Blacks in SSI states actually gained more between grade 4 and grade 8 than did White students. On the number and operations scale, the gain by Black and White students was the same. The greatest difference in the White-Black gap between SSI and non-SSI states was on the data analysis scale.
- The White-Hispanic gap in the gain scores between grade 4 and grade 8 was less in SSI states than in non-SSI states on all of the scales except on measurement.

Even though the descriptive trends 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 SSI in those states. There are many factors involved in how students learn over the years. School structures, home environments, state educational policies, and others can affect learning. In the next chapter, we will identify some policy-relevant variables related to SSI states and their relationships with student outcomes.

List of Appendices

Appendix A shows the results of student performance in the composite and the five content strands for all of SSI and non-SSI states participating in any given assessment year.

Table 5A.1
Average Scale Scores, by SSI Status: All Available Samples

Table 5A.2
Average Scale Scores, by Gender and SSI Status: All Available Samples

Table 5A.3
Average Scale Scores, by Race and SSI Status: All Available Samples

Table 5A.4
Average Scale Scores in Content Strands, by SSI Status: All Available Samples

Table 5A.5
Average Scale Scores in Content Strands, by Gender and SSI Status: All Available Samples

Table 5A.6
Average Scale Scores in Content Strands, by Race and SSI Status: All Available Samples

Table 5A.7
Gender Differences in Average Scale Scores, by SSI Status: All Available Samples

Table 5A.8
Differences in Average Scale Scores Between Racial Subgroups, by SSI Status: All Available Samples

Appendix B presents the results of cohort student performance in the composite and the five content strands for SSI and non-SSI states that participated in both 1992 grade 4 assessment and 1996 grade 8 assessment.

Figure 5B.1. Cohort growth in average scale scores, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.2. Cohort growth in average scale scores, by gender and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.3. Cohort growth in average scale scores, by race and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.4. Cohort growth in average scale scores on content strands, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.5. Cohort growth in average scale scores on content strands, by gender and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.6. Cohort growth in average scale scores on content strands, by race and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.7. Gender differences in average scale scores, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.8. Differences in average scale scores between White and Black students, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Figure 5B.9. Differences in average scale scores between White and Hispanic Students, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

Appendix A

Appendix A shows the results of student performance in the composite and the five content strands for all of SSI and non-SSI states participating in any given assessment year.

Composite Scores

Total Group

Table 5A.1
Average Scale Scores, by SSI Status: All Available Samples

	1990	1992	1996
Grade 8			
SSI	260.6	265.4	270.2
Non-SSI	263.8	266.9	272.2
Grade 4			
SSI		217.9	221.7
Non-SSI		218.3	223.0

Gender

Table 5A.2
Average Scale Scores, by Gender and SSI Status: All Available Samples

	1990		1992		1996	
	Male	Female	Male	Female	Male	Female
Grade 8						
SSI	261.9	259.3	266.4	264.3	271.3	269.2
Non-SSI	265.1	262.5	267.7	266.0	272.5	272.0
Grade 4						
SSI			218.8	216.9	222.6	220.7
Non-SSI			218.7	217.9	223.7	222.2

Ethnicity

Table 5A.3
Average Scale Scores, by Race and SSI Status: All Available Samples

	1990			1992			1996		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
Grade 8									
SSI	270.0	235.6	237.9	274.8	239.2	242.4	279.4	244.7	248.7
Non-SSI	268.6	236.9	240.5	271.7	239.0	244.0	278.9	243.7	251.9
Grade 4									
SSI				226.4	194.6	203.3	229.9	200.1	206.1
Non-SSI				224.3	194.9	204.4	229.2	200.3	208.7

Subtopic Scores

Total Group

Table 5A.4
Average Scale Scores in Content Strands, by SSI Status: All Available Samples

		1990	1992	1996
Grade 8				
Number and Operations	**SSI**	264.8	269.7	272.1
	Non-SSI	268.3	271.0	274.7
Measurement	**SSI**	256.8	263.2	267.8
	Non-SSI	261.1	266.0	271.0
Geometry	**SSI**	258.0	260.8	267.9
	Non-SSI	261.6	262.5	269.2
Data Analysis	**SSI**	260.7	265.4	270.2
	Non-SSI	263.1	267.0	272.0
Algebra and Functions	**SSI**	259.8	264.9	271.6
	Non-SSI	262.0	265.7	273.0
Grade 4				
Number and Operations	**SSI**		214.9	218.2
	Non-SSI		215.6	219.3
Measurement	**SSI**		222.5	223.9
	Non-SSI		223.5	225.9
Geometry	**SSI**		220.8	223.6
	Non-SSI		220.7	224.7
Data Analysis	**SSI**		218.9	223.0
	Non-SSI		218.4	223.9
Algebra and Functions	**SSI**		216.4	225.1
	Non-SSI		216.6	226.5

Gender

Table 5A.5
Average Scale Scores in Content Strands, by Gender and SSI Status: All Available Samples

		1990		1992		1996	
		Male	Female	Male	Female	Male	Female
Grade 8							
Number and Operations	**SSI**	265.6	264.1	270.1	269.4	273.1	271.2
	Non-SSI	269.4	267.1	271.3	270.6	274.9	274.5
Measurement	**SSI**	260.9	252.6	266.7	259.9	270.1	265.5
	Non-SSI	264.9	257.3	269.1	262.9	272.7	269.4
Geometry	**SSI**	259.6	256.3	262.4	259.2	269.0	266.9
	Non-SSI	262.8	260.4	263.7	261.2	269.3	269.1
Data Analysis	**SSI**	262.2	259.2	266.6	264.3	270.7	269.7
	Non-SSI	264.9	261.4	267.8	266.1	271.4	272.6
Algebra and Functions	**SSI**	259.2	260.4	264.7	265.1	272.3	270.9
	Non-SSI	261.6	262.4	265.2	266.2	273.0	273.1
Grade 4							
Number and Operations	**SSI**			215.8	214.0	218.9	217.5
	Non-SSI			215.9	215.2	219.8	218.8
Measurement	**SSI**			224.8	220.2	225.9	221.9
	Non-SSI			225.3	221.6	227.9	223.9
Geometry	**SSI**			220.7	220.9	223.2	224.0
	Non-SSI			220.0	221.4	224.1	225.4
Data Analysis	**SSI**			219.1	218.7	224.0	222.0
	Non-SSI			218.5	218.4	224.3	223.5
Algebra and Functions	**SSI**			217.1	215.7	226.7	223.4
	Non-SSI			216.4	216.7	227.8	225.0

Ethnicity

Table 5A.6
Average Scale Scores in Content Strands, by Race and SSI Status: All Available Samples

		1990			1992			1996		
		White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
Grade 8										
Number and Operations	**SSI**	273.3	243.1	243.4	278.3	246.2	247.6	280.4	249.2	251.8
	Non-SSI	272.4	243.9	246.4	275.4	246.7	248.4	280.7	250.2	255.0
Measurement	**SSI**	267.5	226.5	232.9	274.8	229.1	238.6	280.3	231.5	240.6
	Non-SSI	266.2	229.9	235.6	272.1	230.0	239.3	279.8	229.9	246.9
Geometry	**SSI**	266.6	233.3	238.3	269.1	236.4	242.0	275.6	245.8	250.4
	Non-SSI	266.3	235.2	239.6	266.8	235.3	243.4	274.7	244.1	253.4
Data Analysis	**SSI**	272.3	231.5	232.6	276.6	235.9	237.2	281.4	239.9	245.2
	Non-SSI	268.9	233.3	237.7	272.8	236.2	241.6	280.5	237.7	247.3
Algebra and Functions	**SSI**	268.7	236.8	236.8	273.7	241.5	241.9	279.7	250.1	251.3
	Non-SSI	266.7	236.2	238.4	270.0	240.2	243.1	279.0	248.9	253.4
Grade 4										
Number and Operations	**SSI**				223.6	192.2	199.2	226.6	197.0	201.3
	Non-SSI				221.7	192.2	200.9	225.7	196.9	204.4
Measurement	**SSI**				231.7	196.3	208.7	233.1	198.7	207.1
	Non-SSI				229.6	197.1	209.9	232.9	199.2	211.0
Geometry	**SSI**				227.9	200.3	209.0	230.6	204.1	211.3
	Non-SSI				225.7	200.5	208.8	230.1	204.8	211.9
Data Analysis	**SSI**				227.8	194.4	204.5	231.3	201.3	208.0
	Non-SSI				225.0	194.9	204.2	230.4	201.0	208.9
Algebra and Functions	**SSI**				224.8	193.8	200.9	232.6	205.5	210.8
	Non-SSI				222.2	194.0	202.5	232.1	205.9	213.8

Gaps Between Different Groups

Gender

Table 5A.7
Gender Differences in Average Scale Scores, by SSI Status: All Available Samples

		1990	1992	1996
Grade 8				
Composite	**SSI**	2.6	2.1	2.1
	Non-SSI	2.6	1.7	0.5
Number and Operations	**SSI**	1.5	0.7	2.0
	Non-SSI	2.2	0.7	0.4
Measurement	**SSI**	8.3	6.8	4.6
	Non-SSI	7.5	6.2	3.4
Geometry	**SSI**	3.3	3.1	2.1
	Non-SSI	2.3	2.5	0.3
Data Analysis	**SSI**	3.0	2.4	1.0
	Non-SSI	3.5	1.6	-1.2
Algebra and Functions	**SSI**	-1.2	-0.4	1.4
	Non-SSI	-0.8	-0.9	-0.1
Grade 4				
Composite	**SSI**		1.9	1.9
	Non-SSI		0.8	1.5
Number and Operations	**SSI**		1.8	1.3
	Non-SSI		0.7	1.1
Measurement	**SSI**		4.6	4.1
	Non-SSI		3.7	4.0
Geometry	**SSI**		-0.3	-0.9
	Non-SSI		-1.4	-1.3
Data Analysis	**SSI**		0.4	2.0
	Non-SSI		0.0	0.7
Algebra and Functions	**SSI**		1.5	3.3
	Non-SSI		-0.3	2.8

Ethnicity

Table 5A.8
Differences in Average Scale Scores Between Racial Subgroups, by SSI Status: All Available Samples

		Black Gap (White – Black)			Hispanic Gap (White – Hispanic)		
		1990	1992	1996	1990	1992	1996
Grade 8							
Composite	**SSI**	34.3	35.4	33.9	32.1	32.1	31.4
	Non-SSI	31.8	31.6	34.5	29.5	28.6	27.0
Number and Operations	**SSI**	30.2	32.0	30.7	29.8	30.5	29.3
	Non-SSI	28.8	27.9	29.7	27.3	27.7	25.7
Measurement	**SSI**	40.7	45.3	47.2	34.6	35.8	40.4
	Non-SSI	36.8	40.2	49.1	32.6	33.8	32.9
Geometry	**SSI**	33.1	32.5	29.0	28.3	26.8	25.7
	Non-SSI	30.7	30.1	29.6	27.7	24.1	21.4
Data Analysis	**SSI**	40.8	40.5	41.0	39.7	39.0	37.1
	Non-SSI	35.4	35.9	42.7	32.7	32.3	33.1
Algebra and Functions	**SSI**	31.7	32.2	28.9	31.9	31.6	29.3
	Non-SSI	30.8	28.9	29.6	29.8	27.8	25.7
Grade 4							
Composite	**SSI**		31.5	29.7		23.1	23.9
	Non-SSI		29.3	29.0		19.7	20.5
Number and Operations	**SSI**		31.2	29.7		24.4	25.3
	Non-SSI		29.6	29.1		20.7	21.3
Measurement	**SSI**		35.0	34.2		23.0	26.1
	Non-SSI		32.7	33.5		19.5	21.9
Geometry	**SSI**		27.2	26.1		18.9	19.4
	Non-SSI		24.3	25.1		16.7	18.2
Data Analysis	**SSI**		33.1	30.0		23.3	23.3
	Non-SSI		29.6	29.6		20.0	21.5
Algebra and Functions	**SSI**		30.7	27.1		23.9	21.8
	Non-SSI		28.4	26.1		19.9	18.3

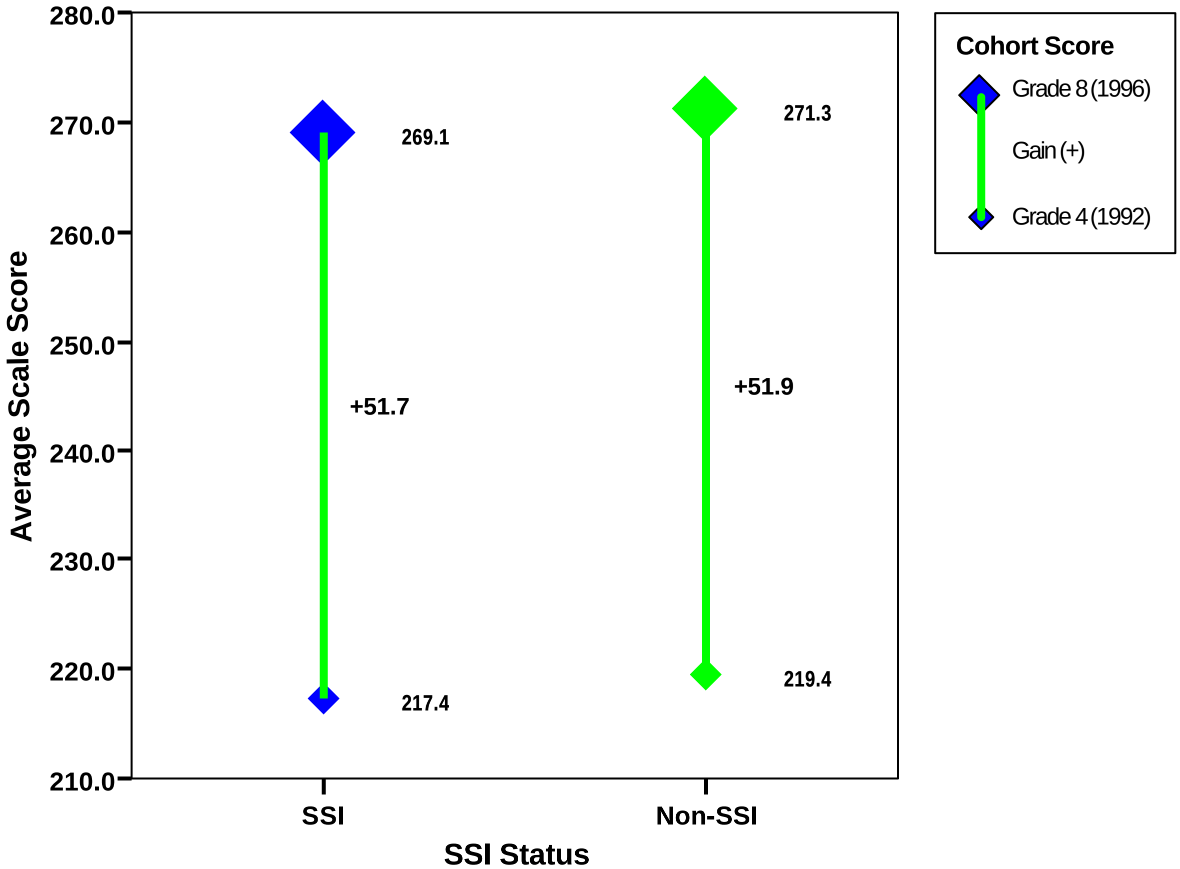
Appendix B

Appendix B presents the results of cohort student performance in the composite and the five content strands for SSI and non-SSI states that participated in both 1992 grade 4 assessment and 1996 grade 8 assessment.

Composite Scores

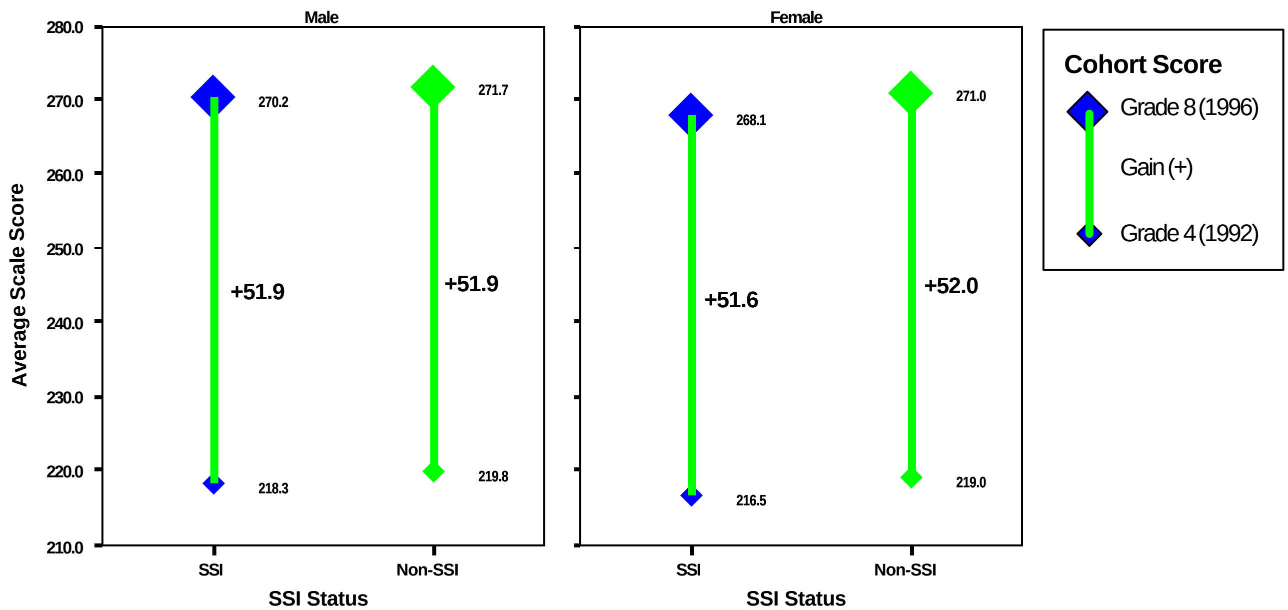
Total Group

Figure 5B.1. Cohort growth in average scale scores, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



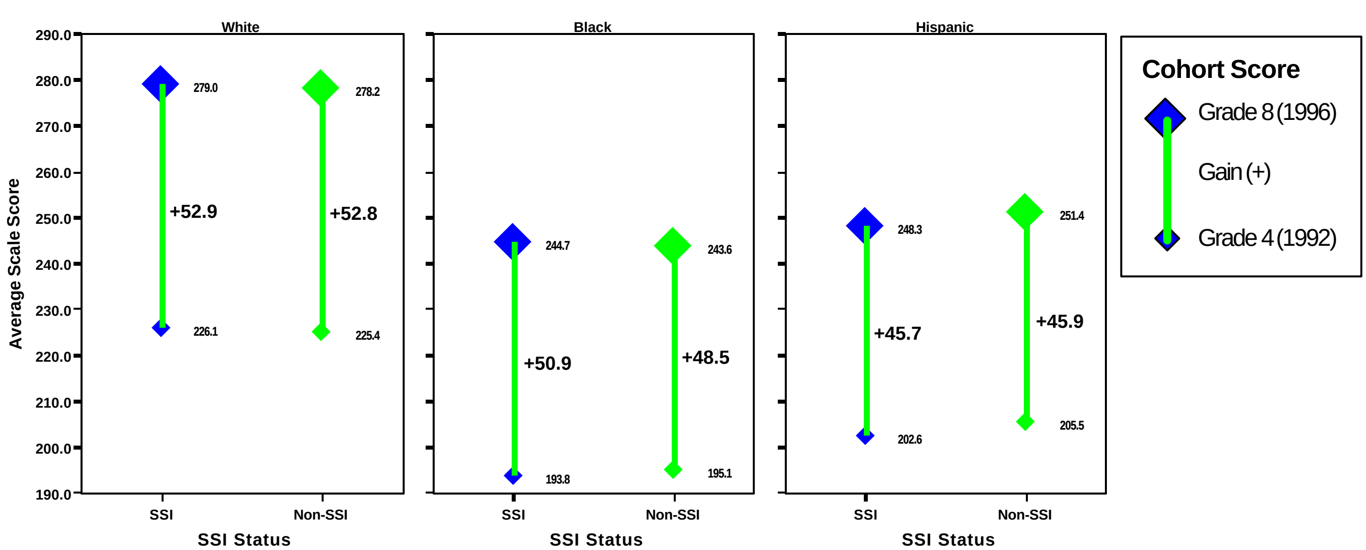
Gender

Figure 5B.2. Cohort growth in average scale scores, by gender and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



Ethnicity

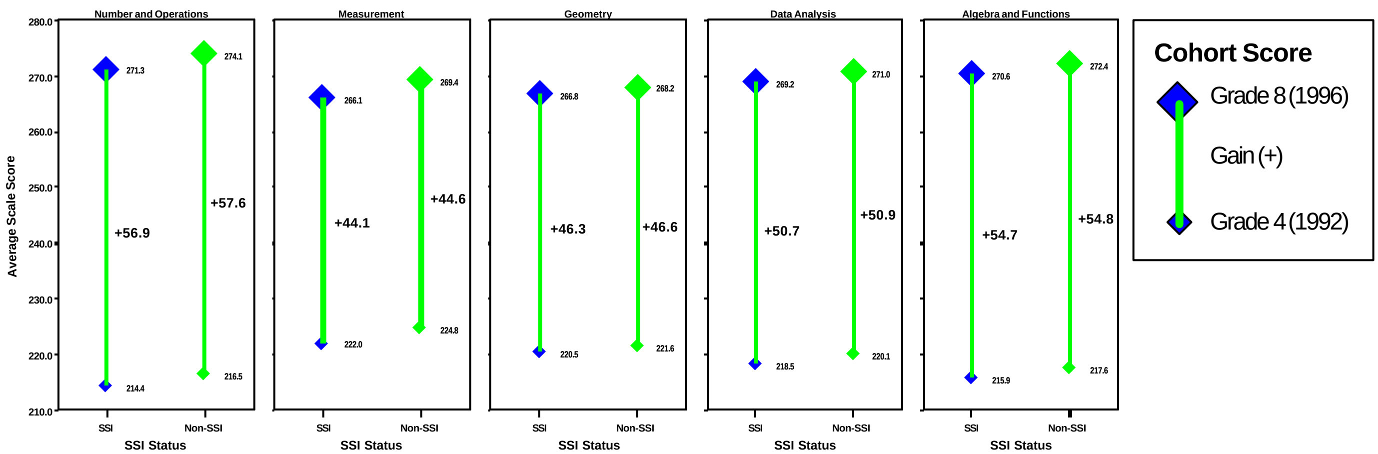
Figure 5B.3. Cohort growth in average scale scores, by race and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



Subtopic Scores

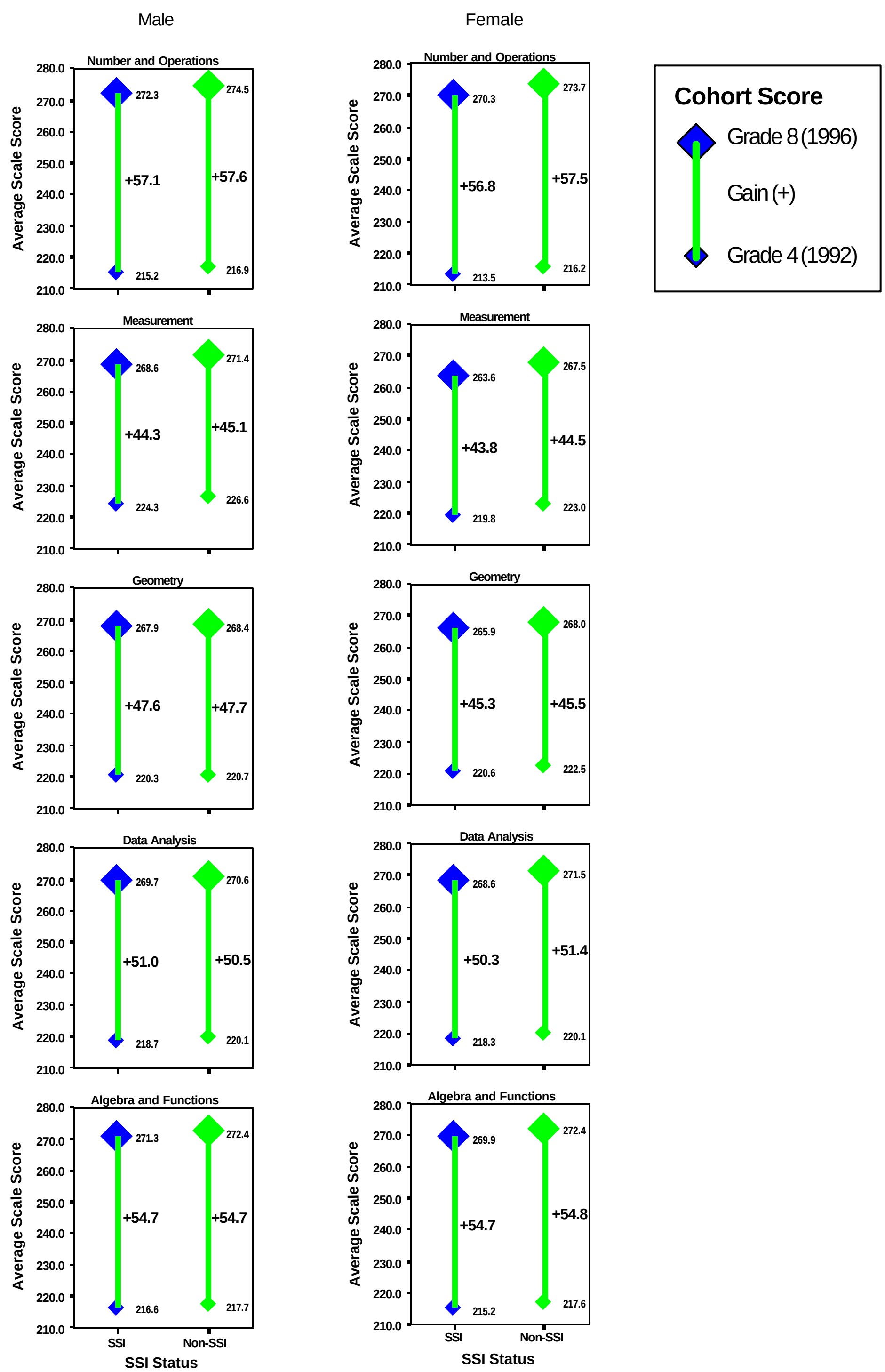
Total Group

Figure 5B.4. Cohort growth in average scale scores on content strands, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



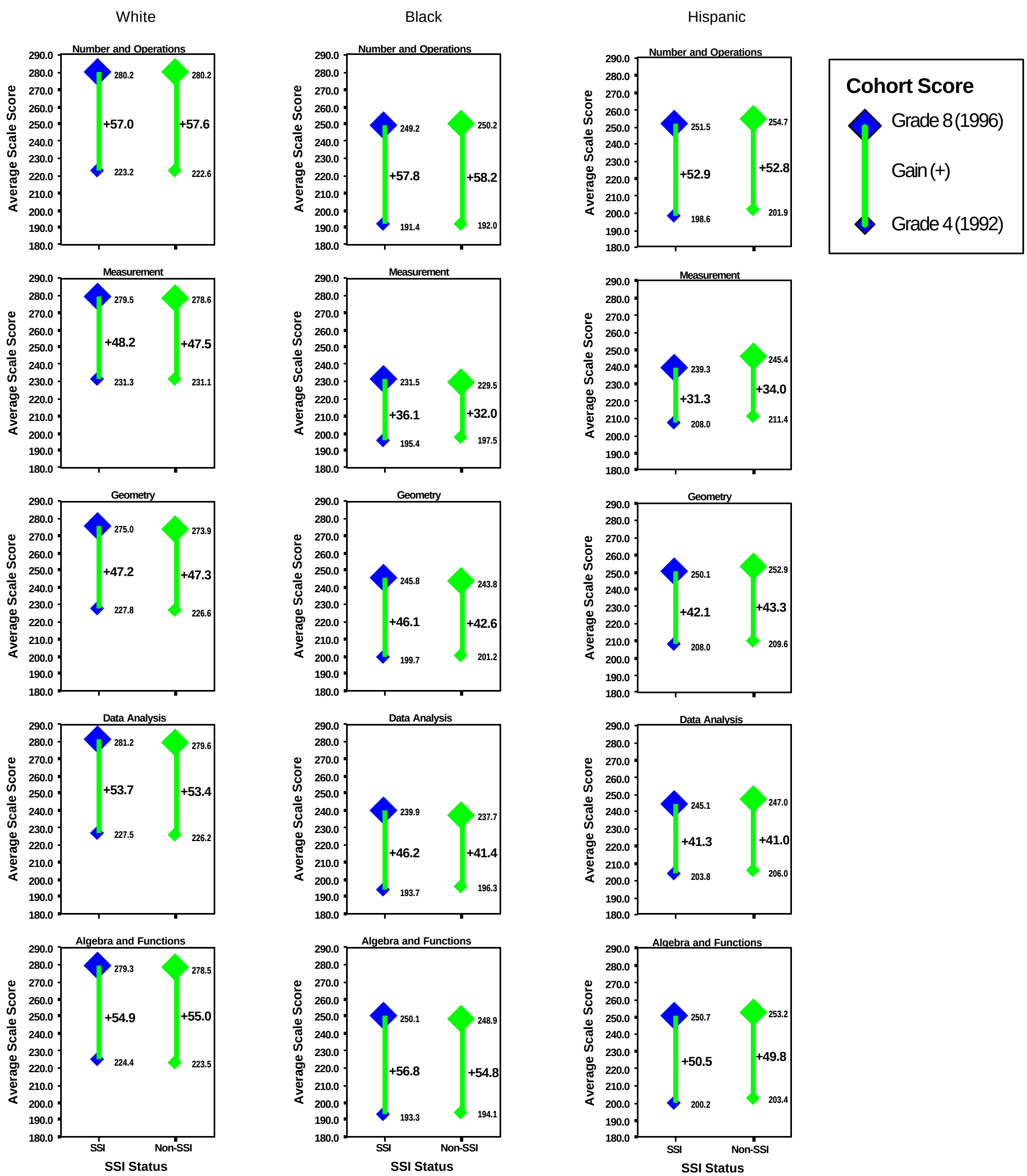
Gender

Figure 5B.5. Cohort growth in average scale scores on content strands, by gender and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



Ethnicity

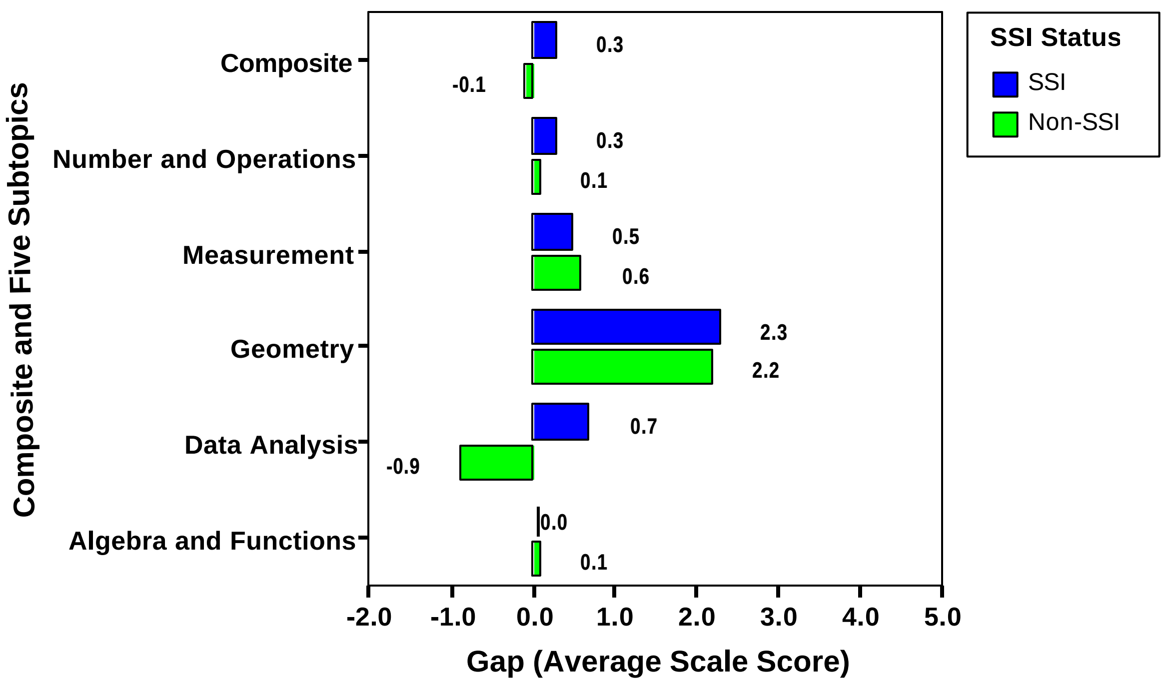
Figure 5B.6. Cohort growth in average scale scores on content strands, by race and SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



Gaps Between Different Groups in Cohort Growth from Grade 4 (1992) to Grade 8 (1996)

Gender

Figure 5B.7. Gender differences in average scale scores, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



Ethnicity

Figure 5B.8. Differences in average scale scores between White and Black students, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).

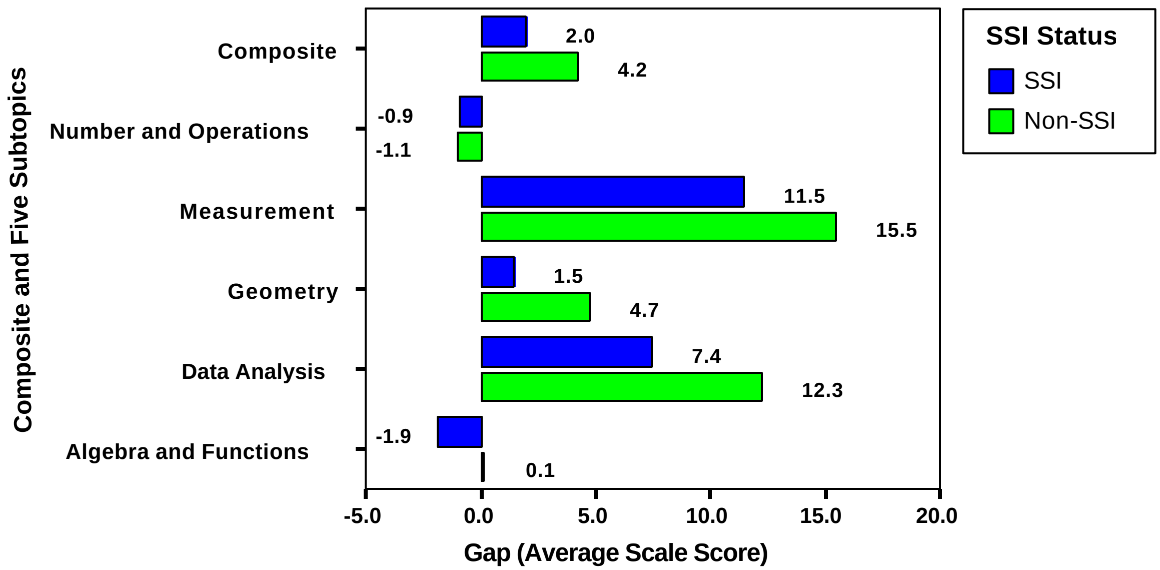
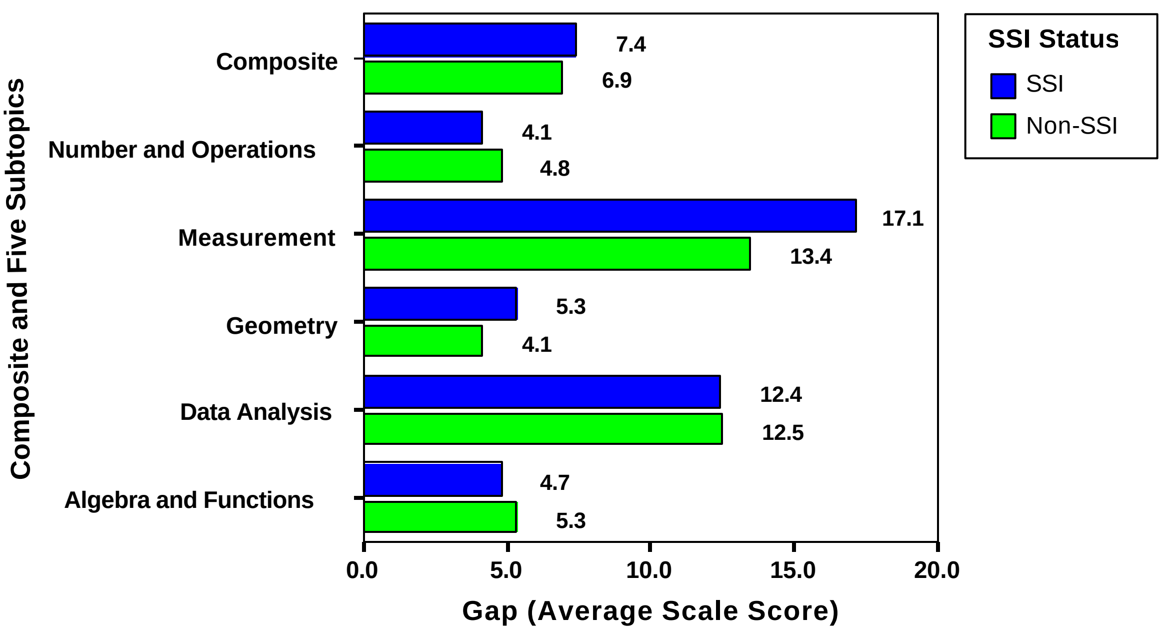


Figure 5B.9. Differences in average scale scores between White and Hispanic Students, by SSI status: Cohort Group 92-96 (20 SSI and 15 non-SSI states).



CHAPTER 6

COMPARING SSI AND NON-SSI STATES USING INDICATORS OF MATHEMATICS CURRICULAR REFORM FROM THE STATE NAEP TEACHER QUESTIONNAIRES

Introduction

This chapter presents six indicators of mathematics reform that were developed from the State NAEP teacher questionnaires. The chapter begins with a general description of the development of the indicators and a discussion of research design issues. The next six sections describe each indicator in detail. Each begins with a discussion of an indicator in the context of curricular reform in mathematics, followed by a description of the 1996 indicator and the 1996 comparisons between SSI and non-SSI states at both grade 8 and grade 4. Then comparisons in 1992 and 1990 are presented, using the information available from NAEP teacher questionnaires in those years. The sections end by examining the effect of SSI status on changes in the indicator across time. After each indicator has been described, relationships among the six indicators are explored, along with their relationship with student achievement. The chapter ends with a general summary of the study's conclusions and limitations.

Research Design

Developing Indicators from the State NAEP Teacher Questionnaires

The State NAEP included a teacher questionnaire for the teachers of participating students. The questionnaire addressed the teachers' backgrounds, general training, and their instructional practices. The teacher questionnaire changed considerably from 1990 to 1996. The 1996 questionnaire included a number of new items, particularly those related to curricular reform in mathematics. Very few items were exactly the same across all three years. There were changes in both the wording of the question as well as the number of response options and their labels.

State NAEP results are often reported in terms of the proportion of students with teachers selecting each response option for each item. (See, for example, Shaughnessy, Nelson, & Norris, 1997.) Nonparametric analyses are used to compare two or more groups on the proportion of responses in each category. Statistical models using questionnaire items frequently create dummy variables, collapsing the response categories into a dichotomous variable and, consequently, reducing the information content of the measure. As an alternative approach, we created scales by combining responses to related items. With a scale, random error is reduced and true score variability is increased. A scale simplifies reporting because the responses to several items are combined into a single measure. Scale scores allow the use of parametric statistics when the distribution of scale scores approximates a normal distribution.

We began with an examination of the teacher questionnaires in order to identify items indicative of the goals of the Statewide Systemic Initiatives Program. We used a model of systemic reform (Clune, 1998) to categorize the items, and then examined the responses to each selected item. Following this review, some selected items were eliminated because almost all respondents chose the same response option, usually the highest or lowest. Either there was an extremely high level of teacher agreement on these items, or the item was not sensitive to differences among teachers.

We then reviewed the individual items of the 1996 grade 8 teacher questionnaire and discussed the “best” answer, from the perspective of mathematics reform. Most response options ranged from a low of “Never” or “None” to a high of “Almost every day” or “A lot.” For most items, responses in the NAEP data set were coded from 1 to N, with N as the number of response options. In our analyses, we reversed the scales when necessary, so the highest value represented the most frequent occurrence. In discussions, project staff generally agreed that with successful Statewide Systemic Initiatives, the use of reform-related practices would increase, but traditional practices that were focused on mastering facts, concepts, and routine procedures would also have a major role. We had concerns about a simple scale implying that “more” of something was necessarily “better.” We explored scoring rubrics that assigned the greatest number of points to the response option that described a moderate frequency of occurrence. The alternative scales were evaluated using Cronbach’s coefficient alpha, a measure of internal consistency (Cronbach, 1951). None of the proposed scoring systems improved on the original 1 to N coding, with 1 indicating the lowest frequency and N the highest. While we continue to be concerned about the possibility of excessive use of some reform-related practices, this concern did not apply to most of the results from 1990, 1992, and 1996.

Through this extensive review and analysis of the State NAEP 1996 teacher questionnaire items, we identified six indicators of mathematics reform:

Relative Emphasis on Reasoning and Communication **–** how much reasoning and communication were addressed, relative to facts and procedures.
Mathematical Discourse **–** a scale of students’ opportunities to discuss, present, and write about mathematical ideas.
NCTM **Standards** – a single item that asked about teachers’ knowledge of the NCTM *Standards*.
Last Year’s Professional Development **–** a single item that asked how much time teachers spent in professional development in mathematics or mathematics education during the last year.
Reform-Related Topics Studied **–** a count of the number of reform-related topics teachers have studied out of the seven topics listed in the NAEP questionnaire.
Calculator Use **–** a scale of the extent to which students used calculators in the classroom and on tests.

The 1996 questionnaire was not the same as the 1992 and 1990 questionnaires. A number of items were added over the years, particularly items related to curricular reform. Wording of individual items was modified from one year to the next, and sometimes the number and labels of the response options changed also. Very few items were exactly the same across all three years.

Comparing SSI and Non-SSI States

In this chapter, the effect of the SSI program is examined by comparing the SSI state means on each indicator with the non-SSI state means. In this comparison, the state is the unit of analysis. SSI states are grouped together as replications receiving the treatment (e.g., the SSI program), and non-SSI states are grouped as replications not receiving the treatment. Unlike experimental research, however, the SSI treatment was not randomly assigned to the states. States participating in the SSI program had to submit a proposal, and then NSF selected the proposals to fund.

The State NAEP is designed to provide information about the state as a whole. The student is the sampling unit, and teachers' responses are matched with each of their students, defining one record in the data file. Each student has an associated weight, based on the sampling plan, and state means are computed using weighted values. In this chapter, the focus is on the state means and the variability among the means, rather than on the within-state variability.

With states as the unit of analysis, sample size is relatively small. In order to reject the null hypothesis, differences have to be fairly large. For comparisons between SSI and non-SSI states, we used an alpha level of .10, following the example of Grissmer (2000).

Cross-sectional comparisons. We examined the effect of the SSI program by comparing all SSI and non-SSI states in a given year. In this approach, all states that participated in the State NAEP that year were included in the comparison.

Longitudinal comparisons. In longitudinal comparisons, state means were compared across two or three different years. Longitudinal comparisons were limited to those states that participated in the State NAEP in consecutive years.

The analytic approach used to examine the effect of SSI on changes over time depended on the comparability of the measures. When the measures were the same across time, a repeated measures analysis of variance was used, with SSI as a between-subjects factor and time as a within-subjects factor. When they were similar, but not exactly the same, a two-step regression model was used. The prior year indicator was entered at Step 1, to assess the relationship among the two measures. At Step 2, SSI status was entered to assess the additional contribution of SSI status on the 1996 indicator. The expectation was that SSI status would be significantly related to the indicator in 1996, but not in 1992 or 1990.

Multiple linear regression models assume that the predictors are independent. Relationships among predictors raise issues about how to estimate the model parameters. In part, this issue is solved by the model specification. If the model must include all predictors, an analytic method is used that will divide the shared variance among the predictors.

Samples

Twenty-five states received funding through NSF's SSI program, and twenty-five did not. Some states had their funding discontinued early. In this study, they were included with the

SSI states, since they did receive some benefits from the SSI program. The SSI program had three cohort groups. The first began in 1991, the second in 1992, and the third in 1993.

Not all states participated in the State NAEP each of the three years. Analyses and conclusions about the effects of the SSI program are limited to those states that chose to participate. The State NAEP also included the jurisdictions of Guam, Puerto Rico, the Virgin Islands, Washington, DC, and Department of Defense Schools. In this study, only state data were used.

Yearly samples – 1990, 1992, and 1996. For each year of the State NAEP, comparisons were made between all participating SSI states and non-SSI states. These cross-sectional analyses used all of the data available in a given year. Table 6.1 presents the number and percentage of SSI and non-SSI states that participated each year at each grade. Appendix A lists the individual states, with their years of participation in the State NAEP, and indicates whether they were funded for the full five years.

Table 6.1
Number and Percent of SSI and Non-SSI States Participating in State NAEP Each Year

		SSI states *N* = 25		Non-SSI states *N* = 25	
Grade	Year	*N*	%	*N*	%
Grade 8					
	1990	20	80%	17	68%
	1992	22	88%	19	76%
	1996	22	88%	18	72%
Grade 4					
	1992	22	88%	19	76%
	1996	23	92%	20	80%

Two-Point Trend Sample, 1992-1996. The two-point trend sample included those states that participated in both 1992 and 1996. (See Table 6.2.) At grade 8, the two-point trend sample included 20 SSI states, or 80%, and 15 non-SSI states, or 60%. At grade 4, it included 21 SSI states (84%) and 16 non-SSI states (64%). These trend samples were used to evaluate change across the four years from 1992 to 1996. By 1996, states in the first cohort were completing their fifth year, and others were well into their fourth or their third year.

The 1992 measure provided a reference point for the 1996 measure, but it was not necessarily independent of a state's SSI status. Funding for the first cohort started in 1991, so by spring of 1992 some states had been funded for about a year. A potentially larger factor, though, was that some states had begun reform initiatives on their own in the late 1980s or early 1990s. Their experiences with statewide reforms may have positioned them to apply and to be selected for the Statewide Systemic Initiatives Program.

Table 6.2
States in the Two- and Three-Point Trend Samples at Grade 8 and the Two-Point Trend Sample at Grade 4

Grade 8		Grade 4	
SSI states	Non-SSI states	SSI States	Non-SSI States
Arkansas[a]	Alabama[a]	Arkansas	Alabama
California[a]	Arizona[a]	California	Arizona
Colorado[a]	Hawaii[a]	Colorado	Hawaii
Connecticut[a]	Indiana[a]	Connecticut	Indiana
Delaware[a]	Iowa[a]	Delaware	Iowa
Florida[a]	Maryland[a]	Florida	Maryland
Georgia[a]	Minnesota[a]	Georgia	Minnesota
Kentucky[a]	Mississippi	Kentucky	Mississippi
Louisiana[a]	Missouri	Louisiana	Missouri
Maine	North Dakota[a]	Maine	North Dakota
Massachusetts	Tennessee	Massachusetts	Pennsylvania
Michigan[a]	Utah	Michigan	Tennessee
Nebraska[a]	West Virginia[a]	Nebraska	Utah
New Mexico[a]	Wisconsin[a]	New Jersey	West Virginia
New York[a]	Wyoming[a]	New Mexico	Wisconsin
North Carolina[a]	New York	Wyoming	
Rhode Island[a]	North Carolina		
South Carolina	Rhode Island		
Texas[a]	South Carolina		
Virginia[a]	Texas		
	Virginia		

[a] Also in the three-point trend sample.

Three-Point Trend Sample, 1990, 1992, and 1996. This sample includes the 17 SSI states and the 11 non-SSI states that participated in the State NAEP all three years. The three-point trend sample is limited to grade 8, since the State NAEP was not administered in grade 4 in 1990. (See Table 6.2 for states in the three-point trend sample). The representativeness of the states in this sample is an important consideration. It includes 68% of the 25 SSI states and 44% of the non-SSI states. Generalizations from this self-selected sample to all SSI and non-SSI states must be made carefully, especially since more than half of the non-SSI states are not included.

Participation rates. The accuracy of the parameter estimates for each state depends on the implementation of NAEP's sampling plan. The National Center for Education Statistics (NCES) established participation rate standards for schools and students. (For the guidelines, see the Appendix of Chapter 3, Methodological Issues). The standards for school participation are discussed in the following paragraph from the *NAEP 1996 Mathematics Cross-State Data Compendium for the Grade 4 and Grade 8 Assessment*, Appendix A.

> NCES standards require weighted school participation rates before substitution of at least 85 percent to guard against potential bias due to school nonresponse. The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that declined to participate in the assessment. However, considerable technical consideration has been given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate the possibility of bias because of the nonparticipation of initially selected schools. Thus, for the weighted school participation rates that included substitute schools, the guideline was set at 90 percent. (p. 282)

The strata-specific participation rate guideline is discussed in the following paragraphs.

> The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, inadequate representation of an important segment of a jurisdiction's population is of concern, regardless of the overall participation rate. (p. 283)
>
> This guideline addresses the concern that if nonparticipating schools were concentrated within a particular class of schools, the potential for substantial bias remained, even though the overall level of school participation appeared to be satisfactory. Nonresponse adjustment cells for schools were formed within each jurisdiction, and the schools within each cell were similar in terms of minority enrollment, degree of urbanization, and/or median household income for public schools. . . . If more than 5 percent (weighted) of the sample schools (after substitution) were nonparticipants from a single adjustment cell, then the potential for nonresponse bias was too great. (p. 283)

States that failed to satisfy the participation rate guidelines are annotated in reports of the State NAEP results. That convention was used in this report also. Table 6.3 on the next page lists the states with annotated results. In this chapter, all analyses were done twice—first with the total sample and then with the subsample of states that met the NCES participaton rate guidelines. The second analysis was a check on the findings from the total sample. Confidence in the results was strengthened if the parameter estimates of the two analyses were similar. With the reduced sample size, results for the subsample may not be statistically significant.

Table 6.3
States Not Meeting the NCES Participation Rate Guidelines

	Grade 8			Grade 4	
	1990	1992	1996	1992	1996
SSI States		Maine Nebraska New Jersey New York	Arkansas Michigan Montana New York South Carolina Vermont	Delaware Maine Nebraska New Jersey New York	Arkansas Michigan Montana New Jersey New York South Carolina Vermont
Non-SSI States	Iowa	Alabama	Alaska Iowa Maryland Wisconsin	-	Alaska Iowa Nevada Pennsylvania

Relative Emphasis on Reasoning and Communication—I_{RC}

The Importance of Reasoning and Communication

Two of the five education goals for students identified by the National Council of Teachers of Mathematics in its first standards document, *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), are for all students to learn to communicate and to reason mathematically. These standards expressed a vision for classrooms based on the premise that what students learn depends to a great degree on how they have learned. The *Standards* stress the need to teach much of the same mathematics being currently taught, but to teach this mathematics with a different emphasis, including, for example, less focus on complex paper-and-pencil computations and memorizing rules and algorithms and more emphasis on development of critical thinking skills and on the understanding of numbers and operations. The Relative Emphasis on Reasoning and Communication indicator is computed to produce a statistic that represents a stress on reasoning and communication over addressing facts, concepts, procedures, and skills.

Since NCTM published the first *Standards* document, there has been a growing acceptance in the mathematics education field of the importance of having students reflect on their own learning and understanding. This requires students to express their understanding of mathematics and to reveal their thinking so that ideas can be discussed and clarified. Students do this by presenting their arguments in debates, discussing their solutions to problems, and making predictions about various phenomena (National Research Council, 2000). Such instructional engagement requires a classroom environment in which reasoning and communication are the norm. In an analysis of the reforms in thirteen Organization of Economic Co-operation and Development (OECD) countries, Black and Atkin (1996) report that most of the international innovations studied, including three in the United States, accepted the basic assumption of constructism, that is, pupils construct meanings for themselves. Black and Atkin observed that new methods succeeded to the extent that they engaged the thinking of students. Carpenter and Lehrer (1999) described in greater detail how students need to learn mathematics with understanding: develop relationships among mathematical ideas, extend and apply these ideas in new situations, reflect on and articulate their thinking, and make mathematical knowledge their own. A logical conclusion from the Third International Mathematics and Science Study (Stigler & Hiebert, 1997) is that classroom emphasis on demonstrating and practicing procedures and ideas has resulted in students learning simple computation procedures, terms, and definitions, the dominant approach to teaching grade 8 mathematics in US schools. For students to learn how to develop their own reasoning skills to solve new problems, engage in more challenging mathematical processes, and increase their capacity for reasoning and communication (Hiebert, 1999) requires the reform curriculum being advocated by NCTM.

1996

In 1996, questions 37-40 of the State NAEP teacher questionnaire were part of a section labeled "skills." The questions were:

> In this mathematics class, how often do you address each of the following?
> Learning mathematics facts and concepts
> Learning skills and procedures needed to solve routine problems
> Developing reasoning and analytical ability to solve unique problems
> Learning how to communicate ideas in mathematics effectively
>
> *Response options*
> A lot
> Some
> A little
> None

Responses were coded from 0 for "None" to 3 for "A lot." Individual variables were:
X_F - Learning facts and concepts (Facts)
X_P - Learning skills and procedures (Procedures)
X_R - Developing reasoning and analytical ability (Reasoning)
X_C - Learning how to communicate ideas in mathematics (Communication)

The first two are often considered to be "basic skills" and the other two "higher order skills." In mathematics reform, all four skills areas are included. Reasoning and communication are often emphasized, but mastery of basic skills in also important in reform curricula.

In order to incorporate all four skills areas into one indicator, responses to the four questions were combined in a relative, or ipsative, measure:

$$I_{RC} = (X_R + X_C)/ (X_F + X_P + X_R + X_C) * 100.$$

I_{RC} ranged from 0 to 100, with 50 indicating an equal balance between basic and higher order skills. A value greater than 50 indicated that the two higher order skills were addressed more than the two basic skills. A value less than 50 indicated that the two basic skills were addressed more than the two higher order skills.

Grade 8

Relative emphasis on reasoning and communication. The mean of $I_{RC(96)}$ for all SSI and non-SSI states in 1996 is presented in Table 6.4. The mean difference of 1.32 was statistically significant (t = 2.86, df = 38, p < .01), with SSI states averaging higher than non-

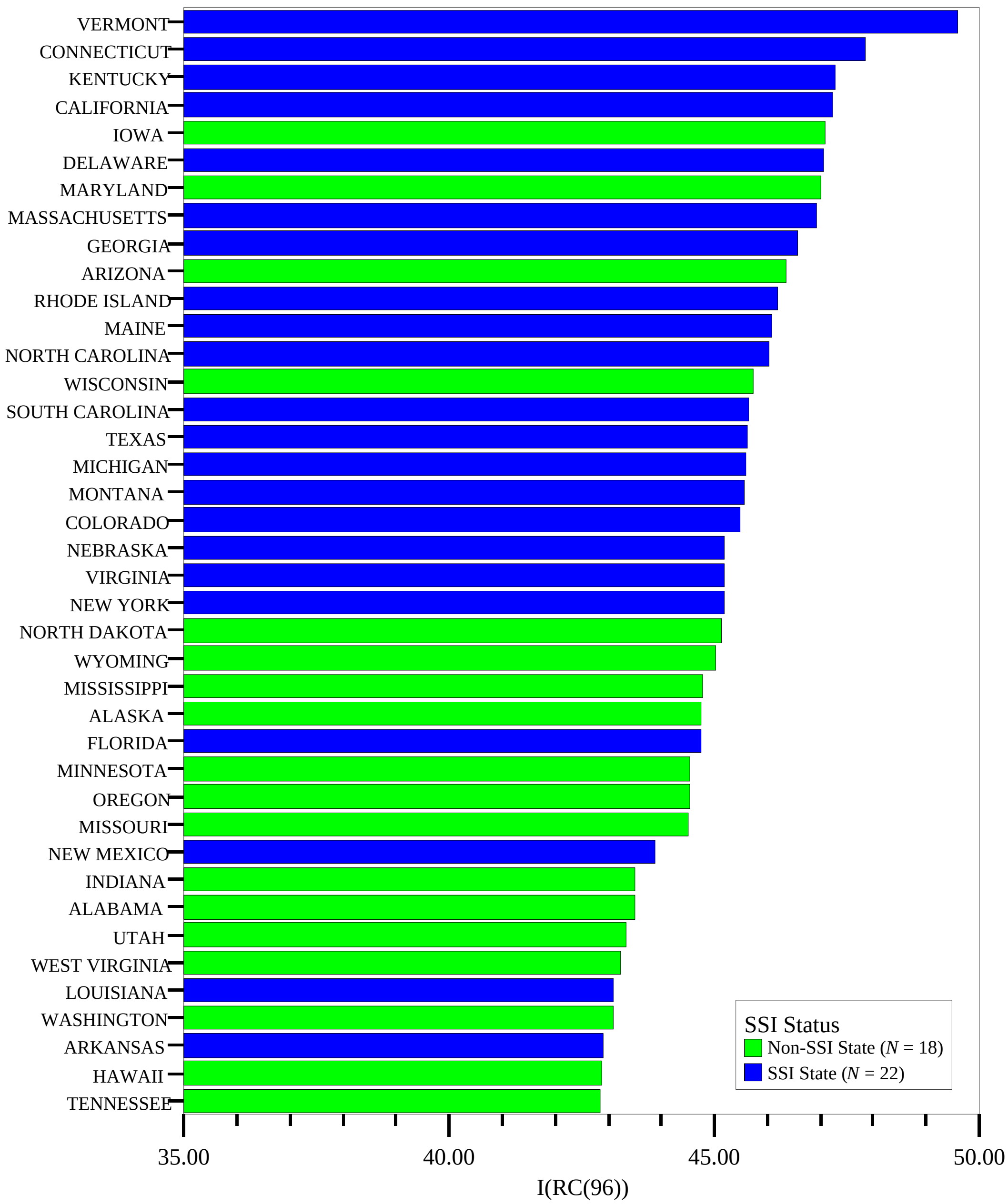
SSI states. The absolute value of the difference is not large, but it is about as large as the standard deviation among the non-SSI states. Comparisons for all samples and subsamples are reported in Table 6B.1 of the Appendix. SSI states averaged significantly higher on I_{RC} for all samples and subsamples except for the total threee-point trend sample, where the difference was not statistically significant.

Table 6.4
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96)}$, *Grade 8*

	SSI States *N* = 22	Non-SSI States *N* = 18
Mean	45.86	44.54
Standard Deviation	1.52	1.35

The individual state means on $I_{RC(96)}$ are presented in Figure 6.1, with states ranked from highest to lowest. Of the ten highest states, seven are SSI states: Vermont, Connecticut, Kentucky, California, Delaware, Massachusetts, and Georgia; of the ten lowest, only three are SSI states: Arkansas, Louisiana, and New Mexico. Individual state means are listed in Table 6B.2 of the Appendix.

Figure 6.1. State means on $I_{RC(96)}$ ordered from highest to lowest, grade 8.



Individual skill areas. Table 6.5 summarizes the results for the four questions that comprise $I_{RC(96)}$. Means for individual states are listed in Table 6B.2 of the Appendix.

Table 6.5
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, *1996*

Skill area	SSI States *N* = 22	Non-SSI States *N* = 18
X_F - Facts		
Mean	2.67	2.68
Standard Deviation	0.09	0.09
X_P - Procedures		
Mean	2.73	2.73
Standard Deviation	0.07	0.05
X_R - Reasoning		
Mean	2.39	2.30
Standard Deviation	0.08	0.05
X_C - Communication		
Mean	2.27	2.13
Standard Deviation	0.12	0.12

A multivariate analysis of variance was used to examine the effect of the state's SSI status on the four skill areas. The overall effect for SSI status was statistically significant ($F = 5.50$, $df = 1,35$, $p < .01$). Follow-up univariate tests showed that SSI states scored significantly higher than non-SSI states on the two higher order skills: X_R ($F = 18.74$, $df = 1,38$, $p < .01$) and X_C ($F = 13.78$, $df = 1,38$, $p < .01$). However, SSI states did not differ from non-SSI states for the two basic skills: X_F and X_P.

The results for the subsample of states that followed the participation rate guidelines were the same as for the full sample. The effect for SSI status was statistically significant ($F = 5.99$, $df = 1,25$, $p < .01$). SSI states had a higher average than non-SSI states for the two higher order skills, X_R and X_C, but there was no significant difference on the basic skills, X_F, and X_C.

Intercorrelations of skill areas. Relationships among the four skill areas were examined by computing their intercorrelations. The results are reported in Table 6.6. Results for all 40 states that participated in the 1996 State NAEP, as well as results for the subsample of 30 states that followed the participation rate guidelines, are included.

In 1996, the two basic skills items were strongly related, as were the two higher order skills items. Other correlations were near 0 and several had a negative sign. The findings for the subsample were comparable to those of the total sample.

Table 6.6
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 8, 1996

	X_F	X_P	X_R
Total sample, *N* = 40			
X_P	.77*		
X_R	-.13	-.11	
X_C	-.01	-.02	.72*
Subsample, *N* = 30			
X_P	.80*		
X_R	-.23	-.10	
X_C	.03	.02	.68*

*$p < .01$

Grade 4

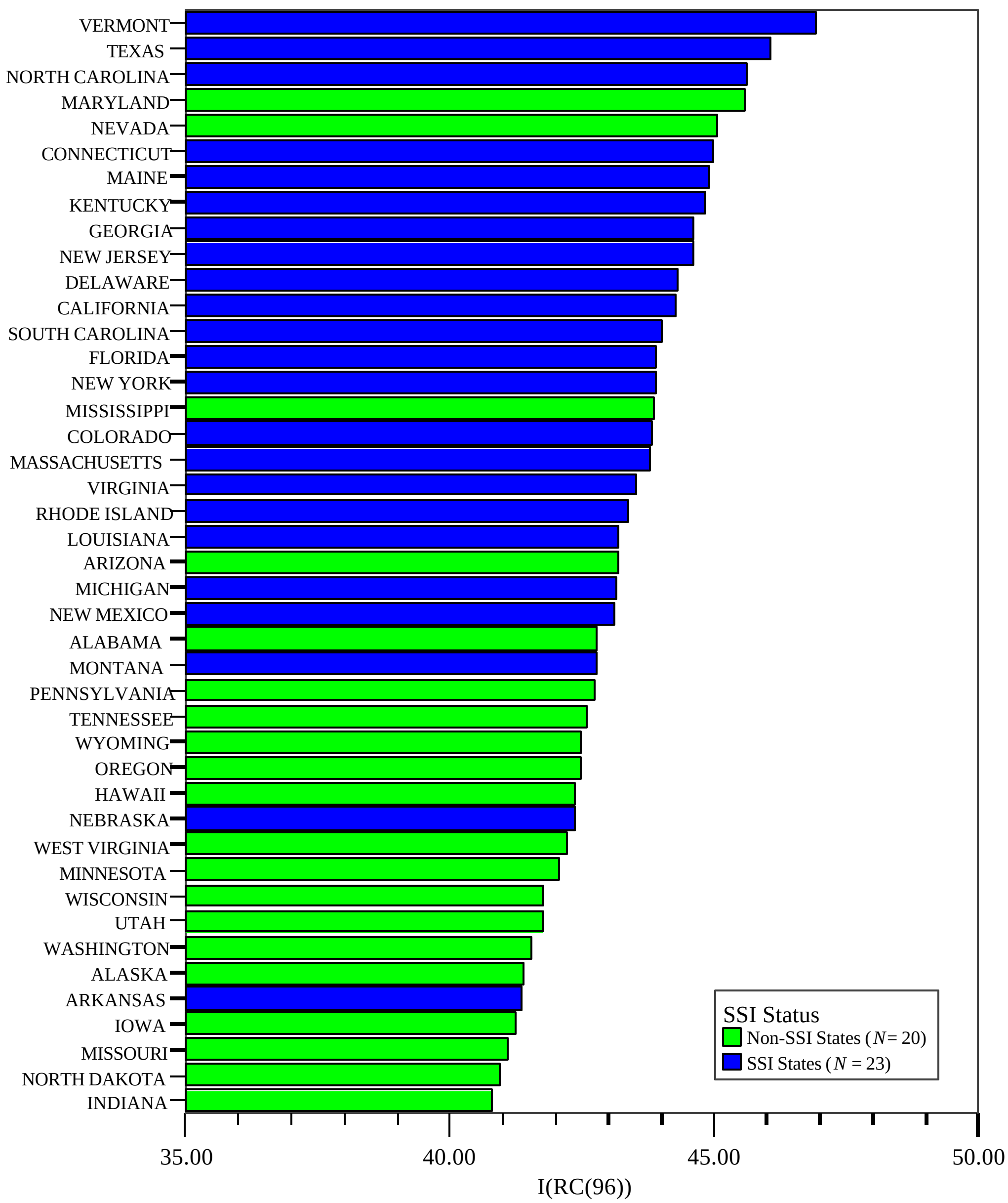
Relative emphasis on reasoning and communication. The grade 4 mean of $I_{RC(96)}$ for all SSI and non-SSI states is presented in Table 6.7. The mean difference of 2.02 was statistically significant ($t = 4.35$, $df = 41$, $p < .01$), with SSI states averaging higher than non-SSI states. While the absolute value of the difference was not large, it was larger than the standard deviation within each group. Comparisons for all samples and subsamples are reported in Table 6B.3 of the Appendix. SSI states averaged significantly higher on I_{RC} for all samples and subsamples.

