



VARC

**VALUE-ADDED RESEARCH CENTER
UNIVERSITY OF WISCONSIN-MADISON**

Education analytics to support students and educators

SAGE Program Evaluation Final Report

August 2015

Evaluation Team

Dr. Robert Meyer (Principal Investigator)

Dr. Emin Dokumaci (Economist)

Grant Sim (Project Lead)

Clarissa Steele (Researcher)

Katerina Suchor (Programmer)

Jenn Vadas (Economist)

Report Authorship

Dr. Emin Dokumaci

Grant Sim

Clarissa Steele

Katerina Suchor

Jenn Vadas

For questions about this report or the evaluation, please contact Grant Sim at grant.sim@wisc.edu.

Executive Summary

The Student Achievement Guarantee in Education (SAGE) program is an initiative of the Wisconsin Department of Public Instruction (DPI). It spans kindergarten to third grade and provides funds to participating Wisconsin schools based on the number of free or reduced price lunch students. To receive the extra funding, classes must have fewer than 18 students when only one teacher is present and fewer than 15 students per teacher when more than one teacher is present.

This report presents the evaluation completed by the Value-Added Research Center (VARC) to measure the impact of the SAGE program on student outcomes. The evaluation includes three separate statistical analyses:

- Analyses of student academic growth comparing students in SAGE schools to students in non-SAGE schools;
- Analyses of student academic growth comparing students in SAGE schools to students in non-SAGE schools over the course of multiple years; and
- Analyses of long-term high school outcomes comparing students from SAGE schools to students from non-SAGE schools.

When comparing characteristics of students in SAGE versus non-SAGE schools, the Value-Added Research Center (VARC) observed large differences in their respective demographic profiles. The selection process into the program explains these differences and precludes simple comparisons across the two groups. Thus, VARC used statistical methods to control for these differences with the goal of estimating the impact of the SAGE program on student growth in mathematics and reading. Results from the statistical analyses of growth yield:

- An estimated positive effect of the SAGE program on reading academic growth in kindergarten as compared to students in non-SAGE schools, and
- An estimated positive effect of the SAGE program on mathematics and reading academic growth from kindergarten through third grade as compared to students in non-SAGE schools.

New to this year's evaluation is an analysis of the effect of SAGE on students' high school completion outcomes and on their choice of dropping out of high school. The results show:

- A positive effect of the SAGE program on students choosing to stay in high school in both ninth and tenth grade, and
- A positive effect of the SAGE program on students completing high school and not dropping out for students from African-American and economically disadvantaged backgrounds.

Finally, this year's SAGE evaluation also included results from the SAGE End-of-Year Report survey to examine non-academic outcomes. The results from the survey show increased flexibility with classroom design, benefits for students unrelated to standardized testing such as having more individual time to work with students, and benefits for teachers in recruitment and retention.

Table of Contents

Executive Summary ii

I. Introduction..... 1

II. Research Questions..... 2

III. Overview of Demographics and Assessment 3

 Characteristics of SAGE Schools and Students 3

 MAP Assessment and Characteristics of MAP Students..... 6

 Characteristic Comparison of SAGE and Non-SAGE Students 10

IV. Methodology 12

 Single-Year MAP Design Model 12

 Multiple-Year MAP Design Model 14

 Long-Term Outcomes Design Model 15

V. Evaluation Results 18

 Single-Year SAGE Effects 18

 Multiple-Year SAGE Effects 19

 Long-Term SAGE Effects..... 20

V. End-of-Year Survey Results 23

 SAGE Implementation..... 23

 Teacher Perceptions of SAGE..... 25

 SAGE Funding and Resources 27

 Non-Achievement Outcomes..... 29

 Classroom Design and Instructional Strategies 31

 Benchmark Assessments 32

VI. Conclusions..... 33

 Summary of Results 33

 Evaluation Context..... 34

Appendix: Technical Specifications 36

I. Introduction

This report presents the tenth year of results of the Student Achievement Guarantee in Education (SAGE) program evaluation by the Value-Added Research Center (VARC) of the University of Wisconsin's Wisconsin Center for Education Research. The SAGE program is an initiative of the Wisconsin Department of Public Instruction (DPI) launched in the fall of 1996. It allocates funds to participating Wisconsin schools with large proportions of students eligible for free or reduced price lunch in exchange for a commitment to maintain small class sizes from kindergarten through third grade. DPI requires SAGE schools to have student-to-teacher ratios of either 18:1, 30:2, or 45:3. Small class size advocates assert that small classes allow the necessary individualized instruction to promote the educational development of students in early grades.

Thinking thoroughly about the limitations in data availability, the program design, and the statistical methodology leads VARC year after year to refine and improve the statistical methods and data collection used to estimate the impact of the SAGE program. This year's evaluation built upon previous methodology as well as additional analyses to determine the SAGE program's impact. For the year 10 evaluation, VARC continued to use a propensity score matching approach for evaluating the impact of SAGE on student growth as measured through Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) assessment. Additionally, the evaluation continues to study the impact of SAGE over the course of multiple years through MAP assessment results.

This year's evaluation was also able to add an analysis of long-term outcomes at the high school level. Schools joined the SAGE programs at different time. The first two cohorts of SAGE had too few schools for any reliable estimates of impact, but the third cohort of SAGE schools starting in fall of 2000 had around 500 schools. The students in the latter schools have reached high school completion age, which allowed VARC to analyze their high school educational outcomes. In particular, VARC examined the impact of SAGE on reducing dropout rates.

Along with the quantitative analysis, the evaluation of the SAGE program in year 10 also included an analysis of the results of the SAGE End-of-Year Report survey that DPI distributed to all SAGE schools at the end of the 2013-14 school year. VARC and DPI designed these survey questions to learn more about possible SAGE benefits outside of those investigated in the quantitative analysis.

This report has six main sections aside from the introduction. The second section poses the main research questions. The third section provides a profile of SAGE students' demographics, a profile of schools using the MAP assessment, and demographic comparisons between SAGE and non-SAGE schools. The fourth section details the methodology of the evaluation including the analysis design and statistical models. The fifth and sixth sections presents the results of the evaluation and of the SAGE End-of-Year Report survey, respectively. Finally, in order to provide decision makers with a full understanding of the meaning of the estimation results, the seventh section includes a summary of findings and a list of data restrictions.

II. Research Questions

The year 10 analysis of the SAGE program evaluated the 2013-14 school year and a combination of the 2010-11, 2011-12, 2012-13, and 2013-14 school years to determine both the single-year impact and the multiple-year impact of the SAGE program on student performance in mathematics and reading.

While SAGE aims to improve learning outcomes in early grades, another important question is whether it helps students have better long-term outcomes. As a start, VARC evaluated high school outcomes of the students who were in SAGE schools in early grades by examining dropout and high school completion.

Lastly, VARC examined the results from an End-of-Year Report survey of SAGE schools.

To provide a foundation for the quantitative and qualitative analyses, this evaluation had seven research questions:

1. What are the characteristics of SAGE students and schools?
2. How does the set of schools using the MAP compare to the whole population of SAGE schools?
3. How does the set of SAGE schools compare to the set of non-SAGE control schools?
4. What is the difference between MAP growth in SAGE schools and non-SAGE schools for the most recent year?
5. What is the difference between MAP growth in SAGE schools and non-SAGE schools over multiple years?
6. What is the difference in dropout and high school completion outcomes between SAGE schools and non-SAGE schools?
7. What are the non-achievement outcomes associated with SAGE?

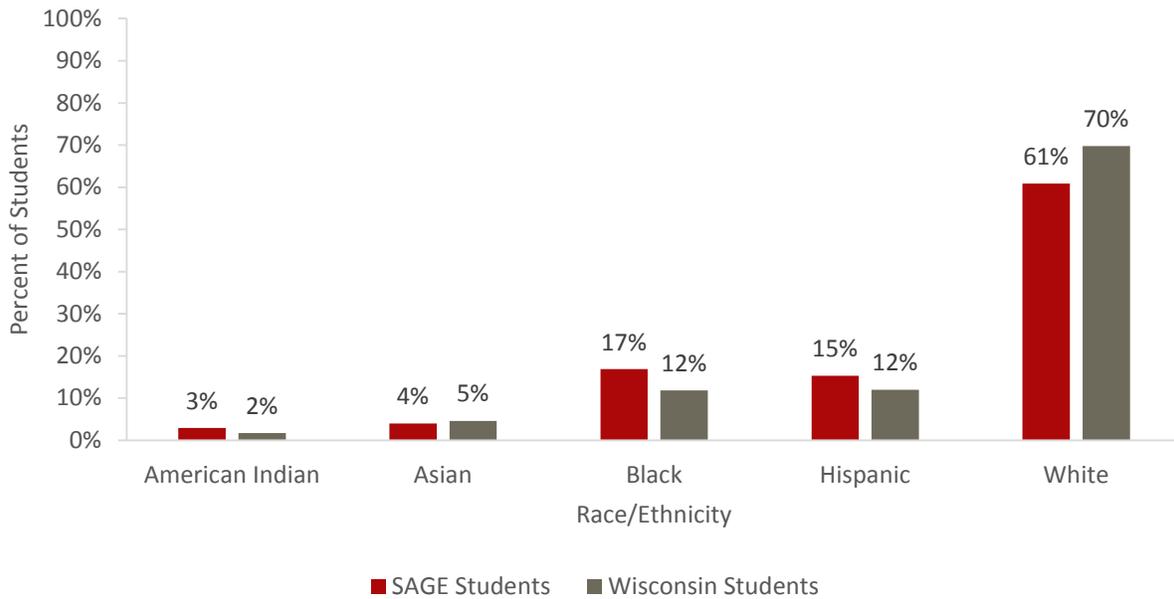
Research questions 1-3 provide the background necessary to design an appropriate method for MAP assessment analyses. For the evaluation results to be generalizable, the set of SAGE schools VARC analyzes should appear similar to the entire population of SAGE schools. Additionally, the analysis will have more power if a large percentage of both SAGE and non-SAGE schools use the MAP assessment. Because it is unlikely that SAGE schools and non-SAGE schools appear identical in their characteristics, this evaluation cannot make a direct comparison and, thus, uses statistical methods to create a control group and design a model to compensate for these differences. With this model, the evaluation can then attempt to answer questions 4-6 related to the impact of the SAGE program. Finally, results of the SAGE End-of-Year Report survey provide information on other possible positive outcomes associated with the SAGE program.

III. Overview of Demographics and Assessment

Characteristics of SAGE Schools and Students

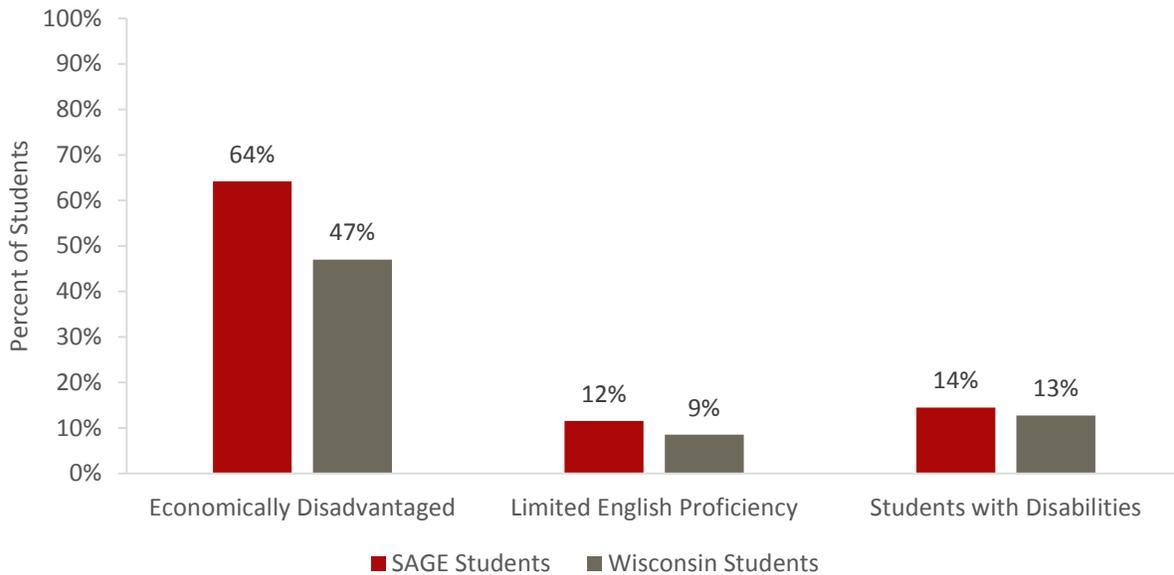
During the 2013-14 school year, 425 schools in 205 districts participated in the SAGE program. The number of students in SAGE schools in kindergarten through third grade was just over 80,000, with roughly equal proportions throughout each grade. Figures 1 – 3 show the demographic breakdown for students in SAGE schools by race/ethnicity, economic status (as determined by free or reduced price lunch eligibility), students with disabilities, ELL, and urban or rural setting as compared to the state as a whole. As seen from these figures, while SAGE schools looked similar to Wisconsin schools in general, they had higher proportions of economically disadvantaged students as the funding mechanism for SAGE relies upon the number of students qualifying for free or reduced price lunch. SAGE students were nearly equally prevalent in urban and rural settings.