Table 6.7
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96)}$, Grade 4

	SSI States *N* = 23	Non-SSI States *N* = 20
Mean	44.06	42.40
Standard Deviation	1.22	1.27

The individual state means on $I_{RC(96)}$ are presented in Figure 6.2 with states ranked from highest to lowest. Of the ten highest states, eight are SSI states: Vermont, Texas, North Carolina, Connecticut, Maine, Kentucky, Georgia, and New Jersey; of the ten lowest, only one is an SSI state: Arkansas. Individual state means are listed in Table 6B.3 of the Appendix.

Figure 6.2. State means on $I_{RC(96)}$ ordered from highest to lowest, grade 4.



Individual skill areas**.** Table 6.8 summarizes the results for the four questions that comprise $I_{RC(96)}$. Means for individual states at grade 4 are listed in Table 6B.3 of the Appendix.

Table 6.8
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 4, 1996

Skill area	SSI States *N* = 23	Non-SSI States *N* = 20
X_F - Facts		
Mean	2.91	2.92
Standard Deviation	0.04	0.04
X_P - Procedures		
Mean	2.87	2.88
Standard Deviation	0.04	0.05
X_R - Reasoning		
Mean	2.42	2.33
Standard Deviation	0.10	0.07
X_C - Communication		
Mean	2.27	2.12
Standard Deviation	0.12	0.14

In the multivariate analysis of variance, the overall effect for SSI status was statistically significant ($F = 4.27$, $df = 1,38$, $p < .01$). Follow-up univariate tests showed that SSI states scored significantly higher than non-SSI states on the two higher order skills: X_R ($F = 13.30$, $df = 1,41$, $p < .01$) and X_C ($F = 15.33$, $df = 1,41$, $p < .01$). SSI states did not differ from non-SSI states for the two basic skills: X_F and X_P.

The results for the subsample of states that followed the participation rate guidelines were the same as for the full sample. The effect for SSI status was statistically significant ($F = 4.82$, $df = 1,27$, $p < .01$). SSI states had a higher average than non-SSI states for the two higher order skills, X_R ($F = 14.65$, $df = 1,30$, $p < .01$) and X_C, ($F = 15.56$, $df = 1,30$, $p < .01$), but there was no significant difference for the basic skills, X_F, and X_P.

Intercorrelations of skill areas. Correlations of the four skill areas at grade 4 in 1996 are reported in Table 6.9. Results for all 43 states that participated in the 1996 State NAEP, as well as results for the subsample of 32 states that followed the NCES participation rate guidelines, are included.

At grade 4, as was found at grade 8, the two basic skills items were strongly related and so were the two higher order skills items. Other correlations were near 0. The findings for the subsample were comparable to those of the total sample.

Table 6.9
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 4, 1996

	X_F	X_P	X_R
Total sample, *N* = 43			
X_P	.66*		
X_R	-.19	.15	
X_C	-.13	.07	.82*
Subsample, *N* = 32			
X_P	.62*		
X_R	-.18	.19	
X_C	-.11	.15	.82*

*$p < .01$

1992

In 1992, questions 23-26 of the teacher questionnaire asked how much the teacher emphasized the four skills areas:

Think about your plans for this mathematics class for the entire year. How much emphasis did you or will you give to each of the following?

- Learning mathematics facts and concepts
- Learning skills and procedures needed to solve routine problems
- Developing reasoning and analytical ability to solve unique problems
- Learning how to communicate ideas in mathematics effectively

Response options

- Heavy emphasis
- Moderate emphasis
- Little or no emphasis

The 1992 questions differed from those in 1996 in several important ways. In 1992, teachers were explicitly asked to think about the entire year. The 1992 question asked, “How much emphasis did you or will you give to each of the skills?” while the 1996 question asked, “How often do you address each . . . ?” With the wording change, the response options were also different from those in 1996. In 1992, there were three options: “Heavy,” “Moderate,” and “Little or no emphasis”; in 1996, there were four: “A lot,” “Some,” “A little,” and “None.”

Computation of $I_{RC(92)}$ used the same formula as in 1996, providing an ipsative measure of the emphasis on the two higher order skills compared to the emphasis on the two basic skills. In 1992, the response options were coded from 0 to 2, with "Little or no emphasis" equal to 0 and "Heavy emphasis" equal to 2.

Grade 8

Relative emphasis on reasoning and communication. In 1992, the 22 SSI states averaged slightly higher on $I_{RC(92)}$ than the 19 non-SSI states, as shown in Table 6.10 This difference was not statistically significant (t = 0.97, df = 40, p = .33). For the subsample that followed the NCES participation rate guidelines, SSI status was not significant either. (See Table 6B.4 of the Appendix.)

Table 6.10
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(92)}$, Grade 8

	SSI States N = 22	Non-SSI States N = 19
Mean	43.50	42.86
Standard Deviation	2.03	2.03

Individual skill areas**.** Table 6.11 summarizes the results for each of the four individual skill areas by state SSI status. In the MANOVA, the effect for SSI was not statistically significant (F = 0.77, df = 4,36). The SSI effect was not significant in the subsample either (F = 1.64, df = 4,31, p = .19). State means of the four individual skill areas are shown in Table 6B.5 of the Appendix.

Table 6.11
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, 1992

Skill area	SSI States N = 22	Non-SSI States N = 19
X_F - Facts		
Mean	1.66	1.67
Standard Deviation	0.07	0.10
X_P - Procedures		
Mean	1.74	1.73
Standard Deviation	0.05	0.08
X_R - Reasoning		
Mean	1.42	1.38
Standard Deviation	0.08	0.07
X_C - Communication		
Mean	1.33	1.28
Standard Deviation	0.11	0.08

Intercorrelations of skills areas (Table 6.12). In 1992 as in 1996, the highest correlations were between the two basic skills, X_F and X_P, and the two higher order skills, X_R and X_C. In 1996, no other correlations were significant; but in 1992, the correlation between X_C, Communication, and X_P, Procedures, was also significant, though not as high as the other correlations.

Table 6.12
Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 8, 1992

	X_F	X_P	X_R
Total sample, *N* = 41			
X_P	.80**		
X_R	-.04	.15	
X_C	.29	.32*	.67**
Subsample, *N* = 36			
X_P	.81**		
X_R	.00	.20	
X_C	.31	.37*	.64**

*$p < .05$; **$p < .01$

Grade 4

Relative emphasis on reasoning and communication. At grade 4, the 22 SSI states averaged slightly higher on $I_{RC(92)}$ than the 19 non-SSI states, as shown in Table 6.13. This difference was statistically significant ($t = 1.79$, $df = 39$, $p < .10$). For the subsample that followed the NCES participation rate guidelines, SSI status was significant also. (See Table 6B.4 in the Appendix).

Table 6.13
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(92)}$, Grade 4

	SSI States *N* = 22	Non-SSI States *N* = 19
Mean	40.08	39.09
Standard Deviation	1.93	1.61

Individual skill areas**.** Table 6.14 summarizes the results for each of the four individual skill areas by state SSI status. In the MANOVA, the effect for the SSI was statistically significant ($F = 3.98$, $df = 4,36$, $p < .01$). The SSI effect was significant in the subsample also ($F = 3.30$, $df = 4,31$, $p < .05$). Table 6B.6 in the Appendix presents the individual state means for the four skills areas.

Table 6.14

Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 4, 1992

Skill area	SSI States $N = 22$	Non-SSI States $N = 19$
X_F - Facts		
Mean	1.93	1.94
Standard Deviation	0.04	0.03
X_P - Procedures		
Mean	1.92	1.90
Standard Deviation	0.03	0.02
X_R - Reasoning		
Mean	1.44	1.41
Standard Deviation	0.09	0.08
X_C - Communication		
Mean	1.32	1.24
Standard Deviation	0.11	0.09

Follow-up univariate tests showed that SSI states scored significantly higher than non-SSI states on one skill: X_C ($F = 5.39$, $df = 1,39$, $p < .05$). SSI states did not differ significantly from non-SSI states in the other three skill areas. Results for the subsample of states that followed the participation rate guidelines matched those for the full sample. The effect for SSI status was statistically significant ($F = 3.30$, $df = 1,31$, $p < .05$), and SSI states had a higher average than non-SSI states for X_C ($F = 5.93$, $df = 1,34$, $p < .05$).

Intercorrelations of skill areas (Table 6.15). In 1992 as in 1996, the highest correlations were between the two basic skills, X_F and X_P, and the two higher order skills, X_R and X_C. In 1996, no other correlations were significant, but in 1992, the correlation between X_C, Communication, and X_P, Procedures, was also significant, though not as high as the other correlations. In addition, at grade 4, X_F, Facts, was negatively related to X_R, Reasoning.

Table 6.15

Intercorrelations of State Means on the Four Mathematics Skills Areas, Grade 4, 1992

		X_F	X_P	X_R
Total sample, $N = 41$				
	X_P	.60**		
	X_R	-.42**	.17	
	X_C	-.01	.49**	.68**
Subsample, $N = 36$				
	X_P	.57**		
	X_R	-.48**	.13	
	X_C	-.13	.40*	.69**

*$p < .05$, **$p < .01$

1990

The wording of the 1990 questions was very similar to the wording in 1992, and the labeling of the response options was almost identical, except that the 1990 questionnaire had four response options. Questions 26 to 29 on the grade 8 teacher questionnaire were:

> Think about your plans for this mathematics class for the entire year. How much emphasis will you give each of the following?
>
> Learning mathematics facts and concepts
> Learning skills and procedures needed to solve routine problems
> Developing reasoning and analytical ability to solve unique problems
> Learning how to communicate ideas in mathematics effectively
>
> *Response options*
> Heavy emphasis
> Moderate emphasis
> Little emphasis
> None

Computation of $I_{RC(90)}$ was the same as in previous years, with response options coded from 0 for 'None" to 3 for "Heavy emphasis."

Grade 8

Relative emphasis on reasoning and communication. In 1990, the mean of the SSI states on $I_{RC(90)}$ was a bit higher than the mean of the non-SSI states, but this difference was not statistically significant (t = 1.08, df = 36, p = .29). The comparison for the subsample also was not statistically significant (t = 1.01, df = 34, p = .32). Comparisons for all samples and subsamples are shown in Table 6B.7 in the Appendix.

Table 6.16
Mean and Standard Deviation for SSI and Non-SSI States on $I_{RC(90)}$, Grade 8

	SSI States N = 20	Non-SSI States N = 17
Mean	46.09	45.68
Standard Deviation	1.12	1.16

Individual items**.** Table 6.19 summarizes the 1990 results by state SSI status. In 1990, the effect for SSI was not statistically significant in either the total sample (F = 2.13, df =

4,32, p = .10) or the subsample (F = 1.91, df = 4,31, p = .13). Individual state means are in Table 6B.8 of the Appendix.

Table 6.17
Mean and Standard Deviation of SSI and Non-SSI States in Each Skill Area, Grade 8, 1990

Skill area	SSI States N = 20	Non-SSI States N = 17
X_F - Facts		
Mean	2.52	2.48
Standard Deviation	0.07	0.08
X_P - Procedures		
Mean	2.62	2.63
Standard Deviation	0.06	0.06
X_R - Reasoning		
Mean	2.30	2.26
Standard Deviation	0.07	0.08
X_C - Communication		
Mean	2.23	2.16
Standard Deviation	0.09	0.10

Intercorrelations of skill areas (Table 6.18). As in 1992 and 1996, the highest correlations were for the pairs of skills: the two basic skills, X_F and X_P, and the two higher order skills, X_R and X_C. As in grade 8 in 1992 and 1996, X_R, was not significantly related to either of the basic skills. Like 1992, X_C, Communication, was moderately related to X_P, Procedures. In addition, in 1990 X_C was related to X_F, Facts and Concepts.

Table 6.18
Intercorrelations of State Means for Emphasis Given to the Four Skill Areas, Grade 8, 1990

	X_F	X_P	X_R
Total sample, N = 37			
X_P	.73**		
X_R	.10	.09	
X_C	.39*	.35*	.74**
Subsample, N = 36			
X_P	.72**		
X_R	.08	.06	
X_C	.38*	.34*	.73**

*p < .05; **p < .01

Change from 1992 to 1996

A direct comparison of the values of I_{RC} across 1990, 1992, and 1996 is not meaningful, because the number of response options and the labels for the options are not consistent across the three years. Several rescaling approaches were explored, but none seemed satisfactory. Since the measures are conceptually very similar but the scale is not the same, a hierarchical regression model was used to examine changes over time as a function of SSI status.

Grade 8

Two-point trend sample. Results of the two-step linear regression model predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI status are summarized in Table 6.19.

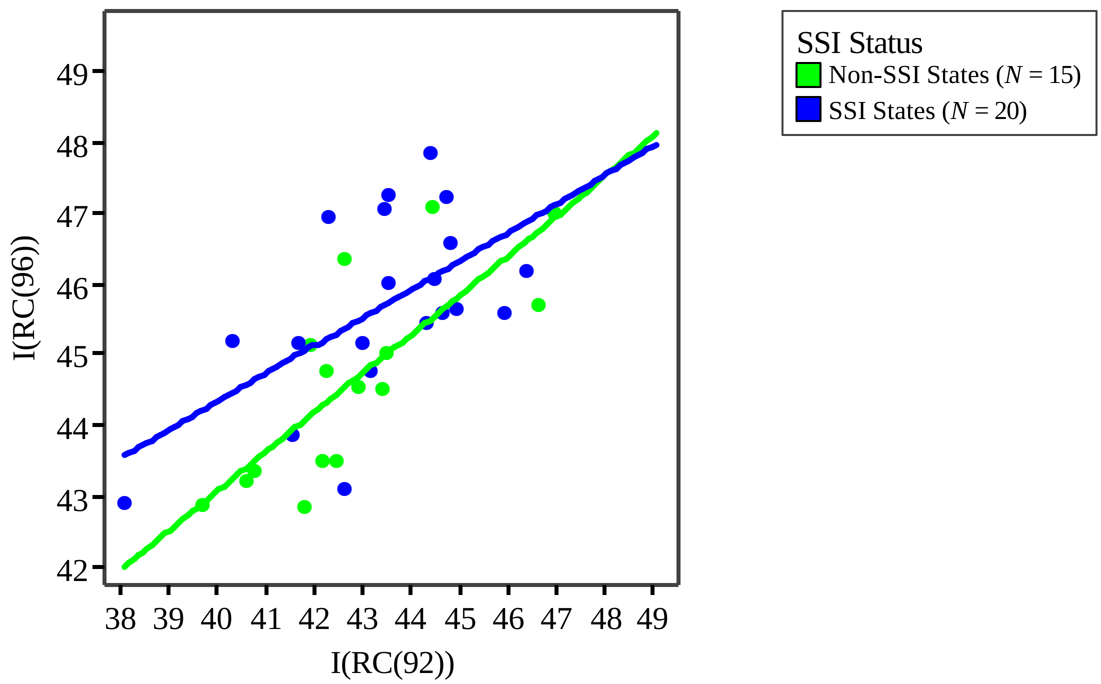
Table 6.19
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{RC(92)}$	0.50	0.10	.67	.45	27.30**		
Step 2							
$I_{RC(92)}$	0.47	0.09	.63				
SSI status	0.78	0.36	.27	.52	17.61**	.07	4.79*

**p* < .05, ***p* < .01

The table shows that $I_{RC(92)}$ was significantly related to $I_{RC(96)}$, and that SSI status also contributed to the prediction of $I_{RC(96)}$. In other words,, the SSI states averaged higher on $I_{RC(96)}$ than the non-SSI states, given $I_{RC(92)}$. The correlation between SSI status and $I_{RC(92)}$ was small ($r = .15$, $p = .19$). The scatterplot for the regression model is shown in Figure 6.3. Regression lines were fitted separately for the SSI and non-SSI states. The graph shows that the SSI effect was largest for those states that were relatively low on $I_{RC(92)}$.

Figure 6.3. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the two-point trend sample, grade 8.



Subsample. For the subsample, the effect for $I_{RC(92)}$ was significant, but SSI status did not add to the prediction of $I_{RC(96)}$. The results are presented in Table 6.20.

Table 6.20
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 8, for the Subsample that Followed the NCES Participation Rate Guidelines

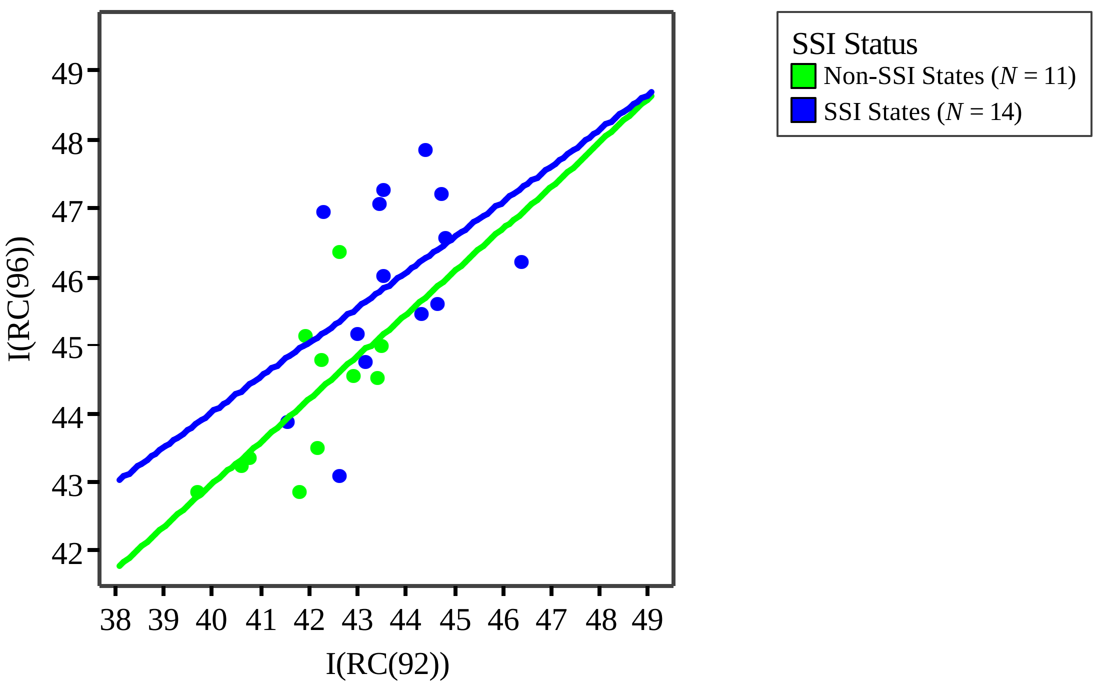
	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{RC(92)}$	0.71	0.15	.70	.49	22.13**		
Step 2							
$I_{RC(92)}$	0.56	0.19	.55				
SSI status	0.75	0.55	.25	.53	12.41**	.04	1.86

*$p < .05$, **$p < .01$

In this subsample, the strong relationship between the predictors may be getting in the way of detecting the effect for SSI. The correlation between SSI and $I_{RC(92)}$ in the subsample is .60 ($p < .01$).

Figure 6.4 presents the scatterplot for the model with the 14 SSI states and 11 non-SSI states in the subsample. The shape is comparable to that for the full model, and the estimate of *b* for the two models is similar. However, with the strong relationship between the two predictors, it is not possible to determine the unique contribution of SSI status to $I_{RC(96)}$.

Figure 6.4. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed the NCES participation rate guidelines, grade 8.



Grade 4

Two-point trend sample. Results of the two-step linear regression model predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI status are summarized in Table 6.21.

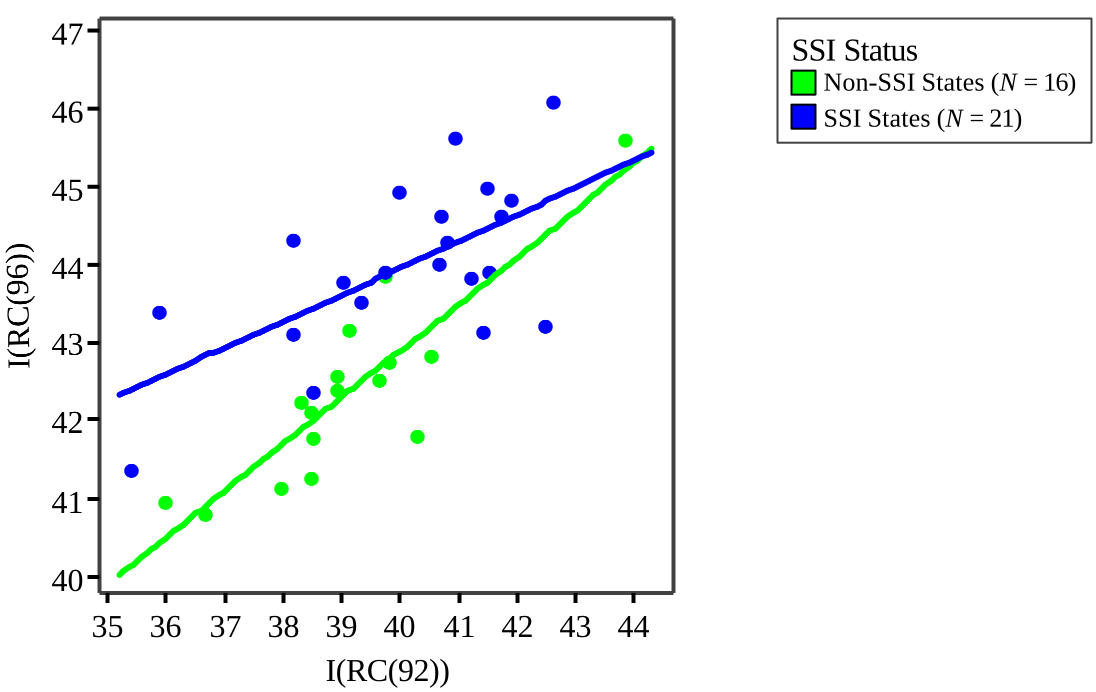
Table 6.21
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 4

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{RC(92)}$	0.52	0.08	.72	.52	38.15*		
Step 2							
$I_{RC(92)}$	0.44	0.07	.61				
SSI status	1.21	0.27	.44	.70	39.69*	.18	20.25*

*$p < .01$

At grade 4, $I_{RC(92)}$ was significantly related to $I_{RC(96)}$, and SSI status also contributed to the prediction of $I_{RC(96)}$. The relationship between SSI status and $I_{RC(92)}$ was moderate ($r = .26$, $p < .10$). The scatterplot for the regression model is shown in Figure 6.5. Regression lines were fitted separately for SSI and non-SSI states. As was found at grade 8, the SSI effect seems largest for those states that were relatively low on $I_{RC(92)}$.

Figure 6.5. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the two-point trend sample, Grade 4.



Subsample. The results for the subsample of states that followed the NCES participation rate guidelines in both 1992 and 1996 are presented in Table 6.22.

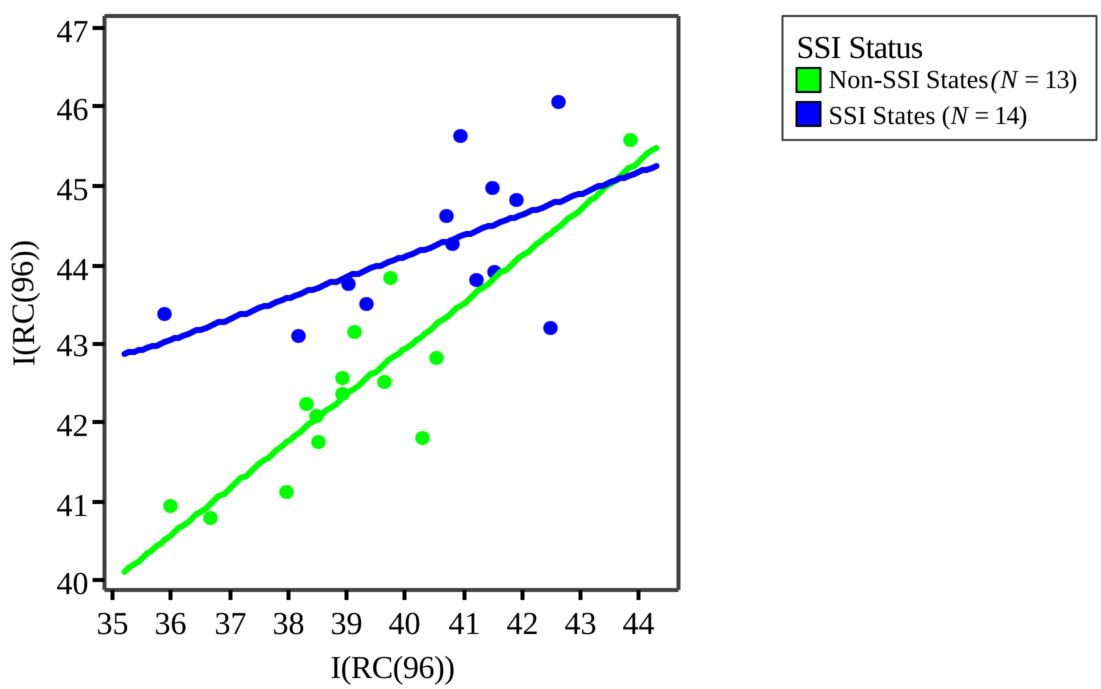
Table 6.22
Predicting $I_{RC(96)}$ from $I_{RC(92)}$ and SSI Status, Grade 4, for the Subsample that Followed the NCES Participation Rate Guidelines

	B	SE B	β	R^2	F	ΔR^2	F_{Δ}
Step 1							
$I_{RC(92)}$	0.54	0.10	.75	.56	32.38*		
Step 2							
$I_{RC(92)}$	0.43	0.08	.59				
SSI status	1.25	0.32	.44	.74	33.21*	.17	15.39*

*$p < .05$, **$p < .01$

In the grade 4 subsample, both $I_{RC(92)}$ and SSI status contributed to the prediction of $I_{RC(96)}$. The relationship between $I_{RC(92)}$ and SSI status was .36 ($p < .05$), a bit higher than the relationship in the full sample, but still moderate. The scatterplot is shown in Figure 6.6.

Figure 6.6. Scatterplot of $I_{RC(92)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed the NCES participation rate guidelines, Grade 4.



Change from 1990 to 1996

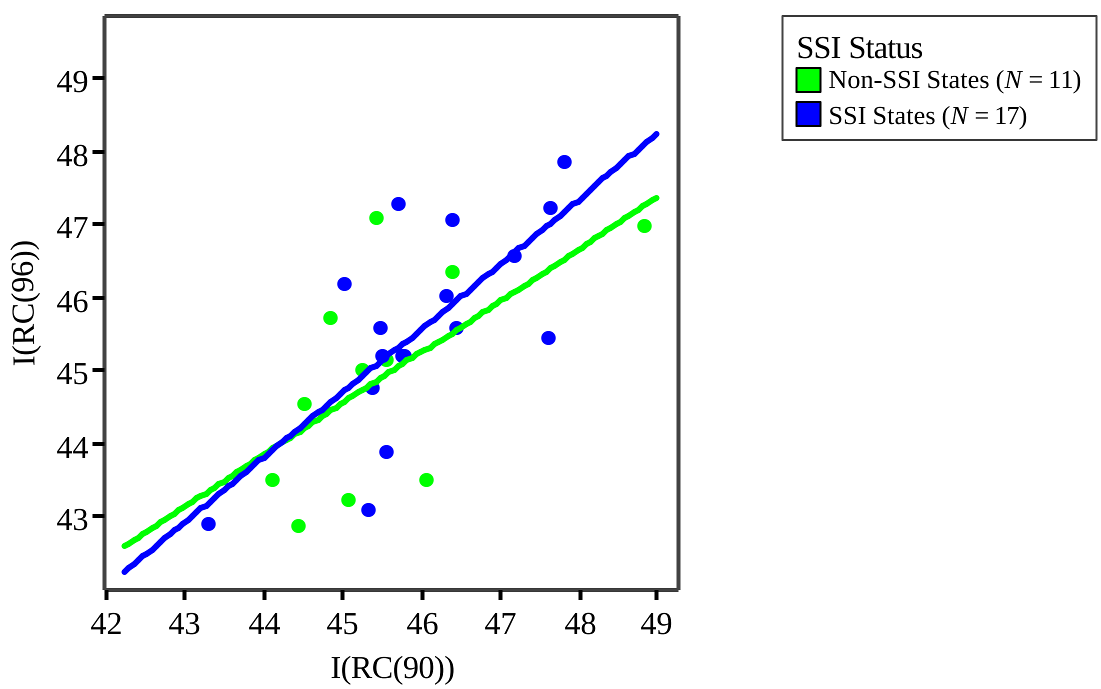
Three-point trend sample. Twenty-eight states are in the three-point trend sample. The results of the linear regression predicting $I_{RC(96)}$ from $I_{RC(90)}$ and SSI status are presented in Table 6.23. In this sample, $I_{RC(90)}$ was related to $I_{RC(96)}$, and SSI status did not add significantly to the prediction. SSI status was not significantly related to $I_{RC(90)}$ (r = .21, p = .14). Figure 6.7 presents the scatterplot for the total three-point trend sample.

Table 6.23
Predicting $I_{RC(96)}$ from $I_{RC(90)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{RC(90)}$	0.83	0.18	.68	.46	21.98*		
Step 2							
$I_{RC(90)}$	0.80	0.18	.66				
SSI status	0.28	0.44	.09	.47	10.93*	.01	0.39

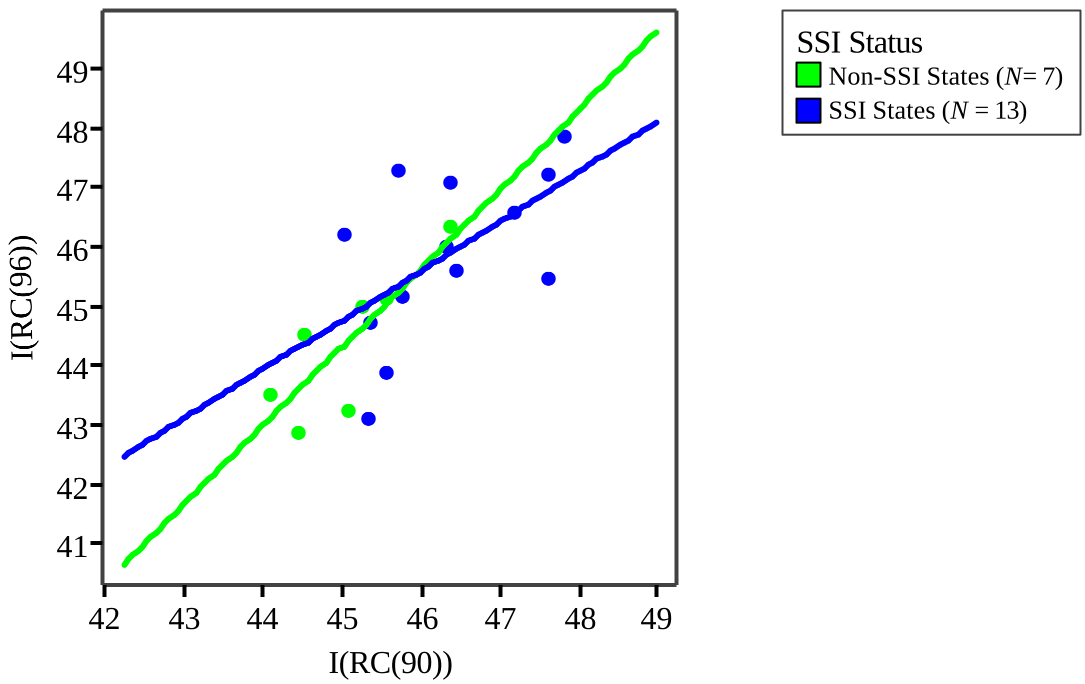
*p < .01

Figure 6.7. Scatterplot of $I_{RC(90)}$ and $I_{RC(96)}$ for all SSI and non-SSI states in the three-point trend sample, grade 8.



Subsample. Twenty states from the three-point trend sample satisfied the NCES participation rate guidelines in all three years. The results of the linear regression model for the subsample are presented in Table 6.23. The results for the subsample are comparable to those of the total sample: the effect for $I_{RC(90)}$ was significant, but the effect for SSI status was not. The relationship between SSI status and $I_{RC(90)}$ was fairly high in this subsample (r = .58, $p < .01$); in other words, states participating in the SSI program were likely to be higher on $I_{RC(90)}$ than the non-SSI states. Given this confounding, the independent effect of SSI cannot be evaluated. Figure 6.8 presents the scatterplot for the compliant subsample.

Figure 6.8. Scatterplot of $I_{RC(90)}$ and $I_{RC(96)}$ for SSI and non-SSI states that followed the NCES participation rate guidelines, Grade 8.



Summary of Results

1996

Grade 8 and Grade 4

In both grades, SSI states averaged significantly higher than non-SSI states on $I_{RC(96)}$, the indicator of the relative emphasis on reasoning and communication.

SSI states also averaged significantly higher than the non-SSI states on the two higher order skills: 1) reasoning and analytic ability, and 2) communicating mathematical ideas. The SSI states did not differ from the non-SSI states on the two basic skills: 1) facts and concepts, and, 2) skills and procedures.

In 1996, the two basic skills items were strongly related to each other, and the two higher order skills were also strongly related, but other correlations were near 0.

1992

Grade 8

SSI states averaged a bit higher than the non-SSI states on $I_{RC(92)}$, but the difference was not statistically significant. SSI states were not significantly different from the non-SSI states on any of the four individual items that were part of $I_{RC(92)}$.

Intercorrelations of skills areas scored highest between the two basic skills and the two higher order skills, consistent with the results in 1996. Unlike 1996, though, X_C was moderately related to X_P.

Grade 4

At grade 4, the SSI states were significantly higher than the non-SSI states on I_{RC}, as well as on one skill area, X_C.

Intercorrelations were highest for the skill pairs: the two basic skills and the two higher order skills. As with grade 8, X_C, Communication, was related to X_P, Procedures. In addition, at grade 4, X_F, Facts, was negatively related to X_R, Reasoning.

1990

Grade 8

In 1990 at grade 8, the difference between SSI and non-SSI states on $I_{RC(90)}$ was small and not statistically significant.

Intercorrelations among the four skills areas again found the highest correlations between item pairs: the two basic skills and the two higher order skills. In 1990, X_C, Communication, was also significantly related to both basic skills, X_F, Facts, and X_P, Procedures.

Change Across Time

Two-point trend: 1992-1996

Grade 8

Both $I_{RC(92)}$ and SSI status were significant predictors of $I_{RC(96)}$ for the total sample, but only $I_{RC(92)}$ was significant for the subsample. In the subsample, the two predictors were fairly strongly related.

Grade 4

Both $I_{RC(92)}$ and SSI status were significant predictors of $I_{RC(96)}$ for the total sample, as well as for the subsample.

Three-point trend: 1990-1996

Grade 8

$I_{RC(90)}$ was significantly related to $I_{RC(96)}$, but SSI status was not, for both the total sample and the subsample. The two predictors, SSI status and $I_{RC(90)}$, were not significantly related to each other in the full sample, but were in the subsample.

Students' Opportunities for Mathematical Discourse—I_{MD}

The Importance of Mathematical Discourse

Discourse in the mathematics classroom refers to the ways of representing mathematical ideas and to thinking, talking, agreeing, and disagreeing about them (NCTM, 1991). It involves the way ideas are exchanged and what the ideas entail. In 1991, NCTM took the position in its *Professional Standards for Teaching Mathematics* that discourse in the mathematics classroom should be based on mathematical reasoning and evidence. The *Professional Standards* claim that classroom discourse founded on mathematical evidence will lead students to develop the ability to formulate problems, to explore, conjecture, and reason logically, all important goals for school mathematics as expressed in the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). Verbalization is important if students are to acquire higher order thinking skills. Such thinking "requires students to manipulate information and ideas in ways that transform their meaning and implications" (Newmann, Secada, & Wehlage, 1995, p. 86). Students manipulate information and test ideas when they have to synthesize from facts and ideas, generalize, or arrive at some conclusion or interpretation.

A classroom environment conducive to discourse and involvement in higher order thinking will encourage students to interact with other students and teachers in extended conversations and in deeper thought processes. The teacher's role is to "translate what is being said into mathematical discourse to help frame discussion, to pose questions, to suggest real-life connections, to probe arguments, and to ask for evidence" (Adler, 1999, p. 51). The objective for teachers is to structure activities that will reveal students' thinking and actively inquire into students' thinking (Bransford, Brown, & Cocking, 2000). Students are able to amend, refine, and discuss ideas when they are engaged in small-group work, collaborative problem solving, and reflective abstract thinking (Shafer, 2001; Wood, 1996; Gravemeijer, 1994; Yackel, Cobb, & Wood, 1991). Through these activities students are more likely to interpret problem situations, express their thinking, react and resolve conflicting points of view, and develop a deeper understanding of mathematics.

1996

The 1996 teacher questionnaires of the State NAEP included several items about student discussion or writing about mathematics.

How often do the students in this class do each of the following?
- Solve mathematics problems in small groups or with a partner
- Write a few sentences about how to solve a mathematics problem
- Talk to the class about their mathematics work
- Write reports or do mathematics projects
- Discuss solutions to mathematics problems with other students
- Work and discuss mathematics problems that reflect real life situations

Response options
- Almost every day
- Once or twice a week
- Once or twice a month
- Never or hardly ever

How much time do the students in this class spend each week working on mathematics with a partner or in a small group?

Response options
- None
- Less than ½ hour
- ½-1 hour
- More than 1 hour

How often do you use each of the following to assess student progress in mathematics?
- Short (e.g., a phrase or sentence) or long (e.g., several sentences or paragraphs) written responses
- Individual or group projects or presentations

Response options
- Once or twice a week
- Once or twice a month
- Once or twice a year
- Never or hardly ever

These items describe instructional practices related to students learning to discuss mathematics and to explain their mathematical reasoning. Items for the scale were selected on the basis of a review of all the questionnaire items by project staff. After the item selection was finalized, project staff considered the scale’s designation. In early conversations, communicating about mathematical ideas seemed to be an organizing principle. Six of the nine items refer to talking, discussing, or working with others, and the other three items refer to writing. We decided to call this a scale of mathematical discourse, since the items represented a variety of opportunities for talking and writing about mathematics.

Response options were scaled from least to most frequent; the option representing the lowest frequency (e.g., Never or hardly ever) equaled 1 and the option representing the highest frequency (e.g., Almost every day) equaled 4. Responses to the nine individual items were added together, creating a scale with a range from 9 to 36. For grade 8, internal consistency of the scale was .76 and at grade 4 it was .79. State means for the mathematical discourse indicator, $I_{MD(96)}$, were computed using the weights provided in the State NAEP database.

Grade 8

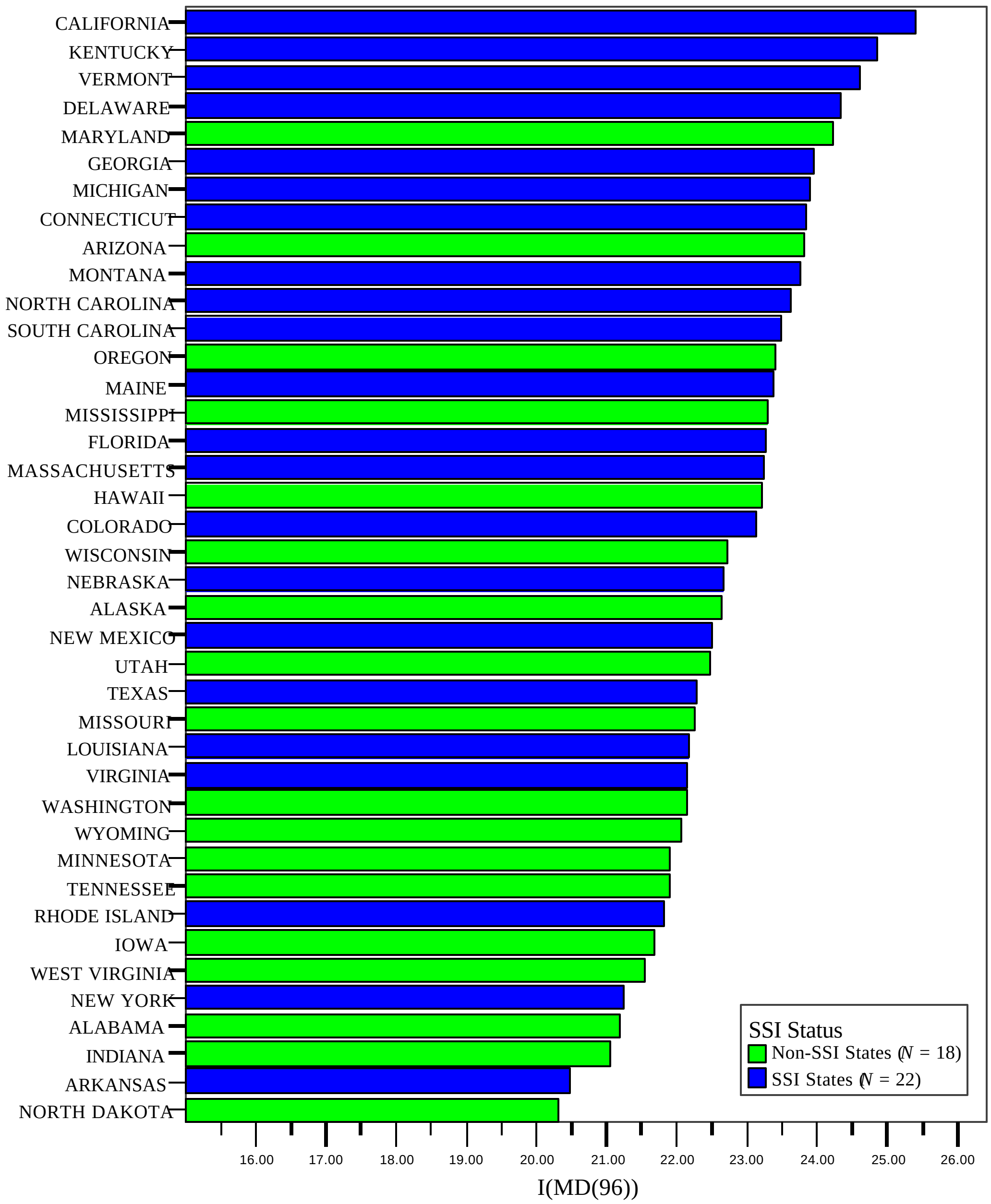
Student opportunities for mathematical discourse. Table 6.24 presents the mean and standard deviation of the state means on $I_{MD(96)}$, the mathematical discourse indicator, for the 22 SSI states and 18 non-SSI states participating in the 1996 State NAEP. The mean for the SSI states was significantly higher than the mean for non-SSI states ($t = 2.43$, $df = 39$, $p < .05$). SSI states averaged significantly higher in 1996 on $I_{MD(96)}$ in all samples and subsamples, as reported in Table 6C.1 of the Appendix.

Table 6.24
Mean and Standard Deviation of $I_{MD(96)}$ in SSI and Non-SSI State, Grade 8, 1996

	SSI States *N* = 22	Non-SSI States *N* = 18
Mean	23.19	22.33
Standard Deviation	1.19	1.01

Individual state means are graphed in Figure 6.9, with states ordered by their means from lowest to highest. The ten states scoring highest on mathematical discourse include eight SSI states (California, Kentucky, Vermont, Delaware, Georgia, Michigan, Connecticut, and Montana) and two non-SSI states (Maryland and Arizona). The ten lowest states include three SSI states (Arkansas, New York, and Rhode Island) and seven non-SSI states (North Dakota, Indiana, Alabama, West Virginia, Iowa, Tennessee, and Minnesota). See Table 6C.2 in the Appendix for the mean values for each state.

Figure 6.9. State means on $I_{MD(96)}$ ordered from highest to lowest, grade 8.



<u>Individual items</u>. Table 6.25 lists the mean for SSI and non-SSI states for each of the nine items in the 1996 mathematical discourse scale. For all participating states, a MANOVA found no overall significant effect for SSI status (F = 1.36, df = 9,30, p = .25). For the subsample, SSI status was statistically significant (F = 2.08, df = 9,20, p < .10). Mean values for each state are listed in Table 6C.2 of the Appendix.

Since the first six items in Table 6.25 have the same set of response options, the means can be compared directly. In grade 8 in both SSI and non-SSI states, two of the practices happened, on average, about once or twice a week:

- Discuss solutions to mathematics problems with other students, and
- Work and discuss mathematics problems that reflect real-life situations.

In the total sample, the means for SSI and non-SSI states were similar for these two items, but in the subsample, the SSI states averaged higher on the second. The item on class time spent in group work showed that there was little difference between SSI and non-SSI states as a group: in both SSI and non-SSI states, students averaged one-half hour to one hour a week working with a partner or in a small group.

The item with the lowest mean at grade 8 was:

- Write reports or do mathematics projects.

Most students never or hardly ever did mathematics reports or projects, though these activities did happen a bit more frequently in the SSI states. The next least frequent item was:

- Write a few sentences about how to solve a mathematics problem.

In both SSI and non-SSI states, students indicated that they wrote a few sentences about how to solve mathematics problems just once or twice a month. This item was one of the two with the largest mean difference between SSI and non-SSI states.

SSI states averaged higher than non-SSI states on items about assessing student progress (the last two items in Table 6.25). In both SSI and non-SSI states, teachers used written responses to evaluate student progress more often than they used individual or group projects or presentations, but both occurred infrequently—somewhere between “Once or twice a year” and “Once or twice a month.”

Table 6.25

Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8, 1996

	SSI States		Non-SSI States		Mean	
	M	*SD*	M	*SD*	Difference	F
Small groups/with partner						
Total sample	2.80	0.17	2.77	0.18	0.03	0.24
Subsample	2.82	0.14	2.76	0.20	0.04	0.77
Write about solution						
Total sample	2.09	0.27	1.92	0.20	0.17	4.96
Subsample	2.09	0.26	1.88	0.19	0.19	6.21
Talk to class						
Total sample	2.74	0.16	2.61	0.13	0.13	5.04
Subsample	2.77	0.16	2.60	0.20	0.17	7.09
Write reports/do projects						
Total sample	1.47	0.17	1.36	0.08	0.11	6.29
Subsample	1.47	0.15	1.32	0.08	0.15	6.53
Discuss with others						
Total sample	3.21	0.11	3.16	0.12	0.05	2.06
Subsample	3.21	0.09	3.16	0.10	0.05	2.41
Discuss real-life situations						
Total sample	2.98	0.14	2.89	0.13	0.09	4.27
Subsample	2.99	0.11	2.87	0.13	0.12	6.92
Class time spent in group work						
Total sample	3.00	0.19	2.98	0.18	0.02	0.06
Subsample	3.02	0.17	2.99	0.20	0.03	0.24
Assess by written responses						
Total sample	2.68	0.25	2.51	0.24	0.17	4.36
Subsample	2.69	0.26	2.47	0.22	0.22	6.04
Assess by individual/group projects						
Total sample	2.25	0.14	2.16	0.11	0.09	4.70
Subsample	2.27	0.12	2.16	0.11	0.07	7.09

Item intercorrelations. The relationships among the state means on each of the nine discourse items are presented in Table 6.26. The first correlation matrix is based on data from all states, and the second is for the subsample of states that met the NCES participation rate guidelines.

Table 6.26
Intercorrelations of State Means on the Nine Mathematical Discourse Items, Grade 8, 1996

	Item number							
	1	2	3	4	5	6	7	8
Total sample, *N* = 40								
1. Work in small groups	-							
2. Write about solution	.60*							
3. Talk to class	.07	.52*						
4. Write reports/do projects	.34	.83*	.36					
5. Discuss with others	.66*	.47*	.19	.22				
6. Real-life situations	.29	.36	.38	.14	.50*			
7. Time in group work	.92*	.54*	.00	.34	.72*	.24		
8. Assess by written responses	.56*	.94*	.54*	.75*	.43*	.35	.48*	
9. Assess by projects	.46*	.76*	.47*	.81*	.26	.19	.42*	.69*
Subsample, *N* = 30								
1. Work in small groups	-							
2. Write about solution	.58*							
3. Talk to class	-.04	.46						
4. Write reports/do projects	.34	.84*	.31					
5. Discuss with others	.67*	.55*	.27	.24				
6. Real-life situations	.12	.33	.44	.26	.49*			
7. Time in group work	.95*	.58*	-.06	.40	.69*	.12		
8. Assess by written responses	.53*	.96*	.50*	.81*	.44	.30	.48*	
9. Assess by projects	.46	.79*	.46	.78*	.40	.28	.50*	.74*

*$p < .01$

The intercorrelations for the 40 states in the total sample are comparable to those for the the 30 states in the subsample. In both samples, correlations were generally positive, and roughly half were statistically significant. Somewhat fewer correlations were statistically significant ($p < .01$) in the subsample, in part because of the smaller sample size.

Relationships among these items may be underestimated because of restrictions in range. For some items, a large proportion of respondents selected the highest or lowest option. If the response options were designed to differentiate among respondents, the correlations might be somewhat different.

Grade 4

Student opportunities for mathematical discourse. The grade 4 state means on $I_{MD(96)}$, the mathematical discourse indicator, are shown in Table 6.27. The mean for the SSI states was significantly higher than the mean for non-SSI states ($t = 2.43$, $df = 22$, $p < .05$). At grade 4, SSI states averaged significantly higher in 1996 on $I_{MD(96)}$ in all samples and subsamples, as reported in Table 6C.1 of the Appendix.

Table 6.27
Mean and Standard Deviation of $I_{MD(96)}$ in SSI and Non-SSI States, Grade 4, 1996

	SSI States *N* = 23	Non-SSI States *N* = 20
Mean	23.83	22.99
Standard Deviation	1.03	1.22

Individual state means are graphed in Figure 6.10, with states ordered by means from lowest to highest. (See Table 6C.3 in the Appendix for the mean values.) The ten states scoring highest on mathematical discourse include seven SSI states (Kentucky, Vermont, Maine, California, North Carolina, Connecticut, and Georgia) and three non-SSI states (Maryland, Nevada, and Mississippi). The ten lowest states include two SSI states (Arkansas and Rhode Island) and eight non-SSI states (North Dakota, Indiana, Missouri, Iowa, Tennessee, Wisconsin, Washington, and Hawaii).

Individual items. Table 6.28 lists the mean for SSI and non-SSI states for each of the nine items in the 1996 mathematical discourse scale. At grade 4, the effect for SSI status was not statistically significant for any of the states ($F = 1.13$, $df = 9,33$, $p = .37$) or for the subsample ($F = 1.22$, $df = 9,22$, $p = .33$).

At grade 4, the two items with the largest differences between SSI and non-SSI states both concerned writing:

- Write a few sentences about how to solve a mathematics problem, and
- Assess student progress with short (e.g., a phrase or sentence) or long (e.g., several sentences or paragraphs) written responses.

The first occurs about once or twice a month, and the second occurs between once or twice a year and once or twice a month.

Figure 6.10. State means on $I_{MD(96)}$ ordered from highest to lowest, grade 4.

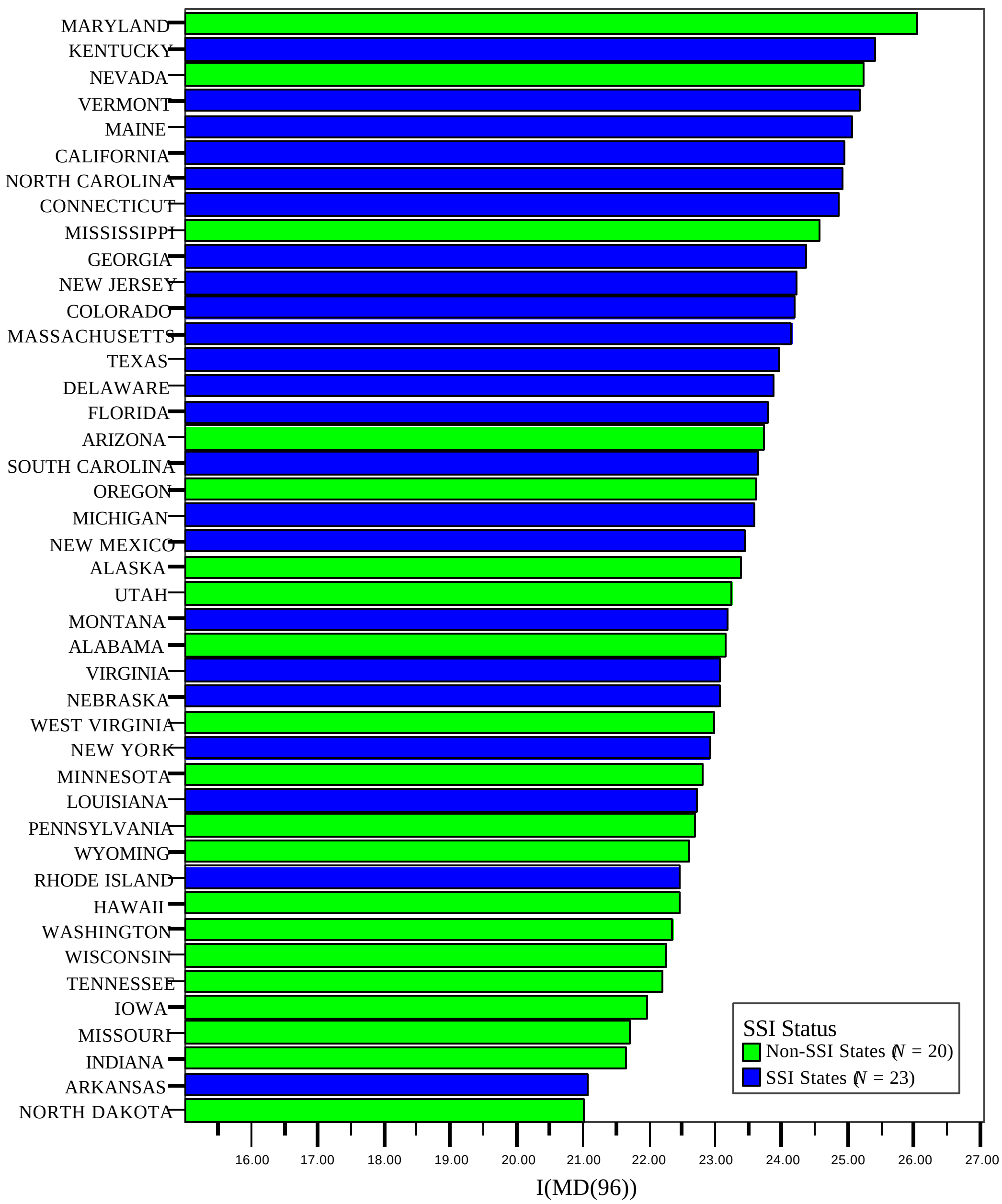


Table 6.28
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 4, 1996

	SSI States		Non-SSI States		Mean	
	M	*SD*	*M*	*SD*	Difference	*F*
Small groups/with partner						
Total sample	2.96	0.11	2.92	0.14	0.04	1.18
Subsample	2.98	0.09	2.89	0.13	0.09	5.32
Write about solution						
Total sample	2.18	0.26	1.99	0.24	0.19	6.35
Subsample	2.21	0.25	1.99	0.25	0.22	5.96
Talk to class						
Total sample	2.91	0.14	2.80	0.22	0.09	3.74
Subsample	2.93	0.13	2.79	0.22	0.14	4.23
Write reports/do projects						
Total sample	1.40	0.10	1.34	0.08	0.06	4.32
Subsample	1.40	0.10	1.34	0.08	0.06	3.55
Discuss with others						
Total sample	3.06	0.12	2.99	0.12	0.07	3.66
Subsample	3.08	0.09	2.98	0.12	0.10	6.84
Discuss real-life situations						
Total sample	2.97	0.13	3.12	0.12	0.15	1.82
Subsample	3.00	0.11	2.92	0.12	0.08	4.80
Class time spent in group work						
Total sample	3.20	0.13	3.12	0.15	0.08	3.39
Subsample	3.23	0.13	3.10	0.14	0.13	8.32
Assess by written responses						
Total sample	2.75	0.24	2.56	0.25	0.19	6.21
Subsample	2.79	0.24	2.57	0.26	0.22	6.23
Assess by individual/group projects						
Total sample	2.38	0.12	2.33	0.15	0.05	2.48
Subsample	2.40	0.13	2.32	0.13	0.08	3.15

Item intercorrelations. The relationships among the state means on each of the nine discourse items are presented in Table 6.29. The first correlation matrix is based on data from all states, and the second is for the subsample of states that met the NCES participation rate guidelines.

Table 6.29
Intercorrelations of State Means on the Nine Mathematical Discourse Items, Grade 4, 1996

	Item number							
	1	2	3	4	5	6	7	8
Total sample, *N* = 43								
1. Work in small groups	-							
2. Write about solution	.65*							
3. Talk to class	.37	.60*						
4. Write reports/do projects	.56*	.76*	.40*					
5. Discuss with others	.70*	.63*	.76*	.43*				
6. Real-life situations	.38	.27	.71*	.18	.70*			
7. Time in group work	.89*	.67*	.34	.63*	.62*	.27		
8. Assess by written responses	.61*	.95*	.64*	.76*	.66*	.30	.64*	
9. Assess by projects	.58*	.54*	.39*	.70*	.38	.31	.57*	.58*
Subsample, *N* = 32								
1. Work in small groups	-							
2. Write about solution	.73*							
3. Talk to class	.30	.62*						
4. Write reports/do projects	.64*	.71*	.44					
5. Discuss with others	.67*	.71*	.78*	.50*				
6. Real-life situations	.29	.28	.68*	.25	.67*			
7. Time in group work	.90*	.66*	.30	.60*	.62*	.25		
8. Assess by written responses	.66*	.95*	.68*	.72*	.72*	.30	.61*	
9. Assess by projects	.61*	.52*	.42	.74*	.42	.36*	.54*	.58*

$p < .01$

Generally, the intercorrelations at grade 4 are similar to those at grade 8. All correlations are positive, and many are above .60. The one item that is the least related to the others is:

- Work and discuss mathematics problems that reflect real-life situations.

1992

In 1992, the teacher questionnaire included four items that matched four from the 1996 mathematical discourse scale. They were:

How often do the students in this class do each of the following things?

- Write a few sentences about how to solve a mathematics problem
- Write reports or do mathematics projects
- Discuss solutions to mathematics problems with other students
- Work and discuss mathematics problems that reflect real life situations

Response options

- Almost every day
- Once or twice a week
- Once or twice a month
- Never or hardly ever

To allow for a direct comparison between 1992 and 1996, the sum of the four identical items, $I_{MD4(92)}$, was used as one indicator of mathematical discourse. The internal consistency of this scale was .54 at grade 8 and .64 at grade 4.

Besides the four identical items, the 1992 questionnaire includes three more items very similar to items that made up $I_{MD(96)}$. The additional items are listed on the next page. The two items about assessment are worded very similarly in 1992 and 1996; however, teachers answered the question separately for each class in 1992, but they answered it just once in 1996.

How often do the students in this class do each of the following things?
Solve mathematics problems in small groups.

Response options
Almost every day
Once or twice a week
Once or twice a month
Never or hardly ever

Think about your plans for this mathematics class for the entire year. How often do you use each of the following to assess student progress in mathematics?
Short (e.g., a phrase or sentence) or long (e.g., several sentences or paragraphs) written responses
Individual or group projects or presentations

Response options
Once or twice a week
Once or twice a month
Once or twice a year
Never or hardly ever

Responses to all seven items were summed for a seven-item scale of mathematical discourse, $I_{MD(92)}$. The internal consistency of the seven-item scale was .69 at grade 8 and .75 at grade 4.

Grade 8

Student opportunities for mathematical discourse. Table 6.30 shows that the mean of the SSI states was a bit higher than the mean of the non-SSI states in 1992, but the difference was not statistically significant for either the four matched items, $I_{MD4(92)}$, ($t = 0.99$, $df = 39$, $p = .33$) or the total scale, $I_{MD(92)}$, ($t = 1.52$, $df = 39$, $p = .14$). See Table 6C.4 in the Appendix for results with other samples and subsamples, Table 6C.5 for the individual state means on $I_{MD(92)}$, and Table 6C.7 for individual state means on I_{MD4}.

Table 6.30
Mean and Standard Deviation of $I_{MD4(92)}$ and $I_{MD(92)}$ in SSI and Non-SSI States, Grade 8

	SSI States *N* = 22	Non-SSI States *N* = 19
$I_{MD4(92)}$		
Mean	9.14	9.04
Standard Deviation	0.37	0.29
$I_{MD(92)}$		
Mean	15.57	14.23
Standard Deviation	0.69	0.66

Individual items. The means for SSI and non-SSI states on each of the seven items on the 1992 discourse scale are listed in Table 6.31. Table 6C.5 of the Appendix lists the values for the individual states. Mean differences between SSI and non-SSI states in 1992 were small. The overall F for the MANOVA comparing SSI and non-SSI states on these seven items was not significant for either the total sample ($F = 1.20$, $p = .33$) or the subsample ($F = 0.82$, $df = 7,28$, $p = .58$).

Table 6.31
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8, 1992

	SSI States *N* = 22	Non-SSI States *N* = 19	Mean Difference
Matching items			
Write about solution			
Mean	1.83	1.77	0.06
Standard Deviation	0.15	0.16	
Write reports/do projects			
Mean	1.24	1.21	0.03
Standard Deviation	0.06	0.07	
Discuss with others			
Mean	3.18	3.18	0.00
Standard Deviation	0.14	0.11	
Discuss real-life situations			
Mean	2.89	2.87	0.02
Standard Deviation	0.14	0.10	
Other similar items			
Solve in groups			
Mean	2.50	2.44	0.06
Standard Deviation	0.16	0.19	
Assess by written responses			
Mean	2.20	2.09	0.11
Standard Deviation	0.16	0.17	
Assess by projects, portfolios, or presentations			
Mean	1.74	1.69	0.05
Standard Deviation	0.12	0.15	

At grade 8 in 1992, in both SSI and non-SSI states, the two most frequent activities were:

- Discuss solutions to mathematics problems with other students, and
- Work and discuss mathematics problems that reflect real-life situations.

The first occurred almost every day, and the second occurred just about as often. The least frequently occurring item was:

- Write reports or do mathematics projects.

On average, reports and projects occurred closer to “Never or hardly ever” than to “Once or twice a month.” The item on which the largest difference between SSI and non-SSI states occurred was:

- How often do you use short or long written responses to assess student progress in mathematics?

Both groups averaged close to “Once or twice a year,” although the SSI mean was a bit higher.

Item intercorrelations. Table 6.32 lists the intercorrelations of the state means on the seven items in the 1992 mathematical discourse scale for the total sample of 41 states as well as for the subsample of 36 states that followed the NCES participation rate guidelines.

Table 6.32
Intercorrelations of State Means on the Seven Mathematical Discourse Items, Grade 8, 1992

	Item number					
	1	2	3	4	5	6
Total sample, *N* = 41						
1. Write about solution	-					
2. Write reports/do projects	.60*	-				
3. Discuss with others	.30	.21	-			
4. Real-life situations	.31	.16	.51*	-		
5. Solve in small groups	.46*	.38	.72*	.28	-	
6. Assess by written responses	.85*	.54*	.24	.40	.40	-
7. Assess by projects/portfolios	.52*	.83*	.25	.21	.46*	.51*
Subsample, *N* = 36						
1. Write about solution	-					
2. Write reports/do projects	.64*	-				
3. Discuss with others	.32	.22	-			
4. Real-life situations	.31	.17	.51*	-		
5. Solve in small groups	.49*	.43	.70*	.25	-	
6. Assess by written responses	.85*	.64*	.34	.42	.50*	-
7. Assess by projects/portfolios	.61*	.85*	.23	.16	.50*	.65*

*$p < .01$

The correlations suggest two groups of items in this scale. One group has items about writing, and includes items 1, 2, 6, and probably 7. The other has items about working and discussing with other students, and includes items 3, 4, and 5. These two subgroups seem to capture the distinction between product and process.

Grade 4

Student opportunities for mathematical discourse. Table 6.33 shows that for both $I_{MD4(92)}$ and $I_{MD(92)}$, the mean of the SSI states was higher than the mean of the non-SSI states. The difference was statistically significant for the total scale, $I_{MD(92)}$, ($t = 1.79$, $df = 39$, $p < .10$), but not for the four matched items, $I_{MD4(92)}$, ($t = 1.60$, $df = 39$, $p = .12$). See Table 6C.4 in the Appendix for results with other samples and subsamples, Table 6C.6 for the individual state means on $I_{MD(92)}$, and Table 6C.7 for individual state means on I_{MD4}.

Table 6.33
Mean and Standard Deviation of $I_{MD4(92)}$ and $I_{MD(92)}$ in SSI and Non-SSI States, Grade 4, 1992

	SSI States *N* = 22	Non-SSI States *N* = 19
$I_{MD4(92)}$		
Mean	9.08	8.86
Standard Deviation	0.41	0.44
$I_{MD(92)}$		
Mean	15.74	15.30
Standard Deviation	0.81	0.76

Individual items. The means for SSI and non-SSI states on each of the seven items on the 1992 discourse scale are listed in Table 6.34. Table 6C.6 of the Appendix lists the values for the individual states at grade 4. Mean differences between SSI and non-SSI states in 1992 were small. The overall *F* for the MANOVA comparing SSI and non-SSI states on these seven items was not significant for either the total sample ($F = 1.36$, $df = 7,33$, $p = .33$), or the subsample ($F = 1.70$, $df = 7,28$, $p = .15$).

Table 6.34
Mean and Standard Deviation of Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 4, 1992

	SSI States *N* = 22	Non-SSI States *N* = 19	Mean Difference
Matching items			
Write about solution			
Mean	1.88	1.81	0.07
Standard Deviation	0.15	0.21	
Write reports/do projects			
Mean	1.25	1.24	0.01
Standard Deviation	0.07	0.04	
Discuss with others			
Mean	2.98	2.93	0.05
Standard Deviation	0.12	0.14	
Discuss real-life situations			
Mean	2.95	2.98	0.05
Standard Deviation	0.14	0.11	
Other similar items			
Solve in groups			
Mean	2.70	2.66	0.04
Standard Deviation	0.15	0.15	
Assess with written responses			
Mean	2.18	2.11	0.09
Standard Deviation	0.19	0.17	
Assess with projects, portfolios, or presentations			
Mean	1.78	1.69	0.09
Standard Deviation	0.16	0.07	

At grade 4 in 1992, as at grade 8, in both SSI and non-SSI states, the two most frequent activities were:

- Discuss solutions to mathematics problems with other students, and
- Work and discuss mathematics problems that reflect real-life situations.

The two occurred almost every day. The least frequently occurring item was:

- Write reports or do mathematics projects.

On average, reports and projects was closer to “Never or hardly ever” than to “Once or twice a month.” The largest difference between SSI and non-SSI states was on the two items about assessment:

- How often do you use short or long written responses to assess student progress in mathematics?
- How often do you use group or individual projects, portfolios, or presentations to assess student progress in mathematics?

At grade 4, the average for the first item was slightly more than once or twice a year, and the average for the second was between “Once or twice a year” and “Never or hardly ever.”

Item intercorrelations. Table 6.35 lists the intercorrelations of the grade 4 state means on the seven items in the 1992 mathematical discourse scale.

Table 6.35
Intercorrelations of State Means on the Seven Mathematical Discourse Items, Grade 4, 1992

	Item number					
	1	2	3	4	5	6
Total sample, *N* = 41						
1. Write about solution	-					
2. Write reports/do projects	.70*	-				
3. Discuss with others	.72*	.52*	-			
4. Real-life situations	.72*	.44*	.73*	-		
5. Solve in small groups	.76*	.67*	.87*	.61*	-	
6. Assess by written responses	.85*	.56*	.55*	.65*	.56*	-
7. Assess by projects/portfolios	.66*	.77*	.59*	.50*	.68*	.67*
Subsample, *N* = 36						
1. Write about solution	-					
2. Write reports/do projects	.69*	-				
3. Discuss with others	.73*	.53*	-			
4. Real-life situations	.72*	.41	.75*	-		
5. Solve in small groups	.78*	.72*	.88*	.65*	-	
6. Assess by written responses	.85*	.55*	.57*	.67*	.59*	-
7. Assess by projects/portfolios	.68*	.78*	.61*	.51*	.72*	.68*

$p < .01$

At grade 4, the seven mathematical discourse items were moderately to strongly related to each other. The items having the strongest relationship with items 7 and 8 show the link between instruction and assessment.

1990

The 1990 teacher questionnaire had two items related to mathematical discourse.

About how often do students in this class do the following types of activities for mathematics?
- Work in small groups
- Write reports or do mathematics projects.

Response options
- Almost every day
- Several times a week
- About once a week
- Less than once a week
- Never

The 1990 questionnaire had five response categories rather than the four in later years. In addition, the 1990 choices provided finer distinctions among activities that occurred at least weekly. Everything else had to be coded as "Never." The1990 response options were coded from 1 for "Never" to 5 for "Almost every day," and the items were summed for the 1990 mathematical discourse scale, $I_{MD(90)}$.

Grade 8

Student opportunities for mathematical discourse. Table 6.36 presents the mean and standard deviation of $I_{MD(90)}$ for the SSI and non-SSI states participating in the 1990 State NAEP. Comparisons for all subsamples are reported in Table 6C.8 of the Appendix, and means for individual states are in Table 6C.9.

Table 6.36
Mean and Standard Deviation of $I_{MD(90)}$ and the Individual Mathematical Discourse Items in SSI and Non-SSI States, Grade 8

	SSI States *N* = 20	Non-SSI States *N* = 17
$I_{MD(90)}$		
Mean	4.18	4.19
Standard Deviation	0.33	0.31
Individual items		
Work in small groups		
Mean	2.62	2.68
Standard Deviation	0.28	0.30
Write reports/do projects		
Mean	1.55	1.51
Standard Deviation	0.09	0.09

Change from 1992 to 1996

Grade 8

Change on the mathematical discourse indicator from 1992 to 1996 was evaluated in two ways on the basis of the available data. The four items that were exactly the same from 1992 to 1996 provided information on the overall change from 1992 to 1996, as well as on the effect of SSI status on any change. In addition, the seven-item mathematical discourse scale, $I_{MD(92)}$, was used to examine whether SSI status was a significant factor in predicting $I_{MD(96)}$, the 1996 measure of mathematical discourse.

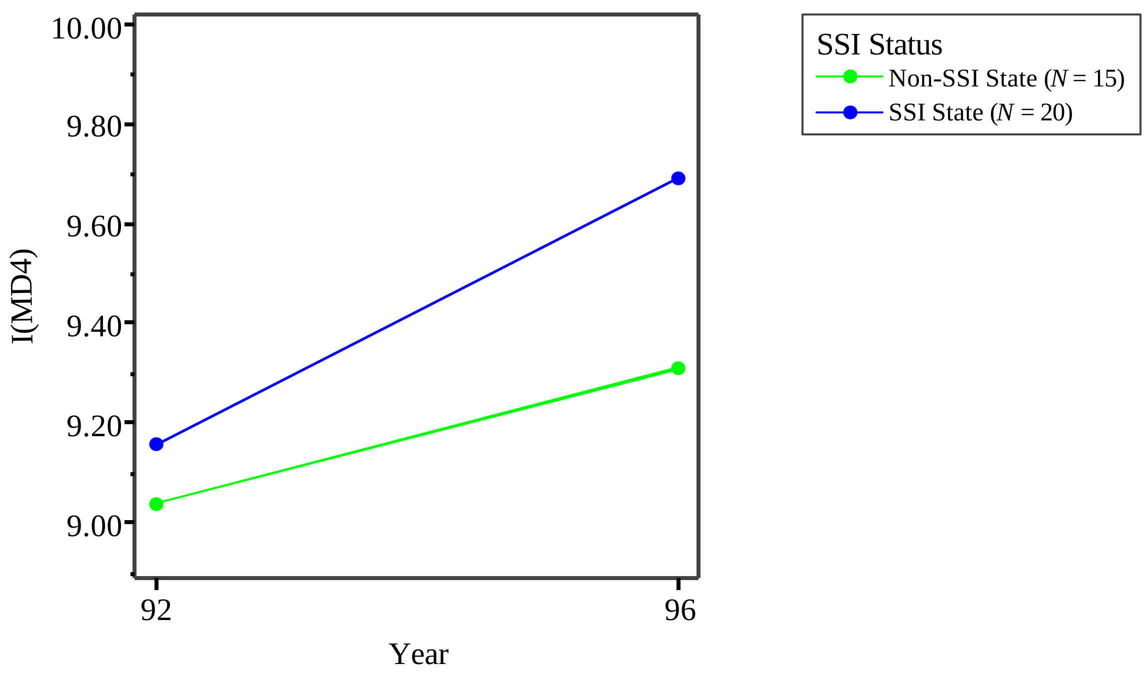
I_{MD4} – Four matching items

Two-point trend sample. For I_{MD4}, the four items that were part of the teacher questionnaire in both 1992 and 1996, a 2x2 repeated measures analysis of variance was used to examine changes in students' opportunities for mathematical discourse in SSI and non-SSI states. The results are presented in Table 6.37 and graphed in Figure 6.11. Overall, I_{MD4} increased from 1992 to 1996 ($F = 31.73$, $df = 1,33$, $p < .01$). In addition, the SSI states scored higher than the non-SSI states across both years ($F = 4.22$, $df = 1,33$, $p < .05$). The increase for SSI states was slightly greater than for the non-SSI states ($F = 3.38$, $df = 1,33$, $p < .10$).

Table 6.37
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States in 1992 and 1996, Grade 8

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI States	20	9.14	0.37	9.69	0.50	0.53	0.46
Non-SSI States	15	9.04	0.29	9.31	0.44	0.27	0.34

Figure 6.11. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, Grade 8.



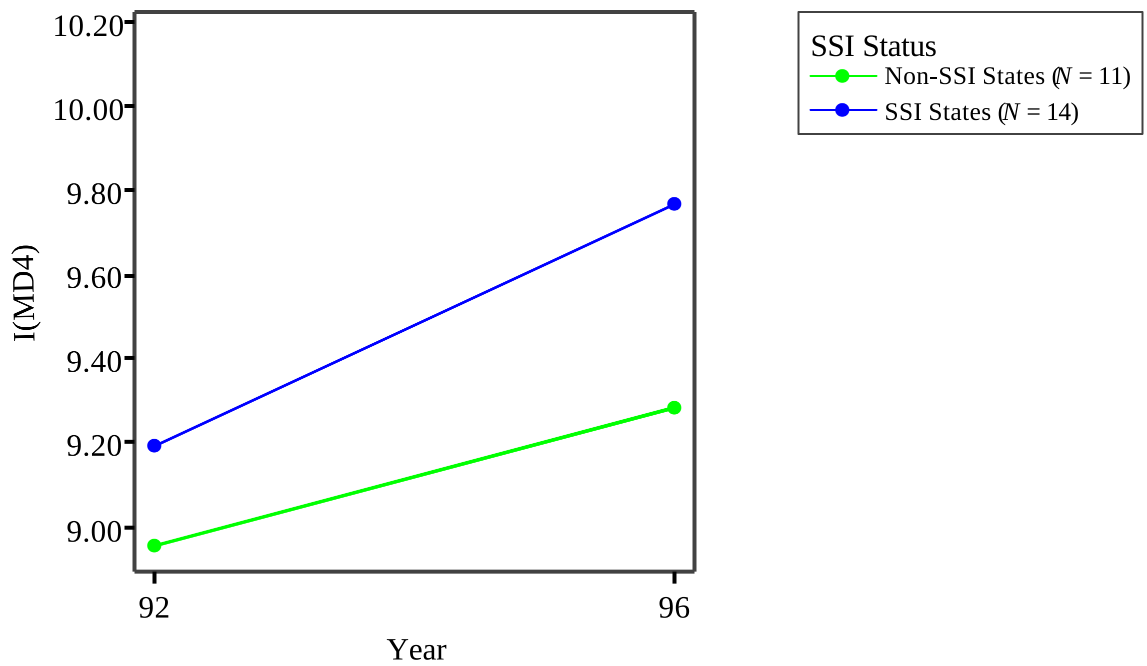
Subsample. The results of the 2x2 repeated measures analysis of variance for 14 SSI states and 11 non-SSI states that followed the NCES participation rate guidelines are presented in Table 6.38 and graphed in Figure 6.12. For the subsample, the increase in I_{MD4} from 1992 to 1996 was statistically significant ($F = 25.08$, $df = 1,23$, $p < .01$). In addition, the SSI states scored higher than the non-SSI states across both years ($F = 8.38$, $df = 1,23$, $p < .01$). Unlike the findings for the total sample, the interaction term of SSI status by year was not statistically significant ($F = 1.84$, $df = 1,23$, $p = .18$).

Table 6.38

Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States that Met the NCES Participation Rate Guidelines in 1992 and 1996, Grade 8

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI States	14	9.12	0.31	9.77	0.49	0.58	0.55
Non-SSI States	11	8.95	0.27	9.29	0.41	0.33	0.28

Figure 6.12. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, Grade 8, for the subsample that followed the NCES participation rate guidelines.



$I_{MD(92)}$ and $I_{MD(96)}$

Two-point trend sample. $I_{MD(92)}$ is not directly comparable to $I_{MD(96)}$, because the scales are of different lengths and contain somewhat different items. Given these constraints, a regression model was used to examine change in mathematical discourse from 1992 to 1996 as a function of a state's SSI status. In the model, the dependent measure was the state's mean on the 1996 nine-item scale, $I_{MD(96)}$, and the predictors were the state's score on the 1992 seven-item scale, $I_{MD(92)}$ and its SSI status. Results are shown in Table 6.39 and graphed in Figure 6.13.

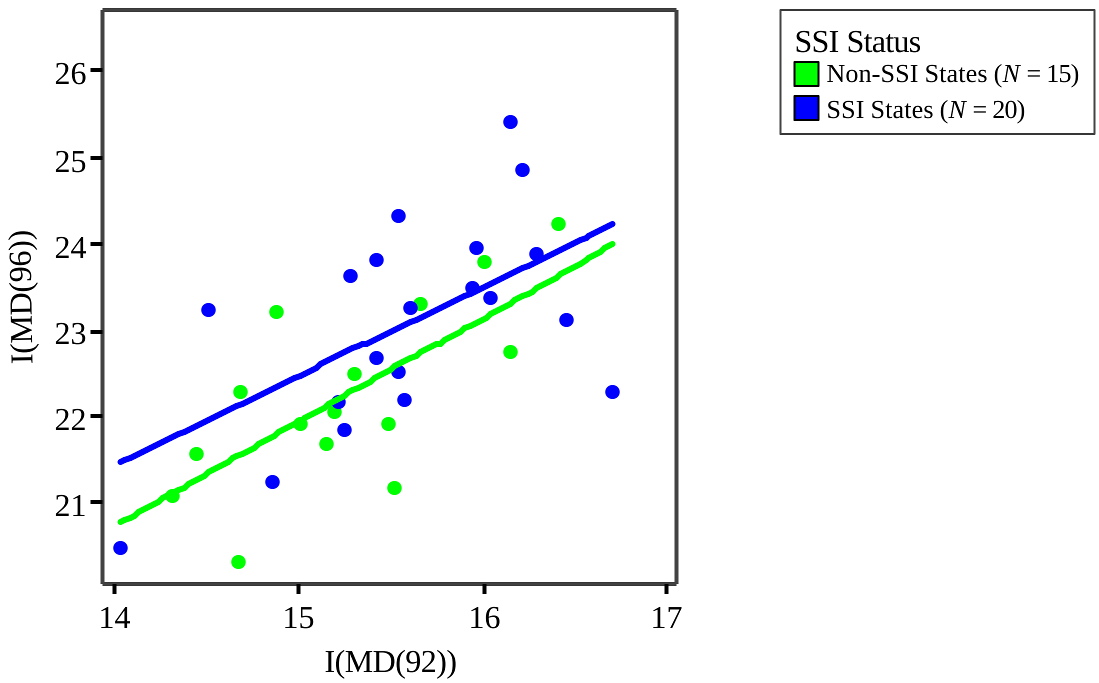
Table 6.39
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{MD(92)}$	1.20	0.24	.65	.42	24.32*		
Step 2							
$I_{MD(92)}$	1.11	0.25	.60				
SSI status	0.46	0.32	.19	.46	13.59*	.04	2.07

*$p < .01$

The regression analysis found that $I_{MD(92)}$ was significantly related to $I_{MD(96)}$ ($\beta = .60$, $t = 4.46$, $p < .01$), but SSI status did not add anything to the prediction of $I_{MD(96)}$ ($\beta = .19$, $t = 1.43$, $p = .16$). In the two-point trend sample, the correlation of the two predictors was .26 ($p < .10$)

Figure 6.13. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for SSI and non-SSI states, grade 8.



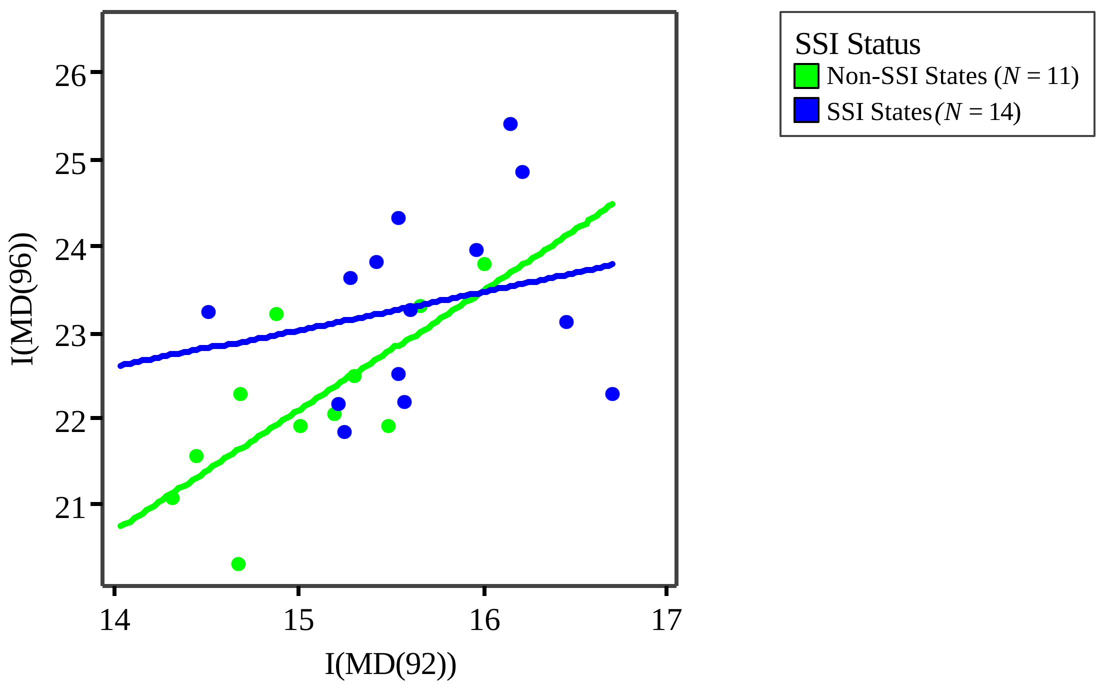
Subsample. The results of the linear regression analysis for the subsample are presented in Table 6.40. As with the total sample, the 1992 mathematical discourse measure was significantly related to the 1996 measure ($\beta = .43$, $t = 2.23$, $p < .05$), but SSI status did not add to the prediction of $I_{MD(96)}$ ($\beta = .28$, $t = 1.45$, $p = .16$). In the subsample, the correlation of the two predictors was .50 ($p < .01$). Figure 6.14 shows the scatterplot.

Table 6.40
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

	B	SE B	*b*	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{MD(92)}$	1.08	0.33	.57	.32	11.09*		
Step 2							
$I_{MD(92)}$	0.82	0.37	.43				
SSI status	0.66	0.45	.28	.38	6.86*	.06	2.10

*$p < .01$

Figure 6.14. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.



With the reduced sample, the relationship between the measures of mathematical discourse in 1992 and 1996 seems stronger for the non-SSI states, based on the slope of the regression lines.

Grade 4

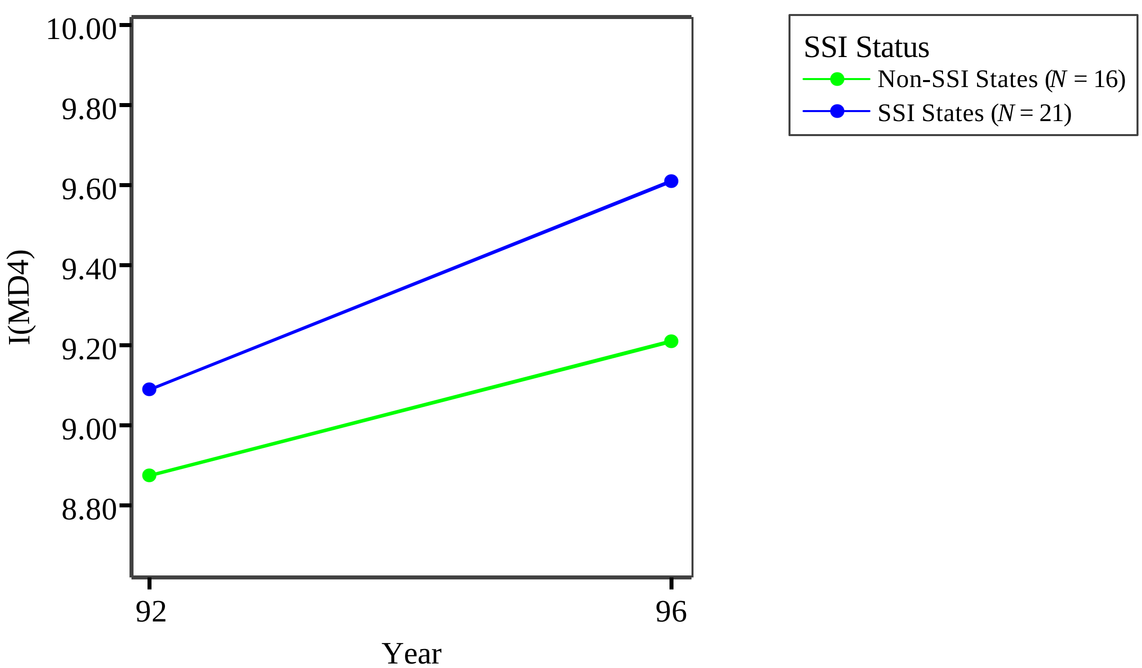
I_{MD4} – Four matching items

Two-point trend sample. For the four items that were part of the teacher questionnaire in both 1992 and 1996, the results of the 2x2 repeated measures analysis of variance found that I_{MD4} increased significantly from 1992 to 1996 ($f = 48.28$, $df = 1, 35$, $p < .01$). The results are presented in Table 6.41 and graphed in Figure 6.15. In addition, the SSI states scored higher than the non-SSI states across both years ($F = 5.00$, $df = 1,35$, $p < .05$). The interaction of SSI status and year was not statistically significant ($F = 2.18$, $p = .149$).

Table 6.41
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States in 1992 and 1996, Grade 4

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI States	21	9.09	0.41	9.61	0.44	0.52	0.37
Non-SSI States	16	8.87	0.48	9.20	0.50	0.34	0.37

Figure 6.15. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, grade 4.

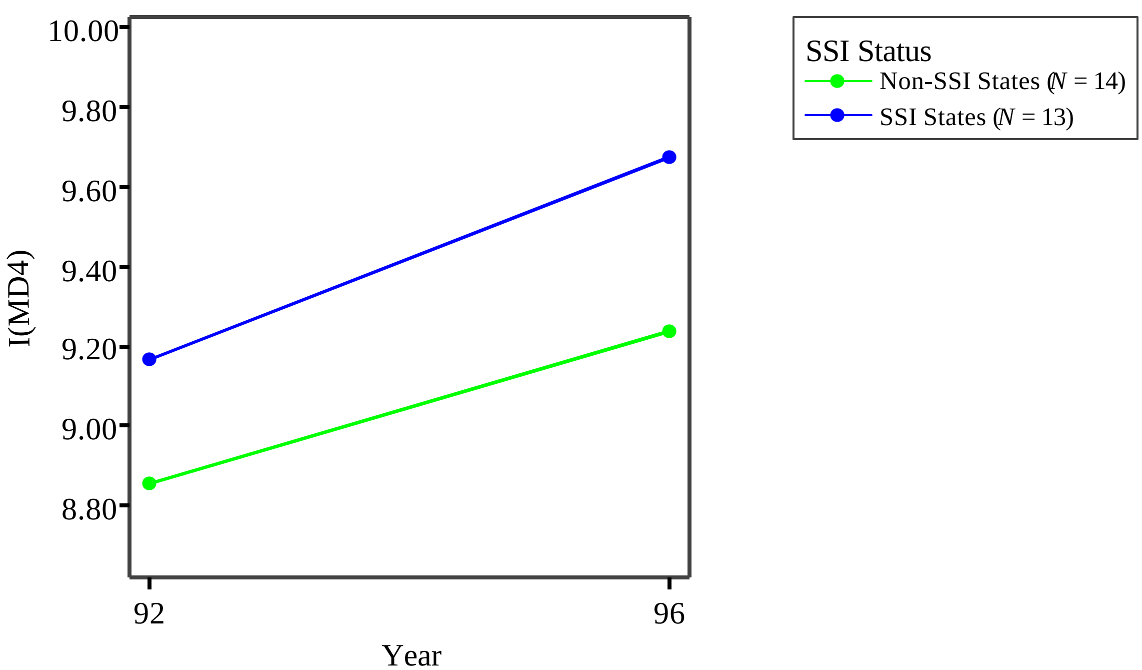


Subsample. Table 6.42 presents the results of the 2x2 repeated measures analysis of variance for those states that followed the participation rate guidelines. For the subsample, the increase in I_{MD4} from 1992 to 1996 was statistically significant (F = 34.02, df = 1,25, p < .01). In addition, the SSI states scored higher than the non-SSI states across both years (F = 5.16, df = 1,25, p < .05). The interaction term of SSI status by year was not statistically significant (F = 0.73, df = 1,25, p = .40). The results are graphed in Figure 6.16.

Table 6.42
Mean and Standard Deviation of I_{MD4} in SSI and Non-SSI States that Met the NCES Participation Rate Guidelines in 1992 and 1996, Grade 4

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI States	13	9.16	0.38	9.67	0.43	0.51	0.42
Non-SSI States	14	8.86	0.52	9.23	0.53	0.38	0.37

Figure 6.16. Change in the mean of I_{MD4} for SSI and non-SSI states from 1992 to 1996, Grade 4, for the subsample that followed the NCES participation rate guidelines.



$I_{MD(92)}$ and $I_{MD(96)}$

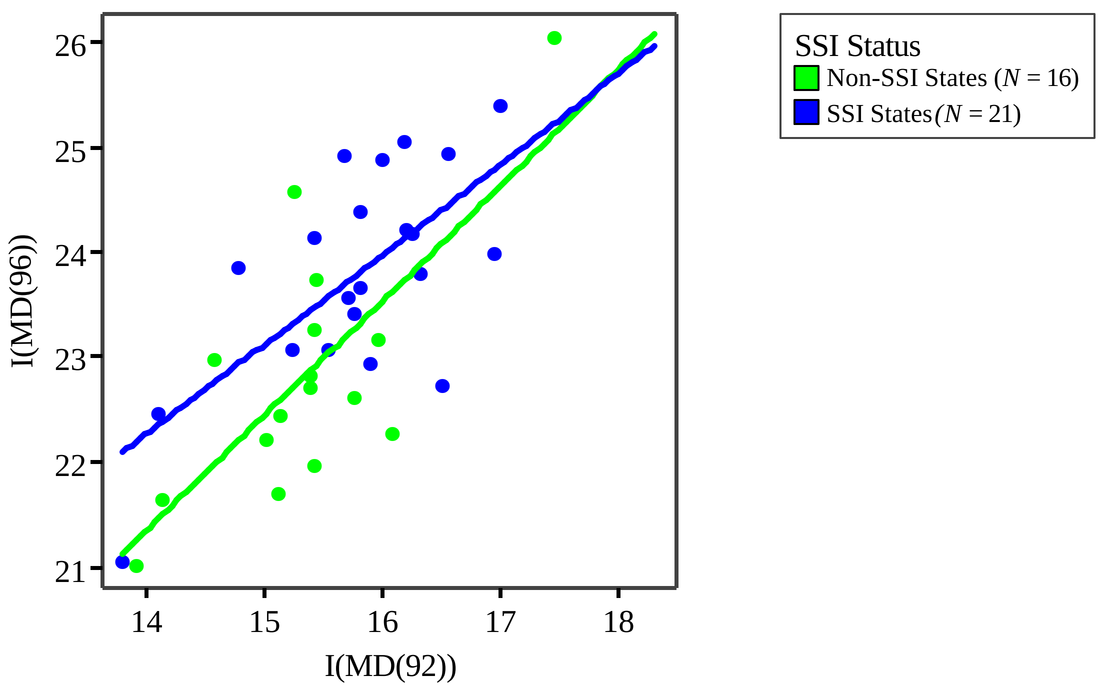
Two-point trend sample. Results of the regression model predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI status are shown in Table 6.43. The results for the total sample show that both $I_{MD(92)}$ and SSI status were significantly related to the prediction of $I_{MD(96)}$. Figure 6.17 presents the scatterplot.

Table 6.43
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status, Grade 4

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{MD(92)}$	1.04	0.17	.72	.52	38.71**		
Step 2							
$I_{MD(92)}$	0.96	0.17	.67				
SSI status	0.55	0.28	.23	.57	22.93**	.05	3.92*

*$p < .10$; **$p < .01$

Figure 6.17. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for SSI and non-SSI states, grade 4.



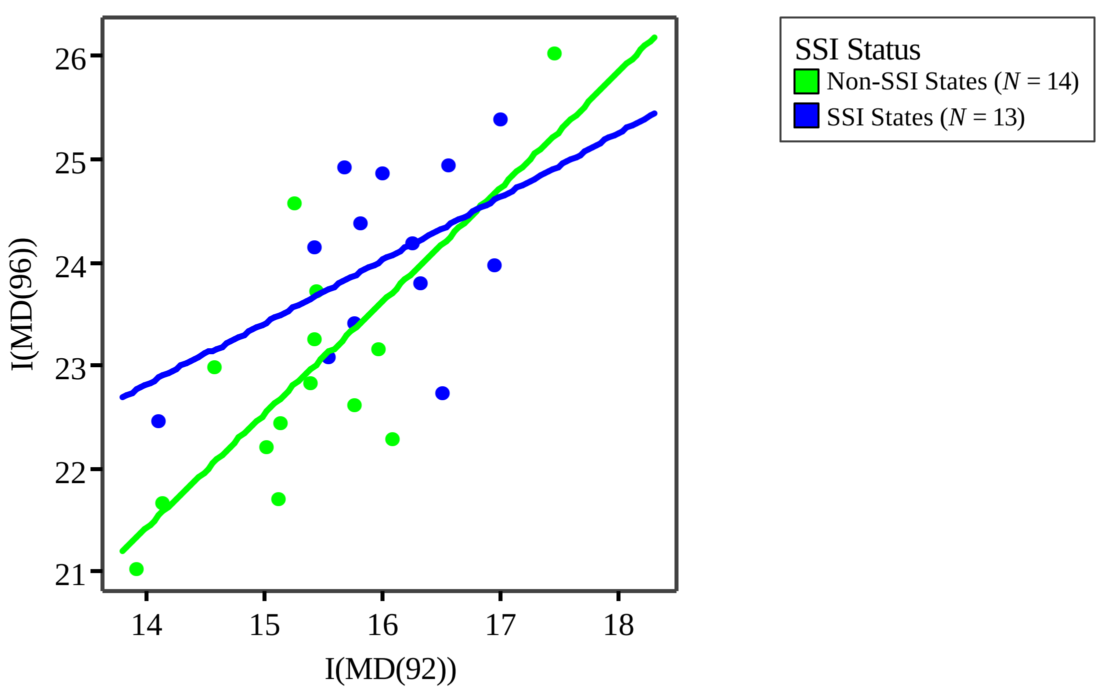
Subsample. The results of the linear regression analysis for the subsample of 13 SSI states and 14 non-SSI states are presented in Table 6.44 and Figure 6.18. As with the total sample, the 1992 mathematical discourse measure was significantly related to the 1996 measure (β = .43, t = 2.23, p < .05). In the subsample, SSI status did not add anything to the prediction of mathematical discourse in 1996 (β = .28, t = 1.45, p = .16). The correlation of the two predictors was .39 (p < .05).

Table 6.44
Predicting $I_{MD(96)}$ from $I_{MD(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 4

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{MD(92)}$	1.02	0.20	.72	.52	26.97*		
Step 2							
$I_{MD(92)}$	0.90	0.21	.64				
SSI status	0.54	0.36	.22	.56	15.33*	.04	2.29

*p < .01

Figure 6.18. Relationship between $I_{MD(92)}$ and $I_{MD(96)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, Grade 4.



Change from 1990 to 1996

Grade 8

Our ability to evaluate change from 1990 to 1996 is limited because of the lack of comparable items across the three years. In addition, the three-point trend sample includes only 17 SSI states and 11 non-SSI states.

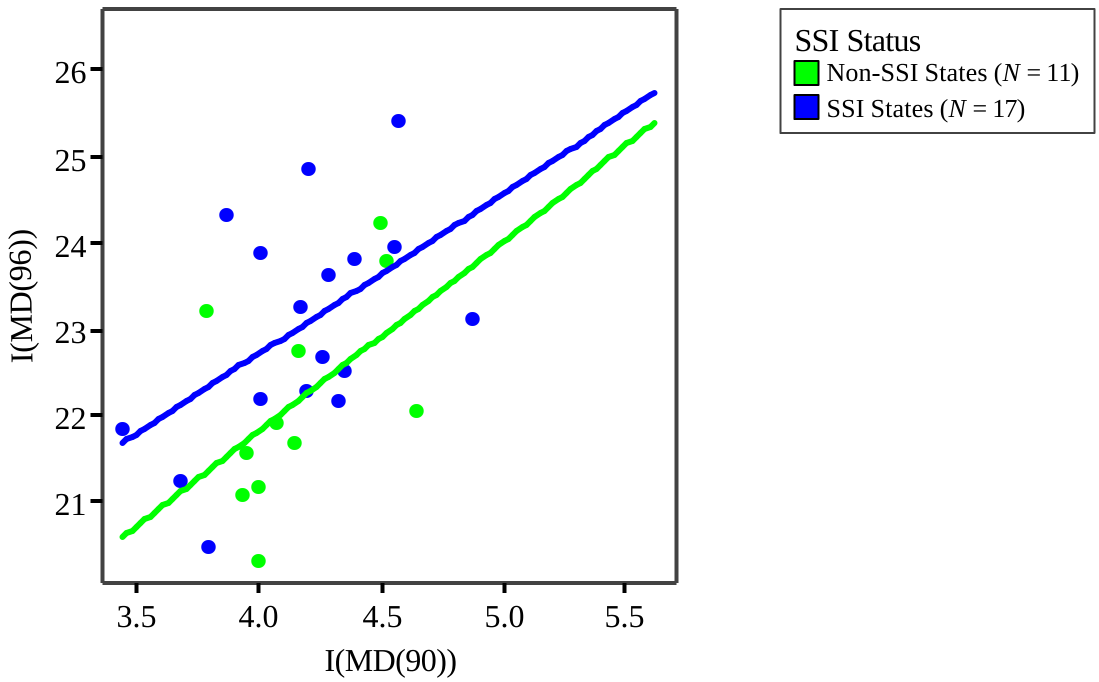
For the total sample, the regression analysis found that $I_{MD(96)}$ was a function of both SSI (β = .32, t = 1.96, p < .10) and $I_{MD(90)}$ (β = .48, t = 2.92, p < .01). The scatterplot is presented in Figure 6.19.

Table 6.45
Predicting $I_{MD(96)}$ from $I_{MD(90)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_{Δ}
Step 1							
$I_{MD(92)}$	2.00	0.71	.48	.24	8.00**		
Step 2							
$I_{MD(92)}$	1.96	0.67	.48				
SSI status	.84	0.43	.32	.34	6.35**	.10	3.83*

*p < .10; **p , .01

Figure 6.19. Relationship between $I_{MD(90)}$ and $I_{MD(96)}$ for SSI and non-SSI states.



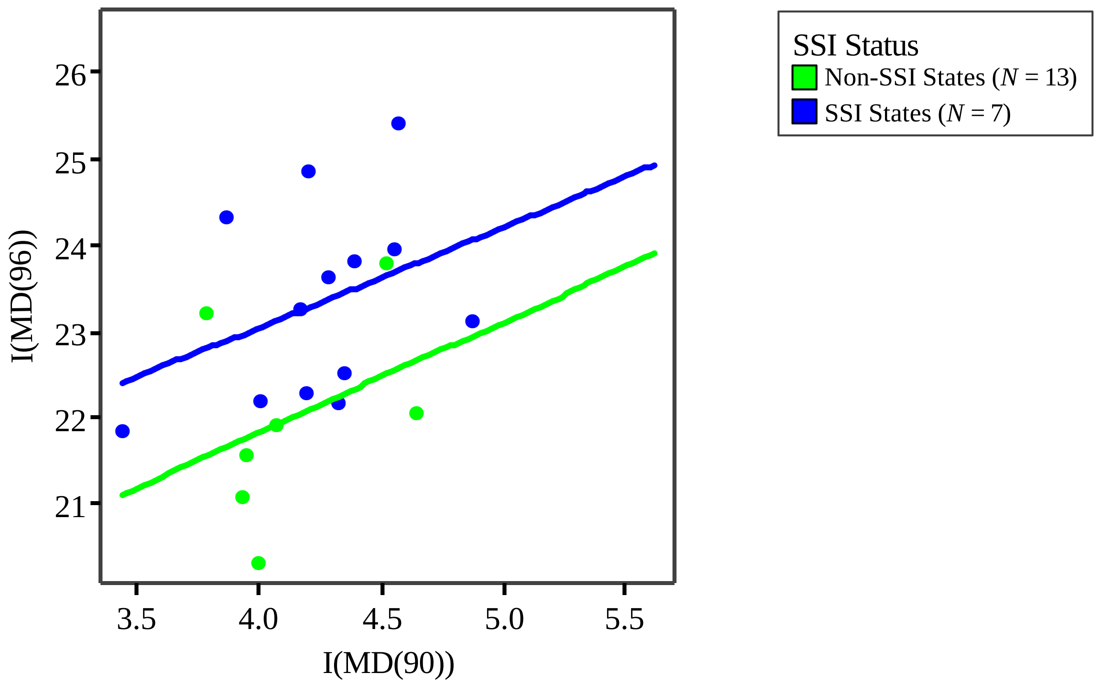
Subsample. For the subsample of 20 states that participated all three years and consistently followed the participation rate guidelines, results of the linear regression are in Table 6.46. In the analysis, $I_{MD(90)}$ is related to $I_{MD(96)}$ in Step 1, but in Step 2, SSI status is the only significant predictor. The correlation of the two predictors is .17. The scatterplot is shown in Figure 6.19.

Table 6.46
Predicting $I_{MD(96)}$ from $I_{MD(90)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_{Δ}
Step 1							
$I_{MD(90)}$	1.50	0.83	.39	.15	3.24*		
Step 2							
$I_{MD(90)}$	1.20	0.76	.31				
SSI status	1.20	0.53	.45	.35	4.64**	.20	5.26**

*p < .10, **p , .05

Figure 6.20. Relationship between $I_{MD(90)}$ and $I_{MD(96)}$ for SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.



Summary of Results

1996

Grade 8

In 1996, SSI states averaged significantly higher than non-SSI states, although the difference was small, on $I_{MD(96)}$, the 9-item mathematical discourse scale. The difference between SSI and non-SSI states on the nine individual items of the mathematical discourse scale was not statistically significant for the total sample, though it was for the subsample at $p < .10$.

Variability among SSI states and non-SSI states was relatively large compared to the difference between the two groups of states.

On average, students had teachers who assessed student progress through written responses and projects and presentations between "Once or twice a year" and "Once or twice a month."

Grade 4

At grade 4, the SSI states averaged significantly higher than non-SSI states on $I_{MD(96)}$, the 9-item mathematical discourse scale. SSI and non-SSI states did not differ significantly on the nine individual items of the scale.

Variability among SSI states and non-SSI states was relatively large compared to the difference between the two groups of states.

1992

Grade 8

In 1992, SSI states averaged slightly, but not significantly, higher than non-SSI states on the $I_{MD(92)}$ and $I_{MD4(92)}$.

Grade 4

At grade 4, SSI states averaged significantly higher than non-SSI states on $I_{MD(92)}$. They also averaged higher on $I_{MD4(92)}$, but the difference was not statistically significant.

1990

The 1990 teacher questionnaire included only two items related to mathematical discourse. The SSI states as a group did not differ from the non-SSI states on $I_{MD(90)}$.

Change Across Time

Two-point trend: 1992-1996

I_{MD4}

A scale of four mathematical discourse items was used to evaluate changes from 1992 to 1996. At grade 8, both SSI and non-SSI states increased in I_{MD4}, and SSI states scored higher than non-SSI across the two years. There was no evidence that SSI states increased more than non-SSI states. At grade 4, both SSI and non-SSI states increased from 1992 to 1996, and the SSI states scored higher across both years. As at grade 8, there was no evidence that the SSI states increased more than the non-SSI states.

$I_{MD(92)}$ and $I_{MD(96)}$

The 1992 measure of mathematical discourse was significantly related to the 1996 measure at both grade 4 and grade 8, but SSI status was not. However, SSI status was moderately related to the mathematical discourse measure in 1992.

Three-point trend: 1990-1996

In the regression analyses, SSI status was related to the 1996 mathematical discourse measure. In 1990, the mathematical discourse scale included only two items, so it was not a very strong indicator of student opportunities for mathematical discourse. In 1990, SSI status was weakly related to the discourse measure.

Teachers' Knowledge of the NCTM Standards—I_S

The Importance of Teacher Knowledge of the National Council of Teachers of Mathematics' *Standards*

On March 21, 1989, the National Council of Teachers of Mathematics (NCTM) publicly released the *Curriculum and Evaluation Standards for School Mathematics* for the first time in a flurry of well-orchestrated press coverage. NCTM, a professional organization, produced a policy document that conveyed a vision for both mathematical content and instruction in response to the need for reform in mathematics education that over the next decade became the "centerpiece of a broad reform movement in education" (McLeod, Stake, Schappelle, Mellissinos, & Gierl, 1996, p. 120). The *Curriculum and Evaluation Standards,* followed by two other documents—*Professional Standards for Teaching Mathematics* (NCTM, 1991) and *Assessment Standards for School Mathematics* (NCTM, 1995)—was a conscientious strategy by the NCTM leadership to advance the learning of more and somewhat different mathematics by all students. The document advocated a substantial change in state curriculum guidelines and at least some documented superficial change by textbook publishers (McLeod, Stake, Schappelle, Mellissinos, & Gierl, 1996; Romberg & Webb, 1993). Four years after the release of the NCTM *Standards,* a national survey indicated that teachers of mathematics who were well aware of the NCTM *Standards* varied by grade—18% in grades 1-4, 28% in grades 5-8, and 56% in grades 9-12 (Weiss, Matti, & Smith, 1994). In a later survey, the 50 state supervisors of mathematics estimated that, on the average, the NCTM *Standards* had had an impact on all grades, but that the greatest impact was on Grades K-4 (McLeod, Stake, Schappelle, Mellissinos, & Gierl, 1996). For teachers simply to declare knowledge of the NCTM *Standards* says very little about their classroom practices and whether what they do in their classrooms is aligned with the vision of the document. However, the percentage of teachers who acknowledged that they know about the NCTM *Curriculum and Evaluation Standards* demonstrates at least the magnitude of their awareness of the most important mathematics education reform document published in recent years. At a minimum, the percentages referred to above reflect efforts within the states toward reform by indicating those states in which teachers are professionally informed.

1996

The 1996 State NAEP teacher questionnaire included an item on teachers' knowledge of the NCTM *Standards*. The item, with responses, was:

> How knowledgeable are you about the National Council of Teachers of Mathematics (NCTM) *Curriculum and Evaluation Standards for School Mathematics*?
>
> *Response options*
>
> Very knowledgeable
> Knowledgeable
> Somewhat knowledgeable
> I have little or no knowledge.

I_S, the indicator of teachers' knowledge of the NCTM *Standards*, presents data on the impact of the *Standards* on student achievement. In our analysis, response options were scaled from 1 to 4, with 1 for "I have little or no knowledge" to 4 for "Very knowledgeable."

Grade 8

Teachers' knowledge of the NCTM *Standards.* Table 6.47 presents the mean and standard deviation on the NCTM *Standards* indicator for all SSI and non-SSI states that participated in the 1996 State NAEP. The mean for the SSI states was significantly higher than the mean for the non-SSI states ($t = 2.44$, $df = 38$, $p < .05$). Comparisons for other samples and subsamples are presented in Table 6D.1 in the Appendix.

Table 6.47
Mean and Standard Deviation of SSI and Non-SSI States on $I_{S(96)}$

	SSI States *N* = 22	Non-SSI States *N* = 18
Mean	2.72	2.59
Standard Deviation	0.21	0.13

The individual state means on $I_{S(96)}$ are presented in Figure 6.20, with states ranked from highest to lowest. Of the ten highest states, eight are SSI states: Vermont, Massachusetts, Connecticut, Delaware, Maine, Rhode Island, Kentucky, and Virginia; of the ten lowest, four are SSI states: Texas, Louisiana, New Mexico, and Michigan. See Table 6D.2 in the Appendix for the state means.

Figure 6.20. State means on $I_{S(96)}$ ordered from highest to lowest, grade 8.

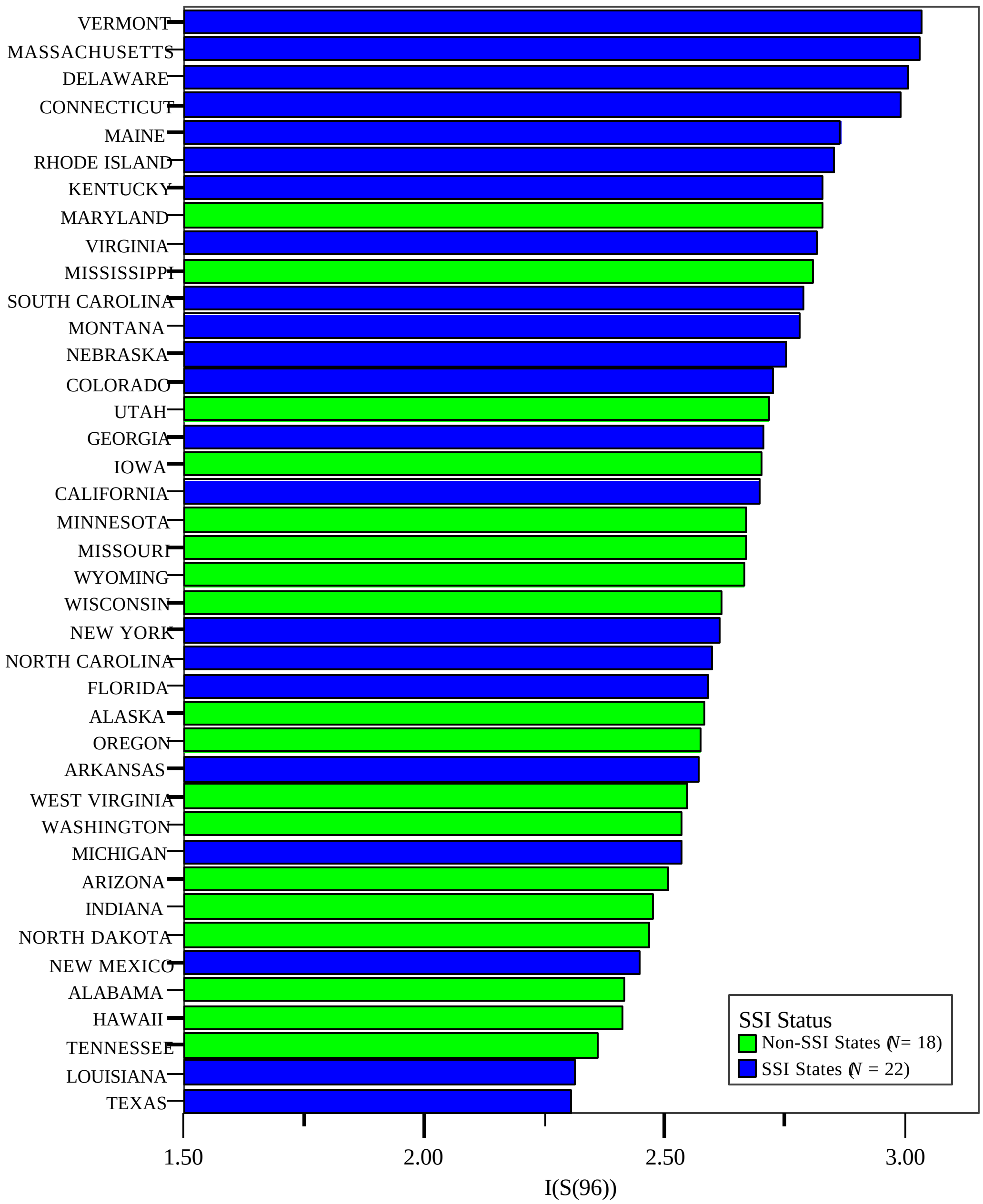


Table 6.48 shows the percentage of students in SSI states and in non-SSI states with a teacher who selected each response. In 1996, SSI states had a larger proportion of students with teachers who were very knowledgeable about the NCTM *Standards*. In both SSI and non-SSI states, roughly one out of ten students had teachers with little or no knowledge of the *Standards.*

Table 6.48
Percentage of Students in Each Category of Teachers' Knowledge of the NCTM Standards, *Grade 8, 1996*

	Little/No Knowledge	Somewhat Knowledgeable	Knowledgeable	Very Knowledgeable
Non-SSI states	11.9%	33.8%	38.9%	16.4%
SSI states	10.7%	29.1%	37.7%	22.5%

Grade 4

Table 6.49 presents the mean and standard deviation on I_S for all the SSI and non-SSI states that participated in the 1996 State NAEP. The mean for the SSI states is slightly higher than the mean for the non-SSI states (t = 1.76, df = 41, p < .10). Comparisons for other samples and subsamples are presented in Table 6D.1 in the Appendix.

Table 6.49
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RC(96)}$

	SSI States N = 23	Non-SSI States N = 20
Mean	1.98	1.88
Standard Deviation	0.21	0.15

The individual state means on $I_{S(96)}$ at grade 4 are presented in Figure 6.16, with states ranked from highest to lowest. Of the ten highest states, seven are SSI states: Vermont, Maine, Delaware, Rhode Island, Massachusetts, Colorado, and Kentucky; of the ten lowest, four are SSI states: Texas, New York, Florida, and California. See Table 6D.2 in the Appendix for the state means.

Figure 6.22. State means on $I_{S(96)}$ ordered from highest to lowest, grade 4.

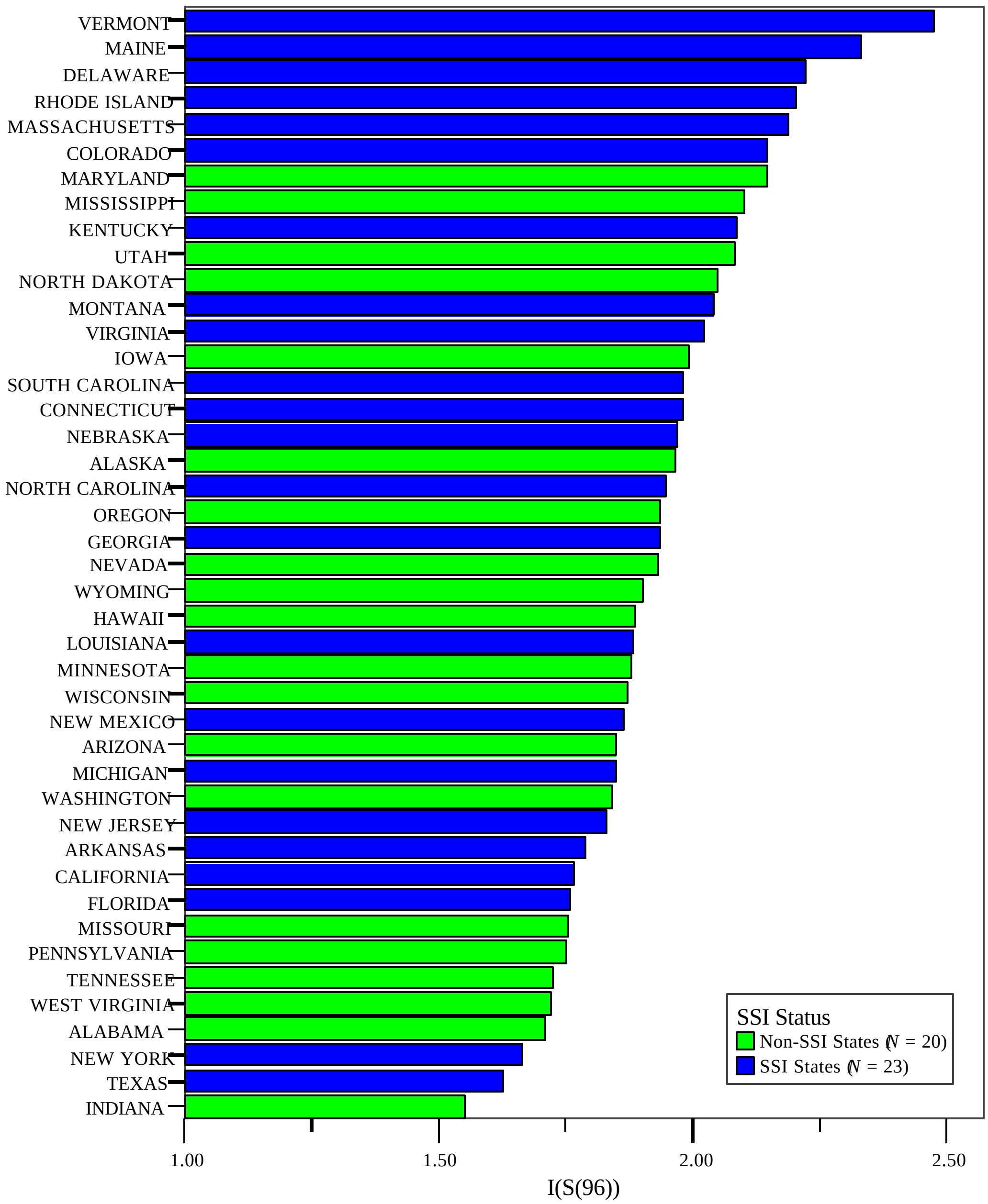


Table 6.50 presents the percentage of students with teachers who selected each response option in SSI and non-SSI states. Percentages for individual states can be found in Table 6D.2 in the Appendix.

Table 6.50
Percentage of Students in Each Category of Teachers' Knowledge of the NCTM Standards, *Grade 4, 1996*

	Little/No Knowledge	Somewhat Knowledgeable	Knowledgeable	Very Knowledgeable
Non-SSI states	36.3%	36.7%	17.9%	5.2%
SSI states	40.2%	36.4%	20.3%	7.1%

In 1996, SSI states had a slightly larger proportion of students with teachers who were very knowledgeable about the NCTM *Standards.* In both SSI and non-SSI states, roughly two out of five grade 4 students had teachers with little or no knowledge of the *Standards.*

Summary of Results

Grades 4 and 8

In the total 1996 sample, the mean for the SSI states is significantly higher than the mean for the non-SSI states in both grade 4 and grade 8.

In both SSI and non-SSI states, grade 8 students have teachers who are more knowledgeable about the NCTM *Standards* than grade 4 students. In grade 8, about 50% of the students had teachers who said they were knowledgeable or very knowledgeable about the *Standards*; in grade 4, just 25% had teachers who said they were knowledgeable or very knowledgeable.

The ten highest scoring states included eight SSI states at grade 8 and seven SSI states at grade 4; the ten lowest scoring states included four SSI states at both grade 8 and grade 4. Six SSI states, Vermont, Massachusetts, Delaware, Rhode Island, Maine, and Kentucky, were in the top ten states at both grade levels; Texas was the only SSI state in the bottom ten at both grade levels.

The item concerning teachers' knowledge of the NCTM S*tandards* was first used on the 1996 NAEP questionnaire, so comparisons with 1992 and 1990 on this indicator are not possible.

Time Spent in Professional Development During the Last Year—I_{PD}

The Importance of Time Spent in Professional Development

It is of vital importance that the professional development of mathematics teachers be an on-going effort. The John Glenn Commission in its report to the nation, *Before It's Too Late*, released September 27, 2000 (http://www.ed.gov/inits/Math/glenn/), described professional development as:

> . . . a planned, collaborative, educational process of continuous improvement for teachers that helps them do five things:
> (1) deepen their knowledge of the subject(s) they are teaching;
> (2) sharpen their teaching skills in the classroom;
> (3) keep up with developments in their fields, and in education generally;
> (4) generate and contribute new knowledge to the profession; and
> (5) increase their ability to monitor students' work, so they can provide constructive feedback to students and appropriately redirect their own teaching (p. 15).

The commission recognized that teachers are rarely afforded extended time periods for engaging in their own educational experiences. Instead, they are subjected to "in-service events that are no more substantive than a broad-brush overview of this semester's teaching fad" (p. 18).

A number of studies indicate that professional development that provides a substantial number of contact hours and is sustained over a long period of time results in a significant impact on teaching practices (Garet, Birman, Porter, Desimone, Herman, & Yoon, 1999; Little, 1993; Loucks-Horsley et.al., 1987; Loucks-Horsley, S., Stiles, K., & Hewson, P., 1996; Sparks, 1994; Loucks-Horsley, Hewson, Love, & Stiles, 1998). Recognizing the importance of professional development, the SSI states allocated the largest single portion of their budgets to professional development (Corcoran, Shields, and Zucker, 1998; Zucker, Shields, Adelman, Corcoran, & Goertz, 1998). However, in 1993, early in the implementation of the program, a national survey of mathematics and science teachers indicated that less than half of science and mathematics teachers had spent more than 15 hours on in-service education in the last three years—32% grades 1-4, 41% grades 5-8, and 55% grades 9-12 (Weiss, 1994). Many other factors in addition to duration, such as the quality of the activity, degree of engagement, and focus on content, contribute to professional development that leads to improved teaching. The amount of time that a teacher has recently spent in professional development produces one indicator that is at least a necessary condition for on-going professional development.

1996

Question 42 of the 1996 grade 8 teacher questionnaire and question 57 of the grade 4 questionnaire asked:

> During the last year, how much time in total have you spent in professional development workshops or seminars in mathematics or mathematics education? Include attendance at professional meetings and conferences, district-sponsored workshops, and external workshops.
>
> *Response options*
> None
> Less than 6 hours
> 6-15 hours
> 16-36 hours
> More than 35 hours

Response options were coded from 1 to 5: "None" equaled 1 and "More than 35 hours" equaled 5. Since the item asked specifically about staff development activities of the previous year, it might have been particularly sensitive to the effects of the SSI program.

Grade 8

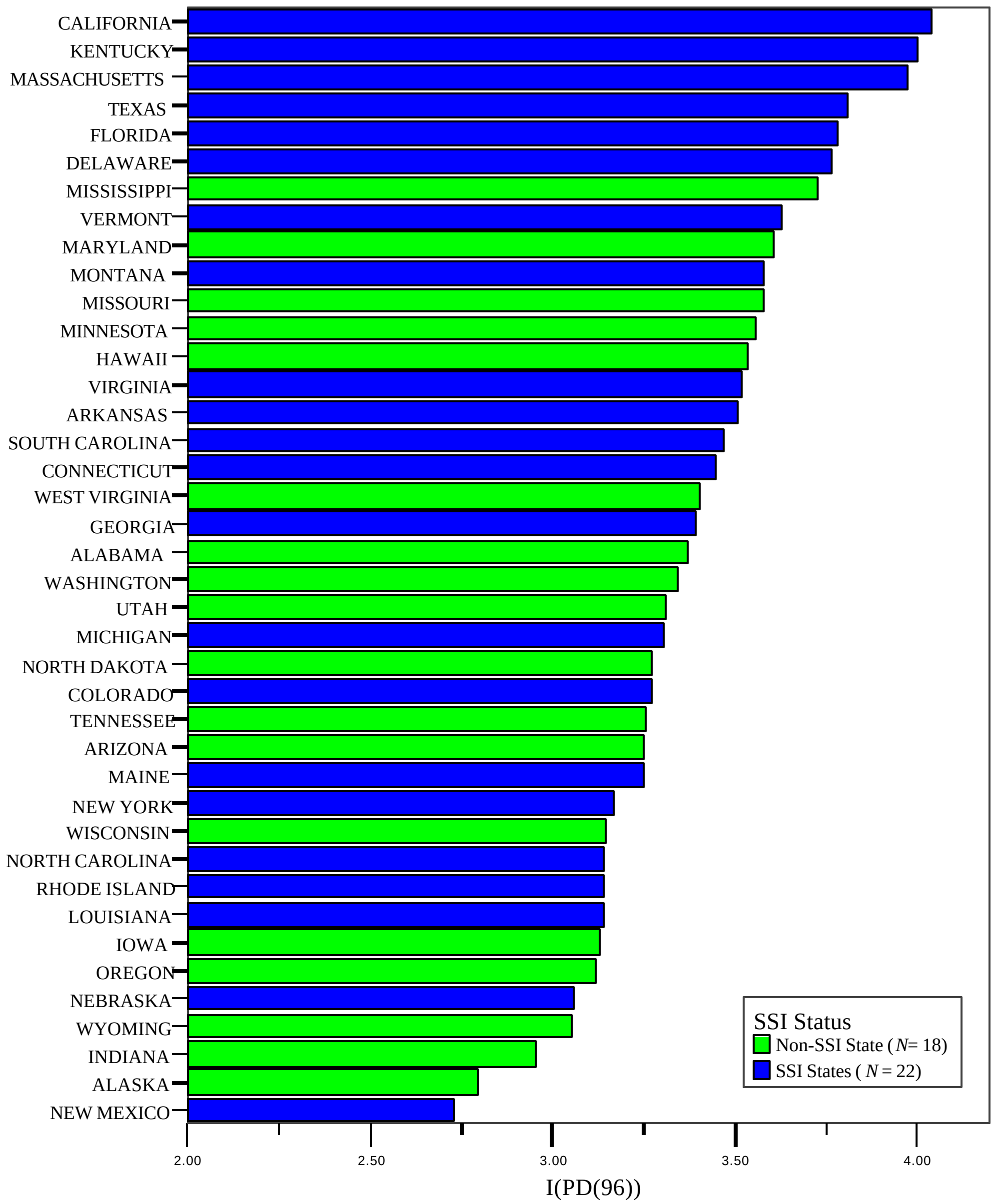
Time spent in professional development during the last year. The mean and standard deviation on $I_{PD(96)}$ for all SSI and non-SSI states in 1996 is presented in Table 6.51. In the total sample, the mean for the SSI states was significantly higher than the mean for the non-SSI states (t = 1.72, df = 38, p < .10). Comparisons for all samples and subsamples are reported in Table 6E.1 in the Appendix. No statistically significant differences were found in any of the other samples or subsamples.

Table 6.51
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(96)}$, Grade 8

	SSI States *N* = 22	Non-SSI States *N* = 18
Mean	3.46	3.30
Standard Deviation	0.34	0.24

The individual state means on $I_{PD(96)}$ are presented in Figure 6.23, with states ordered from highest to lowest. Of the ten highest states, eight are SSI states: California, Kentucky, Massachusetts, Texas, Florida, Delaware, Vermont, and Montana; of the ten lowest, five are SSI states: New Mexico, Nebraska, Louisiana, Rhode Island, and North Carolina. See Table 6E.2 in the Appendix for the individual state means.

Figure 6.23. State means on I(PD(96)) ordered from highest to lowest, grade 8.



Grade 4

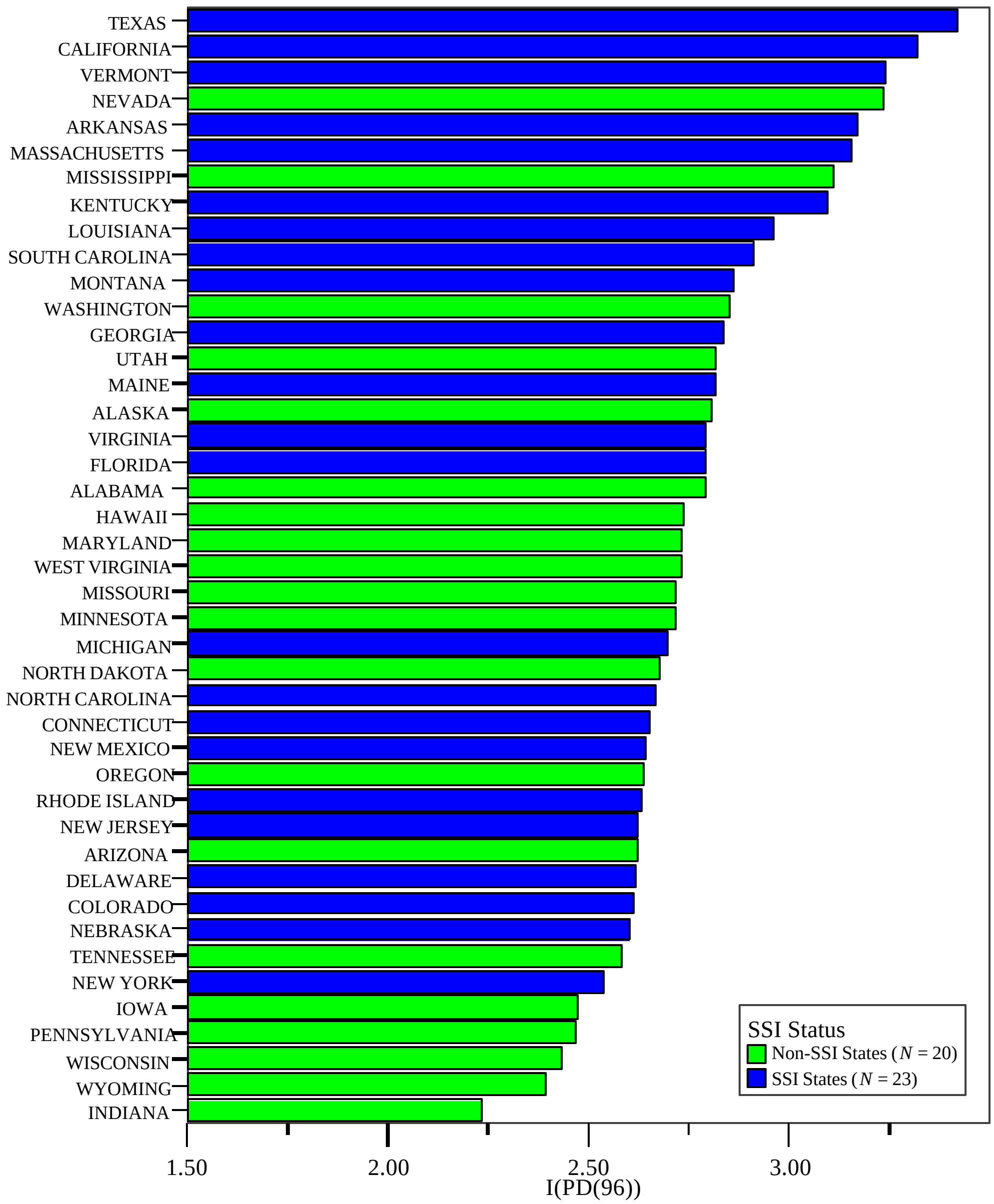
The state mean and standard deviation of $I_{PD(96)}$ by SSI status is presented in Table 6.52. At grade 4, the mean for the SSI states is significantly higher than the mean for the non-SSI states (t = 2.22, df = 41, p < .05). Comparisons for all samples and subsamples are found in Table 6E.1in the Appendix.