Figure 1: Demographic breakdown of SAGE students by race/ethnicity in 2013-14



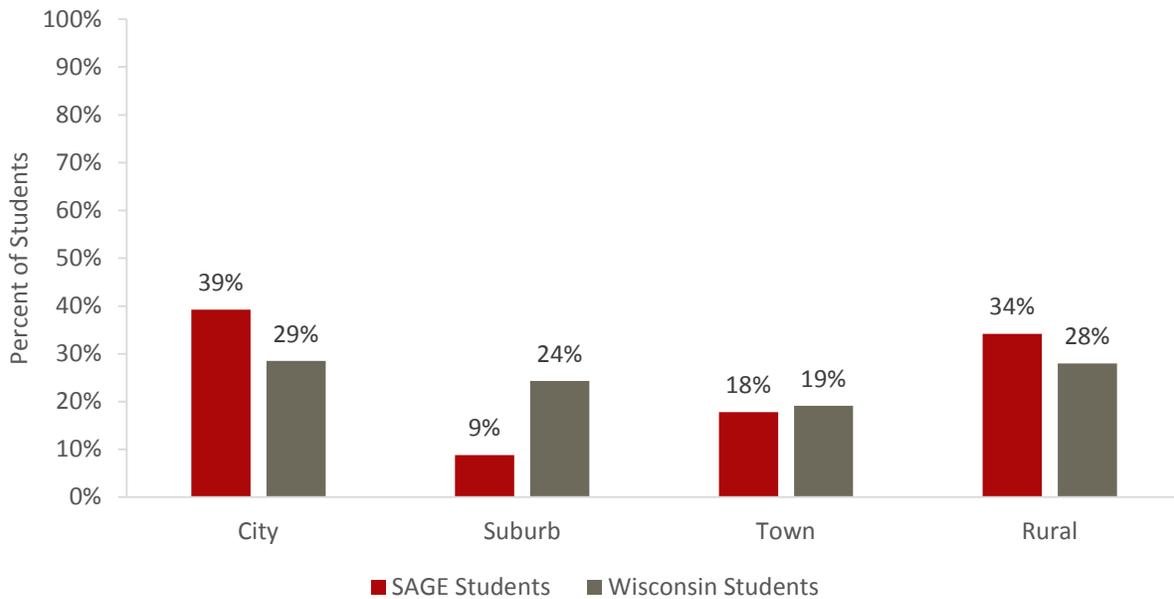
Source: WDPI LDS

Figure 2: Demographic breakdown of SAGE students by economic status, English proficiency, and disability status in 2013-14



Source: WDPI LDS

Figure 3: Demographic breakdown of SAGE students by urban and rural setting in 2013-14



Source: WDPI LDS

Table 1 examines the number of SAGE students in the five most populous SAGE districts: Beloit, Green Bay, Madison, Milwaukee, and West Allis – West Milwaukee. As expected, Milwaukee comprises the largest percentage of the SAGE student population with approximately 16 percent of all SAGE students. When subtotaling these five districts, they comprise approximately 32 percent of all SAGE students.

Table 1: Number and percent of SAGE students by district in kindergarten through third grade in 2013-14

District	SAGE Students	Percent of ALL SAGE Students
Beloit	2107	2.6%
Green Bay	2311	2.8%
Madison	5304	6.4%
Milwaukee	13504	16.4%
West Allis – West Milwaukee	2705	3.3%
Top 5 Subtotal	25931	31.5%
<i>All Other SAGE Districts</i>	<i>56302</i>	<i>68.5%</i>
Total	82233	100.0%

Source: LDS

Another important aspect of SAGE schools is their implementation of the small classroom requirement. As previously mentioned, under SAGE program guidelines, schools must maintain a student-to-teacher ratio of 18:1, 30:2, or 45:3 in kindergarten through third grade. As seen in Table 2, the vast majority, or 96 percent, of SAGE classrooms utilized an 18:1 configuration in 2013-14. To compare the differences across grades, Table 3 shows the average number of students and the standard deviation from this average across the four SAGE grades and by the three classroom configurations. This table shows that little difference exists across grades in the number of students outside of those classrooms with a mixed-grade configuration. By examining the distribution of classrooms with the 18:1 configuration, as seen in Figure 4, the majority of classrooms maintain a ratio of 15:1 or higher. Given this small range for most SAGE classrooms, and that the majority of the classrooms employ an 18:1 ratio, this evaluation does not differentiate impact by ratio or classroom configuration.

Table 2: Number and percent of SAGE classrooms by classroom configuration in 2013-14

Configuration	Number of Classrooms	Percent of Classrooms
18:1	5041	96.1%
30:2	203	3.9%
45:3	3	0.1%
Total	5247	100.0%

Source: SAGE Submission Report for 2013-14

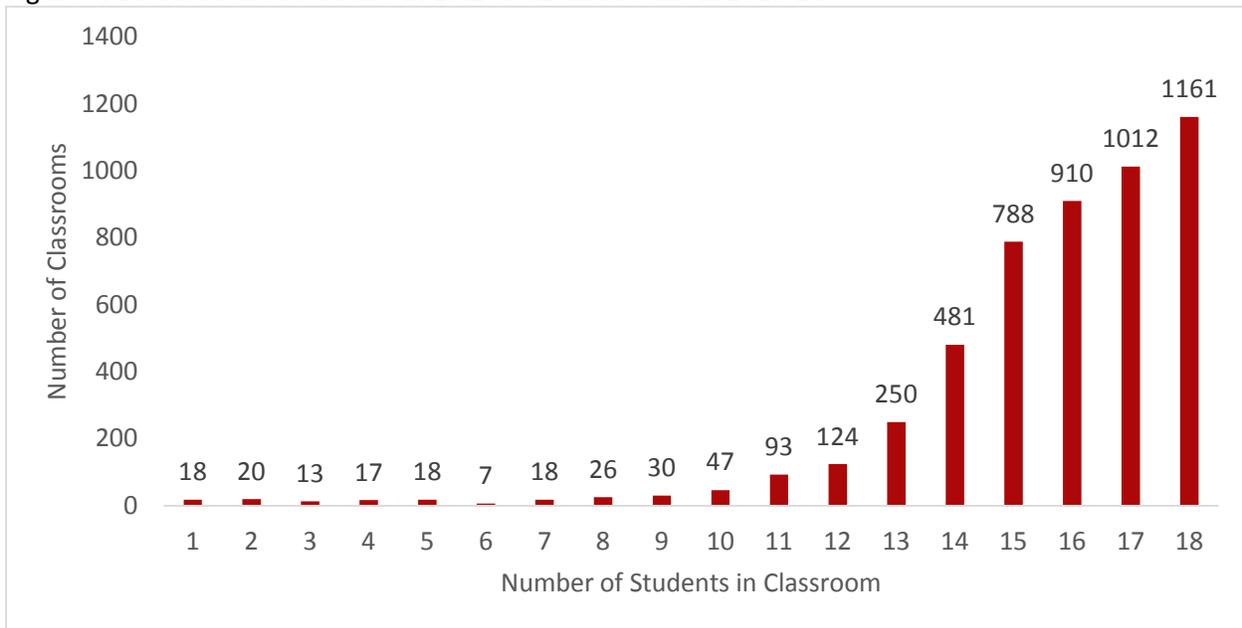
Table 3: Mean number and standard deviation of students in SAGE classrooms by grade and classroom configuration in 2013-14

Grade	18:1 Configuration		30:2 Configuration		45:3 Configuration	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
K5	15.8	2.3	11.4	3.2	4.3	N/A
1 st	15.8	2.0	12.6	2.2	N/A	N/A
2 nd	15.8	2.0	12.7	2.1	N/A	N/A
3 rd	15.7	2.0	12.4	2.1	N/A	N/A
Mixed	13.3	5.7	11.8	3.1	11.0	0.9

Source: SAGE Submission Report for 2013-14

Note: N/A indicates insufficient data

Figure 4: Distribution of students in 18:1 SAGE classrooms in 2013-14



Source: SAGE Submission Report for 2013-14

MAP Assessment and Characteristics of MAP Students

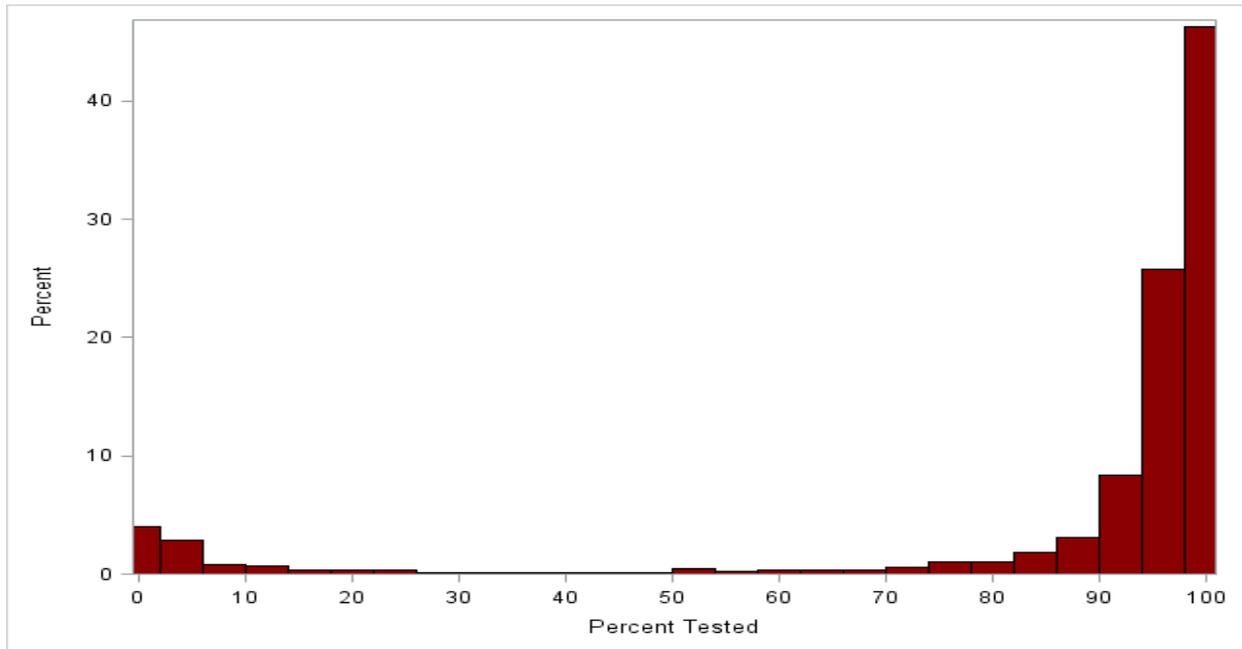
Since VARC uses the MAP assessment as a measure of student achievement, and since the state of Wisconsin does not administer the assessment across the entire population of schools and districts, the number of students who took the MAP reduced the size of the analysis sets.

To keep MAP takers in the analysis, VARC essentially placed two restrictions. First, the evaluation only used students with both a pretest and a posttest score in the value-added analyses. For inclusion in the analysis set, a student needed to take the MAP in both the fall and spring of a given year. Second, VARC needed to ensure that students took the test for benchmarking purposes. VARC assumes that when a school administered the MAP assessments to all or nearly all students in the grade level for benchmarking purposes or other educational purposes, nearly all students within a grade and subject should have a test score. Indeed, a school with a small proportion of students taking the test might have used the assessment for targeted interventions such as enrollment in a gifted program, special education, or grade promotion/retention. Inclusion of schools using the MAP assessment for such purposes would skew the results.

The data did not include a variable indicating the intentions of the school when using the MAP test. A look at the percentages of students taking the MAP in a given school, grade, and subject and their respective means was necessary. Figure 5 shows the distribution of the percentage MAP takers by grade. It indicates that the majority of MAP students were in grades with large proportions of MAP takers. Not all grades followed this pattern, however, as just over five percent of the grades had a small proportion of their students taking the assessment (less than five percent). While this likely means that many of the grades used the assessment for benchmarking purposes, the analysis required a cutoff point to decide which grades in each school to include. The procedure examined four possible cutoff

points and compared the difference in average assessment results as well as the number of schools lost in the analysis due to the cutoff point. The goal of this procedure was to maximize the difference in average assessment results between those students above the cutoff point and those below the cutoff point while at the same time minimizing the number of schools lost. This would allow for differences in the types of students assessed while keeping as many students as possible in the analysis. The results from this procedure determined a cutoff point of 80 percent or more students tested within a grade. The evaluation assumed that schools with grades above this cutoff point were using the MAP assessment for suitable purposes, and, thus, were included in the analysis set.

Figure 5: Distribution of the percentage of students taking the MAP within grades and schools in 2013-14

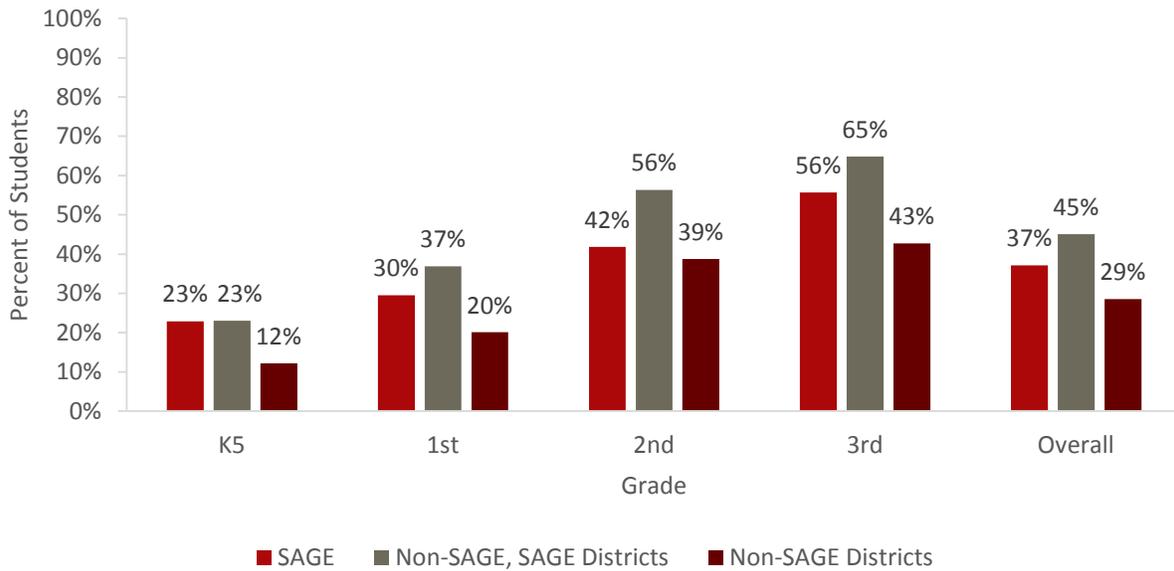


Source: 2013-14 MAP Files

VARC was then able to examine the demographic profiles of students who took the MAP assessment. Figure 6 shows the percentages of students who took the MAP assessment in each grade for three school subgroups: SAGE schools, non-SAGE schools in SAGE districts, and schools in non-SAGE districts.¹ Both SAGE students and non-SAGE students follow a similar trend of having proportions of MAP takers strictly increase with each increase in grade level.