Table 6.52
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(96)}$*, Grade 4*

	SSI States *N* = 23	Non-SSI States *N* = 20
Mean	2.85	2.69
Standard Deviation	0.26	0.23

Individual state means at grade 4 are presented in Figure 6.24. Of the ten highest states, eight are SSI states: Texas, California, Vermont, Arkansas, Massachusetts, Kentucky, Lousiana, and South Carolina; of the ten lowest, four are SSI states: New York, Nebraska, Colorado, and Delaware. See Table 6E.2 in the Appendix for the means for each SSI and non-SSI state.

Figure 6.24. State means on I(PD(96)) ordered from highest to lowest, grade 4



.

1992

In 1992, the wording of the item on time spent in professional development was slightly different from that of 1996, but the response options were the same. The 1992 item is shown below.

> During the last year, how much time in total have you spent on in-service education in mathematics or the teaching of mathematics? Include attendance at professional meetings and conferences, workshops and courses.
>
> *Response options*
> None
> Less than 6 hours
> 6-15 hours
> 16-36 hours
> More than 35 hours

Grade 8

Time spent in professional development during the last year. In 1992, the 22 SSI states averaged just .03 higher on $I_{PD(92)}$ than the 19 non-SSI states, as shown in Table 6.53 (t = 0.97, df = 39, p = .33). See Table 6E.3 in the Appendix for comparisons for all samples and subsamples and Table 6E.2 for individual state means.

Table 6.53
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(92)}$, Grade 8

	SSI States N = 22	Non-SSI States N = 19
Mean	3.28	3.25
Standard Deviation	0.22	0.21

Grade 4

At grade 4 in 1992, the means for SSI and non-SSI states were practically identical, as shown below in Table 6.54. Comparisons for other samples and subsamples can be found in Table 6E.3 of the Appendix and individual state means in 6E.2.

Table 6.54
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(92)}$, Grade 4

	SSI States N = 22	Non-SSI States N = 19
Mean	2.58	2.59
Standard Deviation	0.21	0.22

1990

The wording on the 1990 questionnaire was the same as in 1992:

> During the last year, how much time in total have you spent on in-service education in mathematics or the teaching of mathematics? Include attendance at professional meetings and conferences, workshops and courses.
>
> *Response options*
> None
> Less than 6 hours
> 6-15 hours
> 16-36 hours
> More than 35 hours

Grade 8

In 1990, the mean of the SSI states on $I_{PD(90)}$ was slightly higher than the mean of the non-SSI states, but this difference was not statistically significant (t = 1.08, df = 35, p = .29). See Table 6.55. Results for other samples and subsamples are listed in Table 6E.4 in the Appendix, and individual state means are in Table 6E.2.

Table 6.55
Mean and Standard Deviation of SSI and Non-SSI States on $I_{PD(90)}$, Grade 8

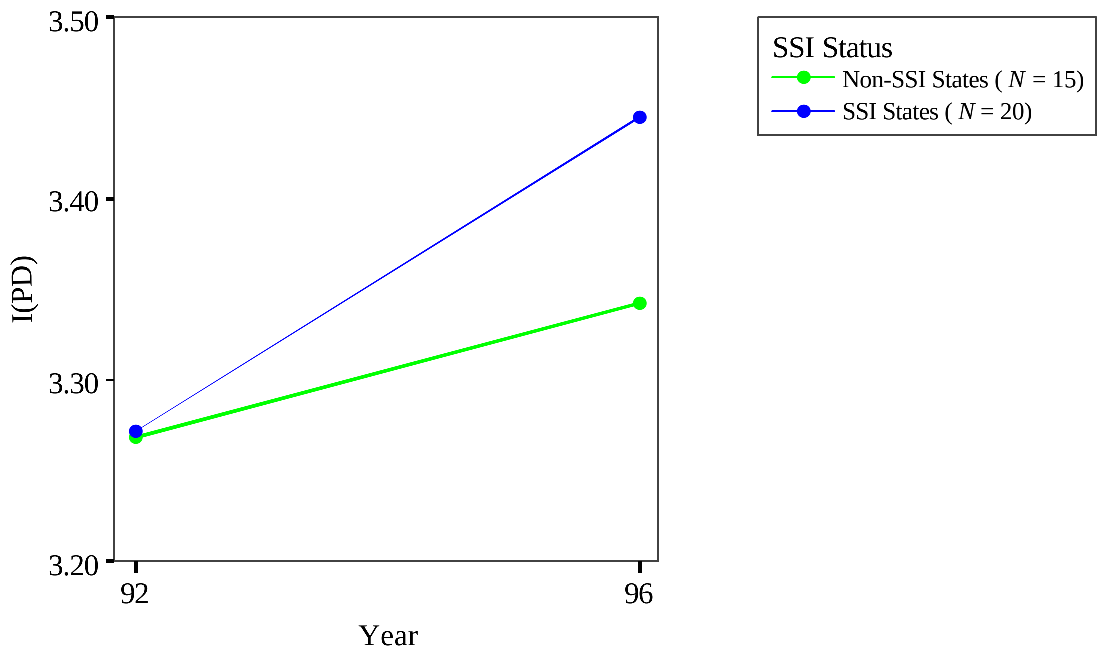
	SSI States N = 20	Non-SSI States N = 17
Mean	2.98	2.89
Standard Deviation	0.31	0.39

Change from 1992 to 1996

Grade 8

Two-point trend sample. Twenty SSI states and 15 non-SSI states participated in the State NAEP in both 1992 and 1996. The 2x2 repeated measures analysis of variance found a significant effect by year ($F = 4.55$, $df = 1,33$, $p < .05$). Neither the effect for SSI ($F = 0.59$, $df = 1,33$, $p = .48$), nor the interaction effect of year by SSI ($F = 0.78$, $df = 1,33$, $p = .38$) were statistically significant. Figure 6.25 presents the graph of the means for the SSI and non-SSI states in 1992 and 1996. The two groups of states are at the same point in 1992, while both are higher in 1996. Although the increase on this item in SSI states is greater than that in the non-SSI states, the difference is relatively small.

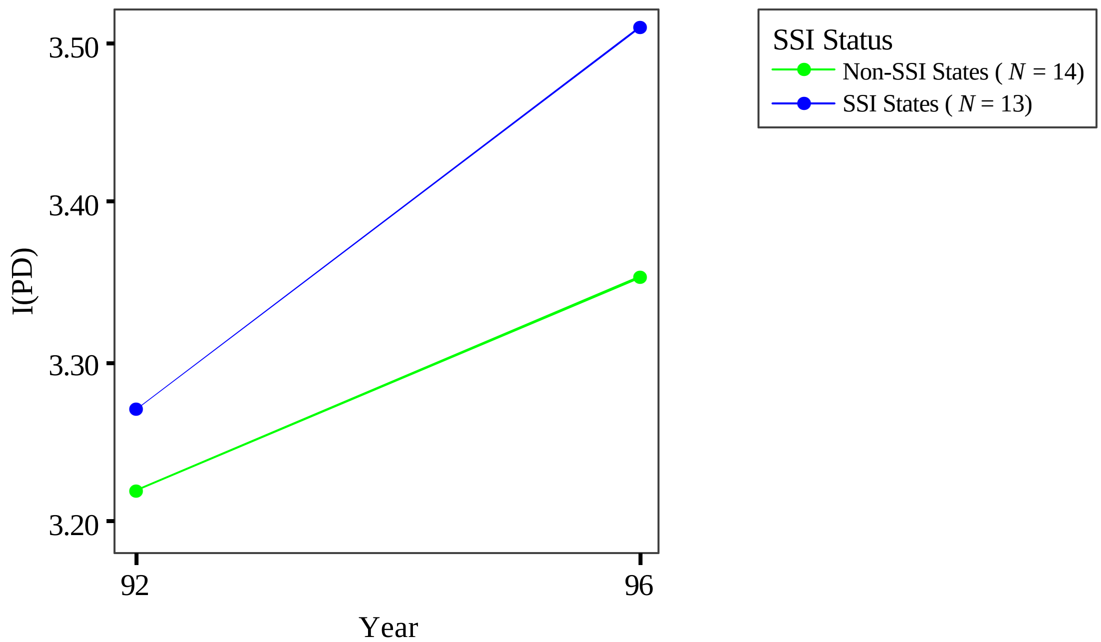
Figure 6.25. Mean for last year's staff development in 1992 and 1996, grade 8.



Subsample. When the analysis is repeated with only those states that followed the NCES participation rate guidelines, the conclusions are the same as for the full sample. The effect for year is statistically significant ($F = 6.28$, $df = 1,23$, $p < .05$), but the effect for SSI and the interaction effect of year by SSI are not.

The graph for the subsample looks similar to that for the total sample, except that the SSI states are slightly above the non-SSI states in 1992. (See Figure 6.26.) Both the SSI states and the non-SSI states increased in staff development participation from 1992 to 1996. While it seems that the SSI states increased slightly more, the difference is not statistically significant.

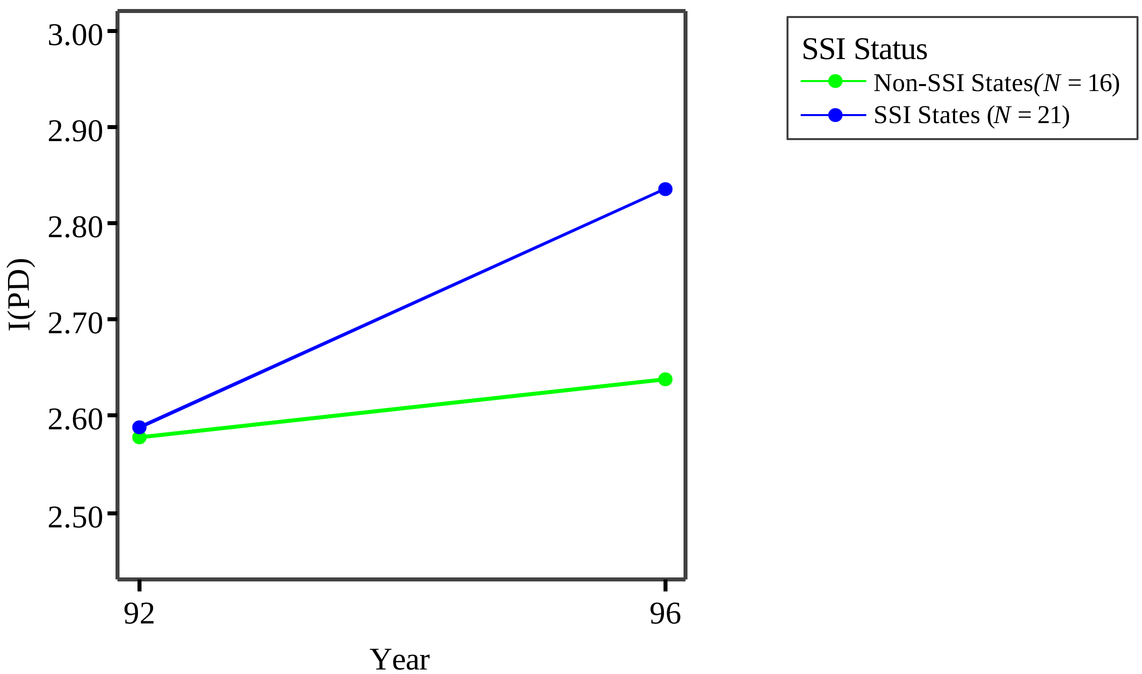
Figure 6.26. Mean for last year's staff development in 1992 and in 1996 for states following the NCES participation rate guidelines, Grade 8.



Grade 4

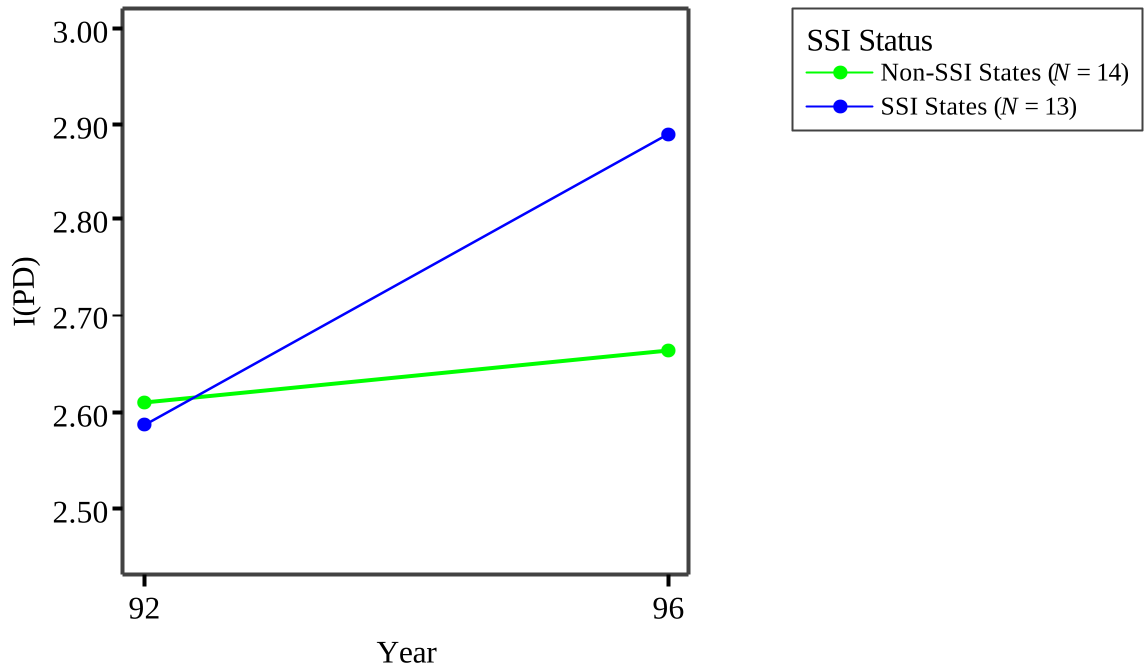
Two-point trend sample. The 2x2 repeated measures analysis of variance for the data from grade 4 indicated a significant effect for year (F = 19.51, df = 1,35, p < .01) and a significant interaction of year and SSI status (F = 7.15, df = 1,35, p < .05). As Figure 6.27 shows, the mean for the SSI states and the non-SSI states was almost the same in 1992. From 1992 to 1996, both groups increased in the amount of staff development time during the last year, and the SSI states, as a group, showed a greater increase. The effect for SSI status was not statistically significant (F = 2.36, df = 1,35, p = .13).

Figure 6.27. Mean for last year's staff development in 1992 and in 1996, grade 4.



Subsample. For the subsample of states that followed the NCES participation rate guidelines, the results of the ANOVA are similar to those for the full sample. Figure 6.28 shows the subsample means for the SSI and non-SSI states. The effect for year was statistically significant (F = 19.04, df = 1,25, p < .01) and the interaction of year by SSI status was also significant (F = 9.37, df = 1,25, p < .01). Overall, SSI status was not significant (F = 1.67, df = 1,25, p = .21).

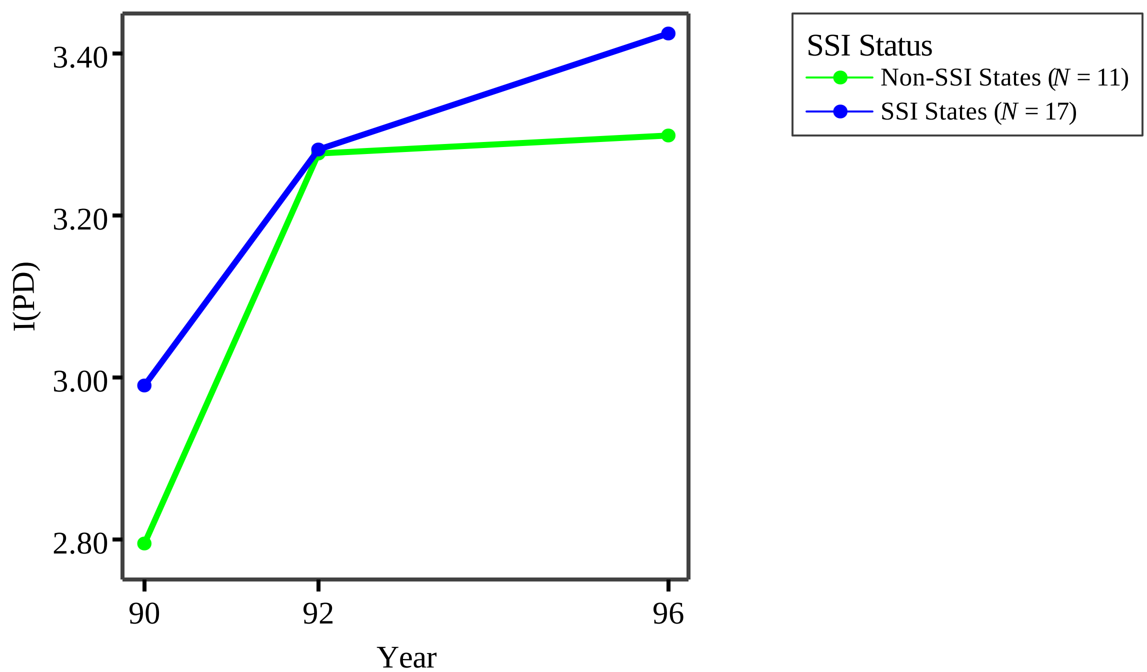
Figure 6.28. Mean for last year's staff development in 1992 and 1996, for states that met the NCES participation rate guidelines, grade 4.



Change Across 1990, 1992, and 1996

Three-point trend sample. A 2x3 repeated measures analysis of variance was used to examine the effect of year and SSI status on the amount of time spent in professional development during the last year. For the 17 SSI states and 11 non-SSI states that participated consistently for all three testing years, the ANOVA found a significant effect for year (F = 28.64, df = 1,25, p < .01), but no significant effects for SSI status (F = 1.66, df = 1, 26, p = .21), or the interaction of year and SSI status (F = 1.20, df = 2,52, p = .31). Figure 6.29 presents the graph of the SSI and non-SSI state means across the three years.

Figure 6.29. Mean for last year's staff development in 1990, 1992, and 1996, grade 8.

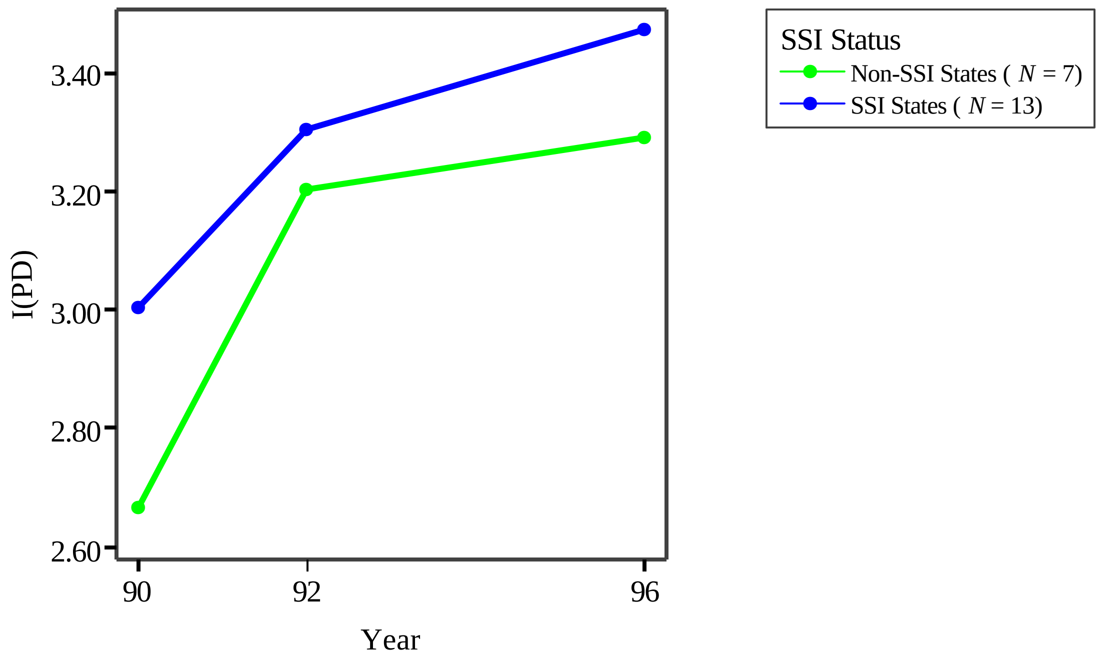


In 1990, the mean for SSI and non-SSI states on this item was slightly under 3, indicating that, on average, students had teachers who spent 6-15 hours in staff development during the last year. In 1996, the mean was around 3.4, indicating that students had teachers who averaged between 6-15 hours and 16-35 hours of staff development over the last year.

Subsample. For the subsample of states that followed the NCES participation rate guidelines, the 2x3 repeated measures analysis of variance found a significant effect for year (F = 20.61, df = 2,17, $p < .01$), replicating the result for the total sample. For the subsample, the main effect for SSI was also statistically significant (F = 4.28, df = 1,18, $p < .10$), but the interaction of year and SSI status was not (F = 1.14, df = 2,36, p = .33). Figure 6.30 presents the means for the SSI and non-SSI states in the subsample over time.

As the graph indicates, the SSI states seemed to make steady progress. In contrast, the non-SSI states showed an increase from 1990 to 1992, and then seemed to level off.

Figure 6.30. Mean for last year's staff development in 1990, 1992 and 1996, for the SSI and non-SSI states following NAEP's participation rate guidelines, grade 8.



Summary of Results

1996

Grade 8

For the total sample, SSI states averaged slightly, but significantly higher than the non-SSI states. In other samples and subsamples, the difference was not statistically significant.

Grade 4

At grade 4, SSI states averaged significantly higher than non-SSI states in all samples and subsamples.

1992 and 1990

The overall means for the SSI and non-SSI states prior to 1990 were very close on this indicator. (Note: Grade 4 is not included in the 1990 State NAEP.)

Change Over Time

Two-Point Trend Sample, 1992-1996

Grade 8

Teachers spent more time in staff development in 1996 than they had in 1992. The interaction of year by SSI status was not statistically significant in either the total sample or the subsample.

Grade 4

Teachers spent more time in staff development in 1996 than they had in 1992. In addition, the SSI states showed a larger increase than the non-SSI states at grade 4.

Three-Point Trend Sample, 1990, 1992, and 1996

As with the two-point trend sample, the grade 8 means increase from 1990 to 1996, but there is no significant effect for SSI and no significant interaction. While both SSI and non-SSI groups increase across the time period, the SSI states seem to increase steadily, while the non-SSI states seem to have increased from 1990 to 1992 and then leveled off.

The Number of Reform-Related Topics Teachers Have Studied—I_{RT}

The Importance of Teachers' Study of Reform-Related Topics

Important if teachers are to make the transition from a traditional approach in teaching mathematics to approaches that are advanced in reform documents, such as those produced by NCTM (1989, 1991, and 1995), is for teachers to have a sound understanding of the pedagogical content knowledge of the reform methods. Effective professional development experiences help teachers develop in-depth knowledge of content, as well as pedagogical content knowledge (listening to students' ideas, posing questions, and attending to different needs of students) (Loucks-Horsley, Hewson, Love, & Stiles, 1998).

For teachers to change requires them to experience a number of factors. Teachers need to experience some dissatisfaction with traditional approaches, to be exposed to new methods, to have these methods modeled for them, to have the opportunity to experiment with them, and to be able to reflect on how the effectiveness of the methods will work with their own students (Webb, Heck, & Tate, 1996). There is some evidence that simply increasing the number of mathematics courses a teacher takes will reach a point of diminishing returns. However, there is a noticeable increase in student performance for each additional college mathematics course a teacher has taken up to five courses (Kilpatrick, Swafford, & Findell, 2001). One explanation for this is that teachers need specialized knowledge of mathematics and teaching practices beyond those taught in advanced mathematics classes.

A very strong driver of reforms advanced by NCTM and the NSF's SSI program is for all students to increase their knowledge of challenging mathematics (NCTM, 1989; National Science Foundation, 1997). Effective professional development for mathematics teachers will focus on mathematics and student thinking by focusing on students' problem-solving strategies and by studying actual examples of their work. Effective professional development for teachers working with students with diverse backgrounds requires them to organize their instruction to accommodate diversity in ways that are beneficial to their students. One important issue for teachers to understand is how learning mathematics is influenced by their students' ethnic, cultural, socioeconomic, and linguistic backgrounds and by gender (Croom, 1997). It is also important for teachers to know mathematics to be effective. But for them to teach challenging mathematics effectively to diverse student populations requires that they have recent pedagogical knowledge about how students learn mathematics.

1996

Questions 9-15 of the grade 8 and grade 4 1996 teacher questionnaires asked:

> Have you ever studied any of the following, either in college or university courses or in professional development workshops or seminars?
>
> Estimation
> Problem solving in mathematics
> Use of manipulatives (e.g., counting blocks of geometric shapes) in mathematics instruction
> Use of calculators in mathematics instruction
> Understanding students' thinking about mathematics
> Gender issues in the teaching of mathematics
> Teaching students from different cultural backgrounds.
>
> *Response options:*
> Yes
> No

For each student, $I_{RT(96)}$ was computed by simply counting the number of "Yes" answers out of the seven topics listed. Since all seven topics could reasonably have been included in the Statewide Systemic Initiatives for mathematics reform, it seemed reasonable to expect that teachers in SSI states would have studied these topics. One weakness of this indicator, however, is that the question did not refer specifically to what teachers had studied recently, but to whether teachers had "ever" studied the topic. With such a broad time-span, this indicator may not be sensitive to the effects of the SSI program. Internal consistency of this scale (i.e., Cronbach's coefficient alpha) was .69 for grade 8 and .70 for grade 4.

Grade 8

Number of reform-related topics teachers have studied. Table 6.56 presents the mean and standard deviation for all SSI and non-SSI states in the 1996 State NAEP on this indicator. The mean difference of 0.16 was statistically significant, with $t = 1.85$ ($df = 38$, $p < .10$). Statistical comparisons for other samples and subsamples are presented in Table 6F.1 in the Appendix. The difference between SSI and non-SSI states was statistically significant for all grade 8 samples and subsamples in 1996.

Table 6.56
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(96)}$, Grade 8

	SSI States *N* = 22	Non-SSI States *N* = 88
Mean	5.16	5.00
Standard Deviation	0.31	0.27

Individual state means on $I_{RT(96)}$ are presented in Figure 6.31, with states ranked from highest to lowest. Of the ten highest states, eight are SSI states: California, Nebraska, Connecticut, Florida, Georgia, Montana, Delaware, and Kentucky; of the ten lowest, three are SSI states: Maine, Vermont, and New York. The values of the individual state means are included in Table 6F.2 in the Appendix.

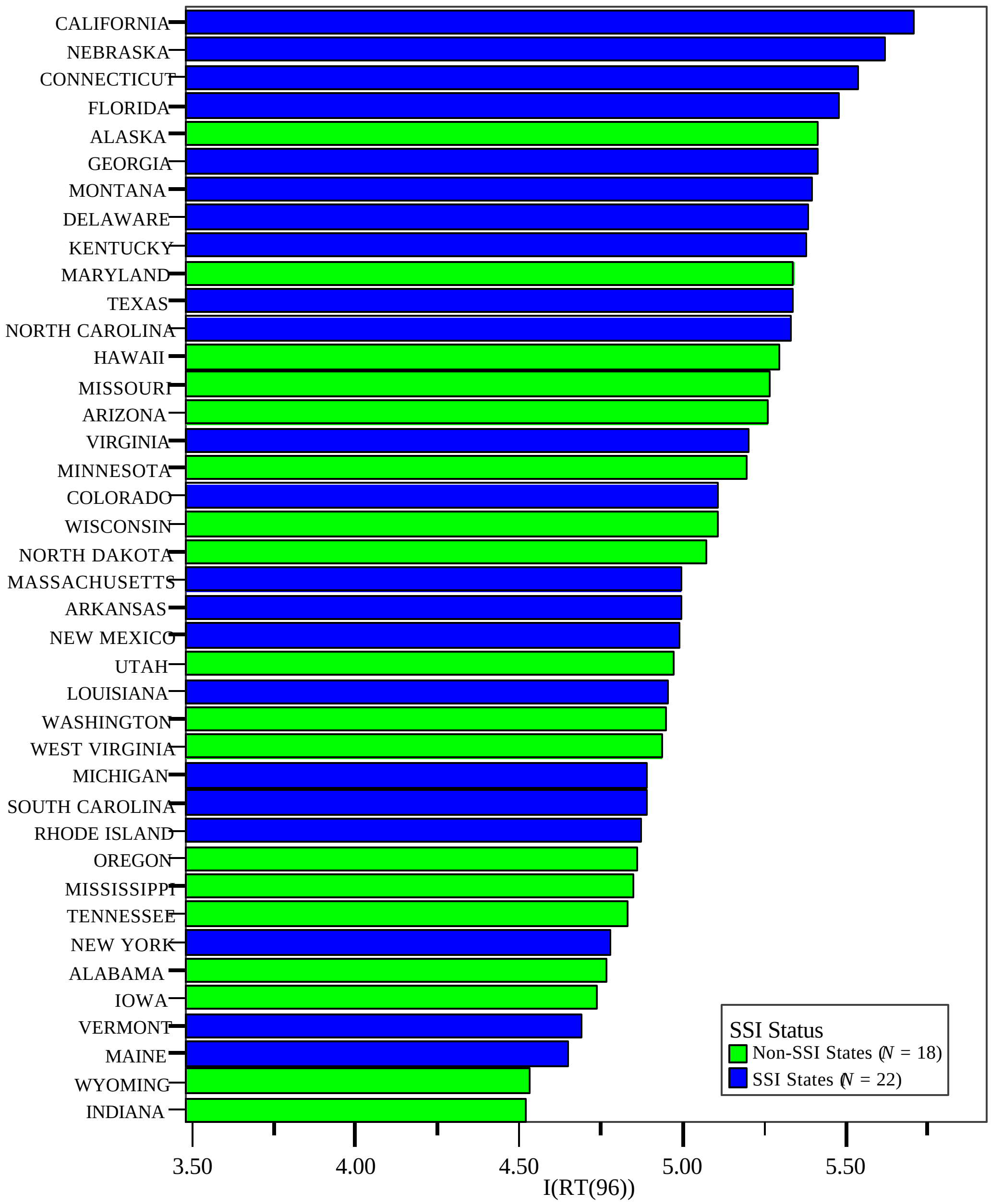
Individual items. Table 6.57 lists each of the seven topics along with the percentage of students with a teacher who had studied the topic at some time. Percentages for individual states are listed in Table 6F.2 in the Appendix.

Table 6.57
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 8, 1996

Topic	Mean Percentage of Students in SSI States *N* = 22	Mean Percentage of Students in Non-SSI States *N* = 18
Estimation	77.9	75.2
Problem solving	94.3	93.4
Manipulatives	91.1	87.9
Calculators	84.6	82.0
Students' thinking	71.2	66.6
Gender issues	53.3	51.9
Cultural differences	49.6	48.1

In 1996 in both SSI and non-SSI states, more than 90% of the students had teachers who had studied problem-solving in mathematics, and around 90% had teachers who had studied the use of manipulatives. Roughly 80 to 85% of the students had teachers who had studied the use of calculators, and about 75% of the students had teachers who had studied estimation. About 66% of the students had teachers who had studied students' thinking about mathematics. Roughly half had studied gender issues and cultural differences.

Figure 6.31. State means on $I_{RT(96)}$ ordered from highest to lowest, grade 8.



For each of the topics, the percentage of teachers in SSI states who responded affirmatively was slightly higher than the percentage in non-SSI states in 1996. However, the difference between the SSI and non-SSI states was very small. The topic with the largest difference was: Understanding students' thinking about mathematics (71% for SSI states, and 67% for non-SSI states).

Grade 4

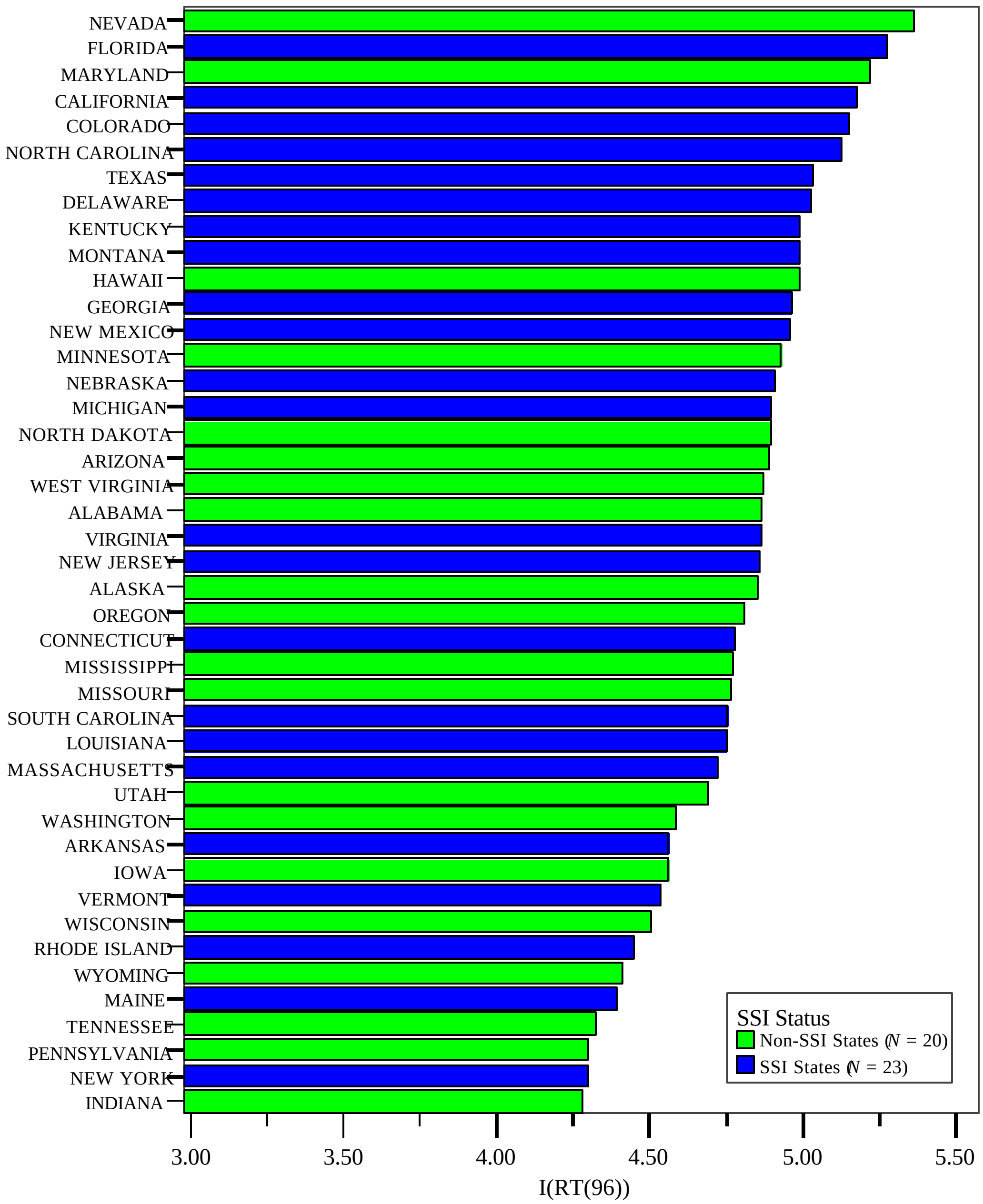
Number of reform-related topics teachers have studied. Table 6.58 presents the grade 4 results on $I_{RT(96)}$ for all SSI and non-SSI states in the 1996 State NAEP. The mean difference of 0.10 for SSI states and non-SSI states was not statistically significant (t = 1.23, df = 41, p = .22). Statistical comparisons for other samples and subsamples are presented in Table 6F.1 in the Appendix. For the subsample of states that followed the NCES participation rate guidelines, SSI states averaged significantly higher on $I_{RT(96)}$ than non-SSI states.

Table 6.58
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(96)}$, Grade 4

	SSI States N = 23	Non-SSI States N = 20
Mean	4.84	4.74
Standard Deviation	0.26	0.29

The individual state means on $I_{RT(96)}$ are presented in Figure 6.32, with states ordered from highest to lowest. Of the ten highest states, eight are SSI states: Florida, California, Colorado, North Carolina, Texas, Delaware, Kentucky, and Montana; of the ten lowest, four are SSI states: New York, Maine, Rhode Island, and Vermont. The values of the individual state means are included in Table 6F.3 in the Appendix.

Figure 6.32. State means on $I_{RT(96)}$ ordered from highest to lowest, grade 4.



Individual items. Table 6.59 lists each of the seven reform-related topics along with the percentage of students with a teacher who had studied the topic. Individual state percentages are included in Table 6F.3 in the Appendix.

Table 6.59
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 4, 1996

Topic	Mean Percentage of Students in SSI States *N* = 23	Mean Percentage of Students in Non-SSI States *N* = 20
Estimation	79.0	77.9
Problem solving	92.0	90.9
Manipulatives	94.8	94.1
Calculators	73.4	70.6
Students' thinking	69.3	68.2
Gender issues	40.0	39.4
Cultural differences	43.3	39.9

In 1996 in both SSI and non-SSI states, almost 95% of grade 4 students had teachers who had studied the use of manipulatives in mathematics instruction and more than 90% of the students had teachers who had studied problem solving in mathematics. Between 70 and 80% of the students had teachers who had studied estimation and/or the use of calculators. About 66% of the students had teachers who had studied students' thinking about mathematics. Roughly two out of five teachers had studied gender issues and cultural differences.

SSI states as a group had a slightly larger proportion of students with teachers who had studied each of the seven reform-related topics, but the difference between the percentages in the SSI and non-SSI states was very small.

1992

In 1992, the question preceding the seven reform-related topics was slightly different from the 1996 question at both grade 8 and grade 4. In addition, the examples in parentheses for the item on manipulatives were different at grade 8, as described below.

In 1992, questions 21-27 of the grade 8 State NAEP teacher questionnaire asked:

> Have you ever had training in any of the following, either in college courses or in in-service education?
>
> Estimation
> Problem-solving in mathematics
> Use of manipulatives (e.g., measuring instruments or geometric solids) in mathematics instruction
> Use of calculators in mathematics instruction
> Understanding students' thinking about mathematics
> Gender issues in the teaching of mathematics
> Teaching students from different cultural backgrounds.
>
> *Response options:*
> Yes
> No

For grade 4, questions 8-14 were the same as at grade 8, except for the parenthetical expression for the item on manipulatives. For grade 4, the examples of manipulatives were "counting blocks or geometric shapes," the same examples used at both grade levels in 1996.

$I_{RT(92)}$ was computed by counting the number of "Yes" answers out of the seven topics listed. While there were slight wording differences between 1992 and 1996, it seemed that these differences would not compromise the comparability of the indicator, since teachers simply indicated whether or not they had studied the topic. The internal consistency of the scale was .70 for both grade 8 and grade 4.

Grade 8

Number of reform-related topics teachers have studied. Table 6.60 presents the mean and standard deviation on $I_{RT(92)}$ for all the SSI and non-SSI states that participated in the 1992 State NAEP. The mean difference of 0.06 was not statistically significant (t = .67, df = 39, p = .52). Statistical comparisons for other samples and subsamples are presented in Table 6F.4 in the Appendix. The difference between SSI and non-SSI states was not significant in analyses that included all of the states, but when the trend samples were limited to the states that had followed the NCES participation rate guidelines, there was a significant difference between the SSI and non-SSI states. The individual state means for all SSI and non-SSI states is included in Table 6F.5 in the Appendix.

Table 6.60

Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(92)}$, Grade 8

	SSI States *N* = 22	Non-SSI States *N* = 19
Mean	4.74	4.68
Standard Deviation	0.28	0.39

Individual items. Table 6.61 lists each of the seven topics along with the percentage of students with a teacher who had studied the topic. Individual state percentages are included in Table 6F.5 in the Appendix.

Table 6. 61

Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 8, 1992

Topic	Mean Percentage of Students in SSI States *N* = 22	Mean Percentage of Students in Non-SSI States *N* = 19
Estimation	75.8	74.1
Problem-solving	93.0	92.0
Manipulatives	84.9	80.9
Calculators	73.5	70.9
Students' thinking	65.2	64.6
Gender issues	41.0	43.5
Cultural differences	44.1	43.6

In 1992 in both SSI and non-SSI states, more than 90% of the grade 8 students had teachers who had studied problem-solving in mathematics, and 80-85% had teachers who had studied the use of manipulatives. About 75% of the students had teachers who studied estimation, and 70-75% had teachers who had studied the use of calculators. About two-thirds of the students had teachers who studied students' thinking about mathematics. Roughly 40-45% had studied gender issues and cultural differences.

In 1992, SSI states had, on average, a slightly higher proportion of students with teachers who had studied six of the seven reform-related topics. The one exception was the topic of gender issues, where the average for non-SSI states was slightly higher. The difference between the SSI and non-SSI states was very small. The item on the use of manipulatives had the largest difference, 84.9% for the SSI states and 80.9% for the non-SSI states.

Grade 4

Number of reform-related topics teachers have studied. Table 6.62 presents the mean and standard deviation of the state means on $I_{RT(92)}$ for all of the SSI and non-SSI states that participated in the 1992 State NAEP. The means are almost identical. Statistical comparisons for other samples and subsamples are presented in Table 6F.4 in the Appendix. The individual state means for all SSI and non-SSI states is included in Table 6F.6 in the Appendix.

Table 6.62
Mean and Standard Deviation of SSI and Non-SSI States on $I_{RT(92)}$, Grade 4

	SSI States *N* = 22	Non-SSI States *N* = 19
Mean	4.70	4.72
Standard Deviation	0.30	0.24

Individual items. Table 6.63 lists each of the seven reform-related topics along with the percentage of students whose teacher had studied the topic. Individual state percentages are included in Table 6F.6 in the Appendix.

Table 6.63
Seven Reform-Related Topics and the Percentage of Students Whose Teacher Had Studied the Topic, Grade 4, 1992

Topic	Mean Percentage of Students in SSI States *N* = 22	Mean Percentage of Students in Non-SSI States *N* = 19
Estimation	79.5	79.8
Problem solving	90.8	92.4
Manipulatives	92.8	93.0
Calculators	61.2	62.0
Students' thinking	68.8	70.6
Gender issues	34.8	35.3
Cultural differences	44.2	40.4

In grade 4 in both SSI and non-SSI states, more than 90% of the students had teachers who had studied problem solving in mathematics, and even slightly more had teachers who had studied the use of manipulatives. About 80% of the students had teachers who studied estimation, about 70% had teachers who had studied students' thinking about mathematics, and about 60% had teachers who had studied the use of calculators. Between 40 and 45% had teachers who had studied cultural differences and about a third who had studied gender issues.

In 1992, non-SSI states had a slightly higher proportion of grade 4 students with teachers who had studied six of the seven reform-related topics, except for the topic of cultural differences, where SSI states were higher. Except for this topic, the difference between the SSI and non-SSI states was quite small

1990

The 1990 grade 8 teacher questionnaire included only one of the seven topics from 1992 and 1996. The 1990 questionnaire asked:

> Have you ever received training in any of the following, either in courses or in-service education?
>
> Teaching students from different cultural backgrounds
>
> *Response options:*
>
> Yes
>
> No

Table 6.64 presents the percentage of students in SSI and non-SSI states who had teachers who had studied this topic

Table 6.64
Percentage of Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds, Grade 8, 1990

Topic	Mean Percentage of Students in SSI States *N* = 20	Mean Percentage of Students in Non-SSI States *N* = 17
Cultural differences	34.6	29.0

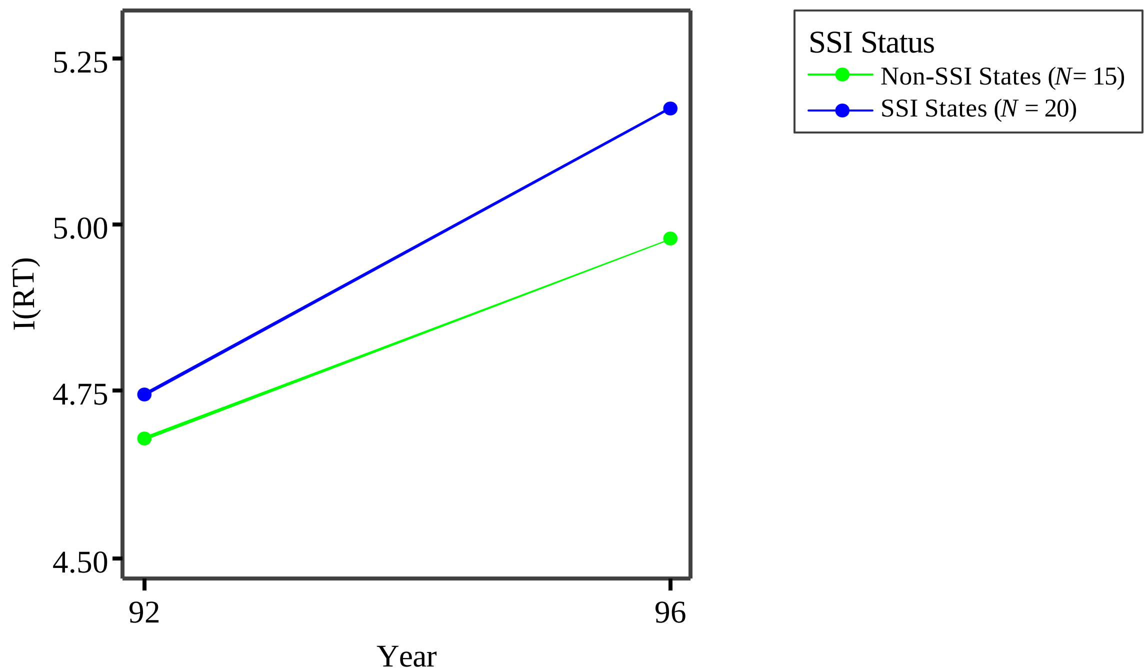
In 1990, the percentage of students with teachers who had studied cultural differences was a bit higher in SSI states than in non-SSI states, but the difference was not statistically significant. See Table 6F.7 in the Appendix and Table 6F.8.

Change from 1992 to 1996

Grade 8

Two-point trend sample. A 2x2 repeated measures analysis of variance found a significant effect for year ($F = 48.21$, $df = 1,33$, $p < .01$) for the 20 SSI and 15 non-SSI states that participated in 1992 and 1996. The effect for SSIs was not statistically significant ($F = 1.82$, $df = 1,33$, $p = .19$), nor was the interaction of year and SSI ($F = 1.60$, $df = 1,33$, $p = .21$). The mean for each year by SSI status is plotted in Figure 6.33.

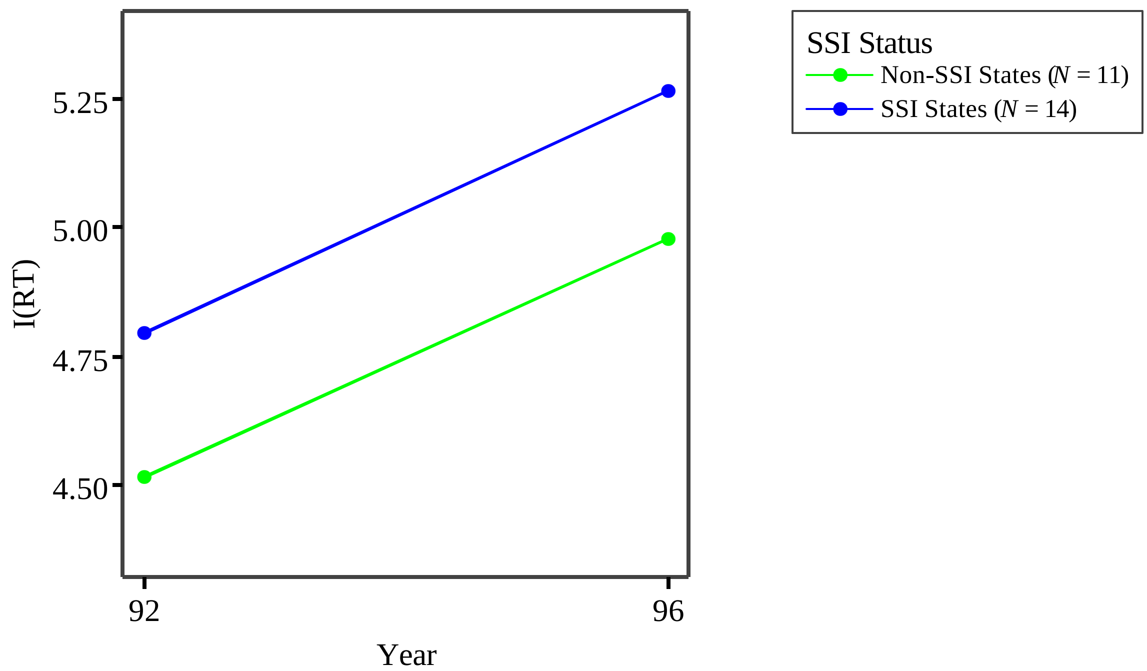
Figure 6.33. Mean of I_{RT} for SSI and non-SSI states at grade 8 in 1992 and 1996.



As Figure 6.32 shows, the mean number of topics studied by teachers in SSI states increased from 4.74 to 5.18 across the four years, while the increase in non-SSI states was from 4.68 to 4.98.

Subsample. For comparison purposes, the 2x2 repeated-measures ANOVA was repeated with the 14 SSI states and the 11 non-SSI states that followed NCES's participation rate guidelines. With this subsample, the analysis again found a significant effect for year ($F = 74.14$, $df = 1, 23$, $p < .01$). In addition, there was a significant effect for SSI status ($F = 8.51$, $df = 1, 23$, $p < .01$). In the subsample, the SSI states averaged higher than the non-SSI states and both groups increased equally from 1992 to 1996. Figure 6.34 shows subsample means.

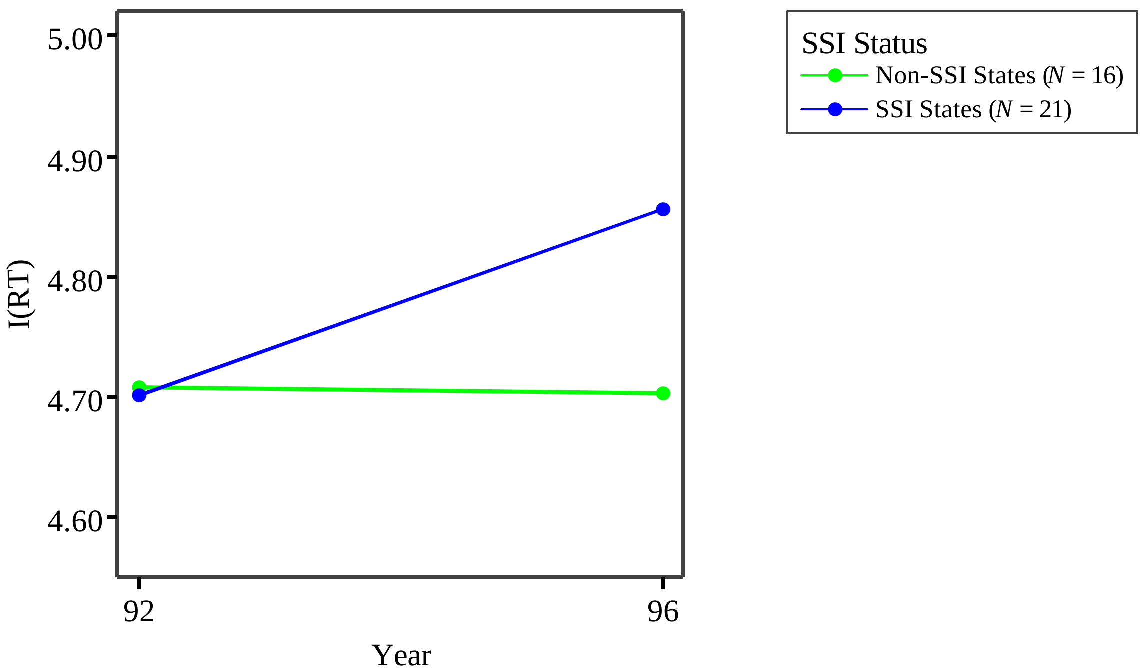
Figure 6.34. Mean of I_{RT} at grade 8 in 1992 and 1996, for SSI and non-SSI states following the NCES participation rate guidelines.



Grade 4

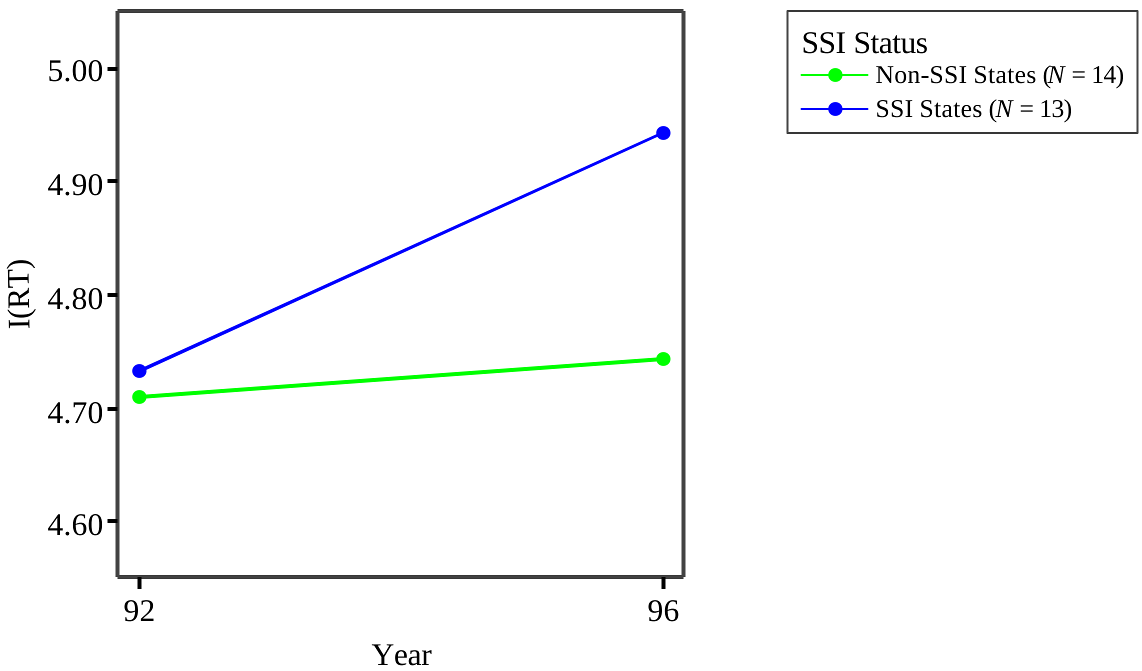
Two-point trend sample. A 2x2 repeated measures analysis of variance found a significant effect for year ($F = 4.05$, $df = 1,35$, $p < .10$) for the 21 SSI and 16 non-SSI states that participated in the State NAEP both years. The effect for SSIs was not statistically significant ($F = 0.75$, $df = 1,35$, $p = .39$), but the interaction of year and SSI was ($F = 4.47$, $df = 1,35$, $p < .05$). The means for each year by SSI status are plotted in Figure 6.35.

Figure 6.35. Mean of I_{RT} for SSI and non-SSI states at grade 4 in 1992 and 1996.



Subsample**.** The results for the subsample of states that met the NCES participation guidelines are shown in Figure 6.36. The 2x2 repeated measures analysis of variance included 12 SSI states and 14 non-SSI states. The main effect for year was statistically significant (F = 6.08, df = 1,24, p < .05), and so was the effect of year by SSI status (F = 3.05, df = 1,24, p < .10). The main effect for SSI status was not statistically significant (F = 1.36, df = 1,24, p = .26).

Figure 6.36. Mean of I_{RT} at grade 4 in 1992 and 1996, for SSI and non-SSI states following the NCES participation rate guidelines.



Change Across 1990, 1992, and 1996

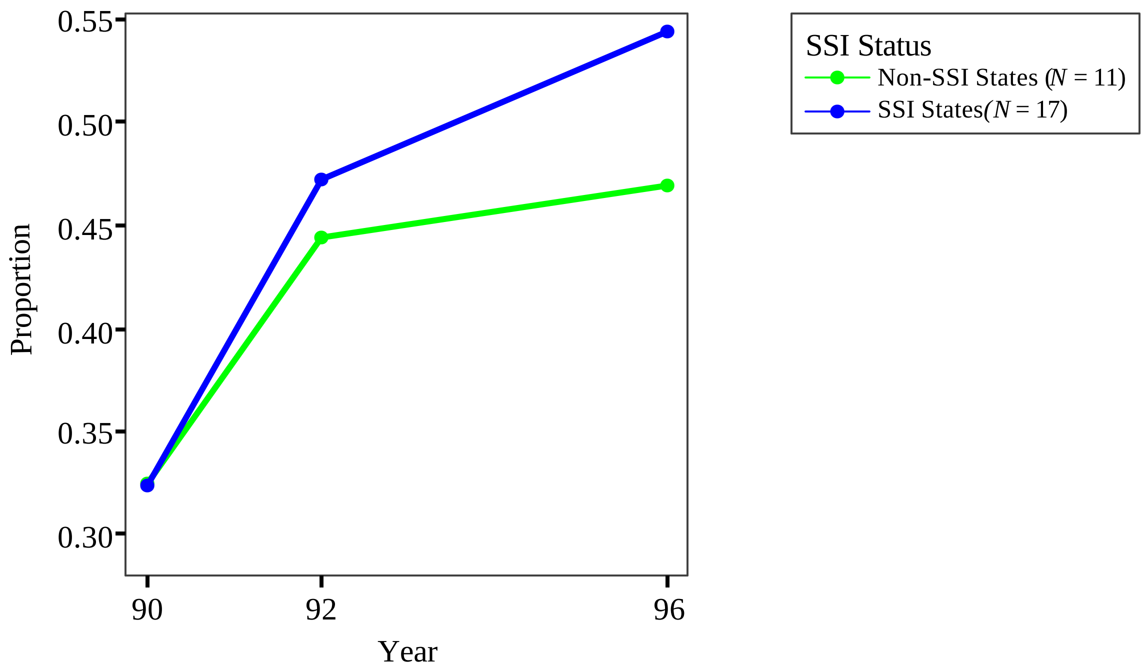
Three-point trend sample. In 1990, just one of the seven topics was included in the teacher questionnaire: i.e., teaching students from different cultural backgrounds. Table 6.65 presents the percentages for the 17 SSI states and 11 non-SSI states in the three-point trend sample.

Table 6.65
Percent of Grade 8 Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds

Year	Mean Percent of Students in SSI States *N* = 17	Mean Percent of Students in Non-SSI States *N* = 11
1990	32.4	32.5
1992	47.3	44.4
1996	54.4	47.0

A 2x3 repeated measures analysis of variance with the percentage of students as the dependent measure found a significant effect for year (F = 36.03, df = 2,25, p < .01), but no significant effect for SSI (F = 0.69, df = 1,26, p = .41), or the interaction term (F = 2.09, df = 2,26, p = .13). Figure 6.36 shows the graph of Table 6.

Figure 6.37. Proportion of students whose teacher had studied teaching students from different cultural backgrounds, grade 8.

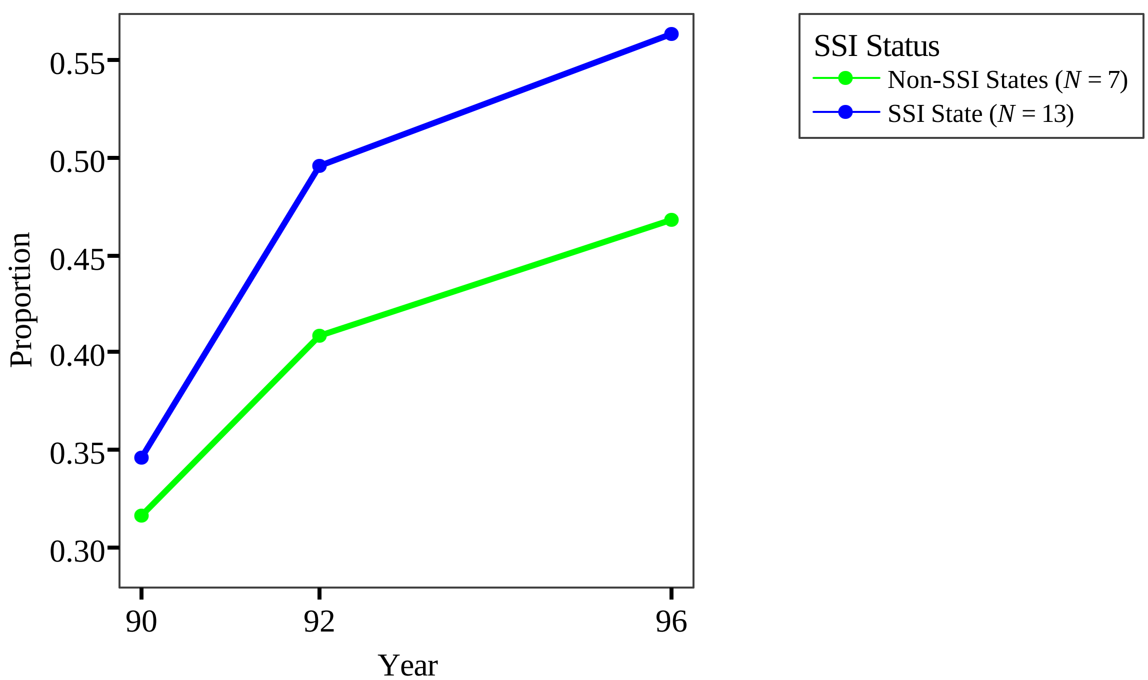


Subsample**.** When the analysis is limited to those states that followed the NCES participation rate guidelines, the results are similar to those for the total sample. The repeated measures ANOVA found a significant effect for year (F = 21.74, df = 2,17, p < .01), but no significant effect for SSI status (F = 2.11, df = 1,18, p = .16), or the interaction (F = 1.38, df = 2,18, p = .27). Table 6.64 below lists the mean percentage of students with teachers who had studied teaching children from different cultural backgrounds in 1990, 1992, and 1996. Figure 6.38 shows the graph of Table 6.66.

Table 6.66
Percent of Grade 8 Students Whose Teacher Had Studied Teaching Students from Different Cultural Backgrounds in the Subsample of States that Followed the NCES Participation Rate Guidelines

Year	Mean Percent of Students in SSI States *N* = 13	Mean Percent of Students in Non-SSI States *N* = 7
1990	34.6	31.6
1992	49.5	40.9
1996	56.4	46.9

Figure 6.38. Proportion of students whose teacher had studied teaching students from different cultural backgrounds in the subsample that followed the NCES participation rate guidelines.



Summary of Results

1996

Grade 8

Students in SSI states had teachers who had studied a slightly greater number of reform-related topics than students in non-SSI states.

Over 90% of the students in both SSI and non-SSI states had teachers who had studied problem solving, and almost that many had teachers who had studied the use of manipulatives in mathematics instruction.

Roughly half of the students in both SSI and non-SSI states had teachers who had studied gender issues or teaching students from different cultural backgrounds.

Grade 4

For the total sample, SSI and non-SSI states did not differ in the number of reform-related topics students' teachers had studied. For the subsample of states that followed NCES participation rate guidelines, students in SSI states had a significantly higher mean than students in non-SSI states.

About 95% of the students in both SSI and non-SSI states had teachers who had studied the use of manipulatives in mathematics instruction, and about 90% had teachers who had studied problem solving.

About 40% of the students in both SSI and non-SSI states had teachers who had studied gender issues or teaching students from different cultural backgrounds.

1992

Grade 8

Students in SSI states averaged a little higher in the number of reform-related topics teachers had studied, but the difference between SSI and non-SSI states was not statistically significant. For the two- and three-point trend samples, the mean of the SSI states that followed the NCES participation rate guidelines was significantly higher than the mean of the non-SSI states.

In 1992, over 90% of the students in both SSI and non-SSI states had mathematics teachers who had studied problem solving, and between 80 and 85% had teachers who had studied the use of manipulatives in mathematics instruction.

About 40-45% of the students had teachers who had studied gender issues or teaching students from different cultural backgrounds.

Grade 4

Students in non-SSI states averaged just .02 higher in the number of reform-related topics their teachers had studied. No statistically significant differences were found for any samples or subsamples in 1992.

Two topics had been studied by teachers of over 90% of the students: the use of manipulatives and problem solving. Gender issues and cultural differences were the least likely to be studied.

1990

Grade 8

In 1990, only one of the seven reform-related topics, "Teaching students from different cultural backgrounds," was included on the teachers' questionnaire. The proportion of students with teachers who had studied the item did not correlate with the state SSI status. In both SSI and non-SSI states, roughly a third of the students had teachers who had studied the topic.

Change Across Time

Two-point trend: 1992-1996

Grade 8

For the total sample of 20 SSI and 15 non-SSI states the number of reform-related topics teachers had studied increased from 1992 to 1996. However, there was no significant effect for SSI and no interaction effect. For the subsample that followed participation rate guidelines, the effects for both time and SSI were significant, with SSI states averaging higher across both years than non-SSI states. The interaction effect of year and SSI status was not significant.

Grade 4

For the total sample of 21 SSI and 16 non-SSI states, the number of topics studied increased from 1992 to 1996, and the interaction effect was significant, with SSI states increasing more than non-SSI states. These results were also found in the subsample of states following the participation rate guidelines.

Three-point trend: 1990, 1992, and 1996

Grade 8

Only one of the seven topics was included on the 1990 teacher questionnaire. From 1990 to 1996, the proportion of students with teachers who had ever studied the topic increased significantly. However, there was no significant effect for SSI status or for the interaction of time and SSI status.

The results were the same for the subsample of states following the participation rate guidelines. The effect for year was significant, but the effect for SSI status and the interaction effect were not significant.

Calculator Use—I_C

The Importance of Calculator Use

For over 20 years, the National Council of Teachers of Mathematics (NCTM) has recommended the thoughtful use of calculators throughout the mathematics curriculum (NCTM, 1980; 1989; and 2000). In 1989, in the *Curriculum and Evaluation Standards for School Mathematics,* NCTM stated its belief that "appropriate calculators should be available to all students at all times" (NCTM, 1989, p. 8). Ten years later, in *Principles and Standards,* NCTM devoted one of its six principles for school mathematics to technology, including calculators. This principle states that calculators are essential tools for teaching, learning, and doing mathematics and that "technology (including calculators) should be used widely and responsibly, with the goal of enriching students' learning of mathematics" (NCTM, 2000, p. 25). NCTM over the years has continued to clarify its position on calculator use, stressing that student use of calculators should not be viewed as a replacement for developing basic understandings, learning basic facts, or using other means for calculating. State educators and others have become more permissive in allowing student use of calculators on state assessments. In 1996, over half of the states permitted student use of calculators on state tests (NCREL, 1996). In March 1994, students who took the Scholastic Aptitude Test I (SAT I) were allowed to use calculators.

An increasing volume of experimental research supports growing evidence that students who regularly use calculators in learning mathematics gain in conceptual understanding and reasoning while becoming as competent as, or more competent in paper-and-pencil computations than, those who are taught in the absence of calculators (Kilpatrick, Swafford, & Findell, 2001; Groves, 1994; Dunham & Dick, 1994; Brolin & Björk, 1992; Hembree & Dessart, 1992). In framing the Statewide Systemic Initiatives Program, the National Science Foundation (NSF) did not explicitly state a position on calculator use. However, one of the six drivers of reform espoused by NSF stressed the need for convergent resources, including materials such as calculators, to focus on the formation of a unitary program to further the learning of all students. At the outset of the SSI program in the early 1990s, a time of increasing availability of hand-held calculators, there was strong support from NCTM and educational researchers for students at all grade levels to use calculators in learning and doing mathematics. As such, indicators of the degree of use of calculators provide at least a partial measure of the implementation of reform practice. These use-indicators, however, do not show whether the calculators are used responsibly and whether calculators are used appropriately in concert with other resources to provide students a deep understanding of mathematics, both important considerations.

1996

Several questions on the 1996 teacher questionnaire focused on calculator use. Five were selected for the indicator, based on the item intercorrelations and the goals of the SSI program. The selected items were:

> How often do the students in this class do each of the following?
> Use a calculator
>
> *Response options*
> Almost every day
> Once or twice a week
> Once or twice a month
> Never or hardly ever
>
> Do you permit students in this class unrestricted use of calculators?
> Do you permit students in this class to use calculators for tests?
> Do the students in this class have access to calculators owned by the school?
> Do you provide instruction to students in this class in the use of calculators?
>
> *Response options*
> Yes
> No

For the first item, responses were coded from 1 for "Never or hardly ever" to 4 for "Almost every day." For the others, "Yes" was coded as 2 and "No"as 1. For each student, $I_{C(96)}$ was computed by adding the responses across the five items, resulting in a scale with a range from 5 to 12. The internal consistency of the scale (i.e., Cronbach's coefficient alpha) was .65 for grade 8 and .61 for grade 4.

Grade 8

Calculator use. Table 6.67 presents the means and standard deviations of $I_{C(96)}$ for all SSI and non-SSI states in the 1996 State NAEP. The mean difference of 0.07 was not statistically significant, with $t = 1.85$ ($df = 38$, $p < .10$). Statistical comparisons for other samples and subsamples are presented in Table 6G.1 of the Appendix.

Table 6.67
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 8

	SSI States N = 22	Non-SSI States N = 18
Mean	9.92	9.85
Standard Deviation	0.72	0.65

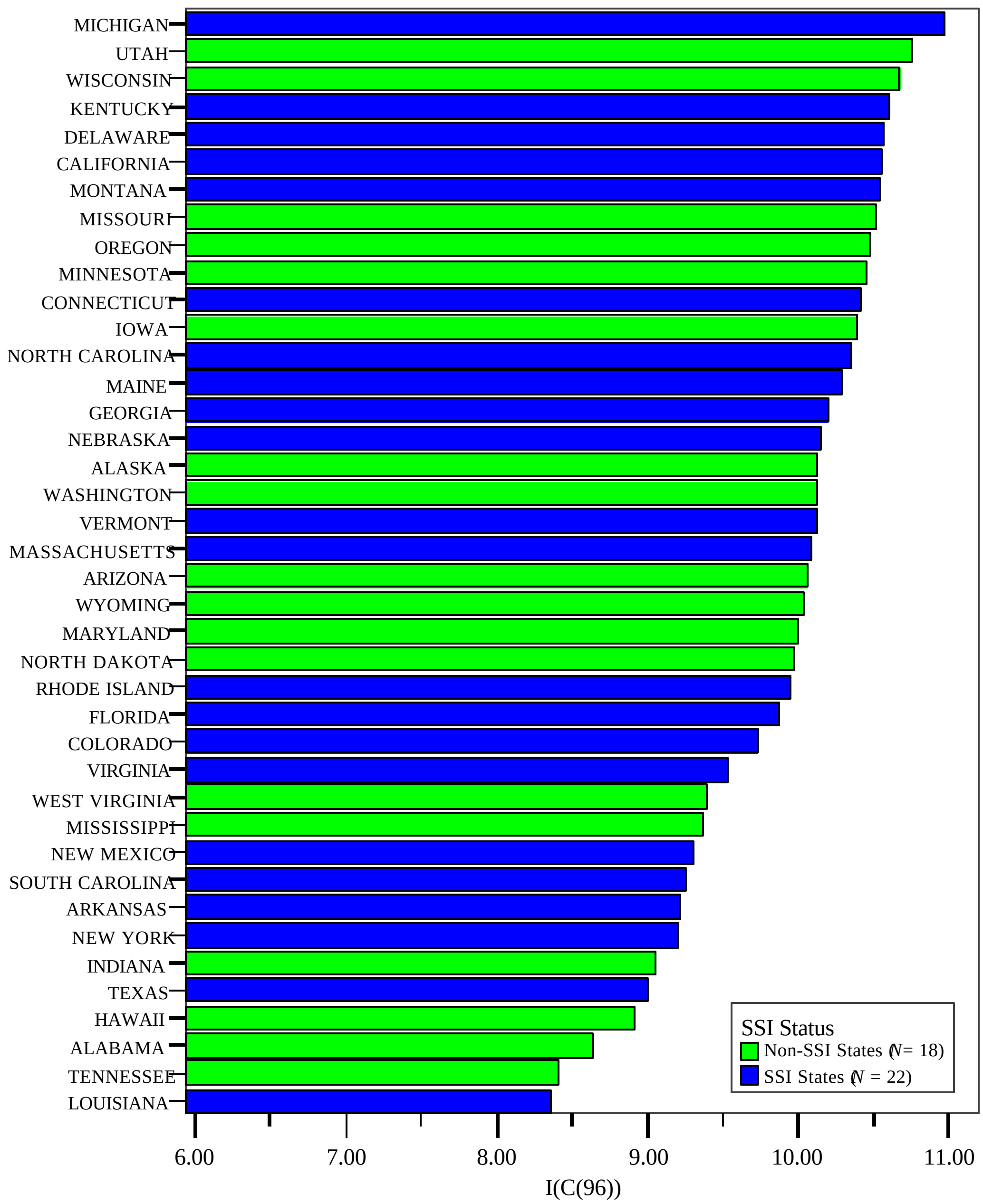
Individual state means on $I_{C(96)}$ are presented in Figure 6.38, with states ranked from highest to lowest. Of the ten highest-ranking states, five are SSI states: Michigan, Kentucky, Delaware, California, and Montana; of the ten lowest, six are SSI states: Louisiana, Texas, New York, Arkansas, South Carolina, and New Mexico. The values of the individual state means are included in Table 6G.2 of the Appendix.

Individual items. Table 6.68 lists the mean and standard deviation for each of the five questions. With the individual items, the effect for SSI was statistically significant ($F = 2.24$, $df = 5,34$, $p < .10$). Post hoc contrasts found a statistically significant difference on one of the five items: Student access to school owned calculators. SSI states, as a group, averaged significantly higher on this item ($F = 6.43$, $df = 1,39$, $p < .01$). Means for individual states are listed in Table 6G.2 in the Appendix.

Table 6.68
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1996

	SSI states N = 22		Non-SSI states N = 18	
Item	Mean	*SD*	Mean	*SD*
Frequency of use	3.15	0.33	3.17	0.41
Unrestricted use	1.41	0.11	1.44	0.15
Use on tests	1.67	0.13	1.66	0.17
School owned calculators	1.84	0.09	1.77	0.10
Instruction in use	1.82	0.08	1.80	0.05

Figure 6.39. State means on $I_{C(96)}$ ordered from highest to lowest, grade 8.



Item intercorrelations. The first three items are very strongly related, with correlation coefficients around .90, as shown in Table 6.69. The item about instruction in calculator use is moderately correlated, around .50, with all other items except for the one on unrestricted use of calculators. The item that seems the most different from the others is the one asking whether students have access to calculators owned by the school. For three of the four comparisons, the relationship is slightly negative, but not significantly different from 0. The only positive significant correlation is between access to school-owned calculators and instruction in calculator use.

Table 6.69
Intercorrelations of State Means on the Five Calculator Use Items, Grade 8, 1996

Total sample, *N* = 40	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.88*			
3. Use on tests	.97*	.90*		
4. School owned calculators	-.15	-.19	-.09	
5. Instruction in use	.51*	.26	.50*	.47*
Subsample, *N* = 30	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.87*			
3. Use on tests	.97*	.89*		
4. School owned calculators	-.21	-.23	-.14	
5. Instruction in use	.47*	.25	.49*	.46*

*$p < .01$

Supplementary information. We expected to find more calculator use in the SSI states than in non-SSI states, given the goal of curricular reform in mathematics. But SSI participation was not the only factor influencing calculator use. For example, the state's testing program might have influenced the use of calculators. The North Central Regional Educational Laboratory has a database of state student assessment programs. The May 1996 database was used to identify states with assessment programs and whether students could use calculators when taking state tests (North Central Regional Educational Laboratory, 1996). Of the 47 states that participated in the State NAEP at least once, 40 answered the question about calculator use on state assessments. Of these, 28 permitted students to use calculators while taking the test and 12 did not (p. 96). Table 6.70 lists the states by their calculator use policy. States included under "Other" either did not have state assessments in 1996 or did not provide information about calculator use on their state assessments.

Table 6.70
Use of Calculators on State Assessments and State's SSI Status, for States Included in State NAEP in 1990, 1992, or 1996

Calculators permitted, *N* = 29
SSI states – Arkansas,* California, Connecticut,* Florida, Georgia, Maryland, Maine,* Michigan, Montana, New Jersey,* New Mexico,* New York, North Carolina,* Ohio, Rhode Island,* Vermont
Non-SSI states – Alabama, Idaho, Illinois, Maryland, Minnesota,* Missouri,* Mississippi,* Oregon, Pennsylvania, South Dakota, Utah, Washington, Wisconsin

Calculators not permitted, *N* = 11
SSI states – Delaware, Louisiana, Massachusetts, Texas, Virginia
Non-SSI states – Hawaii, Indiana, North Dakota, New Hampshire, Oklahoma, Tennessee

Other, *N* = 10
SSI states – Colorado, Kentucky, Nebraska, South Carolina
Non-SSI states – Alaska, Arizona, Iowa, West Virginia, Wyoming

*States reporting that some questions on the state assessment were intentionally designed for calculator use.

The effect of the state's calculator use policy on $I_{C(96)}$ was examined with a 2x2 analysis of variance. The mean and standard deviation for each group are reported in Table 6.69. States with policies permitting calculator use averaged significantly higher on $I_{C(96)}$ ($F = 8.99$, $df = 1,30$, $p < 01$); the mean for states permitting use was 10.09 ($SD = 0.60$) and the mean for states not permitting use was 9.32 ($SD = 0.77$), a difference of about one standard deviation unit. The interaction effect of SSI status and calculator use was not statistically significant, in part because of the lack of power to test the interaction because of the relatively small number of states not permitting calculator use. As Table 6.71 shows, at grade 8, the state means for SSI and non-SSI states on $I_{C(96)}$ are about the same in states where calculator use is permitted, but they differ by about .40 where calculator use is not permitted. These findings about a state's calculator use policy illustrate the importance of considering factors in addition to a state's SSI status in evaluating the effects of SSI and underscores the difficulty of attributing effects to SSI.

Table 6.71
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

	SSI States	Non-SSI States
Permitted	*N* = 13	*N* = 9
Mean	10.07	10.11
Standard Deviation	0.56	0.69
Not permitted	*N* = 5	*N* = 4
Mean	9.51	9.09
Standard Deviation	0.87	0.65

Grade 4

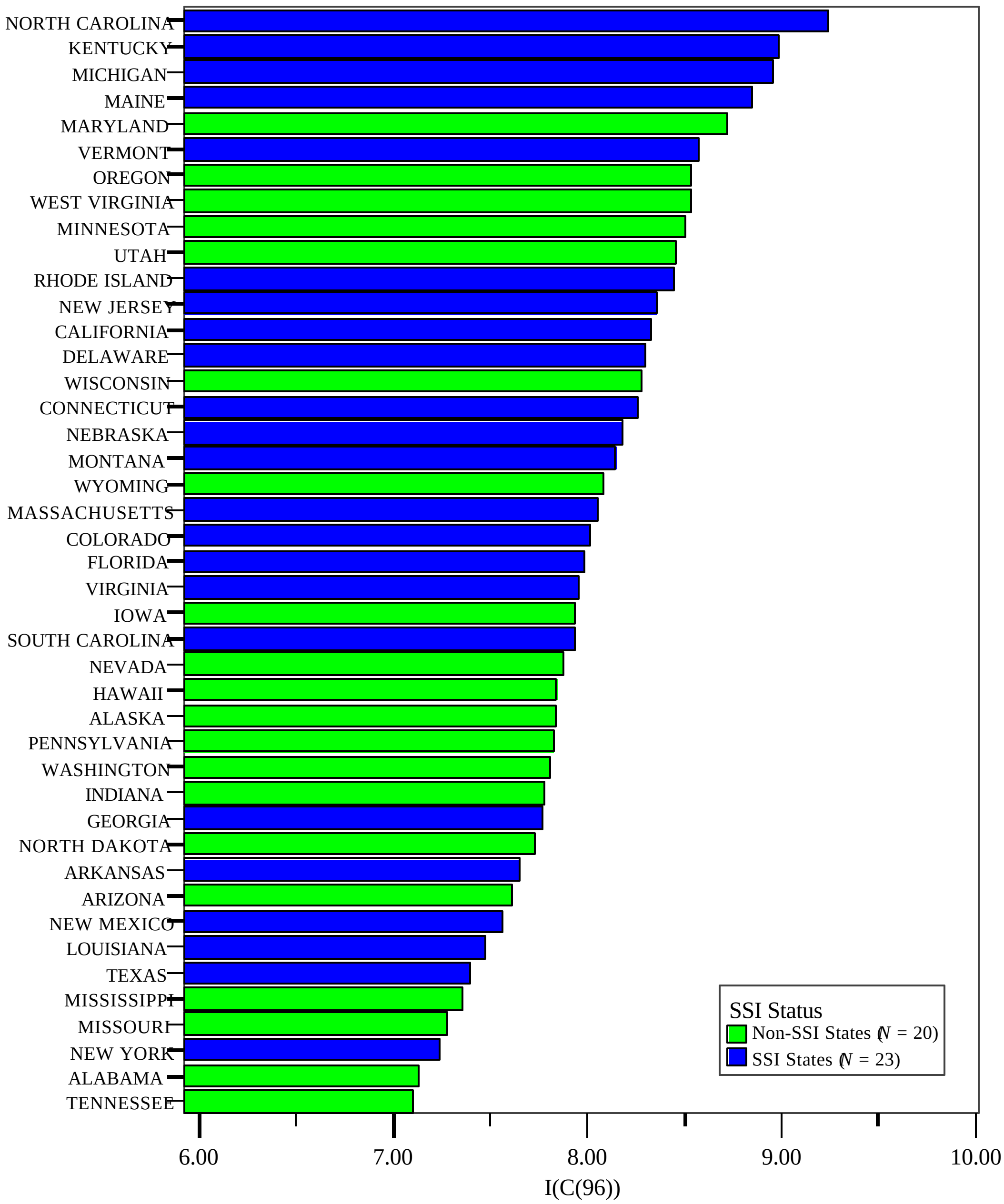
Table 6.72 presents the grade 4 results on $I_{C(96)}$. The mean difference of 0.25 between SSI and non-SSI states was not statistically significant ($t = 1.60$, $df = 41$, $p = .12$). Statistical comparisons for other samples and subsamples are presented in Table 6G.1 in the Appendix.

Table 6.72
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 4

	SSI States *N* = 23	Non-SSI States *N* = 20
Mean	8.16	7.91
Standard Deviation	0.53	0.48

The individual state means on $I_{C(96)}$ at grade 4 are presented in Figure 6.40, with states ordered from highest to lowest. Of the ten highest states, five are SSI states: North Carolina, Kentucky, Michigan, Maine, and Vermont; of the ten lowest, five are SSI states: New York, Texas, Louisiana, New Mexico, and Arkansas. The values of the individual state means are included in Table 6G.3 in the Appendix.

Figure 6.40. State means on $I_{C(96)}$ ordered from highest to lowest, grade 4.



Individual items. Table 6.71 lists the grade 4 mean and standard deviation for each of the five questions in $I_{C(96)}$. The overall effect for SSI was not statistically significant ($F = 1.17$, $df = 5,37$, $p = .33$). Means for individual states are listed in Table 6G.3 in the Appendix.

Table 6.73

Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 4, 1996

	SSI states *N* = 22		Non-SSI states *N* = 18	
Item	Mean	*SD*	Mean	*SD*
Frequency of use	2.21	0.25	2.09	0.24
Unrestricted use	1.12	0.04	1.10	0.03
Use on tests	1.12	0.10	1.10	0.07
School owned calculators	1.86	0.09	1.81	0.11
Instruction in use	1.84	0.09	1.80	0.09

Item intercorrelations. At grade 4, state means on all five items are moderately to strongly related. The highest correlation is between whether students have access to school owned calculators and whether the teacher provided instruction in calculator use.

Table 6.74

Intercorrelations of State Means on the Five Calculator-Use Items, Grade 4, 1996

Total sample, *N* = 40	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.73*			
3. Use on tests	.83*	.86*		
4. School owned calculators	.80*	.50*	.60*	
5. Instruction in use	.86*	.54*	.62*	.90*
Subsample, *N* = 30	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.74*			
3. Use on tests	.84*	.86*		
4. School owned calculators	.80*	.52*	.64*	
5. Instruction in use	.90*	.57*	.68*	.91*

*$p < .01$

Supplementary information. Means and standard deviations for SSI and non-SSI states as a function of the state's calculator use policy are reported in Table 6.75. States with policies permitting calculator use averaged significantly higher on $I_{C(96)}$ ($F = 3.43$, $df = 1,32$, $p < .10$); the mean for states permitting use was 8.13 ($SD = 0.58$) and the mean for states not permitting use was 7.74 ($SD = 0.37$), a difference of .39. Neither the main effect for SSI status nor the interaction of SSI status and calculator use was statistically significant. The mean for the SSI states was about .24 points higher than the non-SSI states, regardless of whether calculator use was permitted on the state assessments.

Table 6.75
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(96)}$, Grade 4 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

	SSI States	Non-SSI States
Permitted	*N* = 14	*N* = 10
Mean	8.24	7.99
Standard Deviation	0.57	0.59
Not permitted	*N* = 4	*N* = 5
Mean	7.84	7.61
Standard Deviation	0.39	0.34

1992

In 1992, the grade 8 teacher questionnaire included seven items about calculator use instead of the five used in 1996. Three questions were the same as in 1996:

How often do the students in this class do each of the following?
Use a calculator

Response options
Almost every day
Once or twice a week
Once or twice a month
Never or hardly ever

Do you permit students in this class unrestricted use of calculators?
Do you permit students in this class to use calculators for tests?

Response options
Yes
No

The other questions at grade 8 were:

Do the students in this class have access to any of the following calculators owned by the school?
- Basic 4-function
- Scientific

Do you provide instruction to students in this class in the use of the following types of calculators?
- Basic 4-function
- Scientific

Response options
- Yes
- No

At grade 4, the questions were the same as in 1996. Scoring for the first item was from 1 for "Never or hardly ever" to 4 for "Almost every day". For the others, "Yes" was coded as 2 and "No" as 1. For each student, $I_{C(92)}$ was the sum of the individual items, resulting in a scale with a range from 7 to 14 at grade 8 and from 5 to 12 at grade 4. Internal consistency of the scale (i.e., Cronbach's coefficient alpha) was .70 for grade 8 and .66 for grade 4.

Grade 8

Calculator use. In 1992, SSI and non-SSI states averaged about the same on $I_{C(92)}$. (See Table 6.76). Statistical comparisons for other samples and subsamples are presented in Table 6G.4 in the Appendix and individual state means are included in Table 6G.5.

Table 6.76
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 8

	SSI States *N* = 22	Non-SSI States *N* = 19
Mean	11.24	11.21
Standard Deviation	0.90	1.11

Individual items. Table 6.77 lists the SSI and non-SSI means on each item in 1992. The effect for SSI was not statistically significant ($F = 1.52$, $df = 7,33$, $p = .19$). Individual state means are included in Table 6G.5 of the Appendix. In 1992, the mean for the non-SSI states was slightly higher on four of the items, but the mean for SSI states was slightly higher on students' access to calculators owned by the school and instruction in the use of a basic 4-function calculator. Results for the subsample were similar to those for the total sample ($F = 1.18$, $df = 7,28$, $p = .34$).

Table 6.77
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1992

Item	SSI states *N* = 22 Mean	*SD*	Non-SSI states *N* = 18 Mean	*SD*
Frequency of use	2.59	0.38	2.68	0.48
Unrestricted use	1.29	0.10	1.31	0.14
Use on tests	1.47	0.14	1.50	0.18
School owned calculators				
Basic 4-function	1.72	0.09	1.61	0.15
Scientific	1.30	0.10	1.27	0.11
Instruction in use				
Basic 4-function	1.67	0.09	1.62	0.08
Scientific	1.31	0.10	1.32	0.13

Intercorrelations. Table 6.76 presents the intercorrelations of the items on $I_{C(92)}$.

Table 6.76
Intercorrelations of State Means on the Seven Calculator Use Items, Grade 8, 1996

Total sample, *N* = 40	1.	2.	3.	4.	5.	6.
1. Frequency of use						
2. Unrestricted use	.93*					
3. Use on tests	.98*	.93*				
4. School owned 4-function calculators	.23	.17	.31			
5. School owned scientific calculators	.56*	.47*	.55*	.48*		
6. Instruction in 4-function calculator	.48*	.38	.52*	.73*	.38	
7. Instruction in scientific calculator	.82*	.75*	.78*	.19	.81*	.36

Subsample, *N* = 30	1.	2.	3.	4.	5.	6.
1. Frequency of use						
2. Unrestricted use	.93*					
3. Use on tests	.98*	.93*				
4. School-owned 4-function calculators	.20	.12	.29			
5. School-owned scientific calculators	.52*	.45*	.52*	.48*		
6. Instruction in 4-function calculators	.46*	.36	.50*	.74*	.37	
7. Instruction in scientific calculator	.82*	.77*	.78*	.17	.78*	.35

*$p < .01$

As Table 6.76 shows, the state means on most items were moderately to strongly related. The main exception was students' access to school-owned 4-function calculators. State means on that item were only weakly related to the states' means for frequency of use, unrestricted use, and use on tests. Access to 4-function calculators was related to instruction in their use, and access to scientific calculators was related to instruction in their use.

Supplementary information. The information from the 1996 NCREL report on state assessments was used to categorize states. A 2x2 ANOVA found that states that permitted calculators use on their tests averaged higher on $I_{C(92)}$ than states that did not permit calculators. The main effect for SSI status was not significant, and neither was the interaction of SSI status and calculator policy. The small number of states in some of the cells limited the power to detect significant effects.

Table 6.79
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

	SSI States	Non-SSI States
Permitted	*N* = 13	*N* = 9
Mean	11.27	11.56
Standard Deviation	0.94	1.26
Not permitted	*N* = 5	*N* = 5
Mean	10.76	10.55
Standard Deviation	0.92	0.89

Grade 4

Calculator use. Table 6.80 presents the grade 4, 1992 descriptive statistics for the SSI and non-SSI states. The means are almost identical. Statistical comparisons for other samples and subsamples are presented in Table 6G.4 in the Appendix. The individual state means for all SSI and non-SSI states are included in Table 6G.6 in the Appendix.

Table 6.80
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 4

	SSI States *N* = 22	Non-SSI States *N* = 19
Mean	7.27	7.21
Standard Deviation	0.59	0.59

Individual items. Table 6.81 lists each of the items along with the means and standard deviations for SSI and non-SSI states. On all five items, the means are very close. Individual state means are included in Table 6G.6 in the Appendix.

Table 6.81
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 4, 1992

	SSI states *N* = 22		Non-SSI states *N* = 19	
Item	Mean	*SD*	Mean	*SD*
Frequency of use	1.81	0.25	1.79	0.24
Unrestricted use	1.08	0.03	1.07	0.03
Use on tests	1.06	0.04	1.06	0.05
School owned calculators	1.64	0.17	1.62	0.18
Instruction in use	1.68	0.14	1.67	0.13

Item intercorrelations. At grade 4, state means on all five items are moderately to strongly related, as shown in Table 6.82. As in 1996 at grade 4, one of the highest correlations is between whether students have access to school-owned calculators and whether the teacher provided instruction in calculator use. In 1992, the next highest correlations were between the frequency of use, access to school-owned calculators, and instruction in use.

Table 6.82
Intercorrelations of State Means on the Five Calculator Use Items, Grade 4, 1992

Total sample, *N* = 41	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.70*			
3. Use on tests	.79*	.73*		
4. School owned calculators	.85*	.44*	.62*	
5. Instruction in use	.93*	.56*	.72*	.91*

Subsample, *N* = 36	1.	2.	3.	4.
1. Frequency of use				
2. Unrestricted use	.77*			
3. Use on tests	.82*	.79*		
4. School owned calculators	.84*	.50*	.65*	
5. Instruction in use	.94*	.63*	.77*	.90*

*$p < .01$

Supplementary information. Table 6.83 presents the means and standard deviations for $I_{C(92)}$ as a function of the states' SSI status and its policy on the use of calculators on state assessments. A 2x2 ANOVA found no significant effect for SSI status, calculator use policy, or the interaction.

Table 6.83
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(92)}$, Grade 4 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

	SSI States	Non-SSI States
Permitted	*N* = 13	*N* = 9
Mean	7.22	7.40
Standard Deviation	0.66	0.51
Not permitted	*N* = 5	*N* = 6
Mean	7.10	6.87
Standard Deviation	0.38	0.74

1990

The 1990 grade 8 teachers questionnaire included four items that were very similar or identical to those in later years. They were:

How often do the students in this class do each of the following?
Use calculators

Response options
Almost every day
Several times a week
About once a week
Less than once a week
Never

Do you permit students in this class unrestricted use of calculators?
Do you permit students in this class to use calculators for tests?
Do the students in this class have access to calculators owned by the school?

Response options
Yes
No

$I_{C(90)}$ was the sum of the four items, with responses to the first question scored from 1 for "Never" to 5 for "Almost every day" and responses to the other three scored 1 for "No" and 2 for "Yes". The scale's range was 4-11, and its internal consistency (coefficient alpha) was .67.

Calculator use. In 1990, SSI and non-SSI states averaged about the same on $I_{C(90)}$. (See Table 6.84.) Statistical comparisons for other samples and subsamples are presented in Table 6G.7 in the Appendix, and individual state means are included in Table 6G.8.

Table 6.84
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(90)}$, Grade 8

	SSI States *N* = 20	Non-SSI States *N* = 17
Mean	6.30	6.44
Standard Deviation	0.65	0.69

Individual items. Table 6.85 lists each of the items along with the means and standard deviations for SSI and non-SSI states. The overall *F* was 0.66 (*df* = 4.32, *p* = .62). On all four items, the means are very close; with the non-SSI average slightly higher on the first three, and the SSI average slightly higher on students' access to school owned calculators. Individual state means are included in Table 6G.8 in the Appendix.

Table 6.85
Mean and Standard Deviation of SSI and Non-SSI States on the Individual Items of the Calculator Use Indicator, Grade 8, 1990

	SSI states *N* = 20		Non-SSI states *N* = 17	
Item	Mean	*SD*	Mean	*SD*
Frequency of use	2.24	0.34	2.34	0.37
Unrestricted use	1.17	0.08	1.20	0.09
Use on tests	1.28	0.13	1.32	0.13
School owned calculators	1.60	0.16	1.58	0.14

Item intercorrelations. At grade 8 in 1990, state means on the first three items are strongly related, as shown in Table 6.86. The lowest correlations are between students' access to school-owned calculators and the other three items. Whether or not a school owns calculators for student use is likely to be related not only to the extent to which calculators are used in classes, but also to other factors, such as the socioeconomic level of the community. In affluent areas, schools may not need to provide calculators for student use.

Table 6.86.
Intercorrelations of State Means on the Four Calculator Use Items, Grade 8, 1990

Total sample, N = 37	1.	2.	3.
1. Frequency of use			
2. Unrestricted use	.89*		
3. Use on tests	.96*	.92*	
4. School owned calculators	.61*	.53*	.56*

Subsample, N = 36	1.	2.	3.
1. Frequency of use			
2. Unrestricted use	.90*		
3. Use on tests	.96*	.92*	
4. School owned calculators	.61*	.53*	.55*

*$p < .01$

Supplementary information. Table 6.87 presents the means and standard deviations for $I_{C(90)}$ as a function of the state's SSI status and its 1994-95 policy on the use of calculators on state assessments. A 2x2 ANOVA found no significant effect for SSI status, calculator-use policy, or the interaction.

Table 6.87
Mean and Standard Deviation of SSI and Non-SSI States on $I_{C(90)}$, Grade 8 as a Function of Whether Students Were Permitted to Use Calculators on State Achievement Tests

	SSI States	Non-SSI States
Permitted	N = 13	N = 8
Mean	6.34	6.65
Standard Deviation	0.71	0.70
Not permitted	N = 4	N = 5
Mean	6.10	6.07
Standard Deviation	0.55	0.61

One limitation of this analysis is that calculator use policy refers to the 1994-95 school year and may be unrelated to policies in 1990. However, the analysis is included as a reference point for the analyses of the 1996 NAEP data.

Change from 1992 to 1996

Grade 8

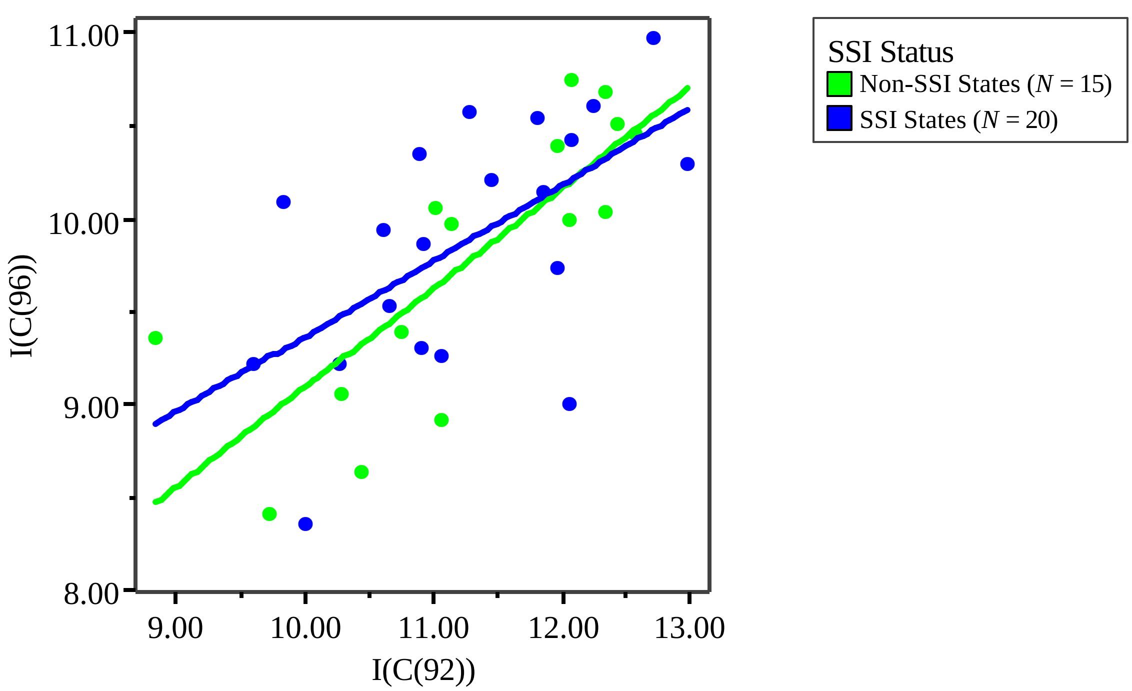
Two-point trend sample. At grade 8, I_C is not directly comparable from 1992 to 1996 because the 1992 measure has seven items while the 1996 measure has five. The two-step regression model predicting $I_{C(96)}$ from $I_{C(92)}$ and SSI status found a significant effect for the 1992 measure, but no effect for SSI status. Results are summarized in Table 6.88, and graphed in Figure 6.41.

Table 6.88
Predicting $I_{C(96)}$ from $I_{C(92)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{C(92)}$	0.47	0.09	.68	.46	28.25*		
Step 2							
$I_{C(92)}$	0.47	0.09	.68				
SSI status	.11	0.18	.08	.47	14.36*	.01	0.40

*$p < .01$

Figure 6.41. Relationship between $I_{C(96)}$ and $I_{C(92)}$ for SSI and non-SSI states, grade 8.



In the cross-sectional analyses in 1996, whether students were allowed to use calculators on the state assessment was significantly related to I_C. A second regression model included calculator use policy at step 1, along with the 1992 indicator. (See Table 6.89.) Both $I_{C(92)}$ and calculator use were related to $I_{C(96)}$, but SSI status did not add to the predictability of $I_{C(96)}$.

Table 6.89
Predicting $I_{C(96)}$ from $I_{C(92)}$, Calculator Use Policy, and SSI Status, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{C(92)}$	0.42	0.10	.60				
Calculator use	0.42	0.23	.26	.53	13.28*		
Step 2							
$I_{C(92)}$	0.42	0.10	.60				
Calculator use	0.40	0.24	.26				
SSI status	0.16	0.22	.11	.54	8.90*	.01	0.58

*$p < .01$

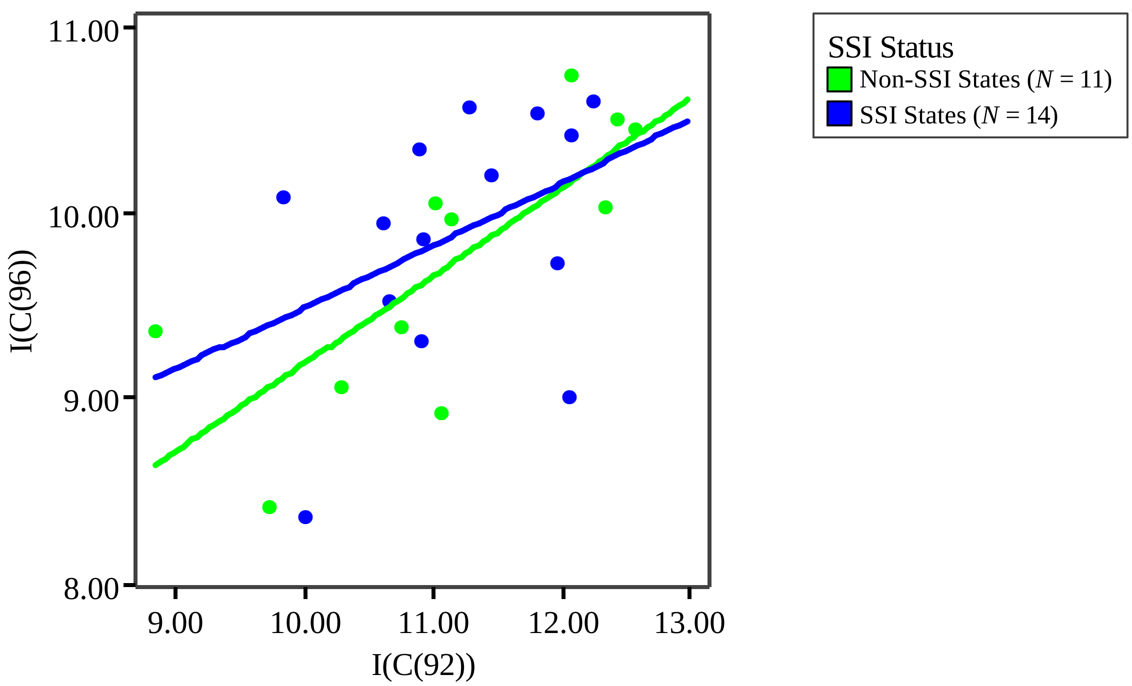
Subsample. For comparison purposes, the analysis was repeated with the subsample of 14 SSI states and 11 non-SSI states. The results were the same as those with the full sample, as summarized in Table 6.90 and illustrated in Figure 6.42.

Table 6.90
Predicting $I_{C(96)}$ from $I_{C(92)}$ and SSI Status for the Subsample of States that Met the NCES Participation Rate Guidelines, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{C(92)}$	0.43	0.12	.60	.36	12.80*		
Step 2							
$I_{C(92)}$	0.43	0.12	.59				
SSI status	0.14	0.23	.10	.37	6.42*	.01	0.38

*$p < .01$

Figure 6.42. Relationship between $I_{C(96)}$ and $I_{C(92)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.



For the subsample as with the full sample, calculator use policy added to the prediction of the 1996 indicator, but SSI status did not.

Table 6.91
Predicting $I_{C(96)}$ From $I_{C(92)}$, Calculator-Use Policy, and SSI Status for the Subsample of States that Met the Participation Rate Guidelines, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_{Δ}
Step 1							
$I_{C(92)}$	0.34	0.14	.45				
Calculator use	0.61	0.26	.42	.51	8.80*		
Step 2							
$I_{C(92)}$	0.34	0.14	.45				
Calculator use	0.60	0.27	.41				
SSI status	0.11	0.26	.08	.51	5.65*	.00	.19

*$p < .01$

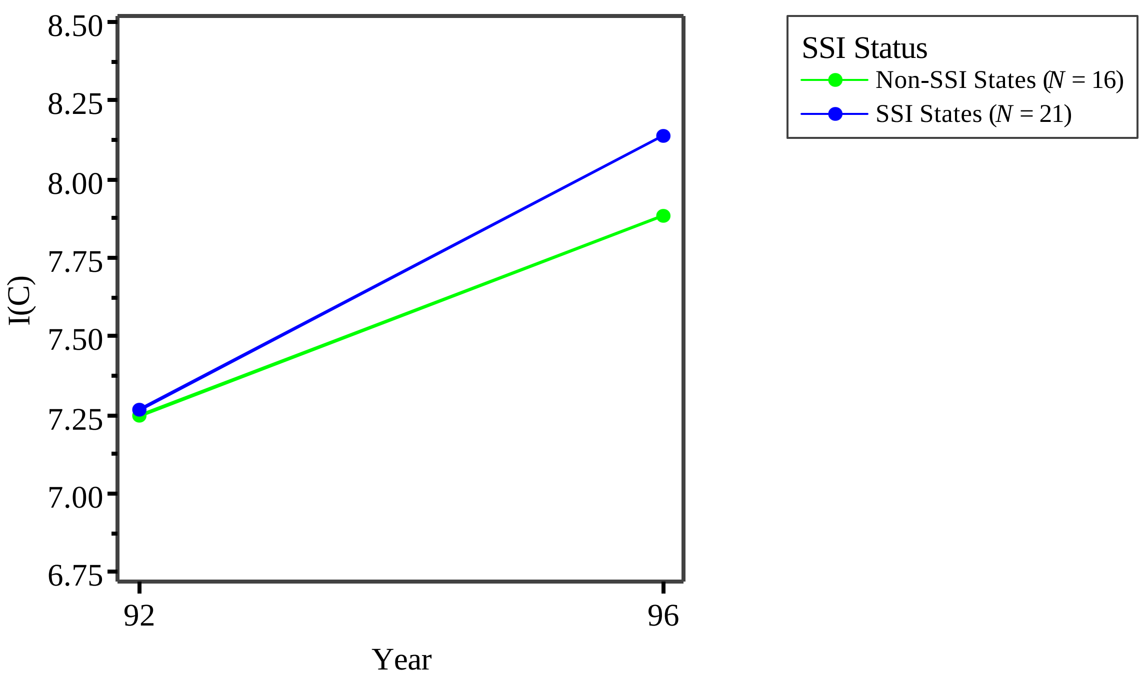
Grade 4

Two-point trend sample. At grade 4, I_C is directly comparable from 1992 to 1996, so a 2x2 repeated measure analysis of variance was used to examine change over time and the effect of SSI status on any change. Results are summarized in Table 6.92, and graphed in Figure 6.43. I_C increased significantly from 1992 to 1996 ($F = 121.91$, $df = 1.35$, $p < .01$). The interaction of year and SSI status was also significant, with SSI states increasing more than non-SSI states ($F = 3.04$, $df = 1,35$, $p < .10$). The main effect for SSI was not statistically significant ($F = .63$, $df = 1,35$, $p = .43$).

Table 6.92
Mean and standard deviation of I_C in SSI and Non-SSI States in 1992 and 1996, Grade 4

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI States	21	7.27	0.60	8.13	0.55	0.86	0.43
Non-SSI States	16	7.25	0.56	7.88	0.51	0.63	0.38

Figure 6.43. Change in the mean of I_C for SSI and non-SSI states from 1992 to 1996, grade 4.



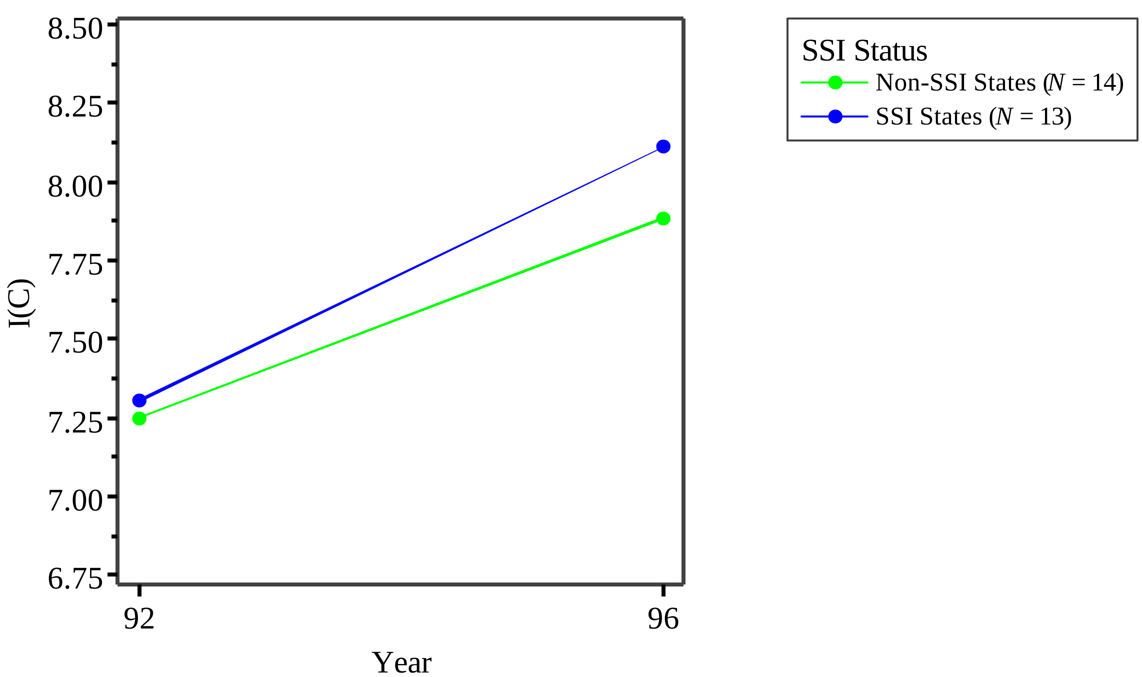
An additional analysis looked at the effect of the state's calculator use policy. In that analysis, the only statistically significant effect was for time. However, the power to detect any differences was small because some of the cells of the design had very few cases.

Subsample. Results for the subsample of 13 SSI states and 14 non-SSI states are summarized in Table 6.93, and graphed in Figure 6.44. As with the total sample, I_C increased significantly from 1992 to 1996 ($F = 67.56$, $df = 1{,}25$, $p < .01$). The interaction of year and SSI status was not statistically significant for the subsample, though SSI states increased more than non-SSI states ($F = 1.02$, $df = 1{,}25$, $p = .32$). The main effect for SSI was not statistically significant ($F = .48$, $df = 1{,}25$, $p = .50$).

Table 6.93
Mean and standard deviation of I_C in SSI and non-SSI states in 1992 and 1996, Grade 4

		1992		1996		Change	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
SSI states	13	7.31	0.58	8.11	0.55	0.80	0.41
Non-SSI states	14	7.25	0.60	7.88	0.55	0.63	0.41

Figure 6.44. Change in the mean of I_C for the subsample of SSI and non-SSI states that followed the NCES participation rate guidelines, grade 4, 1992 to 1996.



The graph for the subsample is similar to that for the total sample.

A supplementary analysis, including the state's calculator-use policy, found a significant effect for time, but no other statistically significant effect. However, the ability to detect differences was limited by the small sample size in some of the cells.

Change from 1990 to 1996

Grade 8

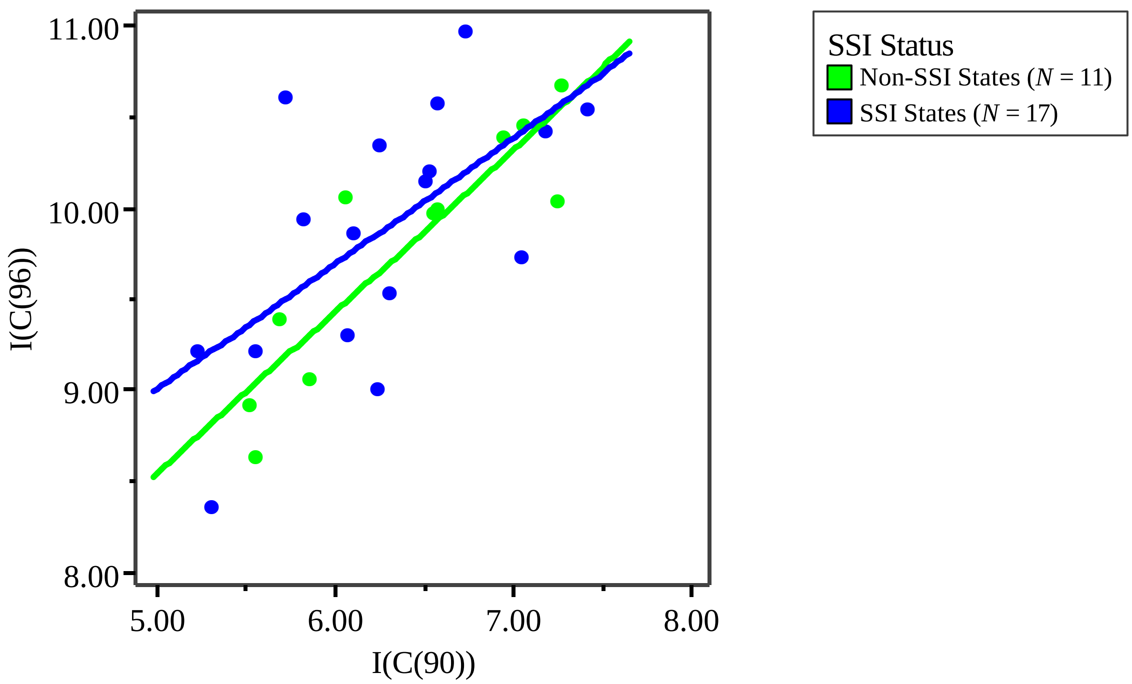
Two-point trend sample. At grade 8, I_C is not directly comparable from 1990 to 1996 because the 1990 measure has four items while the 1996 measure has five and the response options are different for one of the items. The two-step regression model predicting $I_{C(96)}$ from $I_{C(90)}$ and SSI status found a significant effect for the 1990 measure, but no effect for SSI status. Results are summarized in Table 6.91 and graphed in Figure 6.45.

Table 6.91
Predicting $I_{C(96)}$ from $I_{C(90)}$ and SSI Status, Grade 8

	B	SE B	β	R^2	F	ΔR^2	F_Δ
Step 1							
$I_{C(90)}$	0.77	0.14	.72	.52	27.72*		
Step 2							
$I_{C(90)}$	0.78	0.15	.73				
SSI status	0.20	0.19	.14	.54	14.48*	.02	1.11

*p < .01

Figure 6.45. Relationship between $I_{C(96)}$ and $I_{C(90)}$ for SSI and non-SSI states, grade 8.



Predictors of $I_{C(96)}$ included the states' calculator use policy. States that permitted students to use calculators on state tests averaged higher on $I_{C(96)}$ than expected from their average on $I_{C(90)}$, as shown in Table 6.95. However, SSI status did not add to the prediction of $I_{C(96)}$.

Table 6.95

Predicting $I_{C(96)}$ from $I_{C(92)}$, Calculator Use Policy, and SSI Status, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{C(90)}$	0.90	0.14	.78				
Calculator use	0.34	0.19	.22	.75	27.20*		
Step 2							
$I_{C(90)}$	0.92	0.14	.80				
Calculator use	0.31	0.19	.20				
SSI status	0.20	0.18	.13	.77	18.76*	.02	1.22

*$p < .01$

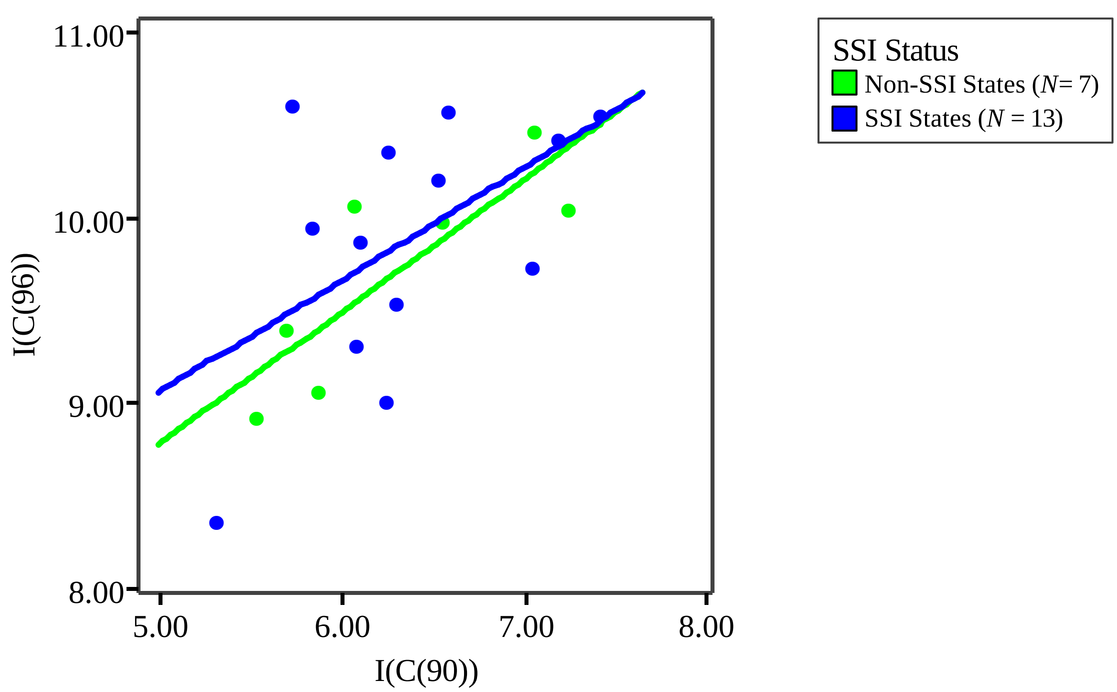
Subsample. For comparison purposes, the analysis was repeated with the subsample of 14 SSI states and 11 non-SSI states. The results were the same as those with the full sample, as summarized in Table 6.96 and illustrated in Figure 6.46.

Table 6.96
Predicting $I_{C(96)}$ from $I_{C(90)}$ and SSI Status for the Subsample of States That Met the Participation Rate Guidelines, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{C(90)}$	0.65	0.19	.62	.39	11.34*		
Step 2							
$I_{C(92)}$	0.65	0.20	.62				
SSI status	0.14	0.25	.11	.40	5.61	.01	0.31

*$p < .01$

Figure 6.46. Relationship between $I_{C(96)}$ and $I_{C(90)}$ for the subsample of SSI and non-SSI states that met the NCES participation rate guidelines, grade 8.



For the subsample as with the full sample, the state's calculator use policy was related to the state's average on $I_{C(96)}$, but SSI status was not (see Table 6.97).

Table 6.97
Predicting $I_{C(96)}$ from $I_{C(90)}$, Calculator Use Policy, and SSI Status, Grade 8

	B	SE B	β	R^2	*F*	ΔR^2	F_Δ
Step 1							
$I_{C(90)}$	0.82	0.20	.70				
Calculator use	0.38	0.22	.28	.74	16.73*		
Step 2							
$I_{C(90)}$	0.82	0.20	.70				
Calculator use	0.38	0.25	.28				
SSI status	0.00	0.25	.00	.73	10.22*	.00	.00

*$p < .01$

Summary of Results

In the cross-sectional comparisons for each of the three years of the State NAEP, there was no statistically significant difference between SSI and non-SSI states on I_C, the indicator of students' use of calculators. At grade 8 in 1996, however, SSI states averaged significantly higher than non-SSI states on one item of the indicator: student access to school-owned calculators.

Information about state policies on the use of calculators on state achievement tests was used to examine the influence of other factors in addition to the state's SSI status. For both grades 4 and 8 in 1996, states permitting students to use calculators on state assessments scored significantly higher on I_C than states that did not permit calculator use. No statistically significant differences were found in earlier years.

Change Across Time

Two-point trend; 1992-1996

Grade 8

$I_{C(92)}$ was significantly related to $I_{C(96)}$, but SSI status was not. The state's calculator use policy added to the predictability of $I_{C(96)}$.

Grade 4

I_C was directly comparable from 1992 to 1996, allowing an evaluation of time-related changes. I_C increased for both SSI and non-SSI states, and the increase was slightly more for the SSI states. Analyses of calculator-use policy did not find a significant effect.

Three-point trend; 1990, 1992, and 1996

Grade 8

$I_{C(90)}$ was significantly related to $I_{C(96)}$, but SSI status was not. The state's calculator use policy added to the predictability of $I_{C(96)}$.

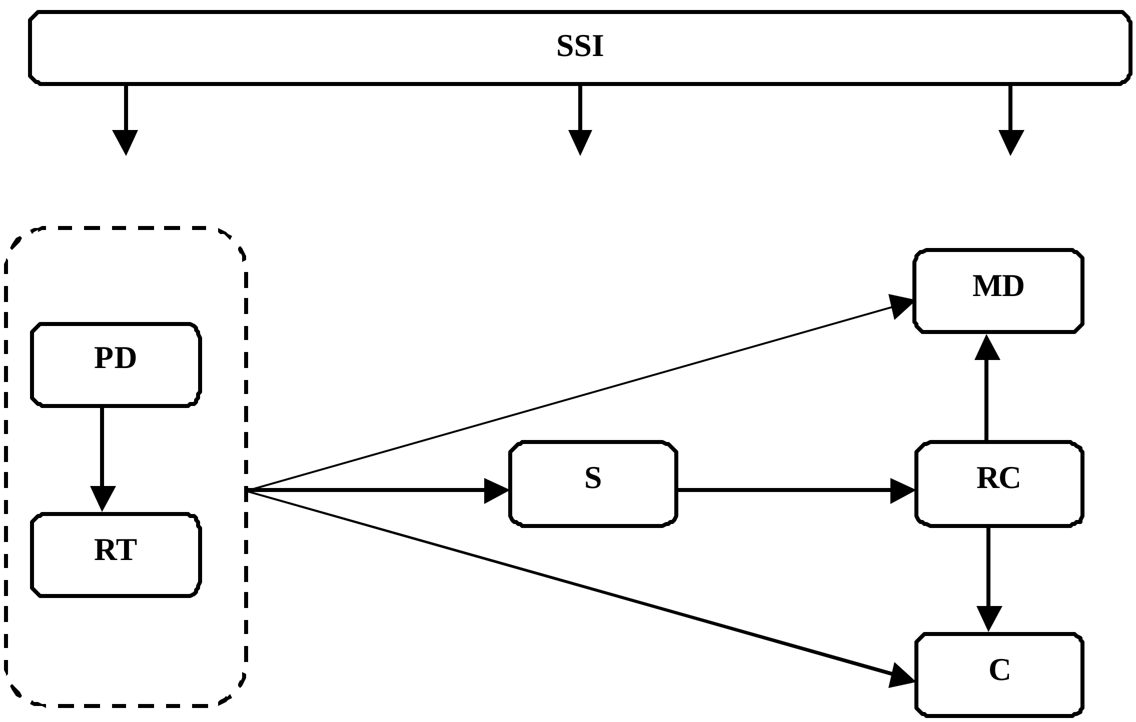
Interrelationships Among the Indicators

The indicators represent important components of a model for implementing curricular reform in mathematics. Clune (1998) describes the complexity of systemic reform. The indicators developed from the State NAEP teacher questionnaire provide limited information on just a few of the components of reform. However, they are useful in examining the impact of the SSI program.

Figure 6.47 illustrates the relationships among the indicators. The indicators are:

I_{PD} – Time in Professional Development Last Year
I_{RT} – Number of Reform-Related Topics Studied
I_S – Teachers' Knowledge of NCTM *Standards*
I_{RC} – Relative Emphasis on Reasoning and Communication
I_{MD} – Students' Opportunities for Mathematical Discourse
I_C – Students' Use of Calculators.

Figure 6.47. Relationships among the indicators within a model of systemic reform.



The indicators fall into three groups, organized according to assumptions of causality implicit in the model of reform. The indicators at the left of Figure 6.47, I_{PD} and I_{RT}, measure teachers' opportunities to learn about mathematics and mathematics instruction. In a complete model, many other factors would be included to describe the teachers' education and training, both preservice and inservice. These two indicators were selected because it seemed plausible that the SSI program could affect them. However, the effect may be somewhat limited because of the constraints of the indicators. I_{PD} asks only about time in staff development during the last year, so the time frame is relatively limited. I_{RT} asks whether teachers have ever studied each of seven topics, so the indicator may be too broad to measure the specific effects of the SSI program. Figure 6.47 shows a relationship between I_{PD} and I_{RT}. With more time in professional development, it might be reasonable to assume that the likelihood of studying the seven reform-related topics would increase.

The middle group of indicators includes just one, I_S, teachers' knowledge of the NCTM standards. In a complete model, the middle group would represent what teachers know and are able to do. I_S is expected to be related to the SSI program, because the NCTM *Standards* were used as a framework for planning the direction of reform. Knowledge of the standards is considered as an important organizing principle of the SSI programs. If nothing else, an effective SSI increase teachers' knowledge of the *Standards*.

The indicators at the right, I_{RC}, I_{MD} and I_C, describe instructional activities. This extends the model a step beyond what teachers know and are able to do to what they typically do in their classes throughout the year. As the diagram shows, $I_{RC,}$ the relative emphasis on reasoning and communication, is considered to result from I_S. In the model, I_{MD}, students' opportunities for mathematical discourse, and I_C, students' use of calculators, result in part from I_{RC}, teacher's knowledge of the standards.

This model proposes that the strongest relationships would be among the 3 indicators at the right side of the diagram, and the next strongest would be between between I_S and I_{RC}. While relationships between I_{PD} and I_{RT} and the other indicators are expected to be positive, they may not be very strong because limitations of the measures as well as the many other factors that affect each of the indicators.

Correlations among the indicators for each year of State NAEP are presented in the following sections.

1996

Table 6.98 presents the intercorrelations at grade 8 and grade 4 for the total sample and the subsample that met NCES participation rate guidelines.

Table 6.98
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8 and Grade 4, 1996

	Grade 8					Grade 4				
	I_{PD}	I_{RT}	I_S	I_{RC}	I_{MD}	I_{PD}	I_{RT}	I_S	I_{RC}	I_{MD}
Total Sample										
I_{RT}	.39					.40*				
I_S	.39	.13				.21	.08			
I_{RC}	.36	.26	.72*			.47*	.36	.38		
I_{MD}	.46*	.48*	.50*	.65*		.43*	.50*	.44*	.86*	
I_C	.14	.32	.55*	.55*	.49*	-.02	.21	.48*	.33	.47*
Subsample										
I_{RT}	.49*					.40*				
I_S	.36	.21				.06	.04			
I_{RC}	.43	.50*	.70*			.50*	.47*	.34		
I_{MD}	.49*	.59*	.48*	.66*		.46*	.50*	.41	.87*	
I_C	.24	.38	.67*	.60*	.49*	-.06	.17	.48*	.33	.48*

*$p < .01$

The correlations support the importance of teachers' knowledge of the NCTM *Standards* at grade 8. I_S was significantly correlated with all three indicators, I_{RC}, I_{MD}, and I_C, and the three indicators were moderately related to each other. One of the classroom practice indicators, I_{MD}, was related to the two indicators of staff development, I_{PD} and I_{RT}. In the subsample, another classroom practice indicator, I_{RC}, was also significantly related to an indicator of staff development, I_{RT}.

At grade 4, I_S was significantly related to one classroom practice indicator, I_C, in both samples and also to another, I_{MD}, in the total sample. The strongest correlation was between I_{RC} and I_{MD}. The two indicators of staff development, I_{PD} and I_{RT}, were related to both I_{RC} and I_{MD} as well as to each other. At grade 4, the calculator use indicator, I_C, was significantly related to I_{MD} but not to I_{RC}.

1992

Five of the six indicators were included in the 1992 NAEP teacher questionnaire. However, all 1992 indicators were slightly different from those in 1996 because of differences in the wording of the questions and/or differences in the response options. The one indicator missing in 1992 was I_S, teachers' knowledge of NCTM *Standards*. The intercorrelations of the 1992 indicators are listed in Table 6.99.

Table 6.99
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8 and Grade 4, 1992

	Grade 8				Grade 4			
	I_{PD}	I_{RT}	I_{RC}	I_{MD}	I_{PD}	I_{RT}	I_{RC}	I_{MD}
<u>Total Sample</u>								
I_{RT}	.39				.39			
I_{RC}	.36	.36			.42*	.49*		
I_{MD}	.41*	.52*	.70*		.30	.46*	.91*	
I_C	.22	.11	.46*	.50*	.34	.43*	.45*	.47*
<u>Subsample</u>								
I_{RT}	.35				.42			
I_{RC}	.36	.36			.46*	.55*		
I_{MD}	.37	.54*	.71*		.34	.55*	.92*	
I_C	.16	.09	.50*	.49*	.32	.40	.49*	.52*

______
*$p < .01$

The intercorrelations of the 1992 indicators were similar to those for 1996 in several ways:

- The highest correlation was between I_{RC} and I_{MD}.
- All three classroom practice indicators were significantly related at grade 8.
- Both staff development indicators were significantly related to I_{MD} at grade 8 and I_{RT} was significantly related at grade 4.
- At grade 4, I_{PD} was significantly related to I_{RC}, but not at grade 8.
- I_C was not related to either staff development indicator in most comparisons.