¹ Approximately 86,000 students took the MAP assessment in 2013-14 for benchmarking purposes. Approximately 31,000 students were in SAGE schools, 23,000 in non-SAGE schools within SAGE districts, and 32,000 in non-SAGE districts.

Figure 6: Proportion of students participating in the MAP assessment by grade for SAGE schools, non-SAGE schools in SAGE districts, and schools in non-SAGE districts in 2013-14



Since less than half of the general population of students in both SAGE and non-SAGE schools took the MAP assessment for benchmarking purposes, it is important for this evaluation to consider the generalizability of any results to the overall population of students. To examine any differences, VARC compared the race/ethnicity; economically disadvantaged status; students with disabilities; English proficiency; and city, suburb, town, or rural setting for MAP and non-MAP students within the general population of both SAGE and non-SAGE schools. Figures 7 – 9 show these comparisons for all students in both SAGE and non-SAGE schools in kindergarten through third grade. As seen, more students who are African-American, economically disadvantaged, and in city or suburban settings took the MAP, and fewer in town or rural settings took the MAP. This indicates that it may be difficult to generalize any results to the entire population of students, especially in rural settings.

Figure 7: Race/ethnicity of all students in kindergarten through third grade by MAP utilization in 2013-14

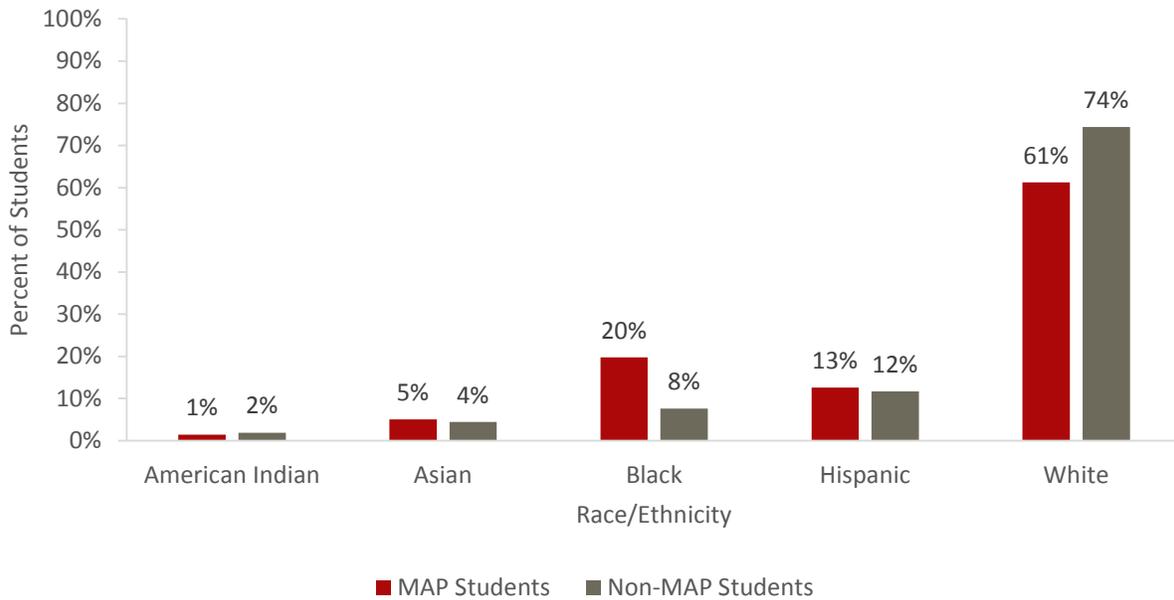


Figure 8: Percent of all kindergarten through third grade students who were economically disadvantaged, students with disabilities, and limited English proficient by MAP utilization in 2013-14

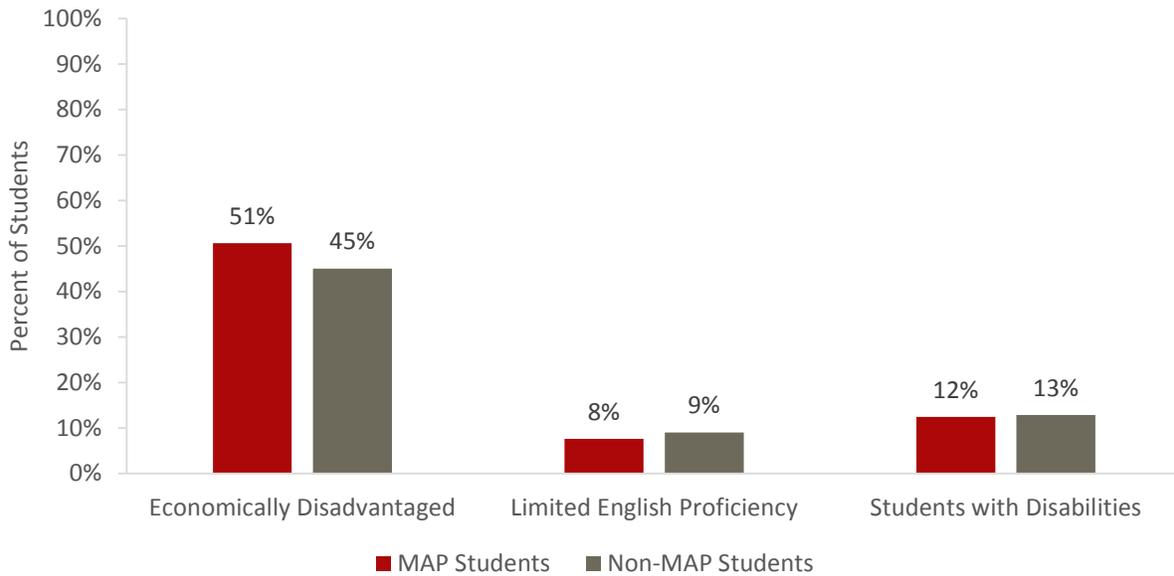
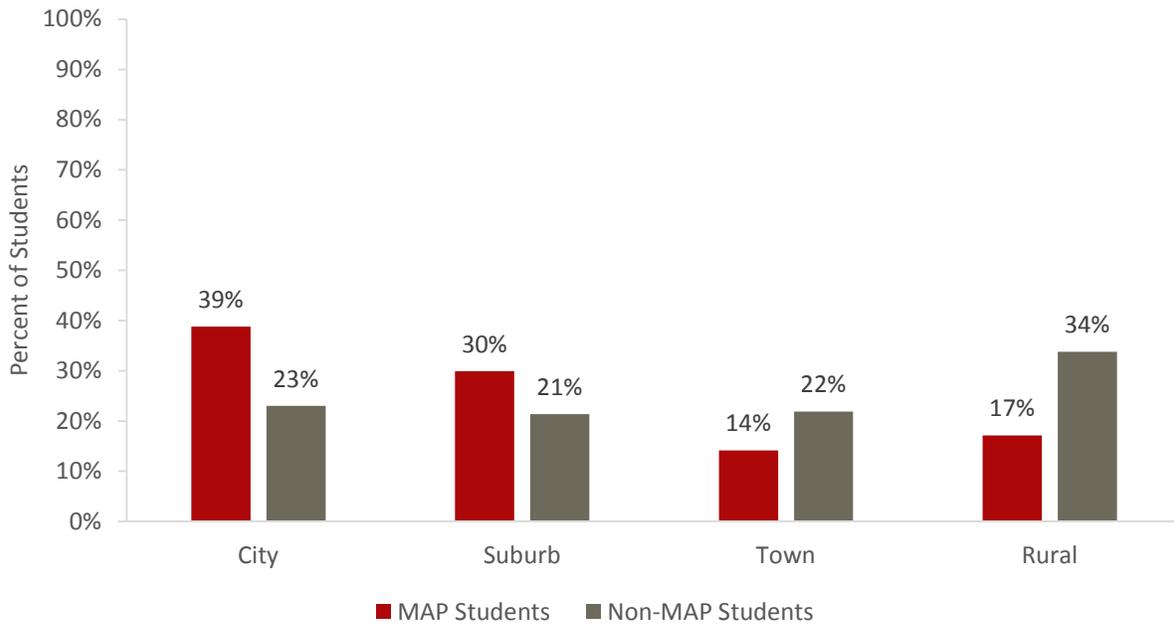


Figure 9: Percent of all kindergarten through third grade students who were in urban or rural settings by MAP utilization in 2013-14



Characteristic Comparison of SAGE and Non-SAGE Students

Another useful comparison in the evaluation of the SAGE program is the characteristics of students in SAGE compared to non-SAGE students. Differences in the types of students that belong to each group would make it difficult to conduct a simple analysis comparing performance of the two groups and require special consideration of a control group. Figures 10 – 12 show the breakdown of the percent of students by race/ethnicity; economically disadvantaged status; students with disabilities; limited-English proficiency; and city, suburb, town, or rural setting for SAGE students, non-SAGE students in SAGE districts, and students in non-SAGE districts across the SAGE grades of kindergarten through third grade. SAGE schools had a higher proportion of African-American students, a higher proportion of economically disadvantaged students, and a lower proportion of white students than non-SAGE schools in the set. The non-SAGE districts had a higher proportion of suburban students than the other two categories.

Figure 10: Race/ethnicity of tested students in kindergarten through third grade by SAGE participation in 2013-14

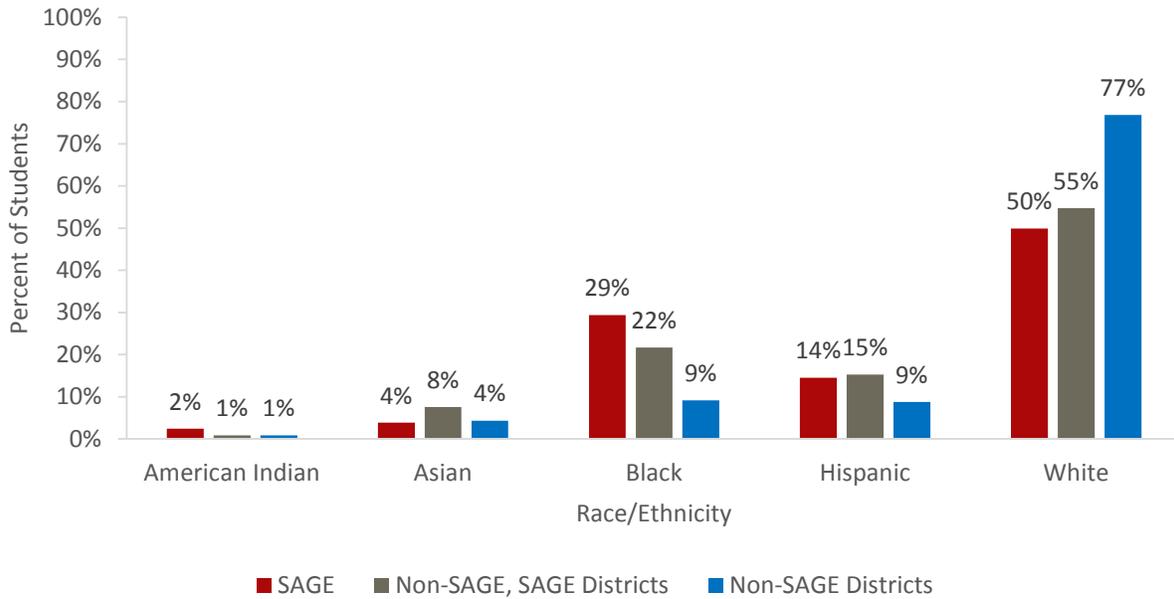


Figure 11: Percent of kindergarten through third grade tested students who were economically disadvantaged, students with disabilities, and limited English proficient by SAGE participation in 2013-14