And there were a few differences:

- In 1996, the two staff development indicators were significantly related, but not in 1992.
- In 1996 at grade 4, I_C was not significantly related to the other two classroom practice indicators, but it was in 1992.

1990

The 1990 NAEP Teacher Questionnaire was substantially different from later questionnaires. I_{RT} included just one of the seven reform-related topics, I_{MD} had 2 items and I_C had 4. Table 6.100 lists the intercorrelations in 1990.

Table 6.100
Intercorrelations Among the Six Indicators of Mathematics Reform at Grade 8, 1990

	I_{PD}	I_{RT}	I_{RC}	I_{MD}
Total Sample				
I_{RT}	.00			
I_{RC}	.43*	.36		
I_{MD}	.43*	.32	.66*	
I_{C}	.43*	.24	.48*	.65*
Subsample				
I_{RT}	.01			
I_{RC}	.43*	.38		
I_{MD}	.43*	.33	.66*	
I_{C}	.44*	.22	.50*	.66*

In 1990, as in the other years, the strongest correlation was between I_{RC} and I_{MD}, and all three classroom practice indicators were significantly related. I_{PD} was significantly related to all classroom practice indicators, suggesting that states where teachers spent more hours in staff development over the year were also states where teachers used reform practices more frequently. In 1990, I_{RT} was not related to other measures, unlike later years. However, only one of the seven reform-related topics was included in the 1990 teacher questionnaire, so $I_{RT(90)}$ is a very limited measure of teachers' study of reform-related topics.

Discriminant Analyses

Descriptive discriminant analysis was used to examine whether the six indicators, as a group, could distinguish between the SSI and non-SSI states. In 1996 at both grade 4 and 8 the indicators differentiate the two groups, but not in earlier years. For grade 8, 1996, the canonical correlation was .56 ($p < .05$) for the total sample and .78 ($p < .01$) for the subsample; for grade 4 it was .66 ($p < .01$) for the total sample and .72 ($p < .01$) for the subsample.

Table 6.101 presents the classification results for grade 8 and Table 6.102 for grade 4. The discriminant functions correctly classified 77.5% of the total sample and 90% of the subsample at grade 8 and 88.4% of the total sample and 90.6% of the subsample at grade 4. A state's SSI status is related to the six indicators of curricular reform at both grade 4 and grade 8.

Table 6.101
Classification Results for the Descriptive Discriminant Function, Grade 8, 1996

	Predicted Status	
Actual status	SSI	Non-SSI
SSI	California[a] Colorado[a] Connecticut[a] Delaware[a] Florida[a] Georgia[a] Kentucky[a] Maine[a] Massachusetts[a] Montana[c] Nebraska[a] New York[c] North Carolina[a] Rhode Island[a] South Carolina[c] Texas[a] Vermont[c] Virginia[a]	Arkansas[c] Louisiana[a] Michigan[c] New Mexico[b]
Non-SSI	Alaska[c] Arizona[a] Hawaii[b] Maryland[c] Mississippi[a]	Alabama[a] Indiana[a] Iowa[c] Minnesota[a] Missouri[a] North Dakota[a] Oregon[a] Tennessee[a] Utah[a] Washington[a] West Virginia[a] Wisconsin[c] Wyoming[a]

[a]Subsample classification the same
[b]Subsample classification different
[c]Not in subsample

Table 6.102
Classification Results for the Descriptive Discriminant Function, Grade 4, 1996

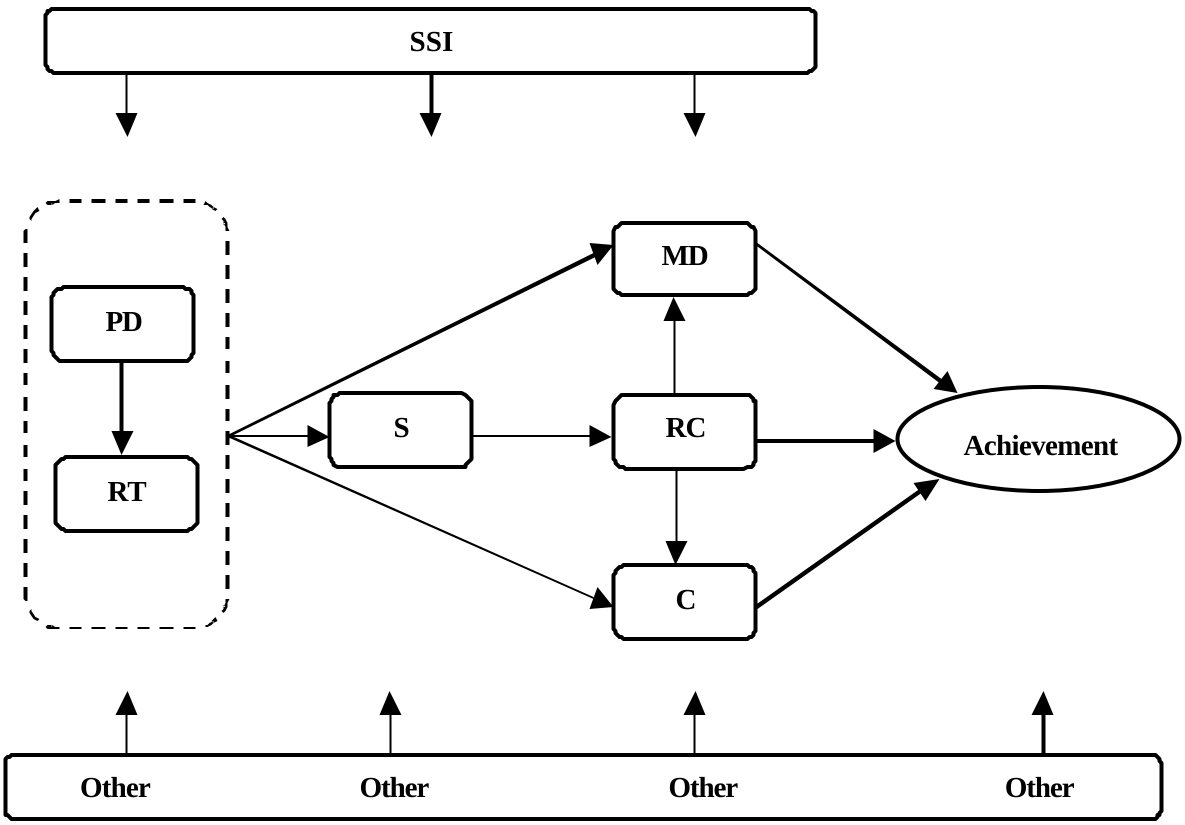
	Predicted Status	
Actual status	SSI	Non-SSI
SSI	Arkansas[c] California[a] Colorado[a] Connecticut[a] Delaware[a] Florida[a] Georgia[a] Kentucky[a] Louisiana[a] Maine[a] Massachusetts[a] Michigan[c] Montana[c] New Jersey[c] New York[c] North Carolina[a] Rhode Island[a] South Carolina[c] Texas[a] Vermont[c] Virginia[a]	Nebraska[a] New Mexico[a]
Non-SSI	Hawaii[b] Maryland[a] Nevada[c]	Alabama[a] Alaska[c] Arizona[a] Indiana[a] Iowa[c] Minnesota[a] Mississippi[a] Missouri[a] North Dakota[a] Oregon[a] Pennsylvania[c] Tennessee[a] Utah[a] Washington[a] West Virginia[a] Wisconsin[a] Wyoming[a]

[a]Subsample classification the same
[b]Subsample classification different
[c]Not in subsample

Relationships Between the Indicators and NAEP Mathematics Achievement

This section examines the relationship between the indicators and the states' mean mathematics composite scores. Figure 6.47 extends the previous diagram to include student achievement and the many other factors that affect each component of the model.

Figure 6.48. Indicators of mathematics curricular reform and their relationship to student achievement.



I_{PD} – Time in Professional Development Last Year
I_{RT} – Number of Reform-Related Topics Studied
I_{S} – Knowledge of the NCTM Standards
I_{MD} – Students' Opportunities for Mathematical Discourse
I_{RC} – Relative Emphasis on Reasoning and Communication
I_{C} – Use of Calculators
Achievement – NAEP Mathematics Composite

As the figure shows, the three indicators of classroom practice are expected to be directly related to student achievement. The other indicators are expected to have their effects primarily indirectly, through the classroom practice indicators.

Univariate analyses

The correlations for 1996, 1992, and 1990 for the total samples and the subsamples are reported in Table 6.103.

Table 6.103
Correlations Between the Individual Indicators and the Mean NAEP Mathematics Composite in 1996, 1992, and 1990

	N	I_{PD}	I_{RT}	I_S	I_{RC}	I_{MD}	I_C
Grade 8							
1996							
Total sample	40	-.20	-.02	.30*	.31*	-.06	.62*
Subsample	30	-.16	-.02	.32*	.21	-.18	.55*
1992							
Total sample	41	-.08	-.13	-	.23	.00	.57*
Subsample	36	-.10	-.16 -	.25	-.05	.55*	
1990							
Total sample	41	.08	-.04	-	.08	.25	.61*
Subsample	36	.10	-.09	-	.10	.27	.60*
Grade 4							
1996							
Total sample	43	-.35*	-.30*	.10	-.16	-.15	.35*
Subsample	32	-.29	-.33*	-.11	-.19	-.18	.35*
1992							
Total sample	41	-.36*	-.09	-	-.14	-.07	.24
Subsample	36	-.38*	-.10	-	-.20	-.14	.24

*$p < .10$

In 1996, at grade 8, the state mean on I_S, teachers' knowledge of the NCTM *Standards*, was positively and significantly related to its mean NAEP mathematics composite. Since NAEP was designed to be aligned with the NCTM *Standards*, the relationship between teachers' knowledge of the standards and students' performance on the test is evidence of validity of the interpretation of both the test and the indicator as a measure of mathematics reform efforts. At grade 4, the relationship between I_S and the NAEP mathematics composite was smaller and not statistically significant.

Also in 1996 at grade 8, the state mean on I_{RC}, the relative emphasis on reasoning and communication, was positively and significantly related to the NAEP mathematics composite for the total sample, but not for the subsample. I_{RC} was not significantly related to the NAEP mathematics composite in any other sample.

For all grade 8 samples and the 1996 grade 4 samples, the state means on I_C, students' use of calculators, was strongly and positively related to NAEP mathematics achievement. The value of the correlation ranged between .55 and .62 for the grade 8 comparisons and was lower at grade 4.

The state mean on I_{MD}, students' opportunities for mathematical discourse, was not significantly related to the NAEP mathematics composite for any year at either grade level.

The other two indicators, I_{PD} and $I_{RT,}$ were not related to achievement in any of the grade 8 samples, but they were significantly and negatively related at grade 4. For grade 4 in both 1992 and 1996, states with the lowest means for NAEP mathematics achievement had students whose teachers had spent more time in professional development. In 1996, states with relatively low NAEP mathematics composites reported that students' teachers had studied more reform-related topics.

Multivariate analyses

Ultimately the goal of any curricular reform effort is to increase student achievement. The SSI program included some assumptions about how to increase student achievement. The indicators used in this study were created to exaine whether SSI states were implementing practices that were expected to result in gains in student achievement. In this section, the relationship between the indicators and student achievement is explored.

Multiple linear regression techniques were used to explore whether the group of six indicators predicted student achievement. Both full models, and reduced, parsimonious models were computed. Intercorrelations among predictors create challenges to estimating regression parameters. Since the indicators were designed to measure various aspects of the SSI program, they are expected to be related. Results of the regression analyses were considered carefully, in light of the limitations of the analytic technique. Both full models, and reduced, parsimonious models were examined.

Grade 8. In the full models with all indicators, the predictors as a set were significantly related to student achievement in all samples and subsamples. In 1992 and 1996, a reduced model of I_{RC}, I_{MD}, and I_C did about as well as the full model. The beta for I_{RC} ranged from .20 to .40 in the models, and the beta for I_C ranged from .66 to .77. The beta for I_{MD} ranged from -.54 to -.69. In 1990, the only significant predictor was I_C, with a beta of around .80 in the multivariate models.

The negative weighting for I_{MD} was unexpected. The result was found in both 1992 and 1996 and was fairly large. The significant negative betas for I_{MD} in the multivariate model contrast with the univariate correlations which were close to .00 (Table 6.103). In all years, I_{MD} was highly correlated with I_{RC}, and moderately correlated with I_C (Tables 6.98 and 6.99). In addition, both I_{RC} and I_C were positively related to the NAEP mathematics composite. In this situation, I_{MD} seems to be acting as a suppressor variable. That is, when I_{MD} was adjusted for I_{RC} and I_C, the residual was negatively related to achievement.

I_{MD} addressed whether students work in small groups, write about problem solutions, and give presentations. However, the measure did not address the content of the discussions, solutions, or presentations. The negative coefficient suggests that mathematical discourse opportunities are most effective when they are linked to an emphasis on reasoning and communication. If there is a mismatch, either with relatively many opportunities for mathematical discourse and a relatively low emphasis on reasoning and communication, or with relatively few opportunities for discourse and a relatively high emphasis on reasoning and communication, the prediction of the NAEP mathematics composite is adjusted by the difference between I_{MD} and I_{RC}.

Grade 4. At grade 4, I_C was a significant predictor in the multivariate regression models for all samples and subsamples. In both 1996 and 1992, the prediction of the State NAEP mathematics composite was increased with another indicator in the model. In 1996, adding either I_{RC} or I_{MD} significantly increased the model R^2, with I_{MD} increasing it a bit more than I_{RC}. In the models, the beta for I_C was positive, and the betas for I_{MD} and I_{RC} were negative. In 1992, adding I_{RC} to the model increased the model R^2 alternatively adding I_{PD} increased R^2 a bit more. The beta for I_C increased when the other indicator was added to the model, and both added indicators had negative betas.

As with the findings for grade 8, the multivariate model included at least one measure with a negative beta. In general, we can conclude that when the indicators are in agreement, student achievement is a linear function of the indicators, but when one indicator is higher or lower than would be expected on the basis of the other indicators, achievement predictions should be adjusted.

Next steps. The models developed here were based on state means on the indicators and state means on the State NAEP mathematics composite. To date, we have not investigated these relationships within the states. In addition, the findings are based on correlational methods. Further research is needed to separate the effects of the indicators, either by sampling to minimize the interdependencies or with experimental methods. Nevertheless, the findings demonstrate the importance of using multivariate as well as univariate approaches.

Summary and Conclusions

The analyses reported here used information from the State NAEP teacher questionnaires to examine the effects of the SSI program. Indicators of curricular reform in mathematics were developed from items on the teacher questionnaire. The analyses were based on state means that were computed using the weights provided in the State NAEP database. Both cross-sectional and longitudinal methods were used to compare the SSI states with the non-SSI states.

In the 1996 cross-sectional comparisons, SSI states, as a group, scored significantly higher than non-SSI states at grade 8 on five of the six indicators:

- I_{PD} – Time in Professional Development Last Year
- I_{RT} – Number of Reform-Related Topics Studied
- I_S – Teachers' Knowledge of NCTM *Standards*
- I_{RC} – Relative Emphasis on Reasoning and Communication
- I_{MD} – Students' Opportunities for Mathematical Discourse

and at grade 4 on four of the six indicators:

I_{PD} – Time in Professional Development Last Year
I_S – Teachers' Knowledge of NCTM *Standards*
I_{RC} – Relative Emphais on Reasoning and Communication
I_{MD} – Students' Opportunities for Mathematical Discourse

For the sixth indicator,

I_C – Students' Use of Calculators,

there were no statistically significant differences between the SSI and non-SSI states. However, I_C was related to whether the state allowed calculator use on their state achievement tests.

The 1992 cross sectional comparisons found that SSI states scored significantly higher than non-SSI states on I_{RC} and I_{MD} at grade 4. There were no statistically significant differences at grade 8.

A descriptive discriminant function analysis used all six indicators to correctly categorize about three-fourths of the total samples in 1996 and about 90% of the subsamples. No statistically significant functions were identified for 1992 or 1990.

The longitudinal comparisons also provide some evidence for the effectiveness of the SSI program, though comparisons are limited because of the reduced sample. From 1992 to 1996, four indicators can be compared directly: I_{PD}, I_{RT}, I_C and I_{MD4}, a subset of four matched items from I_{MD}. From 1990 to 1996, one indicator can be compared directly across all 3 years: I_{PD}. Three other indicators are similar, but not directly comparable: I_{RC}, I_{MD} and I_C.

For the four indicators that allow a direct comparison from 1992 to 1996, there was a statistically significant increase across time for both grade 4 and grade 8. At grade 4, the increase for SSI states was significantly greater than the increase for non-SSI states on I_{PD} and I_{RT}, but not on I_{MD4} or I_C. At grade 8, the SSI states in the total sample increased more than non-SSI states on I_{MD4}.

For the two indicators that can't be compared directly from 1992 to 1996, the 1992 measures were significantly related to the corresponding1996 measures. For I_{MD}, SSI status did not add to the prediction at either grade 4 or grade 8. For I_{RC}, SSI status was a significant predictor, with SSI states averaging higher-than-predicted compared to non-SSI states at both grade 4 and grade 8, although the grade 8 finding was limited to the total sample.

The findings for the three-point trend analyses at grade 8 differed for the three indicators. For I_{MD}, the state's SSI status was significantly related to $I_{MD(96)}$ but $I_{MD(90)}$ was not. For I_{RC}, $I_{RC(90)}$ was a significant predictor of $I_{RC(96)}$, but SSI status was not.

At the beginning of this project, the three-point trend analyses seemed to offer the strongest test for the SSI program. As the study progressed, however, the three-point analyses seemed increasingly limited. First, states did not participate consistently across all years of State NAEP, so the available sample was reduced to just 17 of the 25 SSI states and 11 of the 25 non-SSI states. Second, some states did not meet the participation rate guidelines, raising the concern of bias due to sampling. This further reduced the three-point trend sample to just 13 SSI and 7 non-SSI states. Third, the SSI states in the three-point trend sample started out higher than non-SSI states on some of the indicators, making it difficult to separate the effect of SSI from pre-existing differences between the groups. Finally, the teacher questionnaire changed considerably from one year to the next, so the comparability of the indicators was limited. Given these constraints, the three-year trend sample was small, and comparisons across three years were limited, providing little power to investigate the research questions.

The SSI states varied considerably. Many were among the highest states on the indicators and a few were among the lowest. Future research will examine the specific features of each state's SSI to account for the variability among the SSI states. We will examine the range of the indicators within a state, and explore relationships between the indicators and specific demographic groups. Through this analysis, we hope to refine our understanding of effective curricular reform.

List of Appendices

Appendix A

Table 6A.1.
State Participation in State NAEP

Appendix B

Table 6B.1
Comparison of SSI and Non-SSI States on $I_{RC(96)}$, for the All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6B.2
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1996

Table 6B.3
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 4, 1996

Table 6B.4
Comparison of SSI and Non-SSI States on $I_{RC(92)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6B.5
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1992

Table 6B.6
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 4, 1992

Table 6B.7
Comparison of SSI and Non-SSI States on $I_{RC(90)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6B.8
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1990

Appendix C

Table 6C.1
Comparison of SSI and Non-SSI states on $I_{MD(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines.

Table 6C.2
State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 8

Table 6C.3
State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 4

Table 6C.4
Comparison of SSI and Non-SSI States on $I_{MD4(92)}$ and $I_{MD(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6C.5
State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 8

Table 6C.6
State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 4

Table 6C.7
State Means on I_{MD4} for All States Participating in State NAEP in 1992 and 1996

Table 6C.8
Comparison of SSI and Non-SSI States on $I_{MD4(92)}$ and $I_{MD(92)}$ for All States in Each Sample and for the Subsampls of States that Followed the NCES Participation Rate Guidelines

Table 6C.9
State Means on $I_{MD(90)}$ and Each Mathematical Discourse Item, Grade 8

Appendix D

Table 6D.1
Comparison of SSI and Non-SSI States on $I_{S(96))}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6D.2
State Means on I_S for All States Participating in State NAEP, 1996

Appendix E

Table 6E.1
Comparison of SSI and Non-SSI States on $I_{PD(96)}$ for All States in Each Samples and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6E.2
State Means on I_{PD} for All States Participating in State NAEP

Table 6E.3
Comparison of SSI and Non-SSI States on $I_{PD(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6E.4
Comparison of SSI and Non-SSI States on $I_{PD(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Appendix F

Table 6F.1
Comparison of SSI and Non-SSI States on $I_{RT(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6F.2
State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 8

Table 6F.3
State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

Table 6F.4
Comparison of SSI and Non-SSI States on $I_{RT(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6F.5
State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 8

Table 6F.6
State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

Table 6F.7
Comparison of SSI and Non-SSI States on Percent of Students Whose Teachers had Studied "Teaching Students from Different Cultural Backgrounds", Grade 8

Table 6F.8
Percent of State Students Whose Teacher had Studied Teaching Students from Different Cultural Backgrounds, Grade 8, 1990

Appendix G

Table 6G.1
Comparison of SSI and Non-SSI States on $I_{C(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6G.2
State Means on $I_{C(96)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

Table 6G.3
State Means on $I_{C(96)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

Table 6G.4
Comparison of SSI and Non-SSI States on $I_{C(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6G.5
State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

Table 6G.6
State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

Table 6G.7
Comparison of SSI and Non-SSI States on $I_{C(90)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

Table 6G.8
State Means on $I_{C(90)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

Appendix A

Table 6A.1.
State Participation in State NAEP

	Grade 8					Grade 4		
	90	92	96	3-point trend	2-point trend	92	96	2-point trend
SSI States								
Arkansas	✓	✓	✓	✓	✓	✓	✓	✓
California	✓	✓	✓	✓	✓	✓	✓	✓
Colorado	✓	✓	✓	✓	✓	✓	✓	✓
Connecticut	✓	✓	✓	✓	✓	✓	✓	✓
Delaware	✓	✓	✓	✓	✓	✓	✓	✓
Florida*	✓	✓	✓	✓	✓	✓	✓	✓
Georgia	✓	✓	✓	✓	✓	✓	✓	✓
Kentucky	✓	✓	✓	✓	✓	✓	✓	✓
Louisiana	✓	✓	✓	✓	✓	✓	✓	✓
Maine	-	✓	✓	-	✓	✓	✓	✓
Massachusetts	-	✓	✓	-	✓	✓	✓	✓
Michigan	✓	✓	✓	✓	✓	✓	✓	✓
Montana	✓	-	✓	-	-	-	✓	-
Nebraska	✓	✓	✓	✓	✓	✓	✓	✓
New Jersey	✓	✓	-	-	-	✓	✓	✓
New Mexico	✓	✓	✓	✓	✓	✓	✓	✓
New York	✓	✓	✓	✓	✓	✓	✓	✓
North Carolina*	✓	✓	✓	✓	✓	✓	✓	✓
Ohio	✓	✓	-	-	-	✓	-	-
Rhode Island*	✓	✓	✓	✓	✓	✓	✓	✓
South Carolina	-	✓	✓	-	✓	✓	✓	✓
Texas	✓	✓	✓	✓	✓	✓	✓	✓
Vermont	-	-	✓	-	-	-	✓	-
Virginia*	✓	✓	✓	✓	✓	✓	✓	✓
Number of SSI states	20	22	22	17	20	22	23	21

*Funded for less than 5 years

Table 6A.1 (continued)
State Participation in State NAEP

	Grade 8					Grade 4		
	90	92	96	3-point trend	2-point trend	92	96	2-point trend
Non-SSI States								
Alabama	✓	✓	✓	✓	✓	✓	✓	✓
Alaska	-	-	✓	-	-	-	✓	-
Arizona	✓	✓	✓	✓	✓	✓	✓	✓
Hawaii	✓	✓	✓	✓	✓	✓	✓	✓
Idaho	✓	✓	-	-	-	✓	-	-
Illinois	✓	-	-	-	-	-	-	-
Indiana	✓	✓	✓	✓	✓	✓	✓	✓
Iowa	✓	✓	✓	✓	✓	✓	✓	✓
Maryland	✓	✓	✓	✓	✓	✓	✓	✓
Minnesota	✓	✓	✓	✓	✓	✓	✓	✓
Mississippi	-	✓	✓	-	✓	✓	✓	✓
Missouri	-	✓	✓	-	✓	✓	✓	✓
Nevada	-	-	-	-	-	-	✓	-
New Hampshire	✓	✓	-	-	-	✓	-	-
North Dakota	✓	✓	✓	✓	✓	✓	✓	✓
Oklahoma	✓	✓	-	-	-	✓	-	-
Oregon	✓	-	✓	-	-	-	✓	-
Pennsylvania	✓	✓	-	-	-	✓	✓	✓
Tennessee	-	✓	✓	-	✓	✓	✓	✓
Utah	-	✓	✓	-	✓	✓	✓	✓
Washington	-	-	✓	-	-	-	✓	-
West Virginia	✓	✓	✓	✓	✓	✓	✓	✓
Wisconsin	✓	✓	✓	✓	✓	✓	✓	✓
Wyoming	✓	✓	✓	✓	✓	✓	✓	✓
Number of non-SSI states	17	19	18	11	15	19	20	16

Appendix B

Table 6B.1
Comparison of SSI and Non-SSI States on $I_{RC(96)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	45.86	1.52	2.90	.006
Non-SSI	18	44.54	1.35		
Subsample					
SSI	16	45.90	1.29	4.26	.000
Non-SSI	14	44.08	1.04		
2-point trend sample					
Total sample					
SSI	20	45.69	1.34	2.23	.034
Non-SSI	15	44.63	1.44		
Subsample					
SSI	14	45.94	1.37	3.51	.002
Non-SSI	11	44.19	1.12		
3-point trend sample					
Total sample					
SSI	17	45.59	1.41	1.20	.243
Non-SSI	11	44.90	1.52		
Subsample					
SSI	13	45.86	1.40	2.44	.029
Non-SSI	7	44.37	1.25		
Grade 4					
1996 sample					
Total sample					
SSI	23	44.06	1.22	4.36	.000
Non-SSI	20	42.40	1.27		
Subsample					
SSI	16	44.17	0.98	4.75	.000
Non-SSI	16	42.34	1.18		
2-point trend sample					
Total sample					
SSI	21	43.98	1.07	4.32	.000
Non-SSI	16	42.34	1.20		
Subsample					
SSI	13	44.24	0.93	4.38	.000
Non-SSI	14	42.39	1.25		

Table 6B.2
State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1996

	X_F	X_P	X_R	X_C	I_{RC}
SSI States (N = 22)					
Arkansas†	2.70	2.78	2.23	2.01	42.89
California	2.60	2.61	2.45	2.31	47.22
Colorado	2.68	2.74	2.36	2.20	45.47
Connecticut	2.63	2.69	2.49	2.45	47.85
Delaware	2.56	2.68	2.43	2.34	47.07
Florida	2.72	2.72	2.37	2.23	44.75
Georgia	2.74	2.75	2.46	2.42	46.58
Kentucky	2.74	2.75	2.41	2.56	47.27
Louisiana	2.85	2.83	2.29	2.19	43.10
Maine	2.58	2.62	2.37	2.13	46.09
Massachusetts	2.69	2.76	2.49	2.39	46.94
Michigan†	2.65	2.73	2.34	2.25	45.60
Montana†	2.71	2.65	2.41	2.18	45.58
Nebraska	2.69	2.75	2.33	2.20	45.20
New Mexico	2.80	2.82	2.35	2.14	43.87
New York†	2.70	2.74	2.42	2.16	45.18
North Carolina	2.73	2.79	2.43	2.35	46.02
Rhode Island	2.53	2.70	2.31	2.26	46.20
South Carolina†	2.71	2.80	2.40	2.33	45.65
Texas	2.71	2.83	2.52	2.21	45.61
Vermont†	2.47	2.57	2.51	2.41	49.59
Virginia	2.73	2.75	2.33	2.30	45.18
Non-SSI States (N = 18)					
Alabama	2.77	2.77	2.22	2.16	43.49
Alaska†	2.75	2.69	2.25	2.20	44.75
Arizona	2.56	2.67	2.35	2.22	46.36
Hawaii	2.76	2.77	2.26	2.06	42.87
Indiana	2.66	2.76	2.28	2.00	43.49
Iowa†	2.47	2.61	2.34	2.14	47.10
Maryland†	2.61	2.69	2.33	2.41	47.00
Minnesota	2.65	2.70	2.37	2.00	44.54
Mississippi	2.78	2.80	2.36	2.29	44.77
Missouri	2.68	2.72	2.31	2.13	44.49
North Dakota	2.68	2.76	2.33	2.14	45.13
Oregon	2.59	2.71	2.27	2.10	44.52
Tennessee	2.82	2.80	2.25	2.14	42.84
Utah	2.77	2.79	2.30	2.10	43.35
Washington	2.61	2.68	2.26	1.86	43.09
West Virginia	2.76	2.71	2.24	2.12	43.22
Wisconsin†	2.61	2.68	2.35	2.12	45.73
Wyoming	2.67	2.78	2.39	2.17	45.01

†Did not meet the NCES participation rate guidelines.

Table 6B.3

State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 4, 1996

	X_F	X_P	X_R	X_C	I_{RC}
SSI States (*N* = 23)					
Arkansas†	2.94	2.88	2.19	2.11	41.36
California	2.88	2.76	2.39	2.21	44.28
Colorado	2.88	2.83	2.35	2.19	43.81
Connecticut	2.92	2.90	2.51	2.38	44.97
Delaware	2.89	2.86	2.42	2.30	44.31
Florida	2.93	2.91	2.42	2.30	43.91
Georgia	2.97	2.92	2.49	2.37	44.62
Kentucky	2.91	2.88	2.43	2.43	44.82
Louisiana	2.96	2.93	2.34	2.30	43.20
Maine	2.90	2.88	2.50	2.33	44.92
Massachusetts	2.91	2.88	2.43	2.25	43.78
Michigan†	2.84	2.86	2.39	2.07	43.14
Montana†	2.95	2.87	2.38	2.12	42.77
Nebraska	2.95	2.89	2.32	2.11	42.35
New Jersey†	2.94	2.92	2.54	2.31	44.61
New Mexico	2.85	2.81	2.27	2.17	43.11
New York†	2.92	2.91	2.50	2.25	43.90
North Carolina	2.87	2.89	2.55	2.40	45.62
Rhode Island	2.84	2.86	2.40	2.12	43.39
South Carolina†	2.93	2.90	2.44	2.29	44.01
Texas	2.91	2.92	2.67	2.41	46.06
Vermont†	2.84	2.78	2.51	2.51	46.91
Virginia	2.94	2.88	2.38	2.27	43.52
Non-SSI States (*N* = 20)					
Alabama	2.96	2.94	2.36	2.22	42.80
Alaska†	2.87	2.83	2.26	1.99	41.38
Arizona	2.90	2.76	2.31	2.16	43.17
Hawaii	2.93	2.84	2.30	2.12	42.36
Indiana	2.97	2.96	2.28	1.99	40.79
Iowa†	2.93	2.87	2.26	2.03	41.25
Maryland	2.87	2.87	2.46	2.48	45.59
Minnesota	2.83	2.84	2.28	1.99	42.08
Mississippi	2.95	2.92	2.40	2.32	43.84
Missouri	2.97	2.92	2.27	2.03	41.11
Nevada†	2.88	2.86	2.51	2.31	45.08
North Dakota	2.94	2.90	2.24	1.98	40.93
Oregon	2.88	2.89	2.35	2.09	42.49
Pennsylvania†	2.97	2.95	2.39	2.18	42.73
Tennessee	2.95	2.92	2.34	2.21	42.56
Utah	2.96	2.87	2.29	2.08	41.76
Washington	2.91	2.83	2.26	1.97	41.53
West Virginia	2.95	2.90	2.31	2.10	42.22
Wisconsin	2.93	2.89	2.32	2.03	41.79
Wyoming	2.96	2.88	2.39	2.06	42.50

†Did not meet the NCES participation rate guidelines.

Table 6B.4

Comparison of SSI and Non-SSI States on $I_{RC(92)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
<u>Grade 8</u>					
1992 sample					
Total sample					
SSI	22	43.50	2.09	0.99	.329
Non-SSI	19	42.86	2.03		
Subsample					
SSI	18	43.50	1.89	0.92	.364
Non-SSI	18	42.89	2.09		
2-point trend sample					
Total sample					
SSI	20	43.39	1.94	0.87	.392
Non-SSI	15	42.80	2.01		
Subsample					
SSI	14	43.74	1.23	3.63	.001
Non-SSI	11	41.97	1.20		
3-point trend sample					
Total sample					
SSI	17	43.30	2.04	0.27	.789
Non-SSI	11	43.07	2.23		
Subsample					
SSI	13	43.85	1.21	3.19	.008
Non-SSI	7	41.91	1.34		
<u>Grade 4</u>					
1992 sample					
Total sample					
SSI	22	40.08	1.93	1.79	.082
Non-SSI	19	39.09	1.62		
Subsample					
SSI	17	40.21	2.08	1.79	.083
Non-SSI	19	39.09	1.62		
2-point trend sample					
Total sample					
SSI	21	40.08	1.98	1.62	.114
Non-SSI	16	39.08	1.75		
Subsample					
SSI	13	40.47	1.90	1.92	.067
Non-SSI	14	39.07	1.86		

Table 6B.5

State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1992

	X_F	X_P	X_R	X_C	I_{RC}
SSI States (*N* = 22)					
Arkansas	1.69	1.76	1.18	1.12	38.10
California	1.50	1.67	1.44	1.28	44.73
Colorado	1.67	1.70	1.45	1.32	44.29
Connecticut	1.64	1.68	1.39	1.36	44.40
Delaware	1.69	1.74	1.43	1.31	43.46
Florida	1.75	1.78	1.43	1.43	43.16
Georgia	1.74	1.82	1.47	1.49	44.81
Kentucky	1.73	1.75	1.45	1.36	43.51
Louisiana	1.75	1.77	1.39	1.39	42.63
Maine†	1.55	1.64	1.44	1.25	44.48
Massachusetts	1.71	1.76	1.39	1.31	42.30
Michigan	1.59	1.67	1.49	1.34	45.91
Nebraska†	1.65	1.76	1.32	1.07	40.31
New Jersey†	1.66	1.74	1.61	1.50	47.57
New Mexico	1.71	1.80	1.38	1.31	41.55
New York†	1.70	1.71	1.33	1.23	41.68
North Carolina	1.67	1.71	1.42	1.32	43.54
Ohio	1.59	1.74	1.29	1.22	41.64
Rhode Island	1.53	1.69	1.44	1.39	46.37
South Carolina	1.67	1.77	1.43	1.47	44.91
Texas	1.67	1.79	1.55	1.39	44.66
Virginia	1.79	1.78	1.40	1.40	42.99
Non-SSI States (*N* = 19)					
Alabama†	1.77	1.76	1.37	1.35	42.44
Arizona	1.68	1.82	1.44	1.31	42.63
Hawaii	1.63	1.63	1.20	1.19	39.68
Idaho	1.64	1.73	1.45	1.23	42.90
Indiana	1.68	1.69	1.37	1.21	42.16
Iowa	1.59	1.70	1.39	1.30	44.42
Maryland	1.50	1.62	1.44	1.38	46.94
Minnesota	1.62	1.72	1.42	1.16	42.92
Mississippi	1.88	1.89	1.48	1.49	42.25
Missouri	1.59	1.66	1.33	1.21	43.40
New Hampshire	1.50	1.62	1.41	1.28	45.88
North Dakota	1.65	1.76	1.41	1.21	41.90
Oklahoma	1.82	1.76	1.29	1.29	39.99
Pennsylvania	1.72	1.80	1.46	1.36	43.59
Tennessee	1.74	1.81	1.36	1.35	41.82
Utah	1.73	1.82	1.40	1.22	40.75
West Virginia	1.77	1.78	1.38	1.26	40.61
Wisconsin	1.58	1.66	1.47	1.35	46.60
Wyoming	1.58	1.61	1.31	1.19	43.49

†Did not meet the NCES participation rate guidelines

Table 6B.6

State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 4, 1992

	X_F	X_P	X_R	X_C	I_{RC}
SSI States (*N* = 22)					
Arkansas	1.99	1.91	1.21	1.14	35.40
California	1.88	1.89	1.44	1.28	40.80
Colorado	1.88	1.86	1.44	1.32	41.21
Connecticut	1.89	1.92	1.54	1.32	41.48
Delaware †	1.95	1.93	1.33	1.28	38.16
Florida	1.96	1.94	1.50	1.43	41.52
Georgia	1.93	1.95	1.44	1.42	40.69
Kentucky	1.95	1.93	1.50	1.44	41.89
Louisiana	1.95	1.94	1.51	1.50	42.49
Maine †	1.83	1.86	1.45	1.17	39.98
Massachusetts	1.91	1.91	1.42	1.23	39.01
Michigan	1.91	1.90	1.52	1.34	41.41
Nebraska †	1.95	1.89	1.34	1.23	38.52
New Jersey†	1.95	1.97	1.53	1.46	41.71
New Mexico	1.95	1.89	1.30	1.27	38.16
New York†	1.97	1.95	1.44	1.31	39.76
North Carolina	1.92	1.89	1.49	1.33	40.94
Ohio	1.97	1.92	1.48	1.31	40.05
Rhode Island	1.96	1.93	1.30	1.09	35.89
South Carolina	1.95	1.93	1.42	1.41	40.67
Texas	1.92	1.93	1.58	1.42	42.62
Virginia	1.97	1.93	1.45	1.26	39.34
Non-SSI states (*N* = 19)					
Alabama	1.97	1.93	1.46	1.40	40.53
Arizona	1.94	1.90	1.39	1.24	39.14
Hawaii	1.88	1.88	1.38	1.21	38.94
Idaho	1.91	1.88	1.43	1.13	38.66
Indiana	1.96	1.88	1.29	1.14	36.68
Iowa	1.94	1.91	1.40	1.22	38.48
Maryland	1.88	1.89	1.63	1.43	43.87
Minnesota	1.89	1.88	1.40	1.13	38.49
Mississippi	1.96	1.93	1.39	1.37	39.76
Missouri	1.97	1.91	1.35	1.21	37.97
New Hampshire	1.92	1.88	1.41	1.26	39.91
North Dakota	1.99	1.93	1.29	1.11	35.99
Oklahoma	1.99	1.93	1.44	1.25	38.78
Pennsylvania	1.96	1.93	1.44	1.33	39.81
Tennessee	1.97	1.92	1.35	1.32	38.93
Utah	1.96	1.89	1.39	1.21	38.52
West Virginia	1.95	1.89	1.36	1.22	38.30
Wisconsin	1.93	1.90	1.52	1.25	40.28
Wyoming	1.95	1.91	1.48	1.22	39.66

†Did not meet the NCES participation rate guidelines.

Table 6B.7

Comparison of SSI and Non-SSI States on $I_{RC(90)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1990 sample					
Total sample					
SSI	20	46.09	1.12	1.08	.287
Non-SSI	17	45.68	1.16		
Subsample					
SSI	20	46.09	1.12	1.00	.324
Non-SSI	16	45.70	1.20		
2- and 3-point trend samples					
Total sample					
SSI	17	46.01	1.13	1.07	.296
Non-SSI	11	45.49	1.30		
Subsample					
SSI	13	46.31	0.96	3.20	.006
Non-SSI	7	45.04	0.78		

Table 6B.8

State Means on Four Individual Skill Areas and I_{RC} for All States Participating in State NAEP, Grade 8, 1990

	X_F	X_P	X_R	X_C	I_{RC}
SSI States (*N* = 20)					
Arkansas	2.57	2.61	2.09	2.03	43.31
California	2.40	2.57	2.36	2.23	47.61
Colorado	2.44	2.60	2.37	2.30	47.61
Connecticut	2.39	2.57	2.34	2.24	47.81
Delaware	2.53	2.56	2.32	2.17	46.38
Florida	2.54	2.64	2.26	2.23	45.36
Georgia	2.58	2.70	2.41	2.39	47.18
Kentucky	2.67	2.67	2.29	2.30	45.70
Louisiana	2.58	2.64	2.23	2.21	45.32
Michigan	2.52	2.58	2.25	2.13	45.48
Montana	2.41	2.49	2.30	2.19	47.72
Nebraska	2.53	2.58	2.24	2.11	45.51
New Jersey	2.63	2.69	2.38	2.35	46.60
New Mexico	2.53	2.68	2.33	2.17	45.56
New York	2.49	2.59	2.24	2.15	45.76
North Carolina	2.53	2.62	2.27	2.29	46.31
Ohio	2.51	2.64	2.24	2.16	45.32
Rhode Island	2.50	2.62	2.20	2.15	45.03
Texas	2.54	2.65	2.32	2.26	46.45
Virginia	2.57	2.73	2.29	2.30	45.76
Non-SSI States (*N* = 17)					
Alabama	2.56	2.68	2.31	2.29	46.05
Arizona	2.48	2.56	2.28	2.17	46.38
Hawaii	2.53	2.64	2.23	2.05	44.44
Idaho	2.54	2.63	2.23	2.19	45.36
Illinois	2.50	2.64	2.36	2.16	46.48
Indiana	2.54	2.64	2.13	2.13	44.09
Iowa†	2.46	2.59	2.19	2.07	45.41
Maryland	2.41	2.56	2.42	2.36	48.83
Minnesota	2.38	2.58	2.18	1.98	44.52
New Hampshire	2.43	2.57	2.29	2.20	46.56
North Dakota	2.40	2.64	2.19	2.07	45.55
Oklahoma	2.58	2.65	2.25	2.23	45.35
Oregon	2.39	2.49	2.35	2.12	47.21
Pennsylvania	2.59	2.72	2.34	2.24	45.17
West Virginia	2.55	2.66	2.26	2.16	45.07
Wisconsin	2.43	2.60	2.21	1.99	44.84
Wyoming	2.34	2.58	2.11	2.12	45.25

†Did not meet the NCES participation rate guidelines.

Appendix C

Table 6C.1

Comparison of SSI and Non-SSI states on $I_{MD(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines.

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	23.19	1.19	2.47	.018
Non-SSI	18	22.33	1.01		
Subsample					
SSI	16	23.29	1.02	3.00	.006
Non-SSI	14	22.19	1.00		
2-point trend sample					
Total sample					
SSI	20	23.09	1.20	2.19	.036
Non-SSI	15	22.25	1.07		
Subsample					
SSI	14	23.33	1.08	2.75	.012
Non-SSI	11	22.17	1.02		
3-point trend sample					
Total sample					
SSI	17	23.04	1.30	1.83	.081
Non-SSI	11	22.16	1.21		
Subsample					
SSI	13	23.34	1.12	2.45	.031
Non-SSI	7	21.99	1.20		
Grade 4					
1996 sample					
Total sample					
SSI	23	23.83	1.03	2.43	.020
Non-SSI	20	22.99	1.22		
Subsample					
SSI	16	24.02	0.89	2.97	.006
Non-SSI	16	22.90	1.21		
2-point trend sample					
Total sample					
SSI	21	23.80	1.03	2.60	.015
Non-SSI	16	22.82	1.21		
Subsample					
SSI	13	24.02	0.91	2.67	.013
Non-SSI	14	22.89	1.28		

Table 6C.2

State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 8

	I_{MD}	Work in small groups	Write about solution	Talk to class	Write reports/ do projects	Discuss with others	Discuss real-life situations	Time in group work	Assess by written responses	Assess by projects
SSI States, *N* = 22										
Arkansas†	20.49	2.38	1.61	2.48	1.20	3.05	2.76	2.78	2.27	1.97
California	25.41	3.15	2.56	2.70	1.65	3.29	2.95	3.41	3.09	2.58
Colorado	23.13	2.97	1.97	2.64	1.43	3.20	2.87	3.21	2.62	2.28
Connecticut	23.83	2.76	2.32	2.99	1.51	3.19	3.04	2.87	3.05	2.31
Delaware	24.33	2.90	2.24	3.06	1.58	3.28	3.17	3.09	2.80	2.43
Florida	23.27	2.83	2.00	2.84	1.40	3.27	3.08	3.09	2.44	2.29
Georgia	23.96	2.81	2.19	3.01	1.43	3.31	3.17	3.03	2.79	2.22
Kentucky	24.85	2.78	2.66	2.78	1.91	3.17	2.94	3.07	3.22	2.40
Louisiana	22.17	2.73	1.84	2.81	1.32	3.19	2.82	2.98	2.37	2.19
Maine	23.39	2.88	2.17	2.74	1.48	3.24	2.92	3.01	2.77	2.22
Massachusetts	23.23	2.77	2.11	2.82	2.52	3.12	2.89	2.98	2.78	2.26
Michigan†	23.89	3.01	2.21	2.67	1.40	3.27	3.15	3.20	2.80	2.16
Montana†	23.75	2.94	2.11	2.50	1.40	3.49	3.24	3.18	2.67	2.17
Nebraska	22.67	2.75	1.92	2.62	1.32	3.29	3.01	2.89	2.57	2.25
New Mexico	22.50	2.88	1.91	2.51	1.39	3.28	3.03	3.09	2.37	2.17
New York†	21.25	2.52	1.83	2.54	1.42	3.03	2.80	2.54	2.47	2.11
North Carolina	23.63	2.93	2.12	2.91	1.36	3.25	3.00	3.12	2.74	2.24
Rhode Island	21.83	2.52	1.75	2.69	1.52	2.97	2.85	2.74	2.47	2.29
South Carolina†	23.49	2.76	2.17	2.85	1.50	3.18	3.11	2.91	2.76	2.24
Texas	22.29	2.72	1.84	2.59	1.35	3.16	3.06	2.88	2.41	2.14
Vermont†	24.62	2.84	2.53	2.81	1.89	3.25	2.77	3.10	2.89	2.50
Virginia	22.16	2.73	1.87	2.69	1.34	3.19	2.94	2.84	2.52	2.08

†Did not meet the NCES participation rate guidelines

Table 6C.2, continued

State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 8

	I_{MD}	Work in small groups	Write about solution	Talk to class	Write reports/ do projects	Discuss with others	Discuss real-life situations	Time in group work	Assess by written responses	Assess by projects
Non-SSI States, *N* = 18										
Alabama	21.18	2.48	1.66	2.84	1.27	3.09	2.76	2.75	2.22	2.12
Alaska†	22.64	2.83	1.86	2.47	1.34	3.39	2.86	3.20	2.63	2.80
Arizona	23.80	3.03	2.25	2.68	1.42	3.23	2.97	3.23	2.86	2.32
Hawaii	23.22	2.91	2.08	2.69	1.44	3.19	2.69	3.07	2.79	2.35
Indiana	21.07	2.70	1.76	2.44	1.29	2.96	2.73	2.79	2.34	2.07
Iowa†	21.68	2.69	1.94	2.52	1.27	3.01	2.89	2.82	2.42	2.11
Maryland†	24.23	2.87	2.34	3.02	1.44	3.24	3.10	2.99	3.01	2.17
Minnesota	21.91	2.88	1.68	2.43	1.28	3.11	3.00	3.11	2.39	2.01
Mississippi	23.30	2.66	2.09	2.76	1.51	3.28	3.15	2.98	2.72	2.28
Missouri	22.26	2.76	1.80	2.77	1.35	3.16	2.99	2.96	2.35	2.21
North Dakota	20.32	2.41	1.57	2.30	1.30	2.99	2.90	2.67	2.09	2.08
Oregon	23.43	3.05	2.07	2.63	1.45	3.24	2.86	3.24	2.70	2.20
Tennessee	21.91	2.60	1.83	2.82	1.32	3.13	2.86	2.85	2.37	2.23
Utah	22.49	2.91	1.98	2.46	1.39	3.26	2.89	3.13	2.42	2.13
Washington	22.13	2.95	1.89	2.27	1.36	3.24	2.68	3.23	2.40	2.13
West Virginia	22.55	2.55	1.85	2.85	1.22	3.13	2.83	2.73	2.50	1.99
Wisconsin†	22.73	2.78	2.02	2.57	1.46	3.01	3.00	2.93	2.65	2.35
Wyoming	22.05	2.80	1.84	2.93	1.34	3.22	2.89	3.07	2.42	2.10

†Did not meet the NCES participation rate guidelines

Table 6C.3

State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 4

	I_{MD}	Work in small groups	Write about solution	Talk to class	Write reports/ do projects	Discuss with others	Discuss real-life situations	Time in group work	Assess by written responses	Assess by projects
SSI States, *N* = 23										
Arkansas†	21.06	2.64	1.76	2.64	1.28	2.69	2.67	2.88	2.26	2.31
California	24.94	3.19	2.50	2.84	1.60	3.05	2.84	3.43	2.92	2.54
Colorado	24.19	3.05	2.24	2.82	1.38	3.07	3.01	3.39	2.77	2.55
Connecticut	24.87	3.02	2.55	3.05	1.37	3.14	3.07	3.33	3.03	2.31
Delaware	23.87	2.91	2.29	2.99	1.38	3.10	2.99	3.12	2.78	2.25
Florida	23.81	2.92	1.96	3.02	1.44	3.07	3.07	3.25	2.61	2.46
Georgia	24.38	2.99	2.07	3.12	1.41	3.21	3.11	3.25	2.79	2.44
Kentucky	25.39	3.01	2.59	3.02	1.64	3.10	3.06	3.21	3.25	2.61
Louisiana	22.72	2.83	1.93	3.00	1.26	3.01	3.05	3.00	2.49	2.27
Maine	25.05	3.07	2.59	3.11	1.38	3.22	2.89	3.30	3.16	2.29
Massachusetts	24.14	3.07	2.28	2.80	1.43	3.13	2.90	3.29	2.83	2.49
Michigan	23.58	2.94	2.11	2.91	1.38	3.16	2.96	3.16	2.68	2.27
Montana†	23.18	2.92	2.04	2.72	1.43	3.05	2.82	3.14	2.70	2.44
Nebraska	23.06	2.94	1.93	2.78	1.39	3.01	3.12	3.07	2.48	2.39
New Jersey†	24.22	3.07	2.16	3.11	1.31	3.19	3.05	3.20	2.64	2.50
New Mexico	23.43	2.93	2.03	2.77	1.39	2.91	2.87	3.23	2.74	2.55
New York†	22.93	2.84	2.06	2.84	1.38	3.07	2.97	3.08	2.56	2.19
North Carolina	24.92	3.03	2.41	2.95	1.48	3.14	3.08	3.25	3.09	2.47
Rhode Island	22.46	2.86	2.04	2.69	1.29	2.87	2.83	3.18	2.58	2.19
South Carolina†	23.66	2.96	2.09	2.95	1.40	3.02	3.09	3.12	2.66	2.45
Texas	23.98	2.94	2.02	2.99	1.33	3.15	3.22	3.17	2.66	2.38
Vermont†	25.18	3.01	2.76	2.89	1.63	3.01	2.79	3.39	3.12	2.54
Virginia	23.07	2.88	1.93	2.94	1.31	3.08	2.97	3.17	2.56	2.27

†Did not meet the NCES participation rate guidelines

Table 6C.3, continued

State Means on $I_{MD(96)}$ and Each Mathematical Discourse Item, Grade 4

	I_{MD}	Work in small groups	Write about solution	Talk to class	Write reports/ do projects	Discuss with others	Discuss real-life situations	Time in group work	Assess by written responses	Assess by projects
Non-SSI States, *N* = 20										
Alabama	23.18	2.76	1.94	3.11	1.37	2.98	3.00	3.08	2.56	2.35
Alaska†	23.40	3.10	1.89	2.76	1.38	3.05	2.88	3.36	2.56	2.41
Arizona	23.73	2.96	2.16	2.76	1.46	3.05	2.94	3.27	2.74	2.41
Hawaii	22.44	2.81	1.99	2.73	1.35	2.97	2.75	2.98	2.60	2.22
Indiana	21.65	2.75	1.72	2.69	1.23	2.94	2.91	2.89	2.36	2.15
Iowa†	21.97	2.97	2.92	2.53	1.26	2.92	2.81	3.08	2.34	2.33
Maryland	26.03	3.09	2.70	3.34	1.46	3.29	3.19	3.25	3.23	2.48
Minnesota	22.82	3.05	1.88	2.65	1.34	2.95	3.01	3.26	2.33	2.27
Mississippi	24.57	2.96	2.30	3.13	1.48	3.03	2.98	3.10	2.90	2.68
Missouri	21.70	2.77	1.81	2.71	1.24	2.92	2.90	2.99	2.27	2.21
Nevada†	25.22	3.14	2.37	3.07	1.47	3.18	3.14	3.35	2.82	2.65
North Dakota	21.02	2.71	1.75	2.48	1.23	2.71	2.74	2.84	2.31	2.24
Oregon	23.61	3.02	2.21	2.76	1.37	3.00	2.79	3.24	2.88	2.35
Pennsylvania	22.70	2.99	1.91	2.94	1.27	3.06	2.99	3.11	2.41	2.11
Tennessee	22.21	2.72	1.76	2.87	1.31	2.89	2.91	2.95	2.52	2.30
Utah	23.27	3.01	1.89	2.79	1.33	3.11	3.01	3.25	2.48	2.38
Washington	22.36	2.98	1.89	2.54	1.35	2.91	2.77	3.23	2.46	2.28
West Virginia	22.97	2.86	2.09	2.82	1.30	2.99	2.89	3.11	2.74	2.23
Wisconsin	22.27	2.80	1.89	2.72	1.40	2.97	2.89	3.05	2.39	2.19
Wyoming	22.61	2.94	1.88	2.68	1.26	2.97	2.96	3.09	2.42	2.41

†Did not meet the NCES participation rate guidelines

Table 6C.4

Comparison of SSI and Non-SSI States on $I_{MD4(92)}$ and $I_{MD(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
$I_{MD4(92}$ – 4 matching items					
Grade 8					
1992 sample					
Total sample					
SSI	22	9.14	0.37	1.00	.322
Non-SSI	19	9.04	0.29		
Subsample					
SSI	18	9.14	0.39	0.95	.351
Non-SSI	18	9.03	0.30		
2-point trend sample					
Total sample					
SSI	20	9.16	0.35	1.07	.293
Non-SSI	15	9.04	0.29		
Subsample					
SSI	14	9.19	0.31	2.00	.058
Non-SSI	11	8.95	0.28		
3-point trend sample					
Total sample					
SSI	17	9.17	0.35	1.14	.268
Non-SSI	11	9.03	0.32		
Subsample					
SSI	13	9.24	0.27	2.66	.021
Non-SSI	7	8.88	0.29		
Grade 4					
1992 sample					
Total sample					
SSI	22	9.07	0.41	1.60	.118
Non-SSI	19	8.86	0.44		
Subsample					
SSI	17	9.07	0.45	1.44	.160
Non-SSI	19	8.86	0.44		
2-point trend sample					
Total sample					
SSI	21	9.09	0.41	1.46	.155
Non-SSI	16	8.87	0.48		
Subsample					
SSI	13	9.17	0.38	1.76	.092
Non-SSI	14	8.86	0.52		

Table 6C.4, continued

Comparison of SSI and Non-SSI States on $I_{MD4(92)}$ and $I_{MD(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
$I_{MD(92)}$ – 7 similar items					
Grade 8					
1992 sample					
Total sample					
SSI	22	15.57	0.69	1.53	.134
Non-SSI	19	15.25	0.66		
Subsample					
SSI	18	15.56	0.72	1.39	.173
Non-SSI	18	15.23	0.67		
2-point trend sample					
Total sample					
SSI	20	15.60	0.65	1.57	.127
Non-SSI	15	15.26	0.62		
Subsample					
SSI	14	15.67	0.57	2.78	.011
Non-SSI	11	15.06	0.53		
3-point trend sample					
Total sample					
SSI	17	15.62	0.64	1.24	.230
Non-SSI	11	15.29	0.69		
Subsample					
SSI	13	15.76	0.49	2.88	.016
Non-SSI	7	15.00	0.60		
Grade 4					
1992 sample					
Total sample					
SSI	22	15.74	0.81	1.79	.081
Non-SSI	19	15.30	0.76		
Subsample					
SSI	17	15.77	0.87	1.70	.099
Non-SSI	19	15.30	0.76		
2-point trend sample					
Total sample					
SSI	21	15.78	0.81	1.66	.107
Non-SSI	16	15.33	0.81		
Subsample					
SSI	13	15.99	0.76	2.10	.046
Non-SSI	14	15.32	0.87		

Table 6C.5

State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 8

	I_{MD}	Work in small groups	Write about solution	Write reports/ do projects	Discuss with others	Discuss real-life situations	Assess by written responses	Assess by projects
SSI States, *N* = 22								
Arkansas	14.04	2.31	1.47	1.19	3.01	2.72	1.89	1.48
California	16.15	2.71	1.97	1.38	3.28	2.70	2.22	1.87
Colorado	16.45	2.68	1.97	1.33	3.34	2.92	2.30	1.90
Connecticut	15.42	2.37	1.85	1.29	3.05	2.88	2.21	1.81
Delaware	15.54	2.48	1.74	1.31	3.16	2.93	2.17	1.84
Florida	15.61	2.55	1.80	1.18	3.26	2.95	2.18	1.73
Georgia	15.96	2.55	1.87	1.23	3.31	3.11	2.18	1.71
Kentucky	16.21	2.56	1.96	1.33	3.15	2.86	2.38	1.99
Louisiana	15.57	2.49	1.84	1.23	3.20	2.89	2.22	1.70
Maine†	16.03	2.78	1.79	1.23	3.27	3.04	2.17	1.79
Massachusetts	14.52	2.30	1.70	1.19	2.97	2.74	1.97	1.66
Michigan	16.28	2.68	2.02	1.26	3.29	3.06	2.28	1.76
Nebraska†	15.42	2.55	1.82	1.22	3.24	2.87	2.12	1.62
New Jersey†	16.23	2.53	2.08	1.20	3.20	3.00	2.54	1.65
New Mexico	15.54	2.68	1.60	1.23	3.33	2.87	2.07	1.73
New York†	14.86	2.21	1.77	1.22	2.92	2.79	2.33	1.61
North Carolina	15.28	2.48	1.79	1.25	3.12	2.75	2.18	1.72
Ohio	14.37	2.32	1.69	1.12	2.90	2.78	1.96	1.61
Rhode Island	15.25	2.25	2.00	1.23	3.11	2.84	2.13	1.69
South Carolina	15.94	2.44	1.83	1.29	3.20	2.98	2.34	1.83
Texas	16.71	2.61	1.95	1.27	3.38	3.24	2.44	1.85
Virginia	15.22	2.49	1.72	1.20	3.20	2.74	2.07	1.79

†Did not meet the NCES participation rate guidelines

Table 6C.5, continued

State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 8

	I_{MD}	Work in small groups	Write about solution	Write reports/ do projects	Discuss with others	Discuss real-life situations	Assess by written responses	Assess by projects
Non-SSI States, *N* = 19								
Alabama†	15.52	2.35	1.74	1.25	3.19	3.04	2.14	1.89
Arizona	16.00	2.66	2.03	1.21	3.28	2.88	2.29	1.69
Hawaii	14.88	2.38	1.79	1.19	3.02	2.62	2.11	1.77
Idaho	15.57	2.77	1.75	1.16	3.37	2.94	1.95	1.61
Indiana	14.32	2.27	1.59	1.19	3.02	2.79	1.81	1.67
Iowa	15.15	2.44	1.76	1.18	3.13	2.98	2.04	1.63
Maryland	16.41	2.61	2.05	1.25	3.25	2.83	2.49	1.89
Minnesota	15.49	2.54	1.82	1.22	3.14	2.89	2.16	1.66
Mississippi	15.66	2.35	1.95	1.28	3.17	2.97	2.28	1.70
Missouri	14.68	2.27	1.62	1.14	3.19	2.93	1.98	1.59
New Hampshire	16.30	2.69	1.93	1.43	3.21	2.83	2.13	2.12
North Dakota	14.67	2.23	1.59	1.19	3.20	2.98	1.90	1.61
Oklahoma	14.39	2.12	1.58	1.17	3.20	2.85	1.88	1.59
Pennsylvania	14.59	2.30	1.75	1.17	2.99	2.77	2.12	1.54
Tennessee	15.01	2.21	1.76	1.23	3.04	2.88	2.10	1.77
Utah	15.30	2.54	1.84	1.19	3.37	2.77	2.02	1.56
West Virginia	14.45	2.33	1.57	1.14	3.20	2.76	1.93	1.53
Wisconsin	16.15	2.60	1.92	1.29	3.25	3.01	2.35	1.73
Wyoming	15.19	2.69	1.58	1.14	3.30	2.84	1.97	1.70

†Did not meet the NCES participation rate guidelines

Table 6C.6
State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 4

	I_{MD}	Work in small groups	Write about solution	Write reports/ do projects	Discuss with others	Discuss real-life situations	Assess by written responses	Assess by projects
SSI States, *N* = 22								
Arkansas	13.79	2.26	1.56	1.09	2.61	2.72	2.02	1.56
California	16.55	2.87	2.02	1.39	3.07	2.94	2.30	1.99
Colorado	16.25	2.91	1.92	1.32	3.03	2.91	2.19	2.02
Connecticut	15.99	2.73	1.98	1.27	3.03	2.94	2.21	1.83
Delaware†	14.78	2.54	1.67	1.20	2.91	2.87	1.91	1.63
Florida	16.31	2.81	1.90	1.32	3.09	3.11	2.22	1.86
Georgia	15.80	2.72	1.92	1.23	2.98	2.97	2.11	1.84
Kentucky	17.00	2.84	2.09	1.37	3.02	3.04	2.48	2.16
Louisiana	16.50	2.69	2.04	1.26	3.11	3.15	2.41	1.82
Maine†	16.17	2.88	2.00	1.23	3.06	2.89	2.31	1.77
Massachusetts	15.41	2.64	1.89	1.22	2.99	2.88	2.06	1.76
Michigan	15.70	2.74	1.90	1.29	3.04	2.98	1.99	1.75
Nebraska†	15.22	2.76	1.80	1.21	3.00	2.91	1.93	1.63
New Jersey†	16.20	2.71	2.08	1.30	3.01	3.08	2.32	1.74
New Mexico	15.76	2.66	1.88	1.28	2.94	2.81	2.32	1.87
New York†	15.89	2.61	1.96	1.27	2.96	2.96	2.35	1.77
North Carolina	15.67	2.73	1.77	1.23	3.00	2.90	2.20	1.86
Ohio	14.92	2.57	1.74	1.20	2.88	2.87	2.01	1.64
Rhode Island	14.10	2.50	1.63	1.19	2.72	2.71	1.86	1.46
South Carolina	15.81	2.75	1.80	1.18	3.09	3.11	2.20	1.75
Texas	16.95	2.80	2.08	1.27	3.06	3.28	2.57	1.89
Virginia	15.53	2.71	1.82	1.18	3.05	2.96	2.16	1.67

†Did not meet the NCES participation rate guidelines

Table 6C.6, continued

State Means on $I_{MD(92)}$ and Each Mathematical Discourse Item, Grade 4

	I_{MD}	Work in small groups	Write about solution	Write reports/ do projects	Discuss with others	Discuss real-life situations	Assess by written responses	Assess by projects
Non-SSI States, *N* = 19								
Alabama	15.95	2.69	2.03	1.27	3.01	2.97	2.30	1.76
Arizona	15.43	2.66	1.80	1.26	2.97	2.83	2.14	1.73
Hawaii	15.13	2.58	1.87	1.27	2.78	2.72	2.20	1.73
Idaho	15.17	2.73	1.68	1.18	3.04	2.85	2.07	1.63
Indiana	14.13	2.46	1.53	1.17	2.65	2.78	1.87	1.65
Iowa	15.42	2.70	1.75	1.20	2.98	2.97	2.03	1.78
Maryland	17.45	3.00	2.44	1.28	3.16	3.13	2.65	1.83
Minnesota	15.38	2.70	1.82	1.24	2.93	2.95	2.06	1.67
Mississippi	15.24	2.64	1.79	1.24	2.93	2.87	2.22	1.75
Missouri	15.11	2.52	1.81	1.24	2.82	2.97	2.12	1.65
New Hampshire	15.53	2.79	1.79	1.27	3.07	2.83	2.01	1.78
North Dakota	13.90	2.39	1.53	1.20	2.60	2.69	1.96	1.54
Oklahoma	14.69	2.47	1.64	1.19	2.91	2.92	1.97	1.61
Pennsylvania	15.38	2.63	1.80	1.21	3.00	2.97	2.17	1.60
Tennessee	15.01	2.60	1.72	1.23	2.85	2.84	2.11	1.70
Utah	15.42	2.77	1.79	1.26	2.93	2.97	2.07	1.65
West Virginia	14.57	2.60	1.62	1.20	2.85	2.76	1.96	1.63
Wisconsin	16.07	2.79	2.04	1.33	2.99	3.03	2.18	1.75
Wyoming	15.75	2.83	1.88	1.26	3.09	2.98	1.99	1.74

†Did not meet the NCES participation rate guidelines

Table 6C.7

State Means on I_{MD4} for All States Participating in State NAEP in 1992 and 1996

	Grade 8		Grade 4	
	1996	1992	1996	1992
SSI States				
Arkansas	8.62†	8.39	8.39†	7.97
California	10.45	9.32	9.97	9.40
Colorado	9.47	9.56	9.66	9.16
Connecticut	10.05	9.07	10.15	9.23
Delaware	10.27	9.14	9.77	8.67†
Florida	9.76	9.19	9.55	9.42
Georgia	10.10	9.52	9.77	9.10
Kentucky	10.68	9.30	10.36	9.53
Louisiana	9.18	9.16	9.24	9.57
Maine	9.81	9.33†	10.10	9.19†
Massachusetts	9.63	8.60	9.71	8.98
Michigan	10.03†	9.62	9.60†	9.24
Montana	10.24†	-	9.33†	-
Nebraska	9.54	9.15†	9.44	8.92†
New Jersey	-	9.48†	9.71†	9.45†
New Mexico	9.61	9.03	9.18	8.92
New York	9.07†	8.70†	9.45†	9.16†
North Carolina	9.73	8.91	10.13	8.91
Ohio	-	8.49	-	8.69
Rhode Island	9.09	9.17	9.00	8.26
South Carolina	9.96†	9.29	9.58†	9.18
Texas	9.42	9.84	9.72	9.68
Vermont	10.44†	-	10.20†	-
Virginia	9.34	8.87	9.29	9.00
Total sample *N*	22	22	23	22
Subsample *N*	16	18	16	17

†Did not meet the NCES participation rate guidelines.