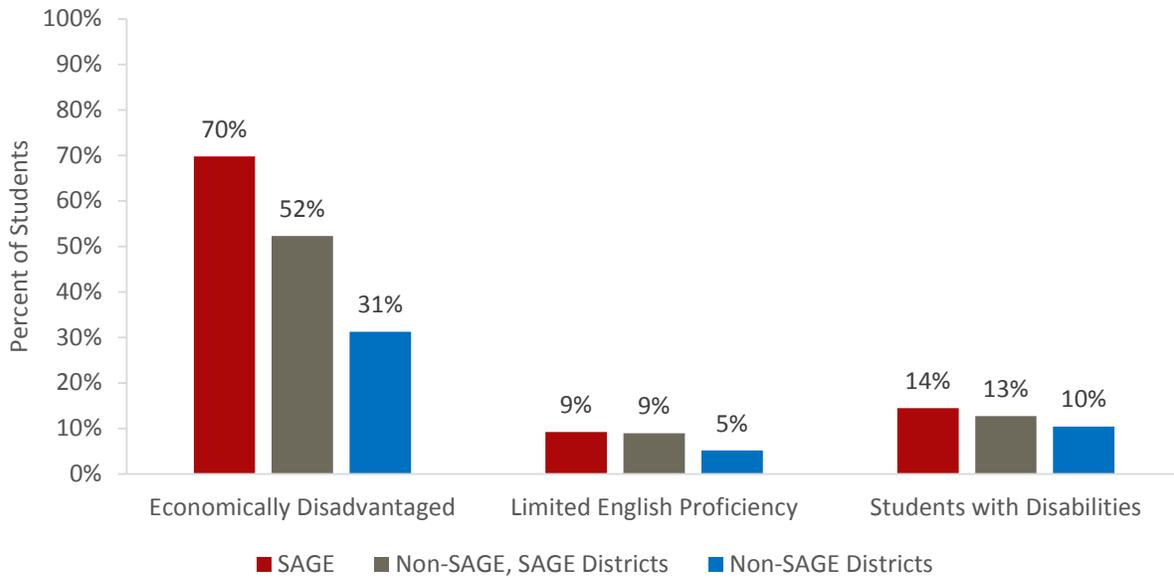
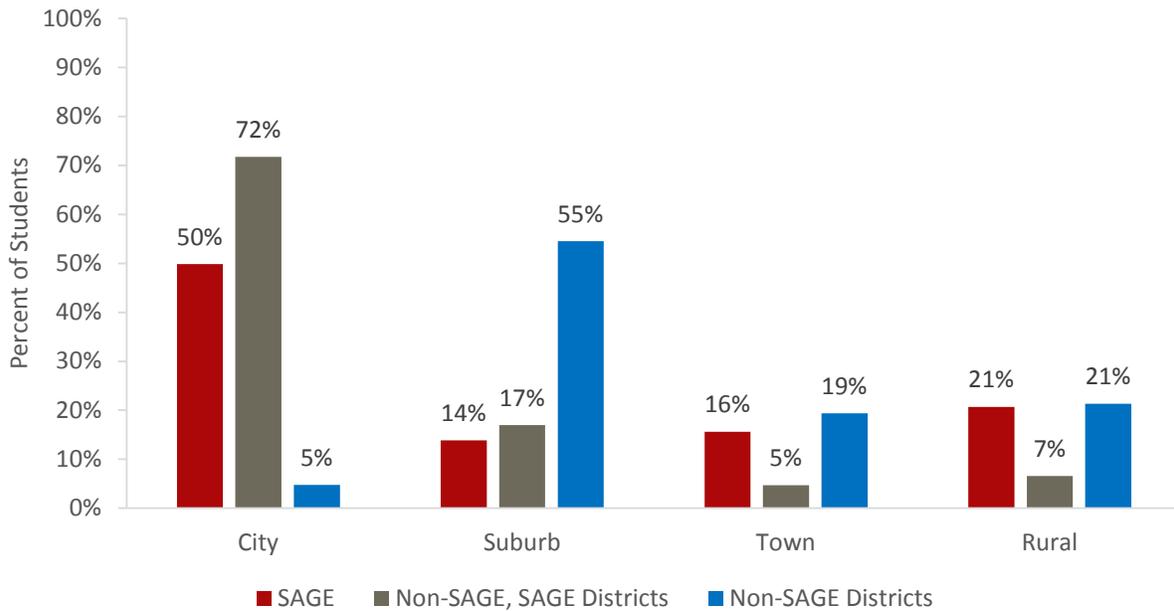


Figure 12: Percent of kindergarten through third grade tested students who were in urban or rural settings by SAGE participation in 2013-14



Overall, these differences between SAGE, non-SAGE within SAGE district, and non-SAGE district students demonstrate the need for statistically advanced methods of analysis to account for these differences. As later sections will show, this evaluation accounts for these factors when attempting to ascertain an effect of the SAGE program on student academic performance.

IV. Methodology

Due to the large differences in student and school characteristics between SAGE schools, non-SAGE schools in SAGE districts, and non-SAGE districts, this evaluation used statistical modeling in an attempt to detect the impact of the SAGE program on MAP growth in mathematics and reading as well as the SAGE impact on long-term outcomes.² This section of the report examines the analysis design models.

Single-Year MAP Assessment Analysis Design Model

Since a direct comparison of SAGE to non-SAGE schools is not feasible due to differences in student and school characteristics, the evaluation of the SAGE program used statistical modeling to detect the impact of the SAGE program on a single year of MAP growth in mathematics and reading. The single-year analysis used a three-step model which consisted of obtaining value-added estimates, propensity score matching, and a weighted linear regression.

² For a detailed and technical description of the design model, refer to the Appendix.

The first step in the model was for VARC to obtain the value-added estimates from student-level information for each school; see the Appendix for technical specifications. The value-added model uses the following information as explanatory student variables for the analysis:

- Fall MAP scores (both mathematics and reading),
- Gender,
- Race/ethnicity,
- Individualized Education Plan (IEP) or disability status,
- Economically disadvantaged status (through the free or reduced price lunch indicator), and
- English language learner status.

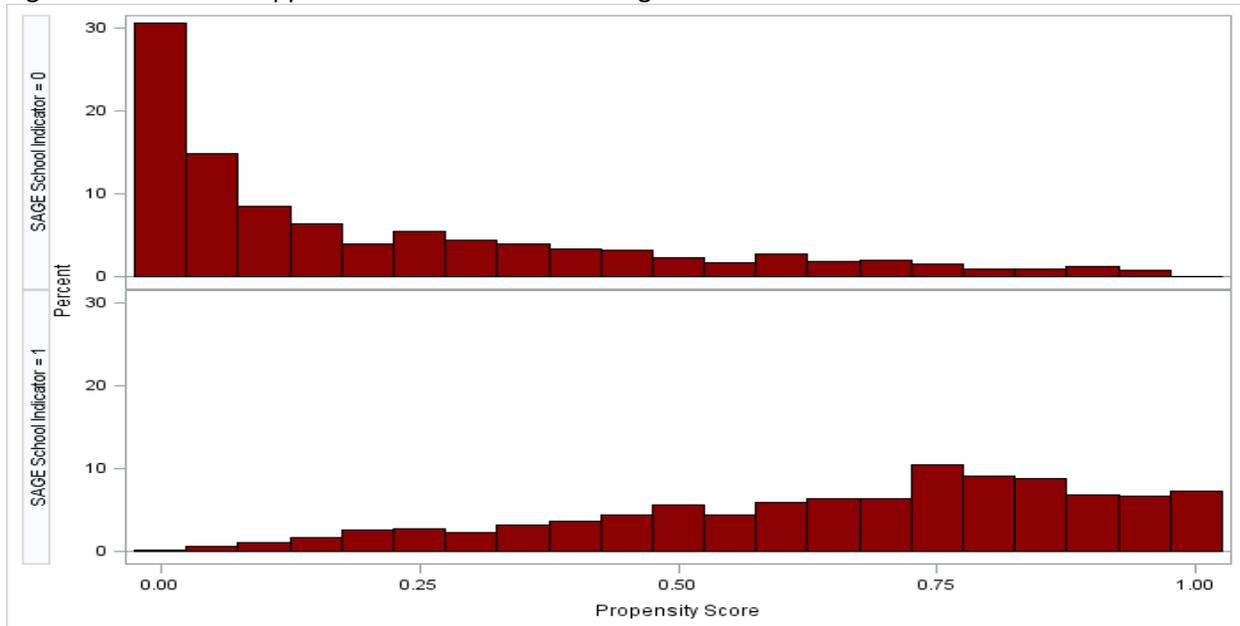
In the second step, using propensity score modeling, VARC estimated the propensity score of each school using a logit specification for the likelihood of being a SAGE school given the characteristics of the school and district. The model includes the following school- and district-level variables:

- Number of students,
- Percent free/reduced price lunch students,
- Percent female,
- Percent in each racial category,
- Percent special education,
- Percent ELL,
- Percent IEP or disability status,
- Average teacher compensation,
- School setting (city, suburb, town, rural), and
- Average Fall MAP scores (both mathematics and reading).

With the propensity scores, VARC matched each SAGE school with a non-SAGE school to obtain a set of similar schools for analysis. The analysis matched SAGE schools to non-SAGE schools using a one-to-one matching method with replacement. This allowed a non-SAGE school to be a match to more than one SAGE school. Allowing for replacement provided for a more similar set of schools to compare.

Examining the results from the propensity score model, the evaluation found common support for SAGE participation between SAGE schools and non-SAGE schools. Common support refers to the common range of propensity scores across control and treatment schools. Figure 13 shows the common support for the estimated propensity scores, which indicate that some non-SAGE schools had a similar likelihood of participation in SAGE across the range of propensity scores.

Figure 13: Common support for the likelihood of being in SAGE in 2013-14



The third step in the model was to use the set of matched schools with similar characteristics in a weighted linear regression. The outcome used in this regression was the value added calculated in the first step. The covariate used in the regression was an indicator of whether the school was a SAGE school. The regression used weights for both the number of students in the school as well as the propensity score. This weighting placed more emphasis on schools with a larger population of students and on schools that had a higher estimated probability of participation in SAGE. This reduced any bias that may occur from small schools or from schools that had characteristics further away from the majority of SAGE schools. The results from this regression would show the impact of SAGE on student growth in both mathematics and reading at each grade level.

Multiple-Year MAP Assessment Analysis Design Model

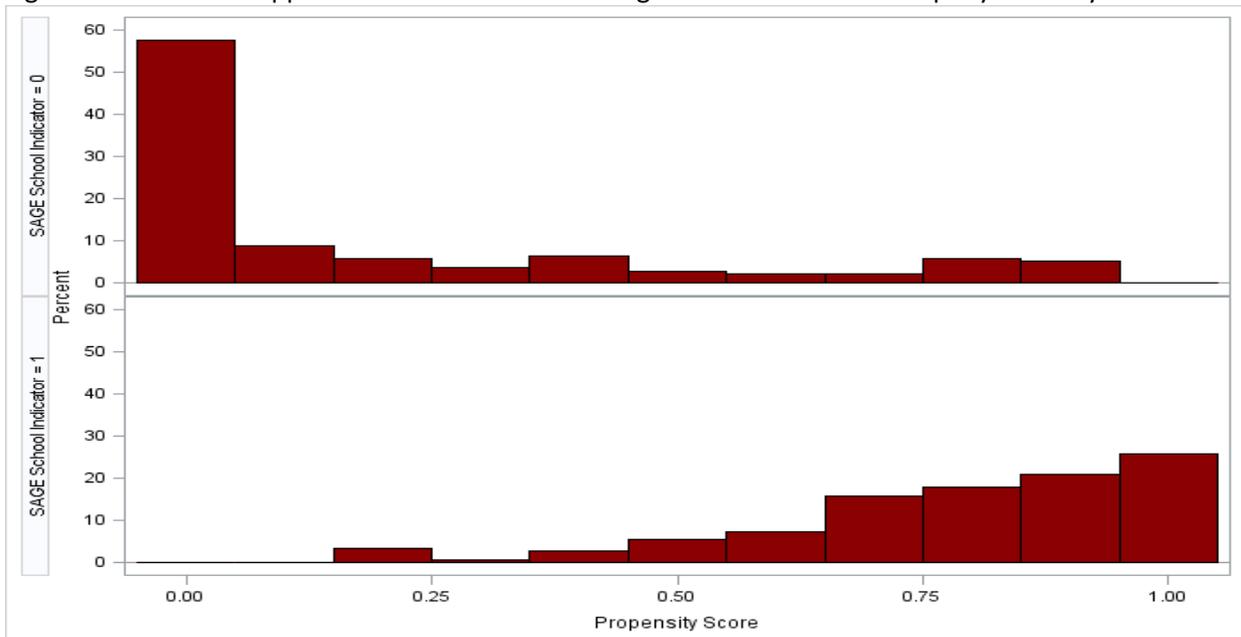
This year’s analysis of the SAGE program also continued to look at the cumulative impact of the program. Since this was the first time the evaluation had access to four years of comprehensive MAP assessment data (2010-11, 2011-12, 2012-13, and 2013-14), VARC worked on examining the effects of the SAGE program over multiple years. This analysis started by looking at the cohort of kindergarten students in 2010-11 and examining their growth over multiple years from the fall of 2010-11 until the spring of 2013-14. The analysis only included students who took the MAP in the fall of kindergarten and the spring of third grade, remained in either a SAGE school for all four years or a non-SAGE school for all four years, and progressed from kindergarten in 2010-11 to first grade in 2011-12 to second grade in 2012-13 and to third grade in 2013-14. The analysis did not restrict students to be at the same school or same district for all four years. With an established cohort set, the analysis used a similar three-step model similar to the one used in the single-year analysis consisting of obtaining value-added estimates, propensity score matching, and a weighted linear regression.

The first step in the model was for VARC to obtain the value-added estimates from student-level information for each school; see the Appendix for technical specifications. The value-added model used the same explanatory student variables for this analysis as the single-year analysis. The difference for the multiple-year analysis was to obtain value-added estimates from the fall of kindergarten to the spring of third grade.

In the second step, VARC estimated the propensity score of each school using a logit specification for the likelihood of being a SAGE school given the characteristics of the school and district from 2013-14 data (the third-grade year). The model includes the same school- and district-level variables as the single-year analysis. The model also matches schools with the same methods as described above.

Examining the results from the propensity score model, the evaluation found common support for SAGE participation between SAGE schools and non-SAGE schools. Figure 14 shows the common support for the estimated propensity scores, which indicate that some non-SAGE schools had a similar likelihood of participation in SAGE across the range of propensity scores.

Figure 14: Common support for the likelihood of being in SAGE in for the multiple-year analysis



The third step in the model was to use the set of matched schools with similar characteristics in a weighted linear regression. This regression used the same specifications as the regression used in the single-year analysis with a different set of schools. The results from this regression show the impact of SAGE on student growth in both mathematics and reading over the four years of SAGE programming.