Table 6C.7, continued
State Means on I_{MD4} for All States Participating in State NAEP in 1992 and 1996

	Grade 8		Grade 4	
	1996	1992	1996	1992
Non-SSI states				
Alabama	8.79	9.22†	9.29	9.27
Alaska	9.45†	-	9.21†	-
Arizona	9.87	9.40	9.61	8.87
Hawaii	9.41	8.63	9.05	8.64
Idaho	-	9.21	8.80	8.74
Indiana	8.74	8.59	-	8.13
Iowa	9.11†	9.04	8.78†	8.92
Maryland	10.13†	9.38	10.64	10.01
Minnesota	9.07	9.07	9.21	8.93
Mississippi	10.03	9.38	9.77	8.83
Missouri	9.30	8.88	8.85	8.84
Nevada	-	-	10.15†	-
New Hampshire	-	9.40	-	8.97
North Dakota	8.75	8.95	8.44	8.02
Oklahoma	-	8.81	-	8.64
Oregon	9.62	-	9.34	-
Pennsylvania	-	8.69	9.21†	8.98
Tennessee	9.13	8.91	8.89	8.63
Utah	9.52	9.16	9.34	8.94
Washington	9.17	-	8.91	-
West Virginia	9.02	8.68	9.26	8.38
Wisconsin	9.49†	9.47	9.15	9.38
Wyoming	9.30	8.86	9.06	9.20
Total sample *N*	18	19	20	19
Subsample *N*	16	18	16	19

†Did not meet the NCES participation rate guidelines.

Table 6C.8

Comparison of SSI and Non-SSI States on $I_{MD4(92)}$ and $I_{MD(92)}$ for All States in Each Sample and for the Subsampls of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
1990 sample					
Total sample					
SSI	20	4.18	0.33	-0.16	.874
Non-SSI	17	4.19	0.31		
Subsample					
SSI	20	4.18	0.33	-0.18	.857
Non-SSI	16	4.20	0.32		
2- and 3-point trend samples					
Total sample					
SSI	17	4.17	0.35	0.16	.876
Non-SSI	11	4.15	0.28		
Subsample					
SSI	13	4.24	0.35	0.76	.464
Non-SSI	7	4.13	0.32		

Table 6C.9
State Means on $I_{MD(90)}$ and Each Mathematical Discourse Item, Grade 8

	I_{MD}	Work in small groups	Write reports/ do projects
SSI States (*N* = 20)			
Arkansas	3.80	2.33	1.47
California	4.57	2.98	1.60
Colorado	4.87	3.25	1.62
Connecticut	4.39	2.76	1.63
Delaware	3.86	2.46	1.41
Florida	4.17	2.62	1.55
Georgia	4.55	2.91	1.65
Kentucky	4.20	2.53	1.67
Louisiana	4.01	2.58	1.43
Michigan	4.00	2.56	1.44
Montana	4.47	2.90	1.57
Nebraska	4.26	2.64	1.61
New Jersey	4.17	2.55	1.62
New Mexico	4.35	2.85	1.50
New York	3.68	2.16	1.52
North Carolina	4.28	2.62	1.67
Ohio	3.97	2.40	1.56
Rhode Island	3.44	2.06	1.39
Texas	4.19	2.55	1.64
Virginia	4.32	2.68	1.64
Non-SSI states (*N* = 17)			
Alabama	4.00	2.37	1.63
Arizona	4.52	3.05	1.48
Hawaii	3.79	2.30	1.50
Idaho	4.36	2.87	1.49
Illinois	4.23	2.65	1.58
Indiana	3.93	2.46	1.48
Iowa†	4.15	2.66	1.49
Maryland	4.50	2.91	1.59
Minnesota	4.07	2.63	1.44
New Hampshire	4.62	2.83	1.79
North Dakota	4.00	2.52	1.48
Oklahoma	4.05	2.55	1.50
Oregon	4.68	3.21	1.49
Pennsylvania	3.67	2.27	1.40
West Virginia	3.95	2.48	1.47
Wisconsin	4.16	2.60	1.56
Wyoming	4.65	3.24	1.41

†Did not meet the NAEP participation rate guidelines.

Appendix D

Table 6D.1
Comparison of SSI and Non-SSI States on $I_{S(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	2.72	0.21	2.44	.020
Non-SSI	18	2.59	0.13		
Subsample					
SSI	16	2.72	0.22	2.44	.022
Non-SSI	14	2.56	0.13		
2-point trend sample					
Total sample					
SSI	20	2.70	0.21	1.85	.074
Non-SSI	15	2.59	0.15		
Subsample					
SSI	14	2.71	0.24	1.77	.091
Non-SSI	11	2.57	0.14		
3-point trend sample					
Total sample					
SSI	17	2.67	0.20	1.49	.149
Non-SSI	11	2.57	0.13		
Subsample					
SSI	13	2.68	0.23	2.02	.059
Non-SSI	7	2.54	0.10		
Grade 4					
1996 sample					
Total sample					
SSI	23	1.98	0.21	1.76	.086
Non-SSI	20	1.88	0.15		
Subsample					
SSI	16	1.99	0.19	1.90	.068
Non-SSI	16	1.88	0.16		
2-point trend sample					
Total sample					
SSI	21	1.95	0.19	1.37	.179
Non-SSI	16	1.87	0.17		
Subsample					
SSI	13	1.95	0.18	1.20	.242
Non-SSI	14	1.87	0.17		

Table 6D.2
State Means on I_S for All States Participating in State NAEP, 1996

	Grade 8	Grade 4
SSI States		
Arkansas	2.57†	1.79†
California	2.70	1.77
Colorado	2.72	2.15
Connecticut	2.99	1.98
Delaware	3.01	2.22
Florida	2.59	1.76
Georgia	2.71	1.94
Kentucky	2.83	2.08
Louisiana	2.31	1.88
Maine	2.87	2.33
Massachusetts	3.03	2.19
Michigan	2.54†	1.85†
Montana	2.78†	2.04†
Nebraska	2.75	1.97
New Jersey	-	1.83†
New Mexico	2.45	1.86
New York	2.61†	1.66†
North Carolina	2.60	1.95
Rhode Island	2.85	2.20
South Carolina	2.79†	1.98†
Texas	2.31	1.63
Vermont	3.03†	2.47†
Virginia	2.82	2.02
Total sample *N*	22	23
Subsample *N*	16	16

†Did not meet the NCES participation rate guidelines.

Table 6D.2, continued
State Means on I_S for All States Participating in State NAEP, 1996

	Grade 8	Grade 4
Non-SSI States		
Alabama	2.42	1.71
Alaska	2.58†	1.97†
Arizona	2.51	1.85
Hawaii	2.41	1.89
Indiana	2.48	1.55
Iowa	2.70†	1.99†
Maryland	2.83†	2.15
Minnesota	2.67	1.88
Mississippi	2.81	2.10
Missouri	2.67	1.76
Nevada	-	1.93†
North Dakota	2.47	2.05
Oregon	2.58	1.94
Pennsylvania	-	1.75†
Tennessee	2.36	1.72
Utah	2.72	2.08
Washington	2.54	1.84
West Virginia	2.55	1.72
Wisconsin	2.62†	1.87
Wyoming	2.66	1.90
Total sample *N*	18	20
Subsample *N*	14	16

†Did not meet the NCES participation rate guidelines.

Appendix E

Table 6E.1

Comparison of SSI and Non-SSI States on $I_{PD(96)}$ for All States in Each Samples and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	3.46	0.34	1.72	.093
Non-SSI	18	3.30	0.24		
Subsample					
SSI	16	3.47	0.39	1.14	.267
Non-SSI	14	3.34	0.21		
2-point trend sample					
Total sample					
SSI	20	3.44	0.36	1.04	.305
Non-SSI	15	3.34	0.22		
Subsample					
SSI	14	3.51	0.40	1.23	.232
Non-SSI	11	3.35	0.23		
3-point trend sample					
Total sample					
SSI	17	3.42	0.36	1.16	.258
Non-SSI	11	3.30	0.22		
Subsample					
SSI	13	3.47	0.39	1.33	.200
Non-SSI	7	3.29	0.23		
Grade 4					
1996 sample					
Total sample					
SSI	23	2.85	0.26	2.22	.032
Non-SSI	20	2.69	0.23		
Subsample					
SSI	16	2.85	0.27	2.12	.043
Non-SSI	16	2.67	0.20		
2-point trend sample					
Total sample					
SSI	21	2.83	0.26	2.57	.015
Non-SSI	16	2.64	0.21		
Subsample					
SSI	13	2.89	0.28	2.38	0.26
Non-SSI	14	2.66	0.21		

Table 6E.2
State Means on I_{PD} for All States Participating in State NAEP

	Grade 8			Grade 4	
	1996	1992	1990	1996	1992
SSI States					
Arkansas	3.51†	3.43	3.35	3.17†	2.65
California	4.04	3.68	3.26	3.32	2.82
Colorado	3.27	3.19	3.05	2.61	2.49
Connecticut	3.44	3.45	3.22	2.65	2.70
Delaware	3.76	3.04	3.24	2.62	2.63†
Florida	3.78	3.45	3.29	2.79	2.92
Georgia	3.39	3.06	3.05	2.84	2.26
Kentucky	4.00	3.07	2.37	3.10	2.64
Louisiana	3.14	3.20	3.08	2.96	2.69
Maine	3.25	3.40†	-	2.82	2.70†
Massachusetts	3.97	2.86	-	3.15	2.64
Michigan	3.31†	3.19	2.69	2.69†	2.90
Montana	3.58†	-	3.17	2.86†	-
Nebraska	3.06	3.35†	2.98	2.60	2.28†
New Jersey	-	3.43†	2.84	2.62†	2.50†
New Mexico	2.73	3.11	2.38	2.65	2.17
New York	3.17†	2.90†	2.75	2.54†	2.27†
North Carolina	3.14	3.28	3.44	2.66	2.61
Ohio	-	3.35	2.69	-	2.51
Rhode Island	3.14	3.64	2.60	2.63	2.43
South Carolina	3.47†	3.40	-	2.91†	2.75
Texas	3.81	3.56	3.12	3.42	2.84
Vermont	3.63†	-	-	3.24†	-
Virginia	3.52	3.19	2.96	2.79	2.44
Total sample *N*	22	22	20	23	22
Subsample *N*	17	18	20	16	17

†Did not meet the NCES participation rate guidelines.

Table 6E.2, continued
State Means on I_{PD} for All States Participating in State NAEP

	Grade 8			Grade 4	
	1996	1992	1990	1996	1992
Non-SSI states					
Alabama	3.37	3.51†	2.83	2.79	2.98
Alaska	2.79†	-	-	2.81†	-
Arizona	3.25	3.14	2.58	2.62	2.42
Hawaii	3.54	3.31	2.63	2.74	2.85
Idaho	-	3.08	2.90	-	2.50
Illinois	-	-	2.67	-	-
Indiana	2.95	3.02	2.40	2.23	2.39
Iowa	3.13†	3.13	2.79†	2.47†	2.39
Maryland	3.61†	3.56	3.43	2.73	2.77
Minnesota	3.56	3.55	3.03	2.72	2.63
Mississippi	3.72	3.28	-	3.11	3.00
Missouri	3.58	3.17	-	2.72	2.33
Nevada	-	-	-	3.23†	-
New Hampshire	-	3.68	3.93	-	2.88
North Dakota	3.27	3.13	2.56	2.67	2.37
Oklahoma	-	3.06	2.71	-	2.60
Oregon	3.12	-	3.39	2.64	-
Pennsylvania	-	3.00	2.75	2.47†	2.31
Tennessee	3.25	3.46	-	2.58	2.59
Utah	3.31	3.08	-	2.82	2.58
Washington	3.34	-	-	2.85	-
West Virginia	3.40	3.19	2.56	2.73	2.65
Wisconsin	3.15†	3.41	3.05	2.43	2.52
Wyoming	3.05	3.07	2.89	2.39	2.44
Total sample *N*	18	19	17	20	19
Subsample *N*	14	18	16	16	19

†Did not meet the NCES participation rate guidelines.

Table 6E.3

Comparison of SSI and Non-SSI States on $I_{PD(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
<u>Grade 8</u>					
1992 sample					
Total sample					
SSI	22	3.28	0.22	0.41	.683
Non-SSI	19	3.26	0.21		
Subsample					
SSI	18	3.29	0.22	0.62	.537
Non-SSI	18	3.24	0.21		
2-point trend sample					
Total sample					
SSI	20	3.27	0.23	0.05	.957
Non-SSI	15	3.27	0.19		
Subsample					
SSI	14	3.27	0.25	0.60	.551
Non-SSI	11	3.22	0.17		
3-point trend sample					
Total sample					
SSI	17	3.28	0.23	0.07	.943
Non-SSI	11	3.28	0.20		
Subsample					
SSI	13	3.30	0.23	1.06	.303
Non-SSI	7	3.20	0.18		
<u>Grade 4</u>					
1992 sample					
Total sample					
SSI	22	2.58	0.21	0.10	.924
Non-SSI	19	2.59	0.22		
Subsample					
SSI	17	2.62	0.21	0.35	.726
Non-SSI	19	2.59	0.22		
2-point trend sample					
Total sample					
SSI	21	2.59	0.22	0.14	.892
Non-SSI	16	2.58	0.22		
Subsample					
SSI	13	2.59	0.22	-0.25	.803
Non-SSI	14	2.61	0.22		

Table 6E.4

Comparison of SSI and Non-SSI States on $I_{PD(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1990 sample					
Total sample					
SSI	20	2.98	0.31	0.76	.456
Non-SSI	17	2.89	0.39		
Subsample					
SSI	20	2.98	0.31	0.68	.505
Non-SSI	16	2.89	0.40		
2- and 3-point trend samples					
Total sample					
SSI	17	2.99	0.33	1.63	.117
Non-SSI	11	2.80	0.30		
Subsample					
SSI	13	3.00	0.34		
Non-SSI	7	2.66	0.22	2.70	.015

Appendix F

Table 6F.1

Comparison of SSI and Non-SSI States on $I_{RT(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	5.16	0.31	1.82	.072
Non-SSI	18	5.00	0.27		
Subsample					
SSI	16	5.25	0.29	2.99	.006
Non-SSI	14	4.95	0.25		
2-point trend sample					
Total sample					
SSI	20	5.18	0.30	2.03	.051
Non-SSI	15	4.98	0.27		
Subsample					
SSI	14	5.26	0.25	2.70	.014
Non-SSI	11	4.98	0.28		
3-point trend sample					
Total sample					
SSI	17	5.23	0.28	2.24	.036
Non-SSI	11	4.98	0.30		
Subsample					
SSI	13	5.28	0.24	2.19	.054
Non-SSI	7	4.97	0.33		
Grade 4					
1996 sample					
Total sample					
SSI	23	4.85	0.26	1.23	.226
Non-SSI	20	4.74	0.29		
Subsample					
SSI	16	4.91	0.25	1.96	.059
Non-SSI	16	4.74	0.26		
2-point trend sample					
Total sample					
SSI	21	4.86	0.26	1.68	.102
Non-SSI	16	4.70	0.28		
Subsample					
SSI	13	4.94	0.23	2.08	.048
Non-SSI	14	4.74	0.27		

Table 6F.2
State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher Had Studied Each Reform-Related Topic, Grade 8

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
SSI States, *N* = 22								
Arkansas†	5.00	80.1	97.1	90.2	90.8	64.5	44.4	36.2
California	5.70	81.6	92.8	92.1	87.2	78.3	62.8	81.6
Colorado	5.11	75.3	92.5	89.1	80.3	63.2	59.7	54.0
Connecticut	5.53	88.1	97.4	93.5	92.3	79.7	57.2	49.1
Delaware	5.39	82.4	95.3	95.6	94.3	78.3	50.5	54.8
Florida	5.48	78.1	93.6	90.6	91.0	71.0	50.3	77.6
Georgia	5.41	85.3	96.5	94.8	86.6	75.2	57.8	53.5
Kentucky	5.38	86.2	98.9	96.6	88.0	71.7	55.4	44.6
Louisiana	4.95	81.8	92.1	90.3	81.0	70.7	46.1	40.5
Maine	4.65	73.2	91.9	91.5	77.0	68.1	56.6	14.6
Massachusetts	5.00	74.7	94.5	85.8	82.3	74.0	52.7	42.8
Michigan†	4.89	82.1	92.7	89.4	82.8	68.6	50.5	30.2
Montana†	5.39	77.8	96.6	90.5	84.7	82.9	64.9	52.5
Nebraska	5.62	79.3	96.1	89.5	79.6	70.6	68.3	78.8
New Mexico	4.99	66.1	89.5	88.9	76.1	64.2	53.5	63.0
New York†	4.78	64.6	93.0	86.5	78.0	65.8	45.2	47.2
North Carolina	5.33	81.7	96.1	95.8	94.0	71.7	49.0	53.8
Rhode Island	4.87	67.9	92.6	93.2	79.4	69.6	44.9	41.2
South Carolina†	4.89	76.3	93.0	91.6	83.4	66.9	43.6	39.8
Texas	5.34	81.6	96.7	93.0	92.3	67.2	51.8	62.2
Vermont†	4.69	75.2	94.9	88.4	71.1	78.6	48.1	16.2
Virginia	5.20	75.1	91.6	86.6	89.5	66.5	58.6	56.6

†Did not follow the NCES participation rate guidelines.

Table 6F.2, continued

State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 8

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
Non-SSI States, *N* = 18								
Alabama	4.77	71.4	89.4	92.9	84.7	57.0	39.4	43.8
Alaska†	5.41	76.5	94.3	95.3	76.5	79.1	54.8	74.7
Arizona	5.26	77.9	93.7	86.9	83.2	74.3	53.2	61.5
Hawaii	5.30	80.2	97.0	88.3	78.9	74.3	48.3	67.0
Indiana	4.52	68.5	90.1	84.3	73.8	63.9	41.3	35.7
Iowa†	4.74	73.3	91.4	83.8	83.4	61.5	47.5	34.7
Maryland	5.34	75.5	95.9	90.2	87.2	73.4	55.0	62.3
Minnesota	5.19	73.0	93.6	88.4	83.4	65.4	61.1	55.8
Mississippi	4.85	82.1	96.5	94.2	85.8	62.6	38.3	35.6
Missouri	5.27	82.0	96.9	88.8	86.6	71.0	61.2	44.2
North Dakota	5.07	76.9	91.9	89.2	88.5	59.3	62.8	42.3
Oregon	4.86	77.4	93.7	95.9	69.9	68.9	51.4	43.8
Tennessee	4.83	75.4	95.3	85.0	81.6	65.4	47.4	37.0
Utah	4.97	70.8	92.2	83.1	89.3	58.1	57.1	58.5
Washington	4.95	72.0	92.9	87.9	72.4	66.2	55.1	56.1
West Virginia	4.94	81.4	94.6	92.5	85.4	60.4	51.7	32.7
Wisconsin†	5.11	73.0	92.7	81.3	85.6	71.6	64.2	47.6
Wyoming	4.53	66.7	90.8	84.0	79.1	65.8	44.2	32.9

†Did not follow the NCES participation rate guidelines.

Table 6F.3

State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

	State Mean	Percent of students with teachers who had studied a reform-related topic						
		Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
SSI States, *N* = 23								
Arkansas†	4.56	77.8	90.3	95.1	76.0	57.8	32.9	36.9
California	5.18	78.6	91.0	94.2	70.9	71.0	48.5	70.0
Colorado	5.15	79.3	92.6	95.8	69.8	74.6	54.2	52.5
Connecticut	4.78	81.8	90.1	95.0	73.5	70.5	36.0	36.7
Delaware	5.03	78.4	94.6	96.6	80.6	72.6	41.1	43.3
Florida	5.27	79.6	92.7	94.2	80.3	72.2	40.4	76.7
Georgia	4.97	81.3	92.7	95.1	75.3	73.2	41.3	45.9
Kentucky	4.99	90.4	97.9	97.9	82.1	68.7	38.3	34.9
Louisiana	4.75	76.3	92.9	93.4	77.7	66.7	37.4	42.1
Maine	4.39	71.3	90.1	92.2	63.8	73.9	42.1	9.6
Massachusetts	4.72	79.7	91.8	94.3	64.6	75.1	38.3	35.1
Michigan†	4.90	83.6	92.7	96.1	79.7	73.5	38.8	30.9
Montana†	4.99	81.4	91.4	93.6	69.9	70.8	52.6	42.4
Nebraska	4.91	77.2	87.4	93.7	75.2	63.1	44.5	58.8
New Jersey†	4.86	75.8	91.2	93.4	77.6	68.3	41.1	42.6
New Mexico	4.96	79.0	87.4	94.8	70.1	66.8	39.5	61.5
New York†	4.30	66.0	87.3	92.4	57.6	65.1	30.6	40.6
North Carolina	5.13	81.6	94.6	97.3	89.9	65.6	42.7	45.6
Rhode Island	4.45	78.5	87.9	91.8	72.0	63.4	28.4	31.6
South Carolina†	4.76	79.1	94.7	94.6	72.6	63.7	37.7	40.3
Texas	5.03	86.8	96.4	95.8	71.1	70.0	32.6	59.5
Vermont†	5.54	75.9	97.0	97.7	61.9	81.8	38.3	13.5
Virginia	4.87	77.9	90.6	94.4	76.4	66.1	43.6	43.7

†Did not meet the NCES participation rate guidelines

Table 6F.3

State Means on $I_{RT(96)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

	State Mean	Percent of students with teachers who had studied a reform-related topic						
		Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
Non-SSI States, *N* = 20								
Alabama	4.87	77.9	93.8	96.7	70.5	73.5	40.8	45.2
Alaska†	4.85	73.3	85.4	91.3	62.4	64.0	54.7	63.6
Arizona	4.89	79.0	89.5	91.3	64.7	73.2	45.2	53.5
Hawaii	4.99	78.4	90.4	94.3	72.7	76.5	37.0	58.4
Indiana	4.28	70.2	92.1	90.6	61.3	70.2	24.3	26.4
Iowa†	4.56	75.8	90.2	93.3	67.1	66.5	38.5	32.3
Maryland	5.22	80.0	92.1	93.5	83.6	75.6	51.0	52.2
Minnesota	4.93	77.3	92.5	93.9	75.3	71.7	46.5	39.3
Mississippi	4.77	85.7	94.7	97.3	72.5	66.7	37.3	34.4
Missouri	4.77	79.2	90.9	96.9	71.7	68.1	35.9	35.0
Nevada†	5.36	87.2	96.7	97.0	77.1	84.6	41.7	61.0
North Dakota	4.90	81.5	90.3	97.4	77.3	62.5	48.4	35.6
Oregon	4.81	82.6	93.3	95.8	68.0	73.7	39.8	37.7
Pennsylvania†	4.30	71.4	90.0	92.7	69.3	61.9	24.2	23.7
Tennessee	4.32	69.4	88.1	93.3	65.1	57.9	33.9	30.5
Utah	4.69	78.3	90.1	97.4	83.2	61.0	32.8	35.6
Washington	4.58	72.1	88.5	89.9	60.4	64.9	43.1	49.1
West Virginia	4.87	86.7	93.6	96.4	83.5	64.9	38.2	28.4
Wisconsin	4.50	78.3	87.3	92.3	64.7	62.9	36.0	29.3
Wyoming	4.41	74.8	88.8	91.4	62.0	64.1	39.3	27.3

†Did not meet the NCES participation rate guidelines

Table 6F.4

Comparison of SSI and Non-SSI States on $I_{RT(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1992 sample					
Total sample					
SSI	22	4.75	0.28	0.65	.519
Non-SSI	19	4.68	0.39		
Subsample					
SSI	18	4.76	0.24	0.81	.424
Non-SSI	18	4.67	0.40		
2-point trend sample					
Total sample					
SSI	20	4.74	0.28	0.50	.623
Non-SSI	15	4.68	0.42		
Subsample					
SSI	14	4.79	0.26	2.31	.032
Non-SSI	11	4.51	0.33		
3-point trend sample					
Total sample					
SSI	17	4.78	0.28	0.59	.564
Non-SSI	11	4.69	0.48		
Subsample					
SSI	13	4.82	0.24	2.66	.026
Non-SSI	7	4.43	0.35		
Grade 4					
1992 sample					
Total sample					
SSI	22	4.70	0.30	0.23	.822
Non-SSI	19	4.72	0.24		
Subsample					
SSI	17	4.72	0.31	0.08	.933
Non-SSI	19	4.72	0.24		
2-point trend sample					
Total sample					
SSI	21	4.70	0.31	0.06	.955
Non-SSI	16	4.71	0.26		
Subsample					
SSI	13	4.73	0.34	0.19	.850
Non-SSI	14	4.71	0.25		

Table 6F.5

State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 8

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
SSI States, *N* = 22								
Arkansas	4.46	65.2	92.3	84.2	78.4	59.7	34.0	34.5
California	5.09	77.3	91.5	88.7	78.3	68.7	41.7	64.5
Colorado	4.96	75.2	94.8	85.1	68.5	65.7	53.9	55.3
Connecticut	5.15	85.7	94.8	87.9	82.5	77.1	45.2	44.1
Delaware	5.65	67.5	90.6	74.4	76.2	64.7	48.9	46.4
Florida	5.00	75.6	92.1	86.9	79.2	67.2	38.7	65.7
Georgia	5.05	83.1	93.3	88.3	72.8	71.7	47.4	54.0
Kentucky	4.50	76.8	91.0	85.2	66.6	61.8	37.1	34.8
Louisiana	4.79	77.9	97.2	87.5	74.1	64.8	38.4	43.1
Maine†	4.43	75.5	94.9	89.1	68.6	61.1	43.4	11.3
Massachusetts	4.37	77.3	91.8	78.6	59.3	67.0	32.2	32.7
Michigan	4.75	81.8	91.3	83.3	78.5	67.2	41.8	33.3
Nebraska†	5.15	80.3	94.0	87.5	76.0	60.0	59.7	58.9
New Jersey†	5.02	81.8	97.5	87.6	78.3	72.8	35.9	49.5
New Mexico	4.72	74.5	91.6	86.4	65.3	58.8	41.2	56.1
New York†	4.22	54.6	91.2	81.0	61.0	66.7	37.2	33.3
North Carolina	4.75	78.5	92.5	87.1	76.9	63.5	38.1	39.5
Ohio	4.55	77.6	93.5	81.7	71.3	58.4	38.2	34.8
Rhode Island	4.35	71.1	86.2	75.2	71.2	63.2	39.4	34.7
South Carolina	4.75	80.3	96.2	90.6	72.0	68.1	33.9	38.3
Texas	4.80	73.8	93.4	86.8	85.0	56.8	31.7	56.6
Virginia	4.93	75.3	94.9	85.2	77.0	68.6	44.5	49.2

†Did not follow the NCES participation rate guidelines.

Table 6F.5, continued

State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform Related Topic, Grade 8

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
Non-SSI States, *N* = 19								
Alabama†	4.86	80.8	95.2	84.2	74.3	74.2	33.9	42.1
Arizona	4.93	75.6	96.0	88.7	69.6	70.4	43.5	57.4
Hawaii	4.35	64.3	89.0	85.1	66.0	55.6	30.9	47.2
Idaho	4.62	73.3	94.0	87.9	76.4	57.7	45.3	42.0
Indiana	4.33	67.4	88.8	74.4	64.6	61.5	35.6	42.1
Iowa	5.09	83.5	92.8	86.9	72.1	70.3	54.6	51.0
Maryland	5.60	80.4	93.7	88.3	88.1	77.3	65.9	65.5
Minnesota	4.52	72.9	92.9	85.2	68.3	55.8	44.4	42.1
Mississippi	5.05	82.2	96.0	87.5	70.2	69.5	47.2	53.4
Missouri	4.59	78.5	91.1	89.1	72.6	67.2	34.4	32.2
New Hampshire	4.87	75.7	94.9	78.6	83.0	74.2	55.6	22.1
North Dakota	4.35	67.8	92.0	83.3	66.5	58.7	44.1	35.7
Oklahoma	4.85	79.1	92.9	87.5	66.1	59.2	39.5	74.5
Pennsylvania	4.31	66.3	85.6	78.0	68.7	67.2	39.1	27.6
Tennessee	4.48	74.7	93.4	84.8	64.5	62.9	33.5	35.7
Utah	4.50	74.0	88.6	73.5	66.1	53.1	45.2	52.1
West Virginia	4.73	78.2	95.7	84.6	73.4	64.7	43.7	34.1
Wisconsin	5.00	79.2	94.8	82.1	75.8	74.9	53.3	43.5
Wyoming	3.82	54.4	80.4	71.3	61.6	52.1	37.2	27.8

†Did not follow the NCES participation rate guidelines.

Table 6F.6
State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
SSI States, *N* = 22								
Arkansas	4.43	76.1	90.4	92.1	57.2	59.5	28.8	41.5
California	5.30	88.4	94.7	95.4	70.4	73.0	42.3	68.8
Colorado	5.00	81.7	93.9	92.6	65.2	73.0	46.2	52.0
Connecticut	4.90	89.5	92.0	95.5	66.1	77.2	33.5	36.4
Delaware†	4.82	78.9	93.9	90.8	64.6	74.0	36.6	48.4
Florida	5.24	84.3	92.4	94.7	73.3	74.0	39.9	67.9
Georgia	4.75	79.4	91.9	93.7	62.3	70.9	35.5	43.3
Kentucky	4.34	81.6	88.3	90.4	55.9	56.0	32.5	30.8
Louisiana	4.64	76.4	90.0	94.1	55.7	70.7	35.9	46.8
Maine†	4.50	81.9	93.1	94.8	59.3	68.8	41.8	12.1
Massachusetts	4.48	81.5	90.0	90.4	51.2	73.3	30.8	31.8
Michigan	4.75	84.1	91.3	90.9	70.8	70.4	31.1	37.8
Nebraska†	4.93	75.5	88.0	97.2	63.1	67.6	43.7	60.2
New Jersey†	4.83	79.7	95.0	92.4	57.2	74.0	38.2	48.0
New Mexico	4.57	73.2	88.3	91.3	53.3	64.0	33.8	52.9
New York†	4.22	72.2	84.7	86.4	41.5	68.0	27.7	43.9
North Carolina	4.62	81.3	88.0	93.9	70.0	60.9	32.5	38.5
Ohio	4.65	76.4	91.9	94.6	63.9	68.7	31.4	38.5
Rhode Island	4.05	69.0	83.5	86.9	53.2	59.0	22.3	32.9
South Carolina	4.75	84.1	94.3	95.3	64.9	68.5	32.8	36.2
Texas	4.76	75.8	92.0	93.3	61.4	66.7	32.1	56.1
Virginia	4.86	77.7	90.0	94.0	64.9	75.1	35.8	48.3

†Did not follow the NCES participation rate guidelines.

Table 6F.6, continued

State Means on $I_{RT(92)}$ and the Percent of Students Whose Teacher had Studied Each Reform-Related Topic, Grade 4

		Percent of students with teachers who had studied a reform-related topic						
	State Mean	Estimation	Problem Solving	Manipulatives	Calculators	Students' Thinking	Gender Issues	Cultural Differences
Non-SSI States, *N* = 19								
Alabama	4.83	79.1	93.5	95.6	57.6	72.2	40.6	45.1
Arizona	4.75	78.8	91.2	91.9	59.0	70.7	33.7	51.4
Hawaii	4.94	79.5	92.8	93.6	74.9	70.3	33.7	50.8
Idaho	4.88	88.1	92.9	94.2	70.4	71.9	36.1	34.8
Indiana	4.35	78.1	92.4	92.1	51.8	65.4	24.7	31.5
Iowa	4.96	82.6	94.6	92.8	65.8	76.0	41.0	44.5
Maryland	5.24	84.4	94.4	95.9	78.8	75.8	41.1	55.5
Minnesota	4.83	77.1	91.6	94.2	62.8	72.8	41.4	43.6
Mississippi	4.99	78.6	92.3	97.0	68.9	73.8	45.2	48.5
Missouri	4.59	79.5	91.5	95.6	64.3	64.8	30.5	33.2
New Hampshire	4.63	82.5	92.6	96.4	55.7	82.8	36.1	23.5
North Dakota	4.48	77.1	93.2	95.3	57.3	68.4	29.2	28.6
Oklahoma	4.84	78.0	91.1	87.4	48.5	74.4	38.9	68.2
Pennsylvania	4.43	76.3	92.6	88.7	59.6	70.2	26.8	29.3
Tennessee	4.42	76.0	91.8	92.9	51.8	65.7	30.6	35.6
Utah	4.50	75.0	89.4	94.1	59.8	62.2	30.3	41.5
West Virginia	4.45	81.6	91.0	90.4	62.6	59.1	30.8	30.7
Wisconsin	4.79	80.6	94.9	87.3	65.4	72.8	39.6	38.4
Wyoming	4.77	84.7	92.9	91.7	63.9	72.1	40.2	32.5

†Did not follow the NCES participation rate guidelines.

Table 6F.7
Comparison of SSI and Non-SSI States on Percent of Students Whose Teachers had Studied "Teaching Students from Different Cultural Backgrounds", Grade 8

	N	Mean Percent	Standard Deviation	*t*	*p*
1996					
Yearly sample					
Total sample					
SSI	22	49.58	17.40	0.30	.765
Non-SSI	18	48.14	12.94		
Subsample					
SSI	16	54.29	16.79	1.55	.132
Non-SSI	14	46.22	11.45		
2-point trend sample					
Total sample					
SSI	20	51.10	16.53	1.04	.308
Non-SSI	15	46.12	11.94		
Subsample					
SSI	14	55.38	12.52	1.90	.071
Non-SSI	11	45.76	12.62		
3-point trend sample					
Total sample					
SSI	17	54.41	14.74	1.42	.170
Non-SSI	11	46.96	12.80		
Subsample					
SSI	13	56.35	12.47	1.47	.169
Non-SSI	7	46.86	14.38		

-continued-

Table 6F.7, continued

Comparison of SSI and Non-SSI States on Percent of Students Whose Teachers had Studied "Teaching Students from Different Cultural Backgrounds", Grade 8

	N	Mean Percent	Standard Deviation	*t*	*p*
1992					
Yearly sample					
Total sample					
SSI	22	44.12	12.96	0.13	.895
Non-SSI	19	43.58	13.29		
Subsample					
SSI	18	45.42	11.01	0.43	.673
Non-SSI	18	43.66	13.67		
2-point trend sample					
Total sample					
SSI	20	44.32	13.40	0.05	.962
Non-SSI	15	44.12	10.36		
Subsample					
SSI	14	48.34	10.78	1.60	.124
Non-SSI	11	41.79	9.67		
3-point trend sample					
Total sample					
SSI	17	47.29	11.19	0.69	.500
Non-SSI	11	44.41	10.68		
Subsample					
SSI	13	49.53	10.20	1.87	.084
Non-SSI	7	40.91	9.66		
1990					
Yearly sample					
Total sample					
SSI	20	32.16	10.91	0.52	.608
Non-SSI	17	30.47	8.94		
Subsample					
SSI	20	32.16	10.91	0.74	.467
Non-SSI	16	29.75	8.72		
2- and 3-point trend samples					
Total sample					
SSI	17	32.39	11.53	-0.03	.976
Non-SSI	11	32.51	8.96		
Subsample					
SSI	13	34.57	12.28	0.61	.552
Non-SSI	7	31.60	9.34		

Table 6F.8

Percent of State Students Whose Teacher had Studied Teaching Students from Different Cultural Backgrounds, Grade 8, 1990

	Mean Percent		Mean Percent
SSI States , *N* = 20		Non-SSI States, *N* = 17	
Arkansas	23.5	Alabama	25.9
California	51.2	Arizona	44.8
Colorado	46.0	Hawaii	37.4
Connecticut	25.0	Idaho	28.3
Delaware	44.8	Illinois	29.3
Florida	26.1	Indiana	23.5
Georgia	27.7	Iowa†	41.9
Kentucky	23.1	Maryland	42.5
Louisiana	28.5	Minnesota	36.8
Michigan	23.5	New Hampshire	19.2
Montana	39.9	North Dakota	34.4
Nebraska	31.5	Oklahoma	37.6
New Jersey	24.2	Oregon	31.2
New Mexico	54.2	Pennsylvania	14.8
New York	22.6	West Virginia	26.3
North Carolina	23.5	Wisconsin	26.1
Ohio	28.5	Wyoming	17.9
Rhode Island	22.0		
Texas	49.0		
Virginia	28.3		

†Did not follow the NCES participation rate guidelines.

Appendix G

Table 6G.1

Comparison of SSI and Non-SSI States on $I_{C(96)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1996 sample					
Total sample					
SSI	22	9.92	0.65	0.32	.753
Non-SSI	18	9.85	0.72		
Subsample					
SSI	16	9.94	0.62	0.82	.418
Non-SSI	14	9.73	0.76		
2-point trend sample					
Total sample					
SSI	20	9.88	0.66	0.43	.669
Non-SSI	15	9.77	0.76		
Subsample					
SSI	14	9.90	0.66	0.61	.547
Non-SSI	11	9.72	0.74		
3-point trend sample					
Total sample					
SSI	17	9.88	0.70	0.38	.708
Non-SSI	11	9.78	0.68		
Subsample					
SSI	13	9.88	0.68	0.63	.541
Non-SSI	7	9.70	0.58		
Grade 4					
1996 sample					
Total sample					
SSI	23	8.16	0.53	1.61	.115
Non-SSI	20	7.91	0.48		
Subsample					
SSI	16	8.17	0.53	1.35	.187
Non-SSI	16	7.92	0.54		
2-point trend sample					
Total sample					
SSI	21	8.14	0.55	1.45	.155
Non-SSI	16	7.88	0.52		
Subsample					
SSI	13	8.11	0.55	1.07	.294
Non-SSI	14	7.88	0.55		

Table 6G.2

State Means on $I_{C(96)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
SSI States, *N* = 22						
Arkansas†	9.21	2.75	30	55	81	80
California	10.54	3.36	61	80	87	87
Colorado	9.73	3.26	46	68	59	75
Connecticut	10.42	3.36	49	77	89	90
Delaware	10.57	3.40	45	84	92	93
Florida	9.87	3.14	42	63	80	84
Georgia	10.21	3.21	40	72	93	90
Kentucky	10.60	3.46	50	81	90	90
Louisiana	8.36	2.36	18	37	75	69
Maine	10.29	3.47	44	74	86	78
Massachusetts	10.09	3.26	54	72	76	80
Michigan†	10.97	3.66	59	86	94	92
Montana†	10.54	3.48	52	78	87	86
Nebraska	10.15	3.42	38	76	77	79
New Mexico	9.31	2.91	38	56	75	72
New York†	9.21	2.78	40	54	78	68
North Carolina	10.35	3.28	35	74	98	98
Rhode Island	9.95	3.08	52	73	90	76
South Carolina†	9.26	2.79	22	51	87	87
Texas	9.00	2.58	28	43	92	79
Vermont†	10.12	3.35	38	68	83	84
Virginia	9.53	2.92	31	58	89	84

†Did not follow the NCES participation rate guidelines.

Table 6G.2, continued

State Means on $I_{C(96)}$ and Frequency of Calculator Use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
Non-SSI States, *N* = 18						
Alabama	8.68	2.44	21	39	82	79
Alaska†	10.12	3.34	46	67	76	86
Arizona	10.06	3.27	46	67	86	82
Hawaii	8.92	2.60	32	48	84	68
Indiana	9.06	2.77	23	48	83	78
Iowa†	10.39	3.48	54	81	73	80
Maryland	9.99	3.12	43	69	91	83
Minnesota	10.46	3.57	63	85	60	81
Mississippi	9.37	2.85	29	51	81	87
Missouri	10.51	3.50	54	76	85	85
North Dakota	9.97	3.47	42	77	49	82
Oregon	10.48	3.53	62	85	68	77
Tennessee	8.41	2.41	17	33	76	74
Utah	10.74	3.66	67	85	72	83
Washington	10.12	3.36	48	74	74	80
West Virginia	9.39	2.84	31	51	89	84
Wisconsin†	10.68	3.55	58	84	81	89
Wyoming	10.04	3.31	52	72	70	77

†Did not follow the NCES participation rate guidelines.

Table 6G.3

State Means on $I_{C(96)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
SSI States, *N* = 23						
Arkansas†	7.65	1.94	7	4	81	79
California	8.32	2.26	17	16	91	83
Colorado	8.01	2.17	9	14	82	81
Connecticut	8.25	2.26	9	8	92	89
Delaware	8.30	2.23	11	12	94	90
Florida	7.98	2.12	12	4	85	84
Georgia	7.77	2.04	9	9	79	79
Kentucky	8.99	2.59	19	28	98	95
Louisiana	7.47	1.91	7	4	68	76
Maine	8.85	2.54	16	26	98	94
Massachusetts	8.05	2.19	12	11	82	79
Michigan†	8.95	2.61	15	25	97	98
Montana†	8.14	2.24	6	8	90	87
Nebraska	8.18	2.22	11	7	90	87
New Jersey†	8.36	2.37	11	6	92	88
New Mexico	7.56	1.96	12	10	69	69
New York†	7.24	1.84	7	4	66	64
North Carolina	9.25	2.72	21	35	100	97
Rhode Island	8.45	2.28	15	22	91	89
South Carolina†	7.93	2.03	10	6	86	88
Texas	7.40	1.82	7	3	78	69
Vermont†	8.57	2.43	17	26	90	81
Virginia	7.96	2.10	6	3	90	85

†Did not meet the NCES participation rate guidelines

Table 6G.3, continued

State Means on $I_{C(96)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
Non-SSI States, *N* = 20						
Alabama	7.13	1.75	10	2	58	70
Alaska†	7.83	2.10	12	11	75	80
Arizona	7.61	1.91	10	10	79	72
Hawaii	7.84	1.94	13	10	89	80
Indiana	7.78	2.01	7	4	85	81
Iowa†	7.94	2.03	10	6	90	87
Maryland	8.72	2.47	16	25	92	93
Minnesota	8.50	2.49	15	17	84	86
Mississippi	7.35	1.83	13	7	63	69
Missouri	7.27	1.73	6	3	76	70
Nevada†	7.88	2.09	11	10	78	79
North Dakota	7.73	2.09	8	9	74	74
Oregon	8.53	2.34	13	27	90	90
Pennsylvania†	7.82	2.07	6	8	84	77
Tennessee	7.10	1.79	5	2	59	65
Utah	8.46	2.32	14	19	92	90
Washington	7.81	2.08	11	9	77	77
West Virginia	8.53	2.47	10	9	91	96
Wisconsin	8.28	2.25	11	13	91	89
Wyoming	8.09	2.10	8	6	93	91

†Did not meet the NCES participation rate guidelines

Table 6G.4

Comparison of SSI and Non-SSI States on $I_{C(92)}$ for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1992 sample					
Total sample					
SSI	22	11.25	0.90	0.10	.924
Non-SSI	19	11.21	1.11		
Subsample					
SSI	18	11.22	0.80	-0.13	.900
Non-SSI	18	11.26	1.13		
2-point trend sample					
Total sample					
SSI	20	11.26	0.95	-0.03	.973
Non-SSI	15	11.27	1.11		
Subsample					
SSI	14	11.19	0.77	-0.19	.856
Non-SSI	11	11.11	1.19		
3-point trend sample					
Total sample					
SSI	17	11.25	0.87	-0.62	.545
Non-SSI	11	11.45	0.82		
Subsample					
SSI	13	11.29	0.69	-0.04	.967
Non-SSI	7	11.31	0.84		
Grade 4					
1992 sample					
Total sample					
SSI	22	7.27	0.59	0.31	.757
Non-SSI	19	7.21	0.59		
Subsample					
SSI	17	7.30	0.63	0.43	.672
Non-SSI	19	7.21	0.59		
2-point trend sample					
Total sample					
SSI	21	7.27	0.60	0.10	.922
Non-SSI	16	7.25	0.56		
Subsample					
SSI	13	7.31	0.58	0.23	.824
Non-SSI	14	7.25	0.60		

Table 6G.5

State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

					Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned 4 function Calculators	School owned Scientific Calculators	Instruction in Use of 4 function	Instruction in Use of Scientific
SSI States, *N* = 22								
Arkansas	10.27	2.20	20	34	61	20	62	20
California	11.80	2.88	39	57	78	25	71	30
Colorado	11.96	3.11	46	65	64	24	64	31
Connecticut	12.06	2.84	37	56	85	37	72	37
Delaware	11.28	2.69	29	50	68	29	63	39
Florida	10.92	2.48	31	44	64	41	59	37
Georgia	11.44	2.59	27	46	84	26	84	26
Kentucky	12.23	3.03	38	64	77	35	77	42
Louisiana	10.01	2.20	19	29	49	20	58	18
Maine†	12.98	3.30	48	71	88	41	79	43
Massachusetts	9.83	2.04	21	29	59	17	49	22
Michigan	12.73	3.17	48	74	83	40	78	45
Nebraska†	11.85	2.90	36	64	66	30	70	32
New Jersey†	11.07	2.49	20	41	71	42	66	39
New Mexico	10.91	2.36	27	40	71	19	73	27
New York†	9.60	1.92	18	24	70	8	64	8
North Carolina	10.89	2.35	19	34	74	33	67	36
Ohio	11.18	2.45	21	44	69	38	64	38
Rhode Island	10.62	2.35	25	42	71	29	50	22
South Carolina	11.05	2.40	20	40	77	29	77	29
Texas	12.05	2.89	38	54	70	47	70	45
Virginia	10.65	2.31	20	34	76	26	61	20

†Did not follow the NCES participation rate guidelines.

Table 6G.5, continued

State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

			Percent of students with					
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned 4 function Calculators	School owned Scientific Calculators	Instruction in Use of 4 function	Instruction in Use of Scientific
Non-SSI States, *N* = 19								
Alabama†	10.44	2.40	21	41	49	20	66	30
Arizona	11.01	2.52	30	45	70	21	65	24
Hawaii	11.06	2.31	25	38	78	41	63	30
Idaho	12.26	3.14	44	62	56	38	69	54
Indiana	10.29	2.18	14	31	72	15	68	15
Iowa	11.96	2.99	31	64	70	31	69	33
Maryland	12.05	2.74	33	54	81	50	69	48
Minnesota	12.58	3.29	57	71	64	33	63	51
Mississippi	8.84	1.85	11	20	40	7	50	13
Missouri	12.44	3.28	44	76	70	26	76	34
New Hampshire	11.70	2.83	31	57	75	24	70	33
North Dakota	11.13	3.00	39	59	45	26	49	40
Oklahoma	9.42	2.03	14	23	43	14	46	17
Pennsylvania	10.66	2.39	25	42	64	31	58	26
Tennessee	9.73	2.08	12	22	41	18	52	20
Utah	12.06	3.24	58	67	38	22	62	41
West Virginia	10.75	2.34	24	39	74	22	65	20
Wisconsin	12.34	3.29	38	64	58	41	61	51
Wyoming	12.35	3.10	43	66	80	34	69	33

†Did not follow the NCES participation rate guidelines.

Table 6G.6

State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
SSI States, *N* = 22						
Arkansas	6.35	1.42	5	3	39	46
California	8.03	2.13	11	12	86	82
Colorado	7.74	2.03	9	9	76	76
Connecticut	7.74	2.01	11	7	81	76
Delaware†	7.39	1.88	3	3	72	73
Florida	7.32	1.80	9	5	66	71
Georgia	6.88	1.62	7	4	53	60
Kentucky	8.31	2.38	13	13	79	89
Louisiana	6.54	1.58	8	7	31	49
Maine†	7.53	1.93	8	9	71	73
Massachusetts	7.08	1.72	9	5	62	61
Michigan	8.44	2.25	10	19	94	96
Nebraska†	7.43	1.89	11	3	69	72
New Jersey†	7.15	1.84	8	4	58	63
New Mexico	6.34	1.45	6	2	37	47
New York†	6.32	1.49	7	4	32	40
North Carolina	7.59	1.85	9	6	78	79
Ohio	7.16	1.74	3	4	66	68
Rhode Island	6.93	1.68	5	4	57	58
South Carolina	7.12	1.66	5	3	70	68
Texas	7.49	1.88	7	6	71	77
Virginia	6.98	1.64	6	4	62	63

†Did not follow the NCES participation rate guidelines.

Table 6G.6, continued

State Means on $I_{C(92)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 4

			Percent of students with			
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators	Instruction in Use
Non-SSI States, *N* = 19						
Alabama	6.87	1.79	7	2	34	65
Arizona	6.78	1.58	6	4	55	55
Hawaii	8.09	2.14	12	14	81	88
Idaho	7.70	1.99	6	8	76	82
Indiana	6.84	1.56	4	1	63	59
Iowa	7.33	1.80	6	2	75	69
Maryland	8.24	2.22	15	20	83	85
Minnesota	7.82	2.03	10	11	83	75
Mississippi	6.86	1.65	9	8	44	62
Missouri	6.93	1.61	4	3	61	62
New Hampshire	7.18	1.83	9	6	63	58
North Dakota	6.87	1.66	5	2	54	61
Oklahoma	6.05	1.33	4	0	28	40
Pennsylvania	7.16	1.73	4	4	69	64
Tennessee	6.19	1.38	5	2	29	47
Utah	7.16	1.77	6	6	61	65
West Virginia	7.34	1.81	8	6	66	71
Wisconsin	7.85	2.11	10	7	79	77
Wyoming	7.73	1.95	6	8	83	82

†Did not follow the NCES participation rate guidelines.

Table 6G.7

Comparison of SSI and Non-SSI States on $I_{C(90)}$, for All States in Each Sample and for the Subsample of States that Followed the NCES Participation Rate Guidelines

	N	Mean	Standard Deviation	*t*	*p*
Grade 8					
1990 sample					
Total sample					
SSI	20	6.30	0.65	-0.62	.538
Non-SSI	17	6.44	0.69		
Subsample					
SSI	20	6.30	0.65	-0.47	.642
Non-SSI	16	6.41	0.70		
2- and 3-point trend samples					
Total sample					
SSI	17	6.27	0.62	-0.48	.633
Non-SSI	11	6.39	0.68		
Subsample					
SSI	13	6.35	0.60	0.22	.829
Non-SSI	7	6.28	0.67		

Table 6G.8

State Means on $I_{C(90)}$ and Frequency of Calculator use and the Percent of Students Whose Teacher Answered Yes to Specific Items about Calculator Use, Grade 8

			Percent of Students		
	I_C	Frequency of Use	Unrestricted Use	Use on Tests	School owned Calculators
SSI States (N = 20)					
Arkansas	5.55	1.97	9	13	35
California	7.41	2.75	31	50	83
Colorado	7.04	2.68	30	45	62
Connecticut	7.18	2.59	26	43	89
Delaware	6.58	2.35	23	33	66
Florida	6.10	2.14	12	23	59
Georgia	6.52	2.39	14	30	69
Kentucky	5.72	2.00	12	20	40
Louisiana	5.30	1.79	5	16	29
Michigan	6.73	2.42	26	37	67
Montana	7.49	2.95	32	57	62
Nebraska	6.51	2.46	21	36	49
New Jersey	5.64	1.84	11	14	55
New Mexico	6.07	2.12	18	20	56
New York	5.23	1.69	5	12	37
North Carolina	6.25	2.16	10	18	81
Ohio	6.40	2.32	15	33	61
Rhode Island	5.83	1.88	19	23	52
Texas	6.23	2.20	12	22	71
Virginia	6.30	2.19	14	27	72
Non-SSI States (N = 17)					
Alabama	5.55	1.88	7	21	40
Arizona	6.06	2.09	17	22	60
Hawaii	5.52	1.75	14	15	49
Idaho	6.49	2.43	28	30	50
Illinois	6.82	2.52	23	36	70
Indiana	5.86	2.00	8	15	63
Iowa†	6.94	2.64	20	42	67
Maryland	6.58	2.31	19	30	77
Minnesota	7.05	2.68	31	47	58
New Hampshire	6.88	2.61	21	38	69
North Dakota	6.55	2.56	24	39	37
Oklahoma	5.54	1.95	10	15	33
Oregon	7.65	2.93	36	53	82
Pennsylvania	5.83	1.97	13	20	54
West Virginia	5.69	1.94	11	20	45
Wisconsin	7.26	2.81	29	50	65
Wyoming	7.24	2.67	36	49	73

†Did not meet the NCES participation rate guidelines.

CHAPTER 7

SSI AND NON-SSI ACHIEVEMENT USING STATE NAEP DATA: EMPIRICAL BAYES AND BAYESIAN ANALYSES

Introduction

State NAEP reported mathematics scale scores of each state that participated in the assessment years for grade 8 in 1990, 1992, and 1996 and for grade 4 in 1992 and 1996. One of the advantages of NAEP scale scores is that they are comparable across data collection years and across grades (Allen et al., 1997). This enables us to compare mathematics score means of states in each assessment year and to assess the trends of states participating in consecutive assessment years. The main purpose of this chapter is to examine the differences between SSI and non-SSI states in mathematics scale scores using longitudinal and cross-sectional analytic approaches. For these analyses, only the 28 states that participated all three years in NAEP are used. This sample consists of 17 SSI states and 11 non-SSI states.

New methods are required for the use of State NAEP data for the longitudinal and cross-sectional analyses because of the unique nature of the State NAEP (i.e., the small number of states participating in each assessment year and states' voluntary participation in the tests). To address these complexities of the NAEP data, we employed empirical Bayes and fully Bayesian methods, suggested by Raudenbush et al. (1999). For the analyses, each state mean and its standard error are needed to estimate the parameters in the models. The input data of state means and jackknife standard errors are listed in Tables 7A.1 and 7A.2 in the Appendix. These estimates are based on the results after taking into account the NAEP sampling design (Allen et al., 1997).

This chapter begins with longitudinal analyses of grade 8 data, grade 4 data, and cohort data in order to compare the overall trends of SSI states and non-SSI states over the assessment years. Then, the next section discusses the results of cross-state analyses using the grade 8 data in 1990, 1992, and 1996, and grade 4 data in 1992 and 1996. The latter analyses allow us to detect the differences between the two groups, the SSI and non-SSI states, in each assessment year. In each section, the statistical models used in the analyses are discussed.

Longitudinal Analysis: Empirical Bayes and Fully Bayesian Methods

In this section, we display the estimated posterior distribution of each parameter from a longitudinal analysis of grade 8 students in mathematics scale scores from 1990 to 1996 (Tables 7.1 to 7.3). Column one describes two different growth models (e.g., unconditional and conditional models) used in our analyses. Columns two and three summarize means, standard deviation, and credibility intervals obtained using the empirical Bayes method and the fully Bayesian method. The upper part of the tables presents the fixed effects or coefficients, and the lower part lists the random effects or variance components.

Overall, the estimates from the two methods appear to be fairly similar. As noted by many researchers, however, the fully Bayesian method has many properties that for this type of analysis are superior to the empirical Bayes method. In particular, the fully Bayesian method takes into account uncertainty regarding the parameters of interest. Thus, even though the results of both methods for the comparison are presented in the next section, we will focus mainly on the results from the fully Bayesian estimates of each parameter.

For all data sets of grade 8, grade 4, and the cohort, the unconditional fully Bayesian model is based on samples of 20,000 iterations with 5,000 burn-in iterations. For the conditional models, the fully Bayesian results were run for 30,000 iterations after a burn-in of 10,000 to approximate the marginal posteriors of the parameters.

Linear Growth Models of Grade 8

Unconditional Model. Table 7.1 displays the results from linear growth models of grade 8 data from 1990, 1992, and 1996 using empirical Bayes and fully Bayesian methods. First, we begin with an unconditional model to estimate the average state mean in 1990 and average state growth rate across 28 states. The fully Bayesian estimate of average state mean in 1990 is 262.800. Average state growth rate per year from 1990 to 1996 is 1.286. This means that grade 8 students across the states are gaining an average 1.286 points per year. Both posterior estimates are statistically significant as the 95% credibility intervals that we obtain exclude a value of zero.

Regarding the variance components, the fully Bayesian posterior means for the variance of average state mean in 1990 and growth rate are 85.430 and 0.186 (Table 7.1). The 95% credibility intervals for these parameters range from 49.960 to 145.300 for average state mean in 1990 and from 0.078 to 0.373 for average growth rate. Since Figure 7.1 displays the posterior distribution of the variances of each estimate, there is evidence of between-state heterogeneity in both average state mean in 1990 and growth rate.

Conditional Model. The next conditional model presents the differences in state mean in 1990 and the growth rate between SSI states and non-SSI states (Table 7.1). On average, SSI states started behind non-SSI states by 5.716 points in 1990 (average SSI differential effect in state mean in 1990 is –5.716). But, the growth rate of grade 8 students in SSI states is 0.200 points per year faster than that of their counterparts in non-SSI states (average SSI differential effect in state growth rate is 0.200). As a result, the learning gap between SSI states and non-SSI states was significantly reduced in 1996. Unfortunately, both SSI differential effects contain zero in their 95% credibility intervals, implying that they are not statistically significant.

Table 7.1 and Figure 7.2 show the posterior distribution of the variance of all estimates in state mean in 1990 and growth rate. All four variance estimates for SSI states and non-SSI states are positive. None of posterior estimates contains a value of zero in the lower boundary of the 95% credibility interval. The results suggest that a substantial variability in state mean in 1990 and growth rate lies across SSI states and across non-SSI states.

Figures 7.3 and 7.4 show the Bayesian posterior estimates with the 95% credibility intervals for each state. Considering the state mean in 1990, each state does vary considerably in its performance (Figure 7.3). Compared to non-SSI states, most SSI states score below the average state mean in 1990. The low-performing states were Louisiana, Alabama, and Hawaii, and the high-performing states were Minnesota, Iowa, and North Dakota. The three states that scored highest are all non-SSI states.

However, when we look at the posterior distribution of annual growth rate of each state, the pattern of state mean in 1990 is reversed (Figure 7.4). This also confirms a relative advantage for the SSI states over non-SSI states in growth rate. Despite the low scores in 1990, grade 8 students in SSI states were more likely to show a gain than their counterparts in non-SSI states. North Carolina, Michigan, and Texas were fast-gaining states, all SSI states. But, on average, the group differences between SSI states and non-SSI states are not statistically different from zero, even though the Bayesian posterior variance estimates provide clear evidence of heterogeneity in growth rate both among SSI states and non-SSI states.

Table 7.1
Longitudinal Analysis of Grade 8 Data over 1990, 1992, and 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

	Empirical Bayes		**Fully Bayesian**			
Fixed Effect	*Coefficient*	*SD*	*Coefficient*	*SD*	*Credibility Interval*	
					2.5%	*97.5%*
Linear Growth Model – Time						
Average state mean in 1990	262.950***	1.421	262.800	1.749	259.300	266.200
Average state growth rate (per year)	1.267**	0.390	1.286	0.098	1.091	1.480
Linear Growth Model – Time, SSI, and Time x SSI						
<u>Non-SSI State</u>						
Average Non-SSI state mean in 1990	266.867***	2.167	266.400	3.124	259.800	272.000
Average Non-SSI state growth rate (per year)	1.141~	0.595	1.166	0.185	0.807	1.537
<u>SSI State</u>						
Average SSI state mean in 1990	260.411		260.454			
Average SSI state growth rate (per year)	1.351		1.372			
<u>SSI Effect</u>						
Average SSI differential effect in state mean	-6.456*	2.781	-5.946	3.652	-12.220	2.145
Average SSI differential effect in state growth rate	0.210	0.764	0.206	0.250	-0.291	0.696
Random Effect						
Linear Growth Model – Time						
Variance (Mean)			85.430	24.600	49.960	145.300
Variance (Time)			0.186	0.077	0.078	0.373
Linear Growth Model – Time, SSI, and Time x SSI						
Variance (Mean)			108.900	46.850	49.450	228.000
Variance (Time)			0.295	0.152	0.114	0.687
Variance (SSI)			13.810	16.750	0.317	60.840
Variance (Time x SSI)			0.377	0.237	0.112	0.983

~ $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Figure 7.1. Posterior distribution of the variance: Unconditional model.

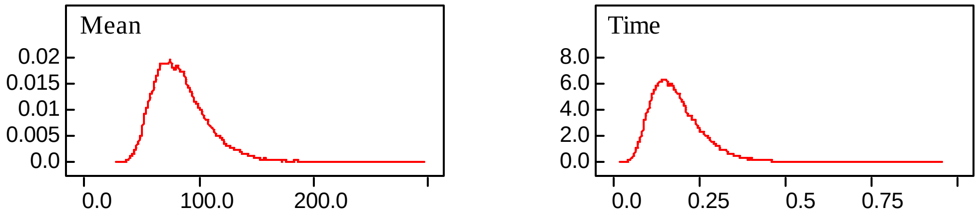


Figure 7.2. Posterior distribution of the variance: Conditional model.

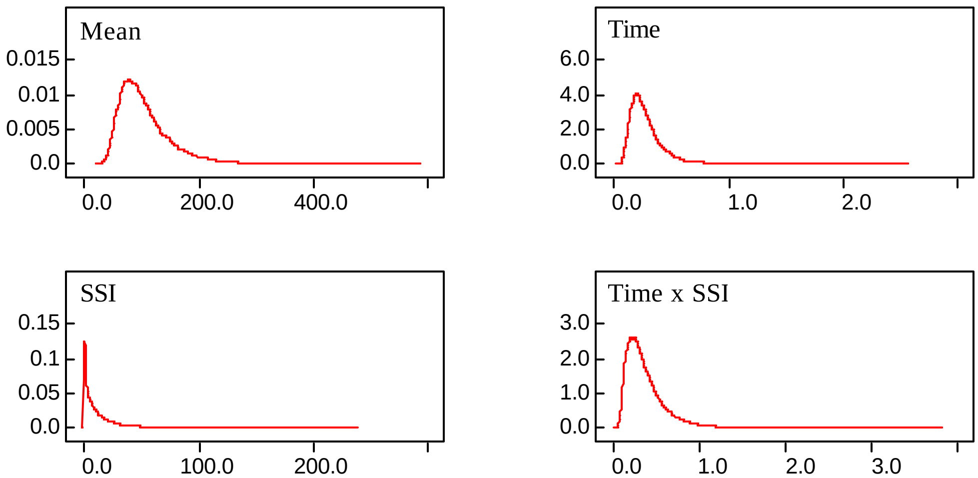


Figure 7.3. Posterior distributions of average scale scores of grade 8 in 1990.