Long-Term Outcomes Analysis Design Model

The new analysis for year 10 was the long-term examination of student outcomes in high school. This analysis used the large cohort of SAGE schools that joined in 2000-01 as an opportunity to examine the before-and-after difference in students' long-term outcomes in schools that joined SAGE, as well as the

difference in outcomes between students in schools that joined SAGE and schools that did not. This analysis examined the differences in students dropping out of high school at each grade and the eventual difference in high school completion between students in pre-SAGE schools, students in SAGE schools, and students in non-SAGE schools over the same periods.

This analysis used data from DPI’s longitudinal data system for the high school outcomes. Since the longitudinal data system did not exist when the third SAGE cohort of schools joined in 2000-01, VARC could not use this as a data source to match students to SAGE or non-SAGE schools. Thus, VARC relied on Wisconsin Reading Comprehension Test (WRCT) records to match students with specific schools from that time. DPI only administered the WRCT as early as third grade, so any information about where a student would have been in kindergarten through second grade was unavailable.

The first step in this analysis was to create seven cohorts of students based on their WRCT year (or third-grade year) both before and after the third group of SAGE schools joined in 2000-01. Table 4 shows the characteristics of each cohort. As seen, three cohorts of students belonged to either non-SAGE schools or pre-SAGE schools before they joined the program. Four other cohorts belonged either to non-SAGE schools or to schools that joined the SAGE program. The analysis excluded schools that joined the SAGE program prior to 2000-01 to maintain an unbiased control group.

Table 4: Cohort descriptions for the long-term analysis

WRCT Cohort	Cohort Description	Possible Grades in SAGE	WRCT 3 rd -Grade Record Year
1998	Pre-SAGE/Non-SAGE	None	1997-98
1999	Pre-SAGE/Non-SAGE	None	1998-99
2000	Pre-SAGE/Non-SAGE	None	1999-00
2001	SAGE/Non-SAGE	3 rd	2000-01
2002	SAGE/Non-SAGE	2 nd , 3 rd	2001-02
2003	SAGE/Non-SAGE	1 st – 3 rd	2002-03
2004	SAGE/Non-SAGE	K – 3 rd	2003-04

The next step in the analysis was to match the students in these cohorts to their high school and exit type records. This allowed VARC to determine the dropout and high school completion status of each student in each cohort. This matching and cleaning process reduced the overall sample of students from these seven cohorts due to student mobility, record keeping, and missing outcomes. Table 5 shows the cleaning process, the number of students lost in each step, and the overall reduction in sample size. As shown from this table, despite the length of time between the WRCT records and their high school outcome records, the analysis sample retained more than three-quarters of the entire sample of tested third-grade students in these cohorts. Table 6 shows the overall reduction in sample size for each cohort.

Table 5: Number of students in the long-term analysis after each cleaning step

Cleaning Step	Number of Students	Total Student Attrition	Percent of Baseline	Percent of Previous Step
Baseline WRCT Cohorts	344,740			
Baseline Cohorts with only 1 School	343,579	1,161	99.7%	99.7%
Single-Student Match with High School Records	294,700	50,040	85.5%	85.8%
Ninth-Grade Record	280,014	64,726	81.2%	95.0%
Outcomes Present	267,103	77,637	77.5%	95.4%

Table 6: Student attrition in the long-term analysis by cohort

WRCT Cohort	Number of Baseline Students	Number of Students with Outcomes	Total Student Attrition	Percent of Baseline
1998	53,569	39,788	13,781	74.3%
1999	54,937	41,478	13,459	75.5%
2000	54,708	41,535	13,173	75.9%
2001	44,530	35,132	9,398	78.9%
2002	44,917	35,532	9,385	79.1%
2003	46,120	36,638	9,482	79.4%
2004	45,959	37,000	8,959	80.5%

With the analysis cohorts, the next step was to use a sequential logistic model to determine the impact of SAGE on the probability of dropping out of school or continuing for each grade level in high school. Specifically, the model examined all students who either completed high school or dropped out, excluding those students who continued after twelfth grade and had not completed or dropped out. Then the model looked at the odds of either dropping out at ninth grade or continuing in high school after ninth grade. Of those continuing after ninth grade, the model then examined the odds of either dropping out at tenth grade or continuing in high school after tenth grade. This analysis continued until twelfth grade where the model examined the odds of either dropping out or completing high school. At each of these grades, the model calculated the difference between SAGE and pre-SAGE, and between SAGE and non-SAGE, while controlling for the same student demographic factors used in the value-added analysis described previously. For more specifics on this model, see the Appendix.

One of the features of a logistic model design is the ability to create predicted probabilities based on the results. This allows for a visual representation of the impact of the program based on assuming certain student characteristics. The final step in this analysis was to create predicted probabilities based on chosen student demographics, then show the impact of SAGE through a visual indication of the probability of dropping out at each grade for SAGE students, pre-SAGE students, and non-SAGE students. These predicted probabilities assumed a white male student without disabilities that qualified for free or reduced price lunch and was not an English language learner. The model used these assumptions to reflect the average characteristics of a SAGE student.

In an attempt to examine the overall long-term high school outcomes, the evaluation also estimated the overall impact of SAGE on completing high school as opposed to dropping out of high school. This analysis used the estimates from the individual high school grade sequential logistic model combined with the same average characteristics of a SAGE student to estimate the overall SAGE impact throughout high school for these students. The analysis also examined the impact of SAGE on students with other varying characteristics to determine if the program was potentially more beneficial in increasing long-term outcomes for different types of students. These characteristics varied by race, special education status, and English language learner status. In all cases, the analysis assumed the student qualified for free or reduced price lunch as this was a main population of focus for the SAGE program.

V. Evaluation Results

Single-Year SAGE Effects

Table 7 shows the number of SAGE schools and matched non-SAGE schools used in the analysis. In first grade, for example, the model matched each of the 101 SAGE schools with one of the 33 distinct non-SAGE schools. As the previous section mentioned, the model allowed non-SAGE schools to match with more than one SAGE school to provide a more similar sample. As a useful comparison, Table 8 shows the total number of SAGE schools and non-SAGE schools throughout Wisconsin. This provides an indication of the proportion of schools used in the main analysis and the analyses that follow.

Table 7: Number of SAGE and matched non-SAGE schools used in main analysis in 2013-14

Grade	SAGE Schools	Non-SAGE Schools
K5	77	18
1 st	101	33
2 nd	159	63
3 rd	151	223

Table 8: Total number of SAGE and non-SAGE schools in Wisconsin in 2013-14

Grade	SAGE Schools	Non-SAGE Schools
K5	409	698
1 st	414	708
2 nd	412	705
3 rd	401	717

Table 9 and Table 10 show the statistically adjusted MAP score growth differences between the students in SAGE and non-SAGE schools from the single-year analysis design model presented in the previous section for mathematics and reading, respectively. In each of these tables, VARC provides both the SAGE coefficient in MAP scale scores and the SAGE coefficient in standard deviations of the posttest results. The coefficient in scale scores shows the estimated impact of the SAGE program for the specific group of students on the number of scale score points of growth. The coefficient in standard deviations

represents a normalized approach to interpreting the results. VARC calculated this by taking the coefficient in scale scores divided by the standard deviation of the whole set’s spring test results.

As seen in Table 9, VARC estimated positive but statistically insignificant effects of the SAGE program on mathematics growth in kindergarten as compared to students in non-SAGE schools. The evaluation found no statistically significant results in first, second, or third grade for mathematics.

Table 9: Results of statistical analysis of the SAGE program effect on *mathematics* MAP growth by grade in 2013-14

Grade	Effect Size (Scale Scores)	Effect Size (Std. Dev.)	Standard Error	T-stat
K5	1.15	0.08	1.26	0.91
1 st	-0.02	0.00	0.64	-0.03
2 nd	-0.36	-0.03	0.50	-0.73
3 rd	-0.43	-0.03	0.32	-1.36

Note: Results in bold indicate a statistically significant effect.

As seen in Table 10, VARC estimated positive and significant effects of the SAGE program on reading growth in kindergarten as compared to students in non-SAGE schools. The evaluation found no statistically significant results in first, second, or third grade for reading. This means that, on average, students in SAGE schools grew at a higher rate than students in non-SAGE schools in kindergarten in reading.

Table 10: Results of statistical analysis of the SAGE program effect on *reading* MAP growth by grade in 2013-14

Grade	Effect Size (Scale Scores)	Effect Size (Std. Dev.)	Standard Error	T-stat
K5	2.50	0.19	1.06	2.35
1 st	-0.12	-0.01	0.62	-0.20
2 nd	0.55	0.04	0.49	1.12
3 rd	0.08	0.01	0.30	0.27

Note: Results in bold indicate a statistically significant effect.

Multiple-Year SAGE Effects

In addition to the single-year effects of the SAGE program on student growth, this report also presents information on the impact of SAGE over the course of the full four years of the program from kindergarten through third grade. In this analysis the propensity score matching model matched 90 SAGE schools to 23 non-SAGE schools. Table 11 shows the statistically adjusted MAP score growth differences between the students in SAGE and non-SAGE schools from the multiple-year analysis design model presented in the previous section for mathematics and reading. In this table, VARC provides both the SAGE coefficient in MAP scale scores and the SAGE coefficient in standard deviations of the posttest results. The coefficient in scale scores shows the estimated impact of the SAGE program for the specific group of students on the number of scale score points of growth. The coefficient in standard deviations represents a normalized approach to interpreting the results. VARC calculated this by taking the coefficient in scale scores divided by the standard deviation of the whole set’s spring test results.

As seen in Table 11, VARC estimated positive and statistically significant effects of the SAGE program on mathematics and reading growth in kindergarten through third grade as compared to students in non-SAGE schools.

Table 11: Results of the cumulative SAGE program effect on *mathematics and reading* MAP growth from 2010-11 to 2013-14

Subject	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
Mathematics	4.35	0.31	1.12	3.88
Reading	6.26	0.41	1.36	4.60

Note: Results in bold indicate a statistically significant effect.

Long-Term SAGE Effects

New to this year’s evaluation was an examination of the long-term outcomes of participation in SAGE. Specifically, the evaluation analyzed the impact of SAGE on the probability of dropping out of school or continuing in school at each high school grade level. Table 12 shows the results from the sequential logistic model described in the previous section for each grade level. This table displays the coefficient on SAGE, the calculated odds-ratio of the SAGE impact, the standard error associated with the coefficient, and an indication of the statistical significance. As seen from Table 12, the analysis found positive and statistically significant impacts of the SAGE program on continuing rather than dropping out at ninth and tenth grade. As indicated from the odds-ratio, this means for instance that students in SAGE schools had 37 percent greater odds of continuing rather than dropping out in tenth grade as compared to students in pre-SAGE schools while controlling for the general trend over time and the general trend of non-SAGE schools over the same time. The analysis found no significant impact at eleventh or twelfth grade; however, the nature of the sequential analysis would suggest that the impacts found in ninth and tenth grade lasted through the end of high school but did not decrease in eleventh or twelfth grade as more students continued on to the later grades.

Table 12: Results of the SAGE program impact on not dropping out at each grade level

Grade	SAGE Coefficient	Odds-Ratio	Std. Err.	P-Value
9 th	0.2717	1.3122	0.1099	0.013
10 th	0.3140	1.3688	0.1030	0.002
11 th	-0.0416	0.9592	0.0686	0.544
12 th	-0.0049	0.9951	0.0515	0.924

Note: Results in bold indicate a statistically significant effect.

One of the benefits of using a logistic model is the ability to visualize the impacts of the SAGE program at ninth and tenth grade using predicted probabilities. This method takes the results from the analysis to predict the probability of continuing rather than dropping out for an individual student based on their known characteristics. This allows for a comparison over time of the probability of not dropping out between students from non-SAGE schools, students from pre-SAGE schools, and students from SAGE schools. Figure 15 shows the impact of the SAGE program on the probability of continuing rather than dropping out in ninth grade. This figure visually represents the probability of not dropping out for each WRCT cohort for students in non-SAGE schools (dotted black line), pre-SAGE schools (solid green line),

and SAGE schools (solid red line). As seen from this figure, after the implementation of SAGE, the probability of not dropping out increases by approximately one-half of a percentage point. The impact of the SAGE program approaches bringing students to the same probability of dropping out as students from non-SAGE schools. The dashed green line seen in Figure 15 shows the model's prediction of the probability of not dropping out for students in SAGE schools if the schools had never joined the SAGE program. Figure 16 shows the same information, only for tenth grade. Since the analysis found no statistically significant results in eleventh or twelfth grade, this report does not include visual representations of these grades.

Figure 15: Results of the SAGE program impact on not dropping out in ninth grade

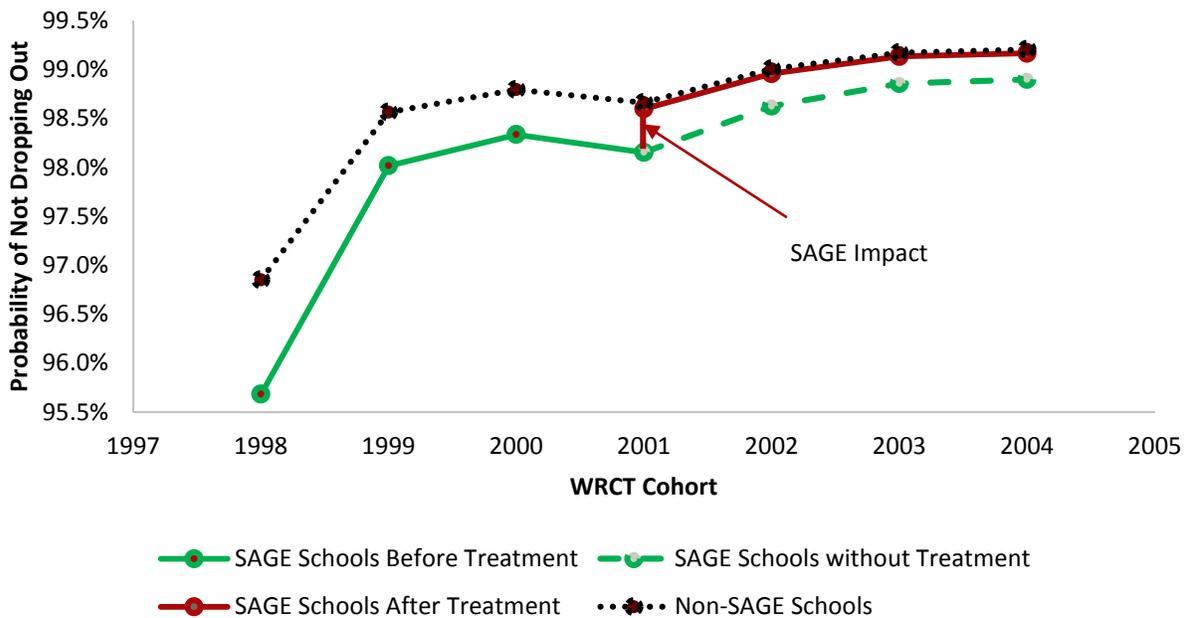


Figure 16: Results of the SAGE program impact on not dropping out in tenth grade

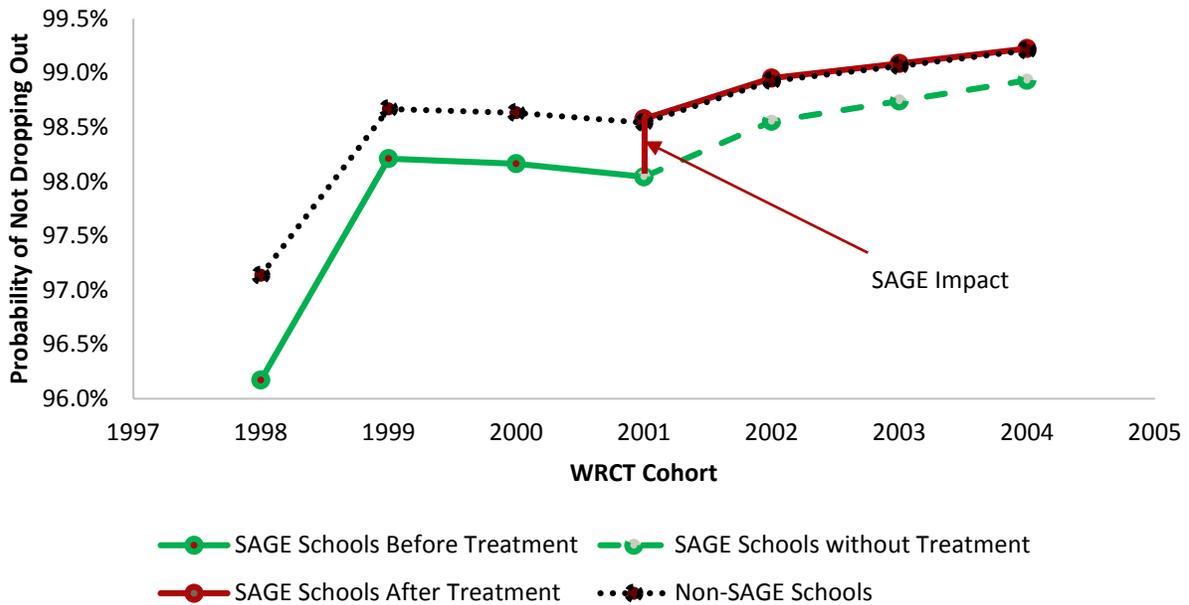


Figure 17 shows the impact of the SAGE program on the probability of completing high school rather than dropping out in high school. This figure includes the predicted probabilities from all years throughout high school to visualize an overall impact.

Figure 17: Results of the SAGE program impact on completing high school rather than dropping out

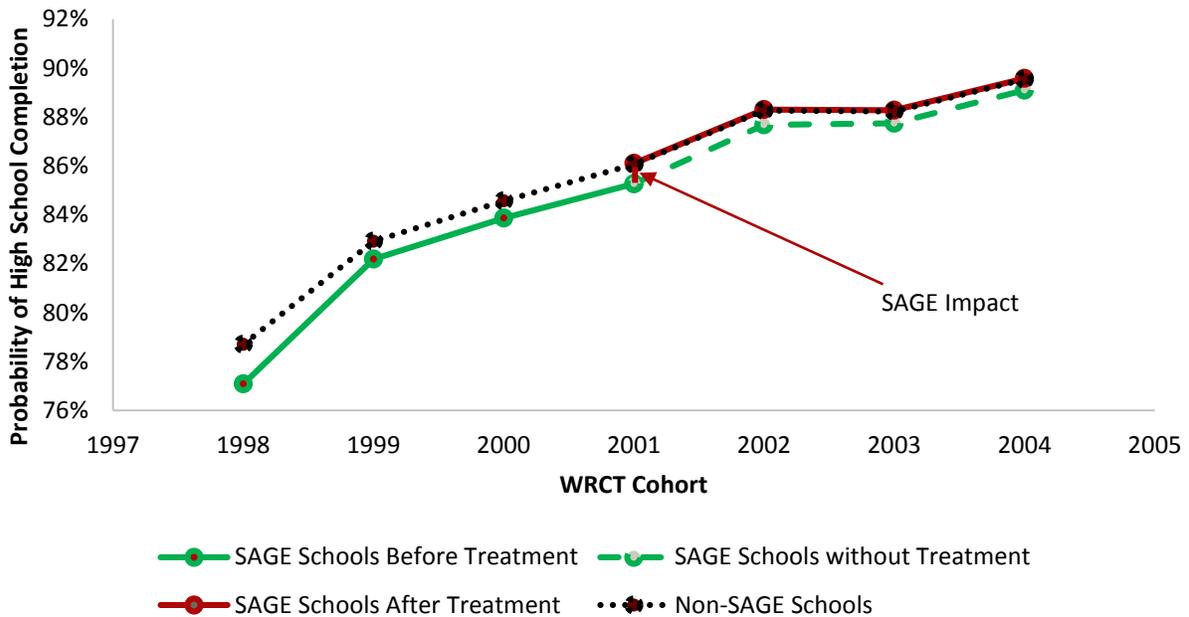


Table 13, below, shows the results from the analysis of the overall impact of SAGE on high school outcomes. This type of evaluation required the use of assumed student characteristics. This table shows

the marginal impact of SAGE for students with these characteristics. For instance, for African-American male students without disabilities that qualified for free or reduced price lunch and were not an English language learners, the SAGE program had a positive and statistically significant impact on increasing high school completion as opposed to dropping out of high school. As this table shows, the SAGE program had positive and statistically significant effects for African-American students regardless of disability or English language learner status. The SAGE program had positive but statistically insignificant effects for white students with similar characteristics. The evaluation also examined the marginal impact of SAGE for female students with the same characteristics. The results showed a similar pattern with significant marginal effects of SAGE for African-American students.

Table 13: Results of the SAGE program on completing high school or dropping out based on student characteristics

Student Characteristics	Marginal Effect of SAGE	Std. Err.	T-Stat
Male, FRL, White, non-SwD, non-ELL	0.00728	0.00440	1.66
Male, FRL, Black, non-SwD, non-ELL	0.03231	0.01128	2.86
Male, FRL, White, SwD, non-ELL	0.00692	0.00543	1.27
Male, FRL, Black, SwD, non-ELL	0.02961	0.01149	2.58
Male, FRL, White, non-SwD, ELL	0.00670	0.00415	1.61
Male, FRL, Black, non-SwD, ELL	0.03084	0.01103	2.80
Male, FRL, White, SwD, ELL	0.00633	0.00513	1.23
Male, FRL, Black, SwD, ELL	0.02828	0.01128	2.51

Note: Results in bold indicate a statistically significant effect. FRL = free or reduced price lunch, SwD = student with disabilities, and ELL = English language learner.

VI. End-of-Year Survey Results

At the end of each school year, the Wisconsin Department of Public Instruction (DPI) sends schools participating in the SAGE program an end-of-year survey. The survey asks schools about how they implement the SAGE program and how the SAGE program impacted students.

This report summarizes the results from the 2014 SAGE End-of-Year Survey fielded in May and June 2014. The responses from this survey cover the 2013-14 school year. All 425 SAGE schools across the state completed the survey, making the responses to this survey representative of all SAGE schools.

SAGE Implementation

The Department of Public Instruction developed the first set of questions on the SAGE End-of-Year Survey to find out more about the classroom configurations schools used to meet the SAGE class size requirements (class sizes no more than 18:1 or 30:2). Table 14 shows that about one-quarter of SAGE schools implemented multi-age classrooms, that is, classrooms with two or more grades of students, in the 2013-14 school year. About 3 percent of schools were planning to implement multi-age classrooms for the first time in the 2014-15 school year, though 72 percent of schools said they were not currently nor were they planning to implement multi-age classrooms.

Table 14: Number of schools with multi-age classrooms (n = 425)

	Number of Respondents
Yes, we will continue with and/or add to our current multi-age classrooms.	108 (25%)
Yes, we will have new multi-age classrooms for the first time at our school.	11 (3%)
No, we will not have any multi-age classrooms.	306 (72%)

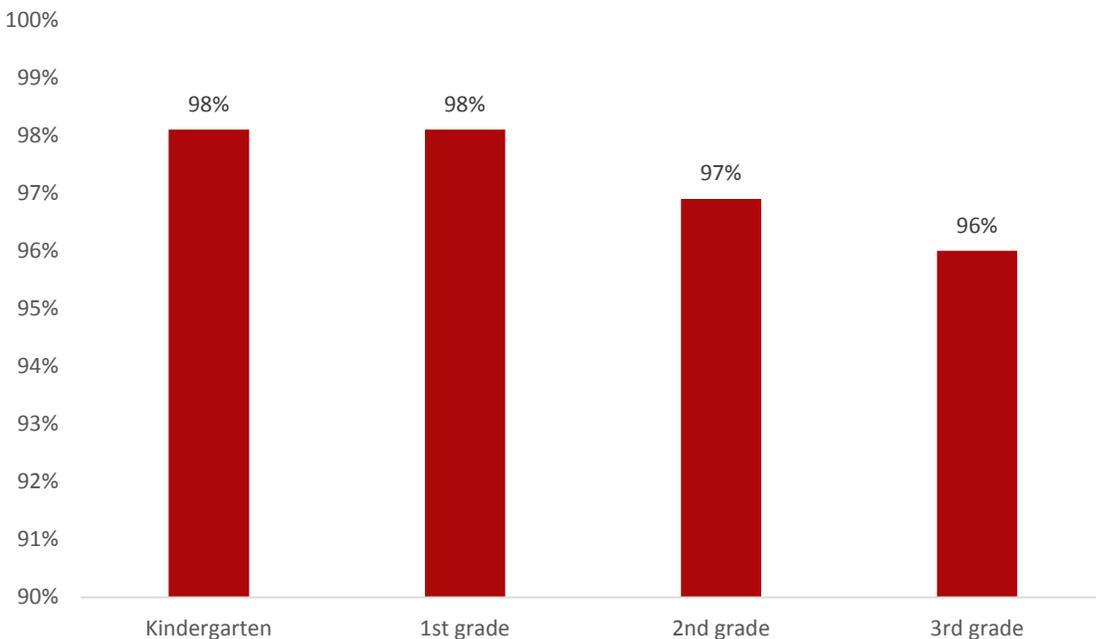
Table 15 displays information about team-teaching in classrooms. Like Table 14 that shows the use of multi-age classrooms in SAGE schools, about one-quarter of SAGE schools used team teaching. Another 4 percent of schools were planning to use team teaching in at least one classroom for the first time in the 2014-15 school year, while 71 percent of SAGE schools were not implementing or were not planning to implement team teaching in their classrooms.

Table 15: Number of schools with team-teaching classrooms (n = 425)

	Number of Respondents
Yes, we will continue with and/or add to our current team-teaching classrooms.	105 (25%)
Yes, we will have new team-teaching classrooms for the first time at our school.	18 (4%)
No, we will not have any team-teaching classrooms.	302 (71%)

Figure 18 shows the results of an item meant to collect information about which grades SAGE schools were implementing the SAGE program. Most SAGE schools, 98 percent, implemented SAGE in kindergarten and 1st grade in the 2013-14 school year, while 97 percent of schools had SAGE in 2nd-grade classrooms and 96 percent in 3rd-grade classrooms.

Figure 18: SAGE grades implementing the SAGE program (n = 425)



To find out why schools did not implement SAGE across all grades kindergarten through three (see Figure 18), DPI asked respondents why they did not have SAGE in each grade level. The most common response (as seen in Table 16) was that SAGE schools did not have grades two and/or three in their school; thus, their answers simply reflect the grades in their school that were eligible for SAGE. Five of the respondents (1 percent) said that the class sizes in their school were projected to be too large to maintain the 18:1 or 30:2 class size requirement in those grades. Most of those who wrote in the “other description” explained they had just a few too many students in their classrooms when they completed the survey (for example, 21) and did not have enough money to hire another teacher to implement SAGE in third grade. Another five respondents (1 percent) mentioned that their budget was too limited to implement SAGE in all grade levels eligible.