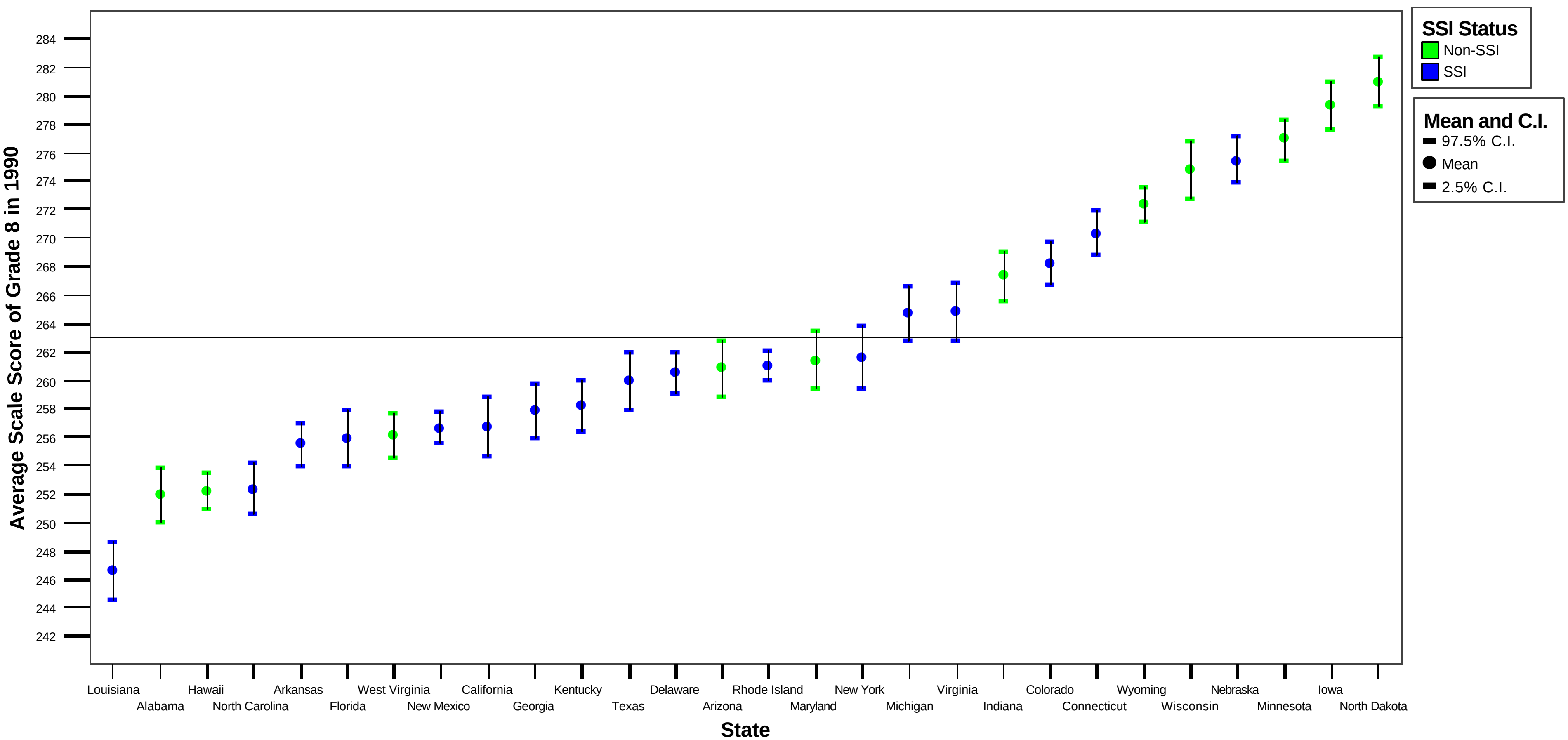
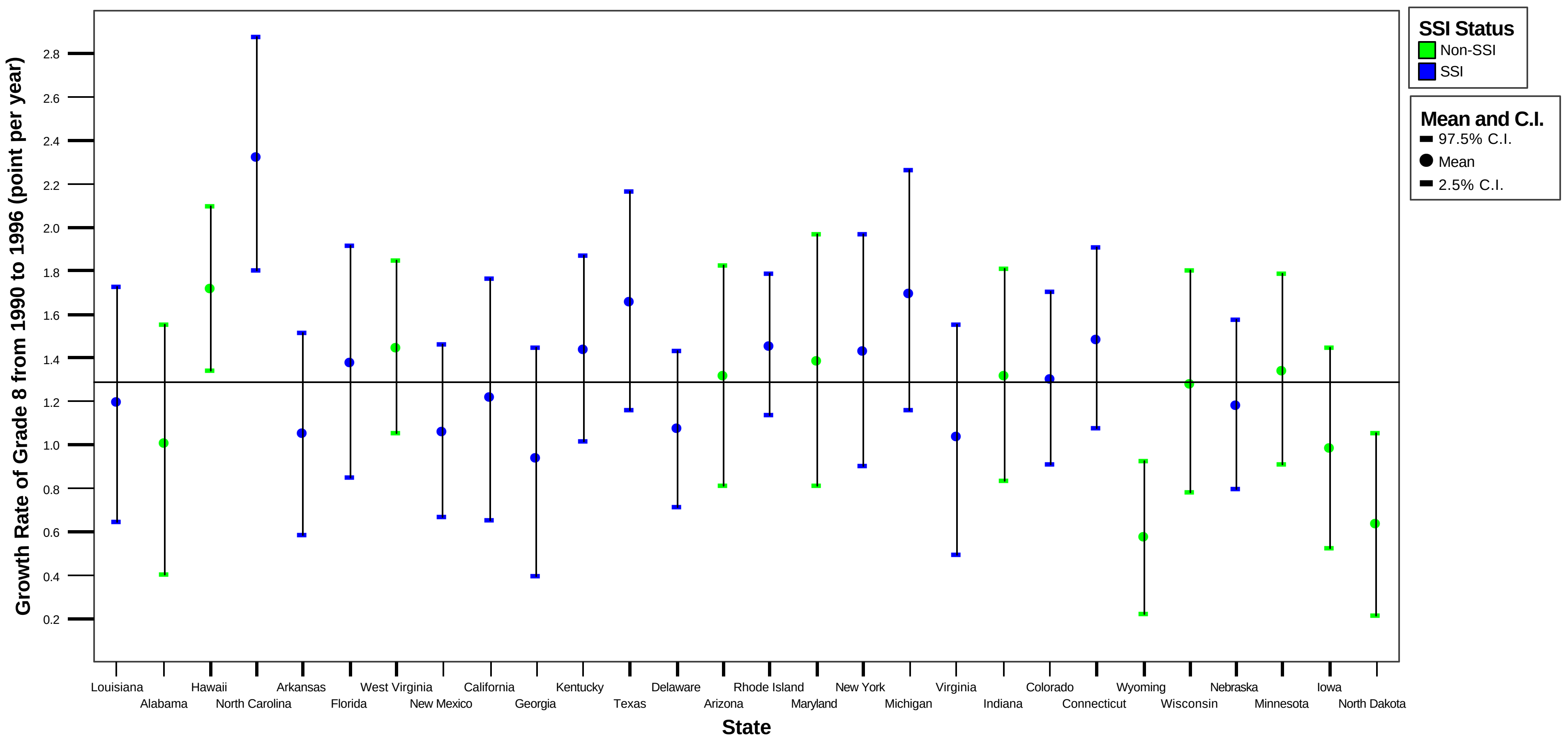


Figure 7.4. Posterior distributions of growth rates of grade 8 from 1990 to 1996.



Linear Growth Models of Grade 4

Unconditional Model. The results from the unconditional linear growth model of grade 4 data are presented in Table 7.2. Both empirical Bayes and fully Bayesian estimates of average state mean in 1992 and average growth rate are identical. The fully Bayesian estimated average state mean in 1992 and average state growth rate for grade 4 mathematics scale scores were 218.200 and 0.889, respectively. This indicates that the average grade 4 mathematics scores in 1992 across SSI and non-SSI states was 218.200 points and students were learning at the rate of 0.889 points per year from 1992 to 1996. As shown by the 95% Bayesian credibility intervals in Table 7.2, both posterior means are significant.

Table 7.2 also shows that there was a true variation in average state mean in 1992 and average growth rate among states. The posterior means of the variance are 46.900 for average state mean in 1992 and 0.604 for average state growth rate. However, the values of the variance of average state mean in 1992 can be as small as 27.110 and as large as 80.690. The posterior distribution of the variance of average growth rate ranges from 0.290 to 1.145. The results imply that a quite substantial variability in these two estimates exists between states.

Conditional Model. Table 7.2 shows the results of a conditional model after adjusting for SSI status. For non-SSI states, the average mathematics score mean in 1992 was 221.100 and average growth rate per year is 0.747. For SSI states, grade 4 students began with 216.276 points (221.100 + (–4.824)), but they gained more, at 0.997 points (0.747 + 0.250), per year. Thus, students in SSI states were more likely to score lower initially but tended to learn faster than their counterparts in non-SSI states. However, it appears that these two SSI differential effects using the fully Bayesian method are not statistically significant because both 95% credibility intervals included zero in their values.

Figures 7.5 and 7.6 display the posterior distributions of the variance of four estimates in average state mean in 1992 and average growth rate. None of the 95% credibility intervals for these estimates included zero, thus implying that each of the groups of SSI and non-SSI states had a considerable heterogeneity in average state mean in 1992 and average growth rate.

This is also confirmed by the line charts in Figures 7.7 and 7.8. The average state mean in 1992 differed significantly from state to state (Figure 7.7). Each state mean can be as low as 205 points and as high as 229 points. The top three states were Minnesota, North Dakota, and Iowa, and the lowest three states were Louisiana, California, and Alabama. All of the higher-performing states are non-SSI states. Figure 7.8 displays an interesting picture regarding state growth rate. While there is some between-state variation in growth rate, both the greatest gaining and least gaining states are SSI states. For example, North Carolina is a high-gaining state and Delaware a low-gaining state. As indicated in Table 7.2, there were no overall differences between SSI and non-SSI states in state mean in 1992 and state gain rate. But, the patterns of Figures 7.7 and 7.8 are consistent with the results of Table 7.2, which show considerable variance across both SSI states and non-SSI states in those two estimates.

Table 7.2
Longitudinal Analysis of Grade 4 Data in 1992 and 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

	Empirical Bayes		**Fully Bayesian**			
Fixed Effect	*Coefficient*	*SD*	*Coefficient*	*SD*	*Credibility Interval*	
					2.5%	*97.5%*
Linear Growth Model – Time						
Average state mean in 1992	218.230***	1.309	218.200	1.311	215.600	218.200
Average state growth rate (per year)	0.890~	0.463	0.889	0.169	0.556	0.889
Linear Growth Model – Time, SSI, and Time x SSI						
Non-SSI State						
Average Non-SSI state mean in 1992	221.192***	2.014	221.100	2.245	216.500	225.200
Average Non-SSI state growth rate (per year)	0.727	0.713	0.747	0.274	0.209	1.290
SSI State						
Average SSI state mean in 1992	216.303		216.276			
Average SSI state growth rate (per year)	0.998		0.997			
SSI Effect						
Average SSI differential effect in state mean	-4.889~	2.588	-4.824	2.856	-10.480	0.480
Average SSI differential effect in state growth rate	0.271	0.916	0.250	0.378	-0.488	1.006
Random Effect						
Linear Growth Model – Time						
Variance (Mean)			46.900	13.780	27.110	80.690
Variance (Time)			0.604	0.224	0.290	1.145
Linear Growth Model – Time, SSI, and Time x SSI						
Variance (Mean)			50.000	23.810	22.100	112.100
Variance (Time)			0.631	0.349	0.223	1.539
Variance (SSI)			5.154	7.916	0.215	27.730
Variance (Time x SSI)			0.708	0.566	0.147	2.224

~ $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Figure 7.5. Posterior distribution of the variance: Unconditional model.

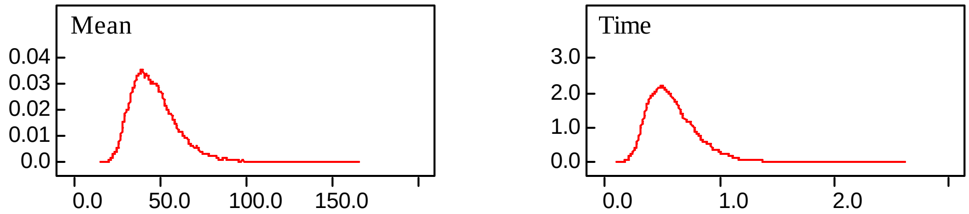


Figure 7.6. Posterior distribution of the variance: Conditional model.

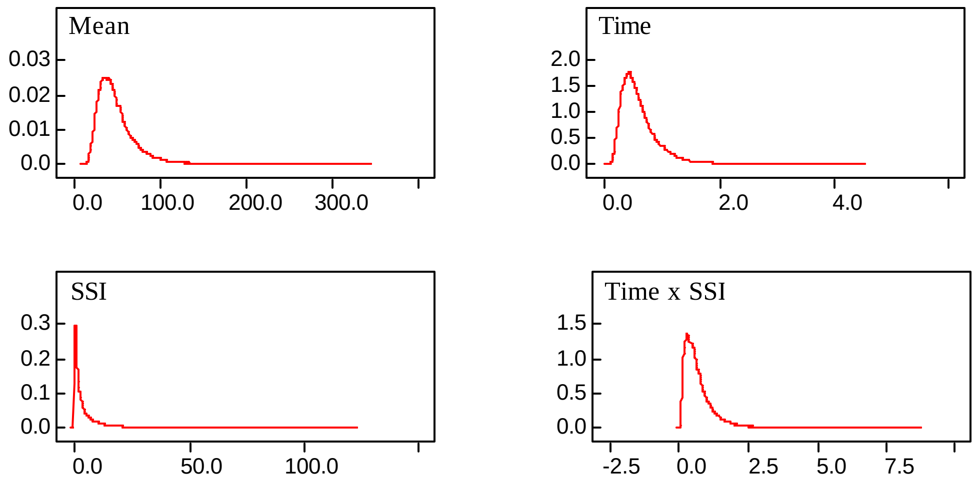


Figure 7.7. Posterior distributions of average scale scores of grade 4 in 1992.

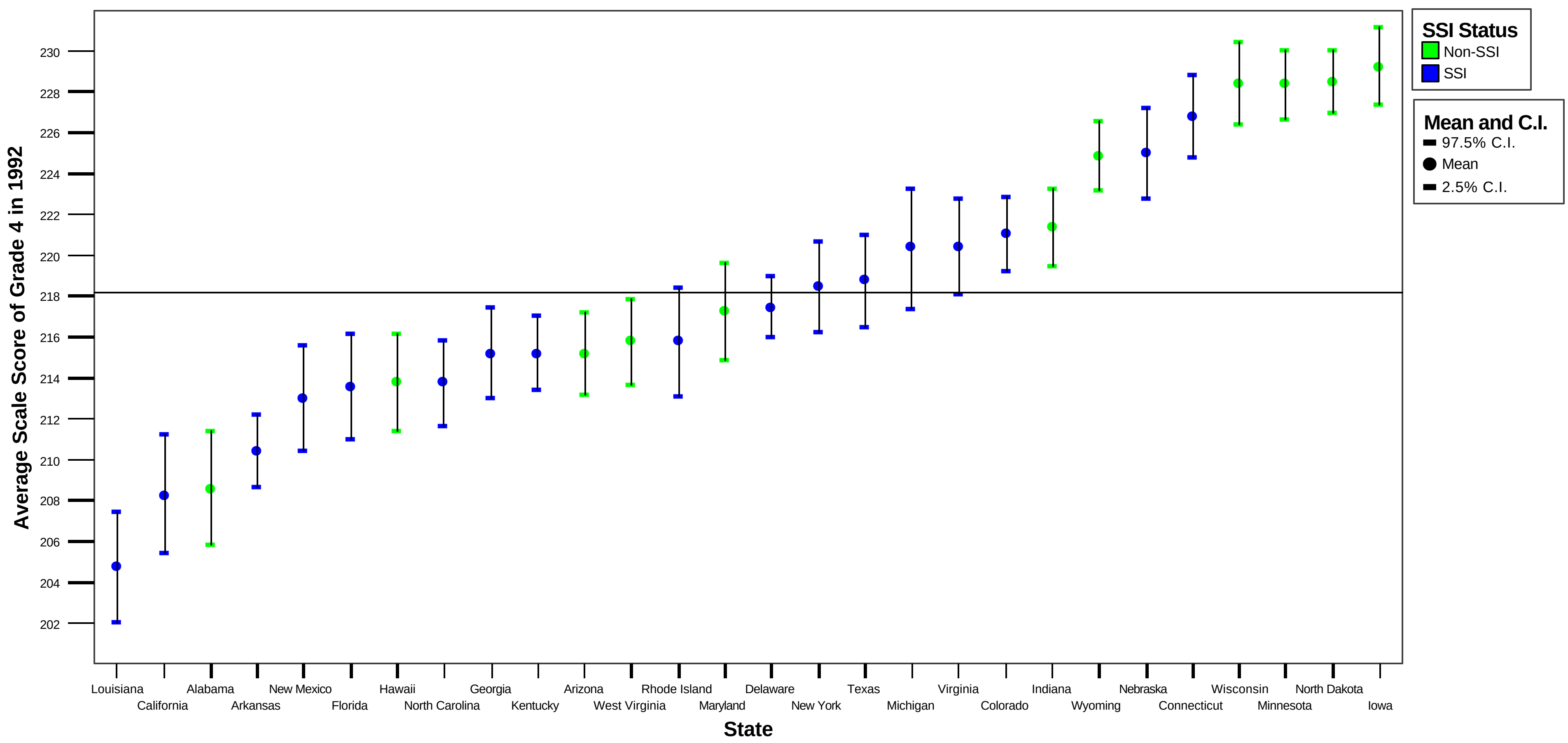
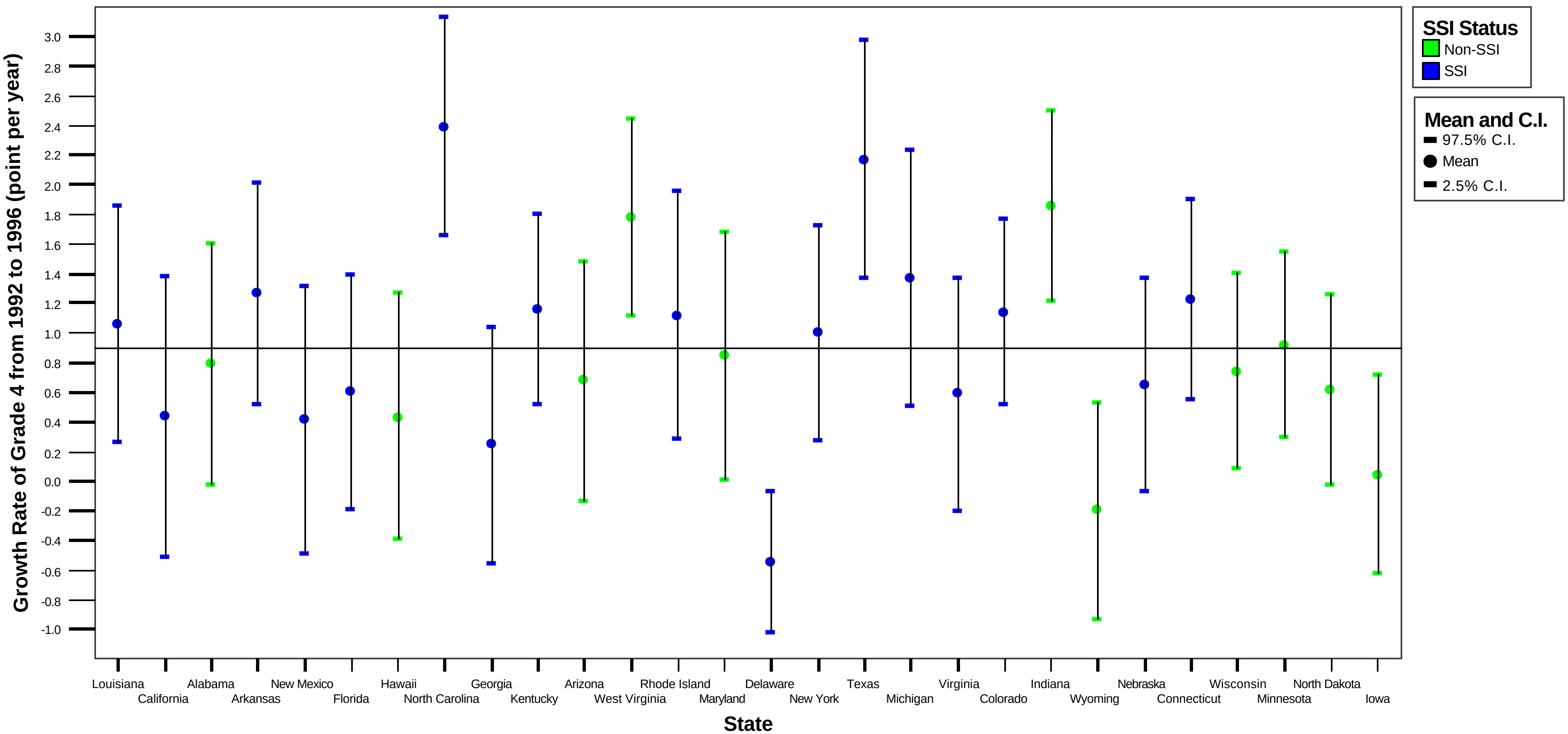


Figure 7.8. Posterior distributions of growth rates of grade 4 from 1992 to 1996.



Linear Growth Models of Cohort

Unconditional Model. Table 7.3 displays the estimates of empirical Bayes and Bayesian methods of cohort data analysis for grade 4 in 1992 and grade 8 in 1996 in the unconditional linear growth model. The estimated posterior mean of average state mean in 1992 is 218.2 points. This is exactly identical to the estimate in the previous result of grade 4 data in Table 7.3 because the base data point for the cohort is grade 4 in 1992. The overall state growth rate of the cohort was 12.96 points per year. That is, the grade 4 cohort among all states was likely to gain 12.96 points in mathematics scale scores in 1992. Each estimate of cohort growth model is statistically significant as indicated by the 95% credibility intervals.

Posterior variations in average state mean in 1992 and average state growth rate are also listed in Table 7.3. The estimated posterior variance mean of average state mean in 1992 is 46.39 and that of average state growth rate is 0.338. As Figure 7.9 shows, the posterior distributions of these estimates indicate that both estimates vary significantly between states. The 95% credibility interval for average state mean in 1992 ranges from 26.690 to 79.490. Average state growth rate in 1992 ranges from 0.114 to 0.138.

Conditional Model. Table 7.3 displays the results from the conditional models. In general, students in non-SSI states scored higher in 1992 and gained more than their counterparts in SSI states. On average, the non-SSI state mean in 1992 was 220.9 points and the growth rate 12.99 points per year. But, the state mean in 1992 and growth rate for SSI states were 216.361 (220.99 + (-4.539)) points and 12.967 (12.990 + (-0.023)) points, respectively. There is no evidence that the differences between SSI states and non-SSI states are statistically significant. Both posterior estimates of SSI differential effects include zero in the 95% credibility intervals.

Next, the posterior estimates of the variance of average state mean in 1992 and average state growth rate for each SSI state and non-SSI state are presented in Table 7.3. All four posterior means of the variance are positive and significant, indicating that each group, SSI states and non-SSI states, is heterogeneous in state mean in 1992 and state growth rate. The posterior distributions of these four fully Bayesian estimates in Figure 7.10 clearly indicate the variability of these estimates.

Figures 7.11 and 7.12 display the line charts describing the posterior distribution of average state mean in 1992 and average state growth rate for each state. As previously noted, the chart of average state mean in 1992 is exactly identical to those shown in grade 4 data. Thus, only Figure 7.12 will be discussed. Each state varies in its posterior mean of state growth rate. Especially, the low-gaining states (e.g., Louisiana and Alabama) appear different from the high-gaining states (Nebraska and North Dakota) in the estimated state growth rate. The distributions of SSI states and non-SSI states do not display a different pattern in the two groups. Instead, the estimated means among SSI states and among non-SSI states were both quite variable, as indicated by the 95% credibility intervals of state growth rate. The distribution of the estimated posterior means supports the results presented in Table 7.3, indicating no significant differences between SSI and non-SSI states in average state growth rate

Table 7.3
Longitudinal Analysis of Cohort Data for Grade 4 in 1992 and Grade 8 in 1996: Empirical Bayes and Fully Bayesian Estimates After Considering Jackknife Standard Errors

	Empirical Bayes		**Fully Bayesian**			
Fixed Effect	*Coefficient*	*SD*	*Coefficient*	*SD*	*Credibility Interval*	
					2.5%	*97.5%*
Linear Growth Model – Time						
Average state mean in 1992	218.220***	1.470	218.200	1.311	215.600	220.800
Average state growth rate (per year)	13.016***	0.521	12.960	0.139	12.680	13.230
Linear Growth Model – Time, SSI, and Time x SSI						
<u>Non-SSI State</u>						
Average Non-SSI state mean in 1992	221.178***	2.258	220.900	2.165	216.600	225.100
Average Non-SSI state growth rate (per year)	13.056***	0.801	12.990	0.272	12.430	13.510
<u>SSI State</u>						
Average SSI state mean in 1992	216.299		216.361			
Average SSI state growth rate (per year)	12.993		12.967			
<u>SSI Effect</u>						
Average SSI differential effect in state mean	-4.879~	2.902	-4.539	2.767	-9.866	0.738
Average SSI differential effect in state growth rate	-0.063	1.029	-0.023	0.362	-0.725	0.721
Random Effect						
Linear Growth Model – Time						
Variance (Mean)			46.390	13.660	26.690	79.490
Variance (Time)			0.338	0.144	0.138	0.694
Linear Growth Model – Time, SSI, and Time x SSI						
Variance (Mean)			48.080	20.840	21.910	101.100
Variance (Time)			0.595	0.333	0.204	1.460
Variance (SSI)			4.006	5.741	0.207	19.910
Variance (Time x SSI)			0.461	0.332	0.118	1.338

~ $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Figure 7.9. Posterior distribution of the variance: Unconditional model.

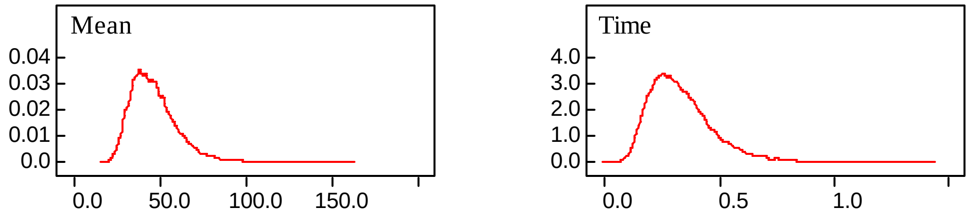


Figure 7.10. Posterior distribution of the variance: Unconditional model.

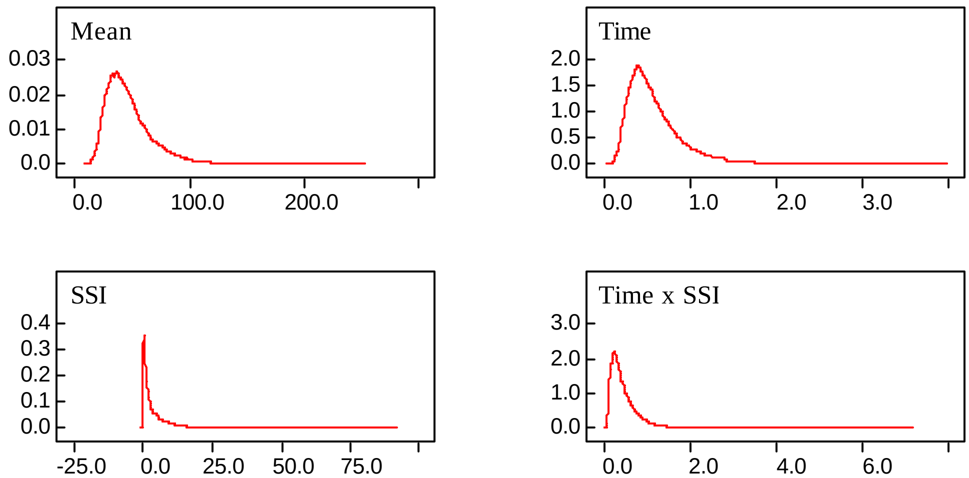


Figure 7.11. Posterior distributions of average scale scores of cohort—grade 4 in 1992 and grade 8 in 1996.

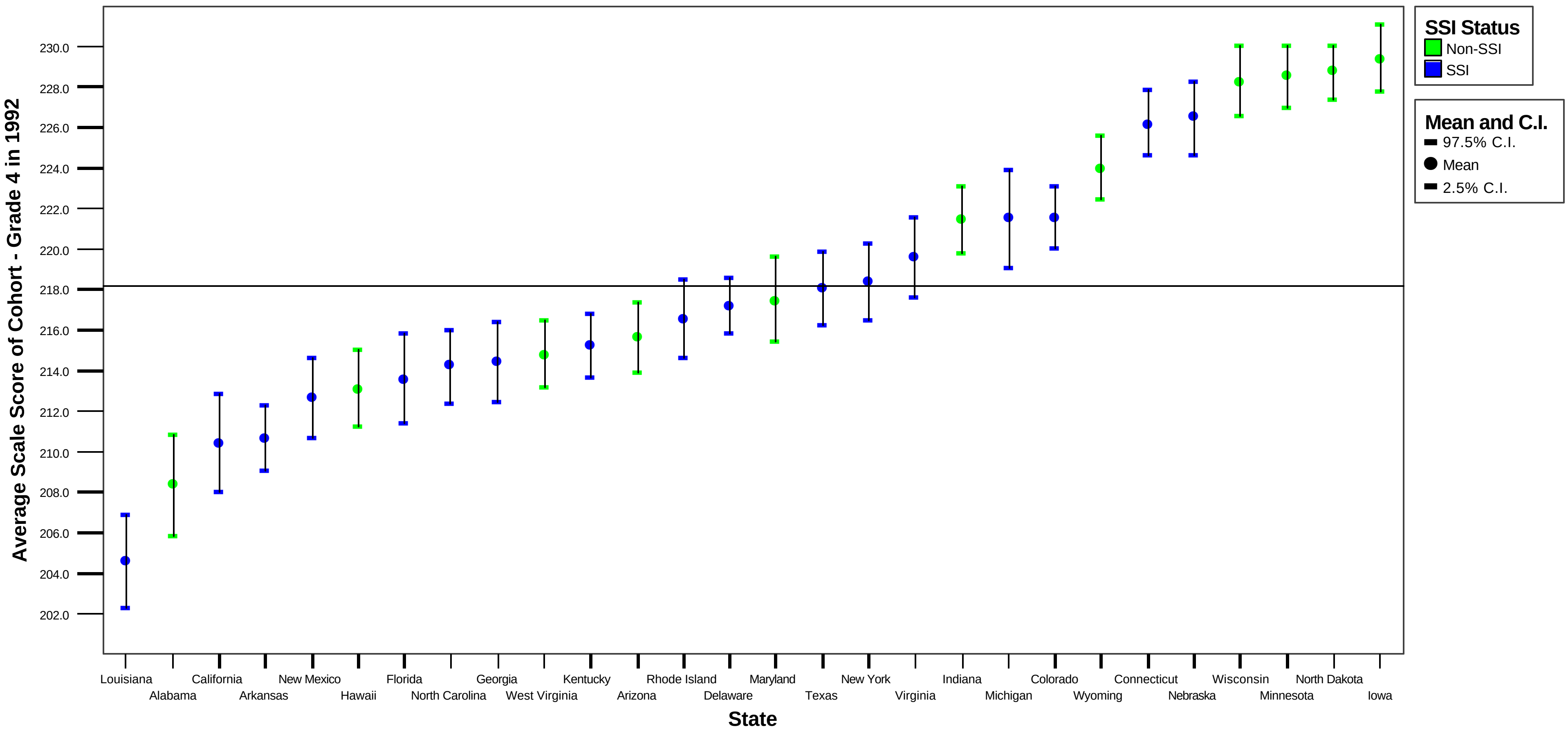
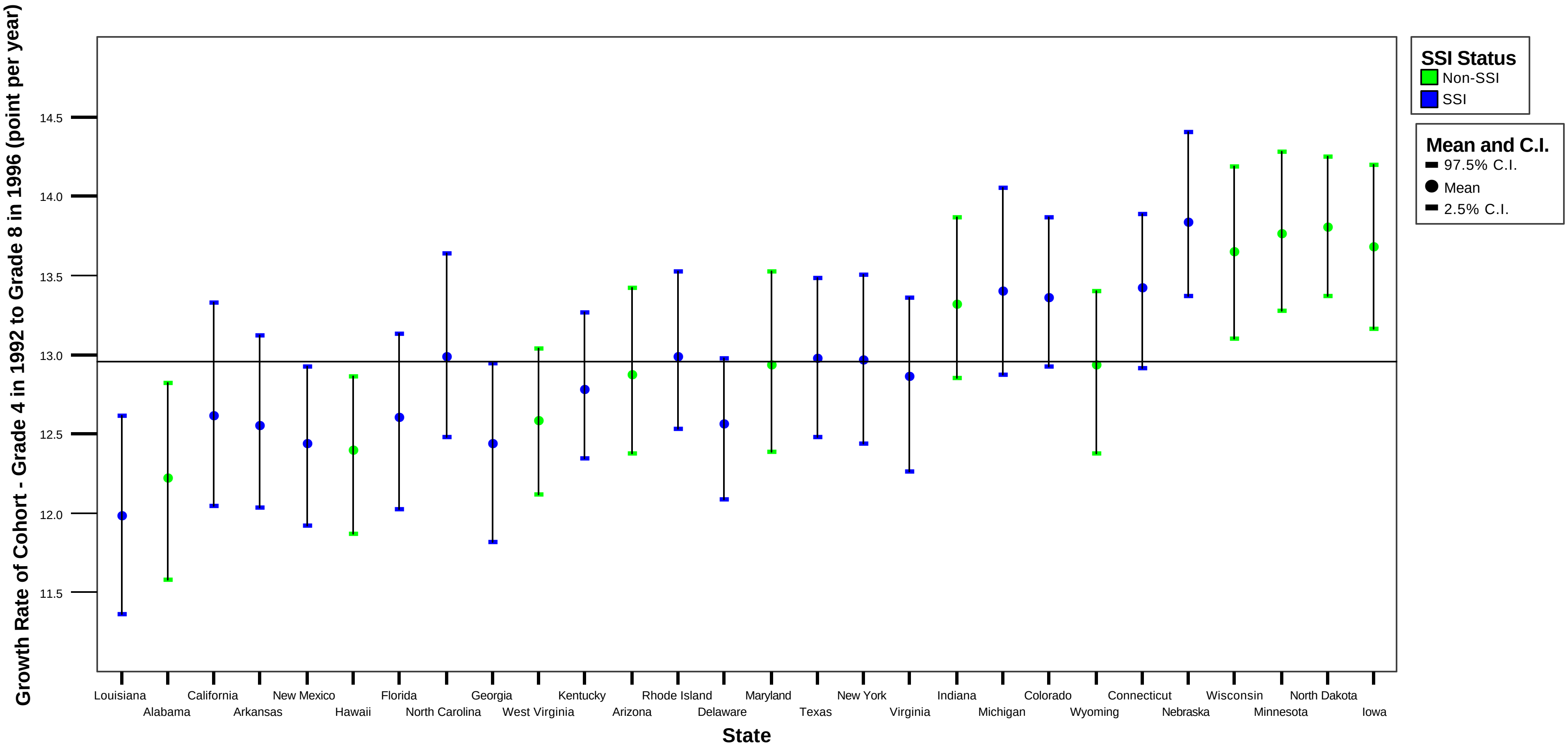


Figure 7.12. Posterior distributions of growth rates of cohort—grade 4 in 1992 and grade 8 in 1996.



Cross-Sectional Analysis: Empirical Bayes Method

Grade 8

Unconditional Model. Table 7.4 presents the results of the unconditional models, using the empirical Bayesian method. Over three assessment years in 1990, 1992, and 1996, average state means increased from 262.389 points in 1990 through 266.328 points in 1992 to 270.274 points in 1996. This was a gain of almost four points over each assessment year. Thus, there was evidence that grade 8 students across the states were likely to gain in mathematics scale scores from 1990 to 1996. The estimated variance of average state mean is also displayed in Table 7.4. Overall, all three estimates of the variance indicate that significant variability between states existed in each of the average state means. The estimated variance ranges from 73.535 to 81.466, which are substantial.

Conditional Model. In the conditional models, we tried to detect any differences between SSI and non-SSI states in average state mean over the period of 1990-1992-1996. The results show that differential effects of SSI states are marginally significant and negative, as large as -6.317 points and as small as -5.124 points. Overall, grade 8 students in non-SSI states outperformed counterparts in SSI states over all three assessment years. However, the gaps between SSI and non-SSI states narrowed by 1.2 points from 1990 to 1996. As the variance components in Table 7.4 indicate, SSI differential effects account for 6 to 8% of between-state variance in average state means. Much of the variability between states still remains to be explained.

Table 7.4
Cross-Sectional Analysis of Grade 8 Data: Empirical Bayes Estimates After Considering Jackknife Standard Errors

	1990		**1992**		**1996**	
Fixed Effect	*Coefficient*	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Unconditional Model						
Average state mean	262.389***	1.705	266.328***	1.722	270.274***	1.643
Conditional Model – SSI						
Non-SSI State						
Average Non-SSI state mean	266.224***	2.599	270.114***	2.629	273.387***	2.552
SSI State						
Average SSI state mean	259.907		263.873		268.263	
SSI Effect						
Average SSI differential effect	-6.317~	3.336	-6.241~	3.376	-5.124~	3.275
Random Effect	*Variance*	*SD*	*Variance*	*SD*	*Variance*	*SD*
Unconditional Model	80.133***	8.952	81.466***	9.026	73.535***	8.575
Conditional Model	73.055***	8.547	74.605***	8.637	69.617***	8.344

~ $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Grade 4

Unconditional Model. The results of grade 4 average state means in 1992 and 1996 are displayed in Table 7.5. As presented earlier in grade 8 results, grade 4 students also show continued increases in mathematics scores from 1992 and 1996. The empirical Bayes estimates of average state means were 218.231 points across states in 1992 data and 221.789 points across states in 1996 data. The overall gain of grade 4 students was about 3.6 points over the period of four test years. As the bottom part of Table 7.5 indicates, both estimated average state means varied significantly between states. The estimates for the variance of average state means in 1992 and 1996 were 44.830 and 48.223, respectively. This suggests that state means in 1996 were more heterogeneous than those in 1992.

Conditional Model. Table 7.5 presents the empirical Bayes estimated results of the conditional models, including SSI status, as a covariate. In general, the mathematics scores for grade 4 students in both SSI and non-SSI states increased substantially over the two assessment years of 1992 and 1996. In 1992, average state means were 221.196 points for non-SSI states and 216.304 (221.196 + (- 4.892)) points for SSI states. In 1996, the average non-SSI state mean was 224.097 points and average SSI state mean was 220.296 (224.097 + (-3.801)) points. Thus, the mathematics score gap between SSI and non-SSI states was reduced by 1.1 points from 1992 to 1996, while non-SSI states were more likely to score higher than SSI states. Table 7.5 also shows how much variation the SSI status indicator explains in average state means. The SSI differential effects explained 10% of between-state variance in 1992 and 4% of the variance in 1996.

Table 7.5
Cross-Sectional Analysis of Grade 4 Data: Empirical Bayes Estimates After Considering Jackknife Standard Errors

	1992		1996	
Fixed Effect	*Coefficient*	*SE*	*Coefficient*	*SE*
Unconditional Model				
Average state mean	218.231***	1.286	221.789***	1.335
Conditional Model – SSI				
Non-SSI State				
Average Non-SSI state mean	221.196***	1.944	224.097***	2.089
SSI State				
Average SSI state mean	216.304		220.296	
SSI Effect				
Average SSI differential effect	-4.892~	2.499	-3.801	2.682
Random Effect	*Variance*	*SD*	*Variance*	*SD*
Unconditional Model	44.830***	6.695	48.223***	6.944
Conditional Model	40.322***	6.350	46.387***	6.811

~ $p \le .1$, * $p \le .05$, ** $p \le .01$, *** $p \le .001$

Summary and Conclusions

In this chapter, we report longitudinal and cross-sectional analyses of the State NAEP data conducted on mathematics achievement for grade 8 students in 1990, 1992, and 1996, for grade 4 students in 1992 and 1996, and for cohort students at grade 4 in 1992 and at grade 8 in 1996. Using the empirical Bayes and fully Bayesian methods, we found that both SSI and non-SSI states showed an overall gain in mathematics scores across the assessment years. The results also revealed that a substantial variation in average state gain existed among states. Comparing the achievement growth between SSI and non-SSI states in average mathematics scale scores, the following summarizes the major findings in this chapter:

- In 1990, grade 8 students in SSI states scored lower by 5.72 points than in non-SSI states, but showed a faster annual growth 0.20 points, from 1990 to 1996, than those in non-SSI states. Among the 28 states, the growth in scores of three SSI states—North Carolina, Michigan, and Texas—was highest.
- Grade 4 students in SSI states started behind those in non-SSI states at 4.82 points in 1992, but learned more, 0.25 points per year, than their counterparts in non-SSI states. North Carolina and Texas made the highest gains.
- For the cohort students in grade 4 in 1992 and grade 8 in 1996, SSI states scored lower than non-SSI states in 1992 by 4.54 points and gained less at 0.02 points per year. The annual gain of Nebraska and North Dakota relative to other states was higher.
- While there was clear evidence of the variance in average growth rate across SSI states and non-SSI states, none of the growth estimates indicated any significant group differences between SSI and non-SSI states.
- In each year of grade 8 data in 1990, 1992, and 1996 and of grade 4 data in 1992 and 1996, the mathematics scores in SSI states were lower than those in non-SSI states. But, the gaps between SSI and non-SSI states were reduced gradually over the assessment years. Much of the between-state variance in average state means within each of the assessment data remains to be further explained.

Our findings regarding the effectiveness of SSI states in improving mathematics achievement over non-SSI states need to be interpreted with care in terms of data limitations. For longitudinal analyses of grade 4 and the cohort, our results were based on only two time points and therefore may not provide adequate data on the overall trends of growth for grade 4 students from 1992 and 1996 and for cohort students. State means used in this study as the input data for empirical Bayes and fully Bayesian analyses were also not adjusted for student socioeconomic and demographic backgrounds, school composition, and other variables reported to be associated with student achievement scores.

In the longitudinal and cross-sectional analyses, we did not attempt to determine which SSI-related factors contributed to achievement growth of the SSI states. In the 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.

List of Appendices

Appendix A
Table 7A.1
Summary of State Means and Jackknife Estimated Standard Errors for Grade 8 Data

Table 7A.2
Summary of State Means and Jackknife Estimated Standard Errors for Grade 4 Data

Appendix A

Table 7A.1
Summary of State Means and Jackknife Estimated Standard Errors for Grade 8 Data

State ID	State Name	1990			1992			1996		
		Stu N	Mean	SE	Stu N	Mean	SE	Stu N	Mean	SE
1	Alabama	2531	252.86	1.10	2522	252.19	1.70	2261	256.59	2.10
4	Arizona	2558	259.59	1.30	2617	265.37	1.30	2136	267.87	1.60
5	Arkansas	2669	256.21	0.90	2556	256.31	1.20	1845	261.65	1.50
6	California	2424	256.32	1.30	2516	260.89	1.70	2290	262.77	1.90
8	Colorado	2675	267.37	0.90	2799	272.40	1.00	2530	275.61	1.10
9	Connecticut	2672	269.87	1.00	2613	273.74	1.10	2485	279.59	1.10
10	Delaware	2110	260.70	0.90	1934	262.87	1.00	1798	266.73	0.90
12	Florida	2534	255.32	1.20	2549	259.91	1.50	2401	263.64	1.80
13	Georgia	2766	258.85	1.30	2589	259.36	1.20	2364	262.47	1.60
15	Hawaii	2551	251.02	0.80	2454	257.41	0.90	2189	262.13	1.00
18	Indiana	2569	267.27	1.20	2659	270.10	1.10	2347	275.53	1.40
19	Iowa	2474	277.97	1.10	2816	283.36	1.00	2169	283.99	1.30
21	Kentucky	2680	257.10	1.20	2756	262.24	1.10	2461	266.59	1.10
22	Louisiana	2572	246.44	1.20	2582	249.98	1.70	2599	252.38	1.60
24	Maryland	2794	260.77	1.40	2399	264.83	1.30	2137	269.68	2.10
26	Michigan	2587	264.40	1.20	2616	267.35	1.40	2155	276.87	1.80
27	Minnesota	2584	275.39	0.90	2471	282.39	1.00	2425	284.05	1.30
31	Nebraska	2519	275.67	1.00	2285	277.65	1.10	2610	282.77	1.00
35	New Mexico	2643	256.42	0.70	2561	259.61	0.90	2371	261.97	1.20
36	New York	2302	260.80	1.40	2158	266.42	2.10	1962	270.23	1.70
37	North Carolina	2843	250.35	1.10	2769	258.41	1.20	2638	267.83	1.40
38	North Dakota	2485	281.10	1.20	2314	283.21	1.10	2602	284.22	0.90
44	Rhode Island	2675	260.04	0.60	2120	265.91	0.70	2055	268.88	0.90
48	Texas	2542	258.19	1.40	2614	264.59	1.30	2245	270.20	1.40
51	Virginia	2661	264.27	1.50	2710	267.86	1.20	2545	269.75	1.60
54	West Virginia	2600	255.90	1.00	2690	259.09	1.00	2578	264.87	1.00
55	Wisconsin	2750	274.49	1.30	2814	277.88	1.50	2165	282.85	1.50
56	Wyoming	2701	272.15	0.70	2444	275.08	0.90	2696	274.78	0.90

Table 7A.2
Summary of State Means and Jackknife Estimated Standard Errors for Grade 4 Data

State ID	State Name	1990			1992			1996		
		Stu N	Mean	SE	Stu N	Mean	SE	Stu N	Mean	SE
1	Alabama				2605	208.33	1.60	2541	211.65	1.20
4	Arizona				2741	215.25	1.10	2113	217.58	1.70
5	Arkansas				2621	210.21	0.90	2047	215.85	1.50
6	California				2412	208.40	1.60	2063	209.13	1.80
8	Colorado				2906	221.02	1.00	2609	225.81	1.00
9	Connecticut				2600	226.80	1.10	2565	232.03	1.10
10	Delaware				2040	217.90	0.80	1984	215.03	0.60
12	Florida				2828	213.69	1.50	2549	215.76	1.20
13	Georgia				2766	215.59	1.20	2542	215.46	1.50
15	Hawaii				2625	214.06	1.30	2578	214.97	1.50
18	Indiana				2593	221.04	1.00	2470	229.39	1.00
19	Iowa				2770	229.88	1.00	2359	229.13	1.10
21	Kentucky				2703	215.05	1.00	2579	219.99	1.10
22	Louisiana				2792	204.14	1.50	2671	209.02	1.10
24	Maryland				2844	217.32	1.30	2465	220.69	1.60
26	Michigan				2412	219.88	1.70	2382	226.26	1.30
27	Minnesota				2640	228.49	0.90	2425	232.19	1.10
31	Nebraska				2327	225.33	1.20	2678	227.54	1.20
35	New Mexico				2342	213.30	1.40	2389	213.84	1.80
36	New York				2284	218.45	1.20	2248	222.63	1.20
37	North Carolina				2884	212.88	1.10	2658	224.33	1.20
38	North Dakota				2193	228.66	0.80	2666	230.90	1.20
44	Rhode Island				2390	215.45	1.50	2461	220.42	1.40
48	Texas				2623	217.92	1.20	2413	228.71	1.40
51	Virginia				2786	220.76	1.30	2586	222.64	1.40
54	West Virginia				2786	215.27	1.10	2530	223.35	1.00
55	Wisconsin				2780	228.69	1.10	2437	231.41	1.00
56	Wyoming				2605	225.38	0.90	2758	223.20	1.40

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

In this interim technical report on our study of the impact of the NSF's Statewide Systemic Initiatives program, we have presented our approach to using State NAEP data as a basis for our impact study. At this stage, we have no conclusive findings, but we have identified promising directions to pursue toward the conclusion of the study. In trying to isolate or discriminate among factors and characteristics of SSI states in the NAEP data that could be attributable to the SSI program, we have had to make a number of choices and assumptions. The purpose of this technical report is to describe these choices and assumptions in some detail and the data that can be used to compare SSI states with non-SSI states.

One major issue we have faced is the change in the number of states that participated in the State NAEP in 1990, 1992, and 1996. What we have had to do is to use different groups of states for different analyses. For example, we can report differences between SSI and non-SSI states in 1996 using the largest group of states that participated in the State NAEP, 22 SSI states (88% of 25 possible states) and 19 non-SSI states (76% of 25 possible states). However, to report trends in student performance, change in classroom indicators, and disaggregated data by ethnicity, we have had to resort to using smaller groupings of states. Another major issue we have faced in analyzing NAEP data is the use of NAEP supplied weights to compute the results for a participating state and to aggregate across several states. In comparing the aggregated results of SSI states with non-SSI states, we have assumed that each state is a replication of the other states in the grouping. Thus we have weighted states equally by taking the mean of all the states in a group to produce a mean for the group. In this way, for example, California has been given the same weighting as Vermont.

Not all states that participated in the State NAEP achieved the 90 percent participation rate of sampled schools as indicated by NAEP's sample selection plan. Not reaching the desired participation rate in the sample increased the likelihood of bias in results for a state, though the extent and direction of the bias cannot be determined from the NAEP data. To address this and the other issues, we have conducted the analyses more than once using different groupings to determine the stability of results.

Demographics

SSI states ($N = 22$) that participated in the State NAEP in 1996 had a higher percentage of Black and Hispanic students than non-SSI states ($N = 18$). The 22 SSI states had about 5% more of the population who were Black students and about 4% more of the population who were Hispanic students in grades 4 and 8 than did the 18 non-SSI states. From this we concluded that the SSI program tended to include those states with a higher proportion of minorities. In the aggregate, SSI states and non-SSI states tested the same percentage of male and female students in the State NAEP. However, four of 17

SSI states compared to only one of the 11 non-SSI states in the 1990-96 trend group tested 4% more female students than male students at grade 8 in 1996. We can only hypothesize about the reasons for this difference between the percentage of female and male students tested in these states. Perhaps male students were not as available on the testing day or, more likely, chose not to be tested.

SSI states and non-SSI states differed very little on the demographic variables related to socio-economic status, such as students' reports of parents' education and enrichment of home environment. The accuracy of student report data such as the education of parents can be questioned, but there is nothing that causes us to suspect that the data for the SSI states would have different sources of variance than the data for the non-SSI states. Thus, it is reasonable to compare the two groups of states on these variables.

A General Comparison of Mathematics Achievement in SSI and Non-SSI States

Student performance on the State NAEP improved for both SSI and non-SSI states from 1990 to 1996 at grade 8 and from 1992 to 1996 at grade 4. At grade 8, prior to joining the SI program, students from the SSI states performed about six points lower on the composite scale than did those from the non-SSI states. At grade 4 between 1992 and 1996, student performance in SSI states also was below student performance in non-SSI states. In 1992, the average composite score for SSI states was five points below that for the SSI states. However, over the duration of the SI program up to 1996, SSI states improved at a slightly faster rate than did the non-SSI states for both grades 4 and 8. This composite of SSI scale scores for both grades of about one point between SSI and non-SSI states is evident in both the mean composite scale scores and the Bayesian analysis. In this technical report, we have described differences between the two groups without taking into consideration the interaction of student performance with demographic and other variables. We will report the relational data in future technical reports.

Findings by Gender

When gender and ethnicity were considered, some interesting differences between SSI states and non-SSI states existed.. Female students posted lower achievement scores than male students in both SSI states and non-SSI states. However, at grade 8 the gap was eliminated in non-SSI states, but remained over two points in the SSI states. Male students in SSI states maintained their advantage in part by achieving higher scores than females on the geometry and data analysis scales.

Findings by Ethnicity

The achievement gap between White and Black students in both SSI and non-SSI states ranged from 30 to 34 points over the three testing times for grade 8 and from 28 to 32 points over the two testing times for grade 4. At grade 8 in the SSI states, the achievement gap between grade 8 White and Black students on the composite scale remained the same over the three testing times, but the extent of the gap varied by subtopic scale. The gap increased for measurement, but narrowed slightly on both the geometry scale and the algebra and functions scale. In contrast, the achievement gap for the six non-SSI states included in the analysis increased on the composite scale and the five subtopic scales. At grade 4, between 1992 and 1996, a similar pattern was observed. The White-Black gap decreased on all six scales for SSI states, but increased on five of the six scales for non-SSI states. Comparing scores of White students with Hispanic students, the gaps were smaller than the gaps between White and Black students. However, both SSI and non-SSI states generally had a decrease in the gaps on all of the scales at grade 8, but generally had an increase in the gaps on all of the scales at grade 4.

Findings for the Same Cohort from 1992 to 1996

There was some evidence that Black students in SSI states gained more between grade 4 and grade 8 than Black students in non-SSI states. When the scale score of grade 4 students in 1992 was compared to the scale score of grade 8 students in 1996, Black students in SSI states gained about three points more than Black students in non-SSI states. This was not statistically significant. But to provide some comparison, the gain by White students in both groups over the same four years was nearly identical. Of particular note, Black students in SSI states gained more on the algebra and functions scale between grade 4 and grade 8 than did White students in SSI states. In non-SSI states, White students gained more than Black students on all six scales. Again, while these results are not statistically significant, they could be the beginning of a trend. These results will be considered more closely in future analyses, including identifying other related classroom practice variables. Hispanic students in SSI states gained more than Hispanic students in non-SSI states over the four years on five of the six scales. Non-SSI Hispanic students only gained more on the measurement scale.

Mathematics Curriculum Reform Indicators

Tracking change in classroom practices is as important as tracking change in student performance in the early years of the Statewide Systemic Initiatives Program. A significant amount of the states' funding for the systemic initiatives went toward teacher professional development with the intent of changing teachers' classrooms practices to be more aligned with the prevalent reforms for mathematics education. Evidence of changed classroom practices that coincides with the current understanding of how students learn mathematics is a positive finding that is consistent with NSF's emphases on making mathematics education more challenging for all students. By 1996, over one-third of the middle school teachers and 15% of the grade 4 teachers in the 26 SSIs had participated in professional development experiences directly funded through the SSIs. About one half of the SSIs had been identified as having had a strong, positive impact on teachers' classroom practices by incorporating more inquiry-based learning, including greater use of hands-on work, greater attention to student inquiry, and greater use of small-group work (Zucker, Shields, Adelman, Corcoran, & Goertz, 1998).

Our analyses of the State NAEP in 1990, 1992, and 1996 allow us to verify findings from other evaluations and to detect statewide changes consistent with NSF's SI 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. 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. This will require that a state-by-state analysis be done, which will be our next step.

In the analyses reported in this technical report, we have tried to identify important variables that can be used to describe changes in instruction coinciding with the duration of the SSIs. Using NAEP teacher questionnaires, we developed six indicators of mathematics reform:

Relative Emphasis on Reasoning and Communication
Students' Opportunities for Mathematical Discourse
Teachers' Knowledge of the NCTM *Standards*
Last Year's Professional Development,
Reform-Related Topics Studied
Calculator Use

The indicators relate to practices advanced as reforms in mathematics in the 1990s. The NAEP data on these indicators allows inferences about the type of classroom experiences students had in a state. The results cannot be interpreted as representative of teachers' practices in a state because teachers were not the sampling unit. Longitudinal comparisons were limited because NAEP varied some of the questions included on teacher and student questionnaires for each testing time.

In general, both SSI and non-SSI states increased on the six indicators of mathematics reform, which corresponds to the change in student achievement. At grade 8 in 1996, SSI states, as a group, scored significantly higher than non-SSI states on five of the six indicators. At grade 4, the SSI states scored significantly higher on four of the six indicators compared to non-SSI states. At both grade levels, there was no significant difference between SSI and non-SSI status in students' use of calculators. In 1990 and 1992, prior to any state's substantial involvement in the SSI program, the SSI and non-SSI states did not differ significantly on these indicators. Thus, this evidence from the cross-sectional analyses of the State NAEP suggests that students in the SSI states in 1996 were more likely to experience instruction that utilized principles of reform mathematics. On longitudinal analyses, we found that from 1992 to 1996 SSI states increased more than non-SSI states on the mathematical discourse and reasoning and communication indicators. At grade 4, SSI states increased more than non-SSI states on the amount of staff development and the number of reform-related topics studied, as well as the relative emphasis on reasoning and communication and calculator use.

Within the group of SSI states and within the group of non-SSI states, there was considerable variation. Some SSI states were among the highest scoring states on each of the indicators and some were among the lowest scoring. In the next phase of the study, we will develop individual profiles for each SSI state so that we will be better able to relate the individual state's SSI emphases to changes in student performance and the indicators. For three selected states, we also will use state assessment data in concert with State NAEP data to determine whether there are consistent patterns on student performance.

Summary

Four years after the first state received funding from the National Science Foundation for a statewide systemic initiatives, State NAEP data on mathematics revealed some differences in performance between states with a SSI and those without. Whereas, student performance in SSI states remained below student performance scores in non-SSI states, the performance in SSI states increased at a slightly higher rate reducing the difference in score by one point. Black students from SSI states made noticeable gains in performance compared to Black students from non-SSI states. Indicators on instructional practices and professional development suggested on the average a greater proportion of students in SSI states were taught using reform practices compared to those in non-SSI states. 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 pointing to the need to consider individual states as will be done in further analyses. The small gains by SSI states between 1990 and 1996 reported here could be the beginning of a trend that can only be determined by analyzing the 2000 State NAEP.

REFERENCES

Adler, J. (1999). The dilemma of transparency: Seeing and seeing through talk in the mathematics classroom. *Journal for Research in Mathematics Education 30*(1), 47-64.

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.

American Educational Research Association. American Psychological Association. National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Authors.

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 Service, Policy Information Center..

Black, P., & Atkin, J. M. (1996). *Changing the subject: Innovations in science, mathematics, and technology education.* London: Routledge.

Bradley, E. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brolin, H., & Björk, L-E. (1992). Introducing calculators in Swedish schools. In J. T. Fey & C. R. Hirsch (Eds.), *Calculators in mathematics education* (1992 Yearbook, pp. 226-232). Reston, VA: National Council of Teachers of Mathematics.

Bryk, A, S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. A. Romberg (Eds.), *Classrooms that promote mathematical understanding* (pp. 19-32). Mahwah, NJ: Erlbaum.

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.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Corcoran, T. B., 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.

Cronbach, L. J. (1971). Test validation. In Robert L. Thorndike (Ed.), *Educational measurement* (2nd edition) (pp. 51-84). Washington, DC: American Council on Education.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrica, 16,* 297-334.

Croom, L. (1997). Mathematics for all students: Assess, excellence, and equity. In J. Trentacosta & M. J. Kenney (Eds.), *Multicultural and gender equity in the mathematics classroom: The gift of diversity. 1997 Yearbook*. Reston, VA: National Council of Teachers of Mathematics.

Dunham, P. H., & Dick, T. P. (1994). Research on graphing calculators. *Mathematics Teacher 87,* 440-45.

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (Eds.). (1999). *Uncommon measures: Equivalence and linkage among educational tests.* Washington, DC: National Academy Press, National Research Council.

Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L., Herman, R., & Yoon, K. S. (1999). *Designing effective professional development: Lessons from the Eisenhower Program (Report).* Washington, DC: U. S. Department of Education.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

Gravemeijer, K. (1994). Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education, 25*(5), 443-471.

Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What NAEP state test scores tell us.* Santa Monica, CA: RAND, MR-924-EDU.

Groves, S. (1994). The effect of calculator use on third and fourth graders' computation and choice of calculating device. In J. da Ponte & J. F. Matos (Eds.), *Proceedings of the eighteenth international conference for the Psychology of Mathematics Education* (Vol. 3, pp. 33-40). Lisbon: Psychology of Mathematics Education.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(41). Available at http://epaa.asu.edu/epaa/v8n41.

Hembree, R., & Dessart, D. J. (1992). Research on calculators in mathematics education. In J. T. Fey & C. R. Hirsch (Eds.), *Calculators in mathematics education* (1992 Yearbook, pp. 23-32). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J. (1999). Relationships between research and the NCTM *Standards. Journal for Research in Mathematics Education, 30*(1), 3-19.

Horizon Research, Inc. (1995, October). *Ohio fifth-year evaluation report outline* [Draft]. Chapel Hill, NC: Author.

Horvay, N. (Ed.). (1999). *The NAEP guide: A description of the contents and methods of the 1999 and 2000 assessments.* Revised edition. Washington, DC: National Center for Education Statistics.

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.

John Glenn Commission. (2000). *Before it's too late.* (http://www.ed.gov/inits/Math/glenn/)

Kenney, P. A., & Silver, E. A. (1998). *Content analysis project—State and NAEP mathematics assessments. North Carolina-NAEP Study: Final Report*. Washington, DC: National Assessment Governing Board.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M., Robin, A., & Burroughs, D. (2000a). *Teaching practices and student achievement: Report of first-year findings from the 'Mosaic' study of systemic initiatives in mathematics and science.* Santa Monica, CA: RAND.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000b). *What do test scores in Texas tell us?* (Issue Paper). Santa Monica, CA: RAND.

Laguarda, K. G., Goldstein, D. S., Adelman, N. E., & Zucker, A. A. (1998). *Assessing the SSIs' impacts on student achievement: An imperfect science.* Menlo Park, CA: SRI International.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.

Linn, R. L., Koretz, D. M., Baker, E. L., & Burstein, L. (1991). *The validity and credibility of the achievement levels for the 1990 National Assessment of Educational Progress in Mathematics.* Los Angeles, CA: Center for Research on Evaluation Statistics.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*(2), 129-151.

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics.* Thousand Oaks, CA: Corwin Press, Inc.

Loucks-Horsley, S., Stiles, K., & Hewson, P. (1996). *Principles of effective professional development for mathematics and science education: A synthesis of standards.* Madison: University of Wisconsin, National Institute for Science Education.

Loucks-Horsley, S., Harding, C. K., Arbuckle, M. A., Murray, L. B., Dubea, C., & Williams, M. K. 91987). *Continuing to learn: A guidebook for teacher development.* Andover, MA/Oxford, OH: The Regional Laboratory for Educational Improvement of the Northeast and Islands/National Staff Development Council.

Loveless, T., & Diperna, P. (2000). *The Brown Center report on American education: How well are American students learning? Focus on math achievement 1*(1). Washington, DC: The Brookings Institution.

McLeod, D. B., Stake, R. E., Schappelle, B. P., Mellissinos, M., & Gierl, M. J. (1997). 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.

Mehrens, W. A. (2000). Defending a state graduation test: GI Forum et al. v. Texas Education Agency, et al: Measurement perspectives from an external evaluator. *Applied Measurement in Education, 13*(4), 387-401.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1993). *NAEP 1992 mathematics report card for the nation and the states. The Nation's Report Card.* (Report No. 23-ST02). Washington, DC: The National Center for Education Statistics.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000, September 27). *Before it's too late: A report to the nation.* Washington, DC: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics in the 1980s.* Reston, VA: Author.

National Research Council. (2000). *How people learn*: *Brain, mind, experience, and school.* In J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Washington, DC: National Academy Press.

National Science Foundation. (1997). *Statewide Systemic Initiatives.* Washington, DC: Author.

Newmann, F. M., Secada, W. G., & Wehlege, G. G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring.* Madison: Wisconsin Center for Education Research, University of Wisconsin.

North Central Regional Educational Laboratory. (1996). *State student assessment programs database.* Oakbrook, IL: Author.

Odden, A. (1991). *Education policy implementation.* Albany, NY: State University of New York Press.

Phelps, R. P., Cullen, A., Easton, J. C., & Best, C. M. (1997). *State indicators in education 1997.* Washington, DC: U.S. Department of Education.

Phillips, M. (2000). Understanding ethnic differences in academic achievement: Empirical lessons from national data. In D. W. Grissmer & J. M. Ross (Eds.), *Analytic issues in the assessment of student achievement* (pp . 103-132). Washington, DC: U.S. Department of Education, RAND.

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 229-272). Washington, DC: Brookings Institution Press.

Raudenbush, S. W. (2000). Synthesizing results from the NAEP Trial State Assessment. In D. W. Grissmer & M. Ross (Eds.), *Analytic issues in the assessment of student achievement* (pp. 3-42). Washington, DC: U.S. Department of Education, RAND.

Raudenbush, S. W., Fotiu, R. P., & Cheong, Y. F. (1998). Inequality of access to educational resources: A national report card for eighth-grade math. *Educational Evaluation and Policy Analysis, 20*(4), 253-267.

Raudenbush, S. W., Fotiu, R. P., & Cheong, Y. F. (1999). Synthesizing results from the Trial State Assessment. *Journal of Educational and Behavioral Statistics, 24*(4), 413-438.

Romberg, T. A., & Webb, N. L. (1993). The role of the National Council of Teachers of Mathematics in the current reform movement in school mathematics in the United States of America. In *Science and mathematics education in the United States: Eight innovations* (pp. 143-182). Paris: Organization for Economic Co-operation and Development.

Shafer, M. C. (2001, April). *Instructional quality in the context of reform.* Paper presented at the Annual Meeting of the American Educational Research Association. Seattle, WA.

Shaughnessy, C. A., Nelson, J. E., & Norris, N. A. (1997). *NAEP 1996 mathematics cross-state data compendium for the grade 4 and grade 8 assessment.* Washington, DC: National Center for Education Statistics.

Shields, P. M., Marsh, J., & Adelman, N. (1998). *Evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program: The SSI's 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.

Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). Bristol, PA: Falmer.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. London: Sage.

Sparks, D. (1994, March 16). A paradigm shift in staff development. *Education Week, 42*.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan, 79*(1), 14-21.

St. John, M. (1999). Understanding the value of NSF's investments in systemic reform. In Norman 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.

Texas Education Agency, National Computer Systems, Harcourt Educational Measurement, Measurement Incorporated, and BETA, In. (1999). *Texas Student Assessment Program Technical Digest for the academic year 1998-1999.* Austin, TX: Authors.

The College Board. (1996). *Mathematics framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Thorndike, R. L. (Ed.). (1971). *Educational measurement.* Washington, DC: American Council on Education.

U.S. Office of Education. (1972). *History of Title I, ESEA.* (Rev. ed.). Washington, DC: U.S. Department of Health, Education, and Welfare Office of Education: For sale by the Supt. of Docs. U.S. G. P. O.

Webb, N. L., Heck, D. J., & Tate, W. F. (1996). The urban mathematics collaborative project: A study of teacher, community, and reform. In S. A. Raizen E. D. Britton (Eds.), *Bold ventures: Case studies of U. S. innovations in mathematics education: Vol. 3* (pp. 246-360). Boston: Kluwer Academic Press.

Weiss, I. R. (1994). *A profile of science and mathematics education in the United States: 1993.* Chapel Hill, NC: Horizon Research, Inc.

Weiss, I. R., Matti, M. C., & Smith, P. S. (1994). *Report of the 1993 national survey of science and mathematics education.* Chapel Hill, NC: Horizon Research.

Wenglinski, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.

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 Initiative (SSI) Program. Washington, D. C.: National Science Foundation.

Wilson, L. D., & Blank, R. K. (1999). *Improving mathematics education using results from NAEP and TIMSS.* Washington, DC: Council of Chief State School Officers, State Education Assessment Center.

Wood, T. (1996). Events in learning mathematics: Insights from research in classrooms. *Educational Studies in Mathematics, 30*(1), 85-105.

Yackel, E., Cobb, P., & Wood, T. (1991). Small group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education, 22*(5), 390-408.

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.