Table 16: Reasons for not implementing SAGE in grades 2 and/or 3 (n = 19)

Reason	Number of Respondents
2 nd and 3 rd grades are separated from kindergarten and/or 1 st -grade classrooms	9 (2%)
Projected class sizes too large to maintain 18:1 or 30:2 class size requirement	5 (1%)
Budget	5 (1%)
Academic needs of students at specific grade levels	0
Open enrollment	0

Teacher Perceptions of SAGE

The SAGE evaluation team at VARC suggested the items in the next section of the survey, and all of the items thereafter. Table 17 shows that most respondents were extremely or mostly satisfied with SAGE, with 73 percent responding that they were extremely satisfied and 21 percent that they were mostly satisfied. No respondents said they were not satisfied with the SAGE program in their school.

Table 17: General satisfaction with SAGE program (n = 425)

	Number of Respondents
Extremely Satisfied	312 (73%)
Mostly Satisfied	90 (21%)
Satisfied	21 (5%)
Somewhat Satisfied	2 (1%)
Not Satisfied	0
I Don't Know	0

Table 18 displays information about teacher satisfaction with the SAGE program. Those who responded to the SAGE survey (generally the SAGE coordinator or the principal of the school) said that their teachers were extremely satisfied (74 percent), mostly satisfied (23 percent), or satisfied (3 percent) with the SAGE program.

Table 18: Satisfaction of teachers with SAGE program (n = 425)

	Number of Respondents
Extremely Satisfied	314 (74%)
Mostly Satisfied	96 (23%)
Satisfied	15 (3%)
Somewhat Satisfied	0
Not Satisfied	0
I Don't Know	0

Almost all SAGE schools mentioned the SAGE program in their teacher recruitment efforts. Table 19 shows that 95 percent mentioned SAGE in recruiting teachers, with only 2 percent explaining that their schools did not mention SAGE in recruitment of teachers. About 3 percent of the respondents did not know if their schools mentioned SAGE in teacher recruiting. Respondents may not have known because they were not part of their school's teacher recruiting efforts or it may be that in some districts, schools did not recruit teachers, which may also explain why some respondents said that their schools did not mention SAGE in their teacher recruitment process.

Table 19: Mentioned SAGE programs in teacher recruitment process (n = 425)

	Number of Respondents
Yes	404 (95%)
No	9 (2%)
I Don't Know/Not Applicable	12 (3%)

Most respondents said that SAGE had been effective in recruiting teachers to their schools. Of all the respondents, 94 percent noted that SAGE was effective for recruiting teachers to their school, as seen in Table 20. Nearly four in 10 respondents said that mentioning SAGE was effective for recruiting teachers to a great extent, about one-third said it was effective to a moderate extent, 17 percent noted that it was effective to some extent, and 5 percent said it was effective to a small extent.

Table 20: Extent of effectiveness of SAGE to recruit effective teachers to respondent's school (n = 425)

	Number of Respondents
To a Great Extent	164 (39%)
To a Moderate Extent	141 (33%)
To Some Extent	74 (17%)
To a Small Extent	21 (5%)
Not at All	7 (2%)
I Don't Know	18 (4%)

More respondents said that SAGE was helpful in retaining effective teachers to their school, as Table 21 shows. Nearly half of all respondents said that SAGE was effective in retaining effective teachers to a great extent, about three in 10 said it was effective to a moderate extent, 13 percent noted that it was effective to some extent, while 2 percent mentioned it was effective to a small extent. In total, 95 percent found SAGE effective in retaining effective teachers.

Table 21: Extent of effectiveness of SAGE to retain effective teachers to respondent’s school (n = 425)

	Number of Respondents
To a Great Extent	211 (50%)
To a Moderate Extent	129 (30%)
To Some Extent	54 (13%)
To a Small Extent	9 (2%)
Not at All	16 (4%)
I Don’t Know	6 (1%)

SAGE Funding and Resources

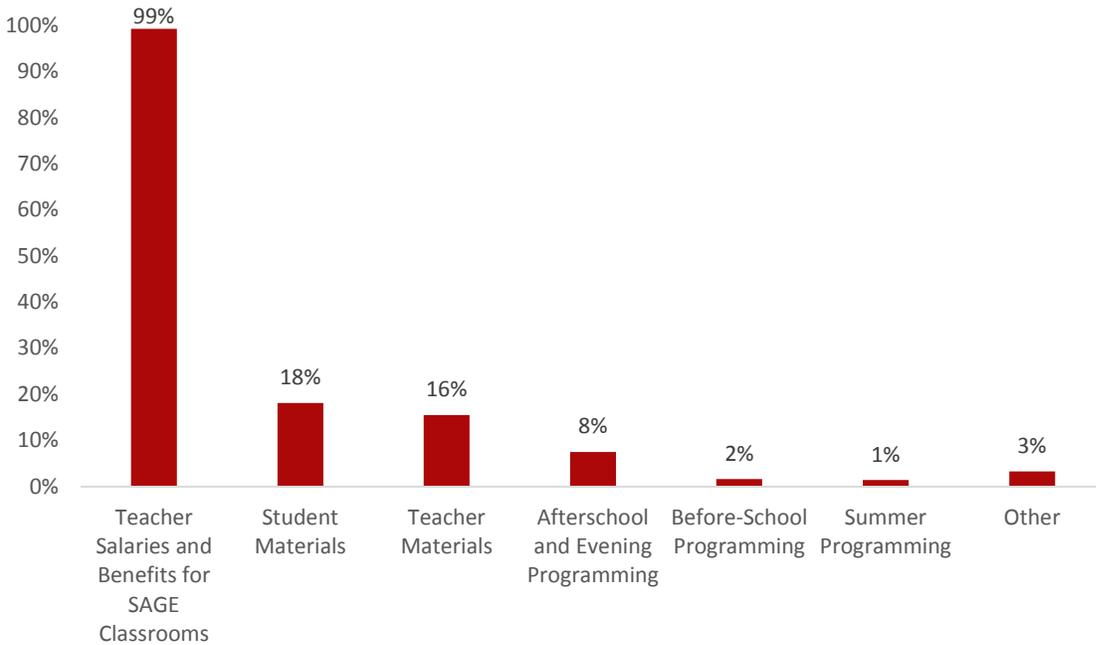
Schools that participate in SAGE receive \$2,250 for each student in grades K-3 who qualifies for free and reduced price lunch. Thus, schools that implement SAGE receive different amounts of funding from DPI depending on the number of eligible students in their SAGE classrooms. Table 22 shows how much of the state funding for the 2013-14 school year covered SAGE program expenses for schools. More than one-third said that SAGE funding from the state covered 76-100 percent of their program expenses, while just over one-quarter said it covered 51-75 percent of the program expenses. About 9 percent of respondents said SAGE state funding covered 26-50 percent of their expenses, and 27 percent said state funding covered 0-25 percent of their program expenses.

Table 22: Proportion of school SAGE expenses covered by state funding for 2013-2014 (n = 425)

	Number of Respondents
76-100%	155 (37%)
51-75%	114 (27%)
26-50%	40 (9%)
0-25%	116 (27%)

So where did SAGE schools spend their SAGE program funding? Figure 19 shows the most common answers to this item on the survey. Nearly all schools spent their SAGE state funding on teacher salaries and benefits for SAGE classrooms. Nearly one in five schools spent their money on student materials and/or teacher materials, whether for SAGE classrooms only or for all classrooms in their school. Respondents also mentioned that they spent SAGE funding on afterschool and evening programming (8 percent), summer programming (1 percent), and before-school programming (2 percent); 3 percent of respondents provided information on other resources funded by state monies for SAGE including fieldtrips (5 respondents), and paraprofessionals and paraeducator support (2 respondents).

Figure 19: Resources SAGE funding supported in respondent's school (n = 425)



Other includes (number of respondents who mentioned the activity/resource):

- Fieldtrip transportation and opportunities (5)
- Paraprofessional and paraeducator support (2)
- Extension of classroom activities (1)
- Transportation (1)
- Mobile unit rental (1)
- SMART Board technology (1)
- Curriculum work for SAGE teachers (1)
- Technology (1)
- Tutor salaries (1)
- Substitutes for planning period (1)

For those schools that spent their SAGE state funding on classroom teacher salaries and benefits, over one-third said that their state funding covered 76-100 percent of SAGE teacher pay. Table 23 shows that fewer than three in 10 respondents said that state funding for SAGE covered 51-75 percent of their teacher salaries and benefits, 14 percent said state funding covered 26-50 percent of their expenses, and 22 percent of respondents said that state funding covered 0-25 percent of their SAGE teacher salaries and benefits. These percentages are similar to those in Table 22.

Table 23: Proportion of SAGE classroom teacher salaries and benefits covered by state funding for 2013-2014 (n = 425)

Number of Respondents	
76-100%	148 (35%)
51-75%	123 (29%)
26-50%	61 (14%)
0-25%	93 (22%)

Table 24 shows responses to an item on the survey about whether some SAGE schools struggled to find space for smaller classrooms. Nearly all respondents, 97 percent, said that their schools had enough classroom space to fulfill SAGE classroom requirements (class sizes with a ratio of 18:1 or 30:2). However, eight respondents (2 percent) said that their schools struggled with finding classroom space for SAGE classrooms.

Table 24: Whether or not respondent’s school had enough space to fulfill SAGE classroom requirements (n = 425)

Number of Respondents	
Yes	415 (97%)
No	8 (2%)
I Don’t Know/Not Applicable	2 (1%)

Of those schools that did not have enough space to fulfill SAGE classroom requirements, five (1 percent) said that the limited space in their school restricted the decisions they could make around student learning activities in their school, as Table 25 shows. Although this problem affected very few schools, for some schools, the SAGE program class size requirements impacted the learning activities that their students participated in at school.

Table 25: (If not enough space) Whether or not limited classroom space restricted decisions for student learning activities (n = 425)

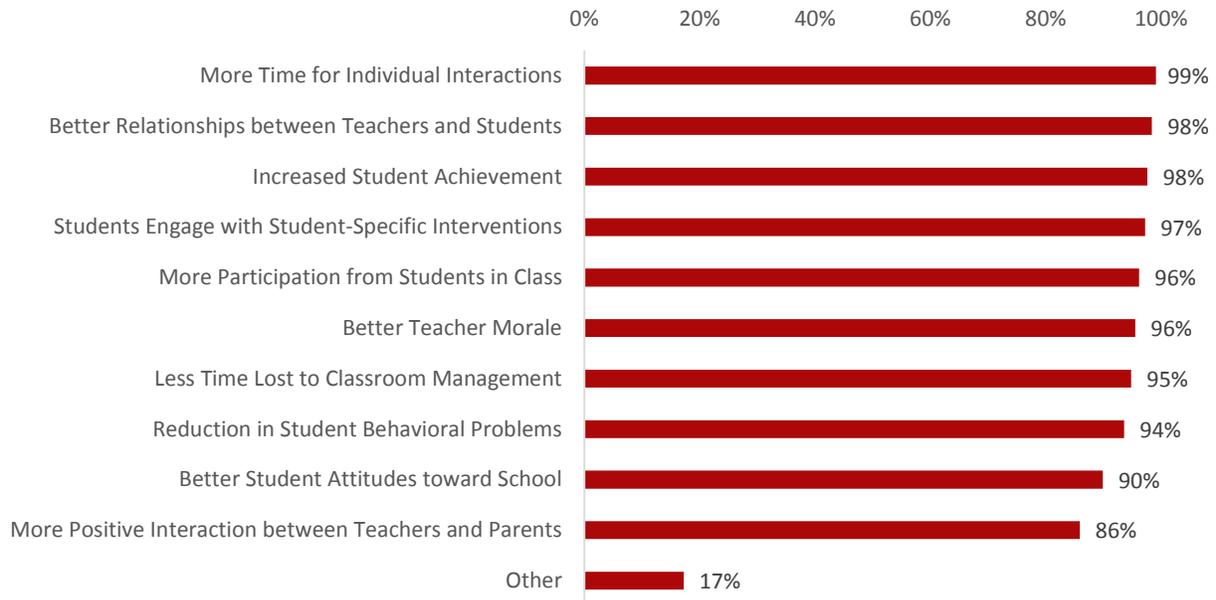
Number of Respondents	
Yes	5 (1%)
No	72 (17%)
I Don’t Know/Not Applicable	348 (82%)

Non-Achievement Outcomes

Much of the SAGE evaluation focuses on learning outcomes for SAGE students on standardized assessments. This section of items on the survey asked about non-achievement outcomes of the SAGE program to better understand how SAGE impacts schools apart from standardized tests. As Figure 20 displays, nearly all respondents said that their schools were positively impacted by the SAGE program in many different ways. The most common impacts of the SAGE program included having more time for individual interactions (9 percent), having better relationships between teachers and students (98

percent), increasing student achievement (98 percent), and engaging students with student-specific interventions (97 percent).

Figure 20: Outcomes respondents saw from SAGE programming in their schools (n = 425)



Other outcomes and benefits (number of respondents who mentioned this outcome/benefit):

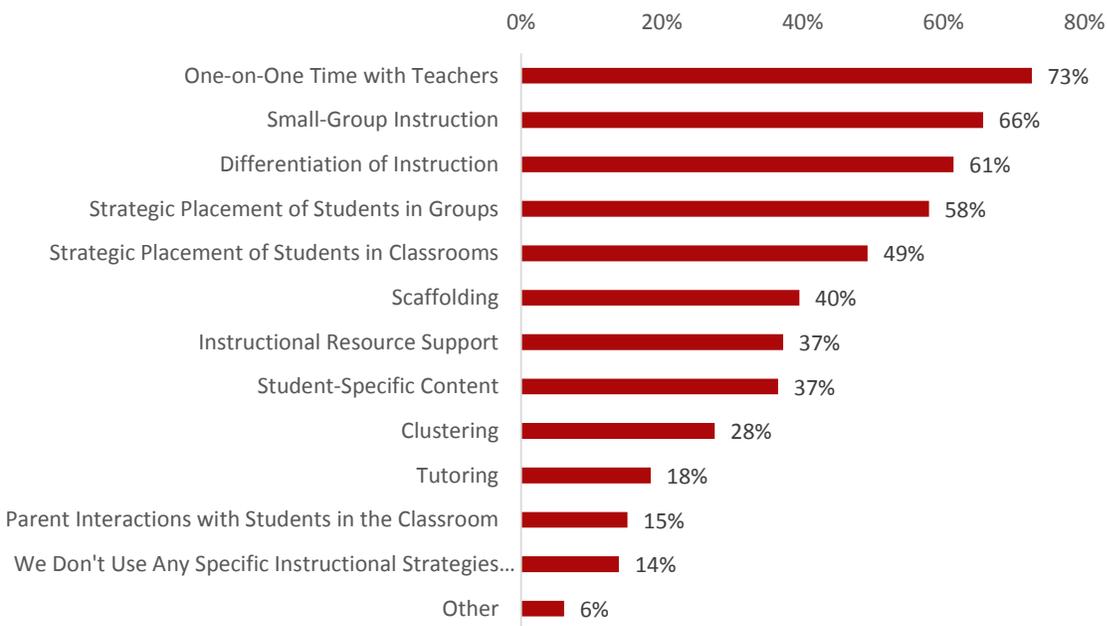
- Improved social interactions among students (5)
- More time for staff to work together (4)
- More leadership opportunities for students (2)
- Teachers know more about their students (2)
- Better intervention grouping of students (2)
- More classroom space (2)
- More time for progress monitoring (2)
- Space-sensory benefits (1)
- SAGE goal progress monitoring (1)
- More positive learning environment (1)
- More intervention time (1)
- More opportunities for activities outside of regular school hours (1)
- Better attendance (1)
- Ease of programming (1)
- More student risk-taking (1)
- More time for students to work with partners or small groups (1)
- More time for English language acquisition for non-native English speakers (1)
- Increased student morale (1)
- Smaller intervention groups (1)
- More space for learning stations (1)
- More paraeducator interaction with students (1)

- Teachers have more planning time (1)
- More time for interventions (1)
- Attracts open-enrollment students (1)
- More flexible scheduling (1)
- More positive public relations with the community (1)
- More opportunities for field trips (1)

Classroom Design and Instructional Strategies

To find out more about the instructional strategies SAGE schools used, this section of items on the survey asked about how smaller classroom sizes had changed the organization of and work within schools. Figure 21 shows common instructional strategies that respondents said their schools used. The most popular responses included one-on-one time with teachers (73 percent), small-group instruction (66 percent), and differentiation of instruction (61 percent) with smaller class sizes. However, Figure 21 shows many other strategies that SAGE schools used to take advantage of their smaller class sizes.

Figure 21: Instructional strategies teachers used with students in respondent’s school (n = 425)

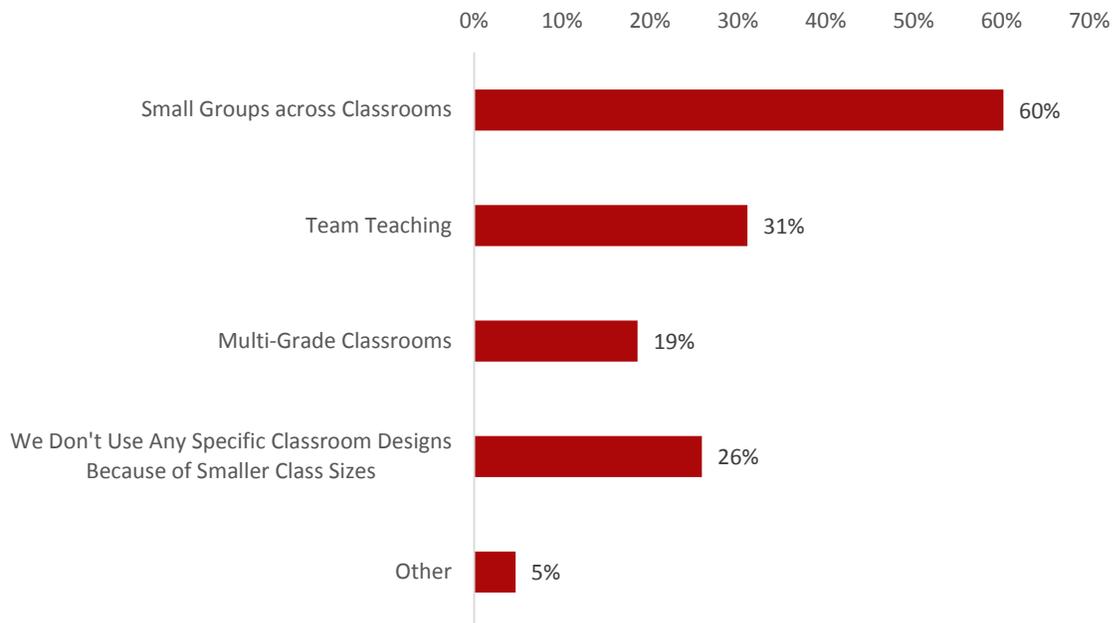


Other instructional strategies (number of respondents who mentioned strategy):

- Schedule Rtl blocks (2)
- Multiple small-group instruction with para support (1)
- Support from special education, speech and language, and ELL staff (1)
- Reading and math coach interventions (1)
- Strategic curriculum instruction and delivery planning (1)
- Intervention grouping (1)
- Gradual release of responsibility (1)
- Multi-age instruction (1)

SAGE schools used a variety of classroom designs with their smaller class sizes. Figure 22 shows that the most common classroom design was to create small groups across classrooms, with 60 percent of respondents mentioning that their schools used this classroom design in at least one grade level. Many also said that they used team teaching and multi-grade classrooms in their school, though it is not clear why more respondents said they used team teaching on this item compared to the second item on the survey (31 percent on this item vs. 25 percent in Table 16). It could be that the team teaching respondents mentioned in this item was more informal; that is, these were not necessarily classroom structures used throughout the day in classrooms, rather some teachers team taught for particular parts of the day more informally. Likewise, it is not clear why fewer respondents said they had multi-age classrooms in their school on this item vs. in Table 17 (where 25 percent said they had multi-age classrooms), though it could be that respondents interpreted these two items differently on the survey.

Figure 22: Smaller class size classroom designs in respondent’s school (n = 425)



Other designs (number of respondents mentioning the design):

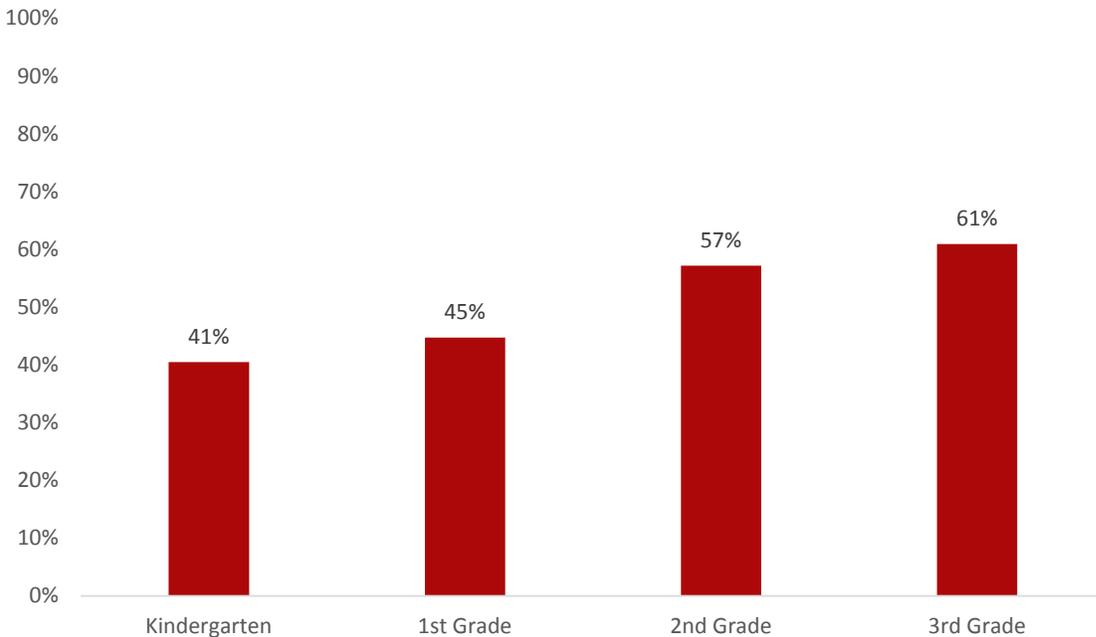
- Centers (2)
- Rtl grouping (2)
- Team teaching with special education teachers (1)
- More assessment time for each student (1)
- Leveled ability instruction (1)

Benchmark Assessments

Finally, in order to aid analysis for the SAGE program evaluation, an item on the survey asked respondents about the use of benchmark assessments. Figure 23 shows that schools use benchmark assessments increasingly with higher grade levels. Only 41 percent of SAGE schools used benchmark

assessments in kindergarten, though that proportion rose to 61 percent by third grade. These results were different from the proportions of students tested within each grade that resulted in the student sample used in the quantitative analysis.

Figure 23: Grades in which students took assessments for benchmarking purposes (n = 425)



VII. Conclusions

This report covers the evaluation work completed by VARC to understand the impact and outcomes of the SAGE program throughout Wisconsin. At the outset of this year 10 evaluation of the SAGE program, VARC’s goal was to answer several quantitative and qualitative research questions. This section of the report summarizes the results from the analyses in an attempt to answer these questions. Following this, VARC provides a summary of the limitations to this year’s analyses.

Summary of Results

This first portion of the quantitative evaluation utilized results from the MAP assessment along with demographic data to analyze the impact of the SAGE program on improving student growth in both mathematics and reading. Because not all schools throughout the state utilize the MAP assessment, VARC first examined how the tested population compared to the general population of SAGE and non-SAGE students throughout the state. VARC found larger proportions of African-American and economically disadvantaged students and fewer proportions of white students in the tested population than in the general population. There were also larger proportions of tested students in city and suburban settings than in town or rural settings. This indicates possible limited generalizability of the MAP analysis results to the larger population of all SAGE students.

VARC also examined the variance in characteristics between SAGE and non-SAGE schools. SAGE schools in the analysis had a higher proportion of African-American students, a higher proportion of economically disadvantaged students, a lower proportion of white students, and a lower proportion of students living in suburban areas. Given these differences, both sets of MAP analyses used statistical modeling to mitigate the differences between the SAGE and non-SAGE populations to create comparable groups of students.

Results from the most recent year of MAP assessment analyses found positive and significant effects of the SAGE program on reading growth in kindergarten and positive but insignificant effects of the SAGE program on mathematics growth in kindergarten. Results from multiple years of MAP assessments examining the impact of the SAGE program over the full four years of possible intervention found positive and significant effects in both mathematics and reading growth.

Results from an examination of the lasting impacts of SAGE into high school found that students who were in SAGE had higher probabilities of not dropping out of school in ninth and tenth grade as compared to similar students from the same schools before SAGE. The evaluation also found that African-American students that qualified for free or reduced price lunch and were in SAGE schools had higher probabilities of completing high school as opposed to dropping out at any point in high school compared to students from the same schools before SAGE.

From the results of the End-of-Year Report survey, it appears that for many SAGE schools, the SAGE program has provided many benefits to students, including some flexibility with classroom design (for example, multi-age classrooms and team teaching) as well as benefits for students unrelated to standardized testing (such as having more individual time to work with students) and for teachers in recruitment and retention. Although most SAGE schools used their state funding to pay for salaries and benefits for SAGE classroom teachers, for many schools, state SAGE funding did not cover the full expense of the program, at least in the 2013-14 school year. Despite this lack of coverage from SAGE funding, many schools continue the program and accrue a variety of benefits from their participation.

Evaluation Context

Each year, VARC and DPI work collaboratively to find ways to improve both the quality of the data and the statistical methods. The goal is always to make the estimates of the SAGE effects as meaningful as possible; this section describes the context of the estimation in order to inform users about these data and provide cautions around the use of the estimates in any decision-making processes.

As explained previously, in an attempt to address the bias arising from the selection of schools receiving SAGE funding, VARC applied a propensity score matching technique for SAGE to estimate the effects of class size reduction. This statistical matching technique accounts for the variables that predict receiving the treatment in order to reduce the bias found when simply comparing learning gains outcomes of schools that received the treatment to the schools that did not.

This bias arises because the learning outcomes between SAGE and non-SAGE schools depend on poverty levels, which also affects whether a district or a school is either eligible for SAGE funding or has financial

incentives to afford SAGE. Only a large-scale randomized experiment would generate an unbiased estimation of SAGE funding. Since the state does not allocate SAGE funds randomly, matching attempts to mimic randomization by creating a set of schools that participate in SAGE that is comparable on observable characteristics to a set of non-SAGE schools.

By design, SAGE is an initiative reducing class size for schools with large proportions of lower income students. The current data available only allow for a comparison in performance between schools receiving SAGE funding and those not receiving SAGE funding without controlling for class size. In reality, some schools categorized as non-SAGE meet the SAGE class size requirements. This can result from such schools making class size reduction a priority when allocating their funds. The presence of these low-class size schools leads to a contamination of the control group and likely a downward bias of the estimate of SAGE effects.

This year's evaluation examined a new set of data from DPI's Coursework Completion Data Collection. These data submitted by districts were sufficiently different from SAGE submission reports from SAGE schools in definitions, purpose, and timing, that the data did not yield useful information for the purposes of this evaluation's analyses.

Overall, it is difficult to estimate the effect of the SAGE program given the fact that SAGE schools receive funding because they are facing adverse conditions that comparison schools are not. While VARC uses a variety of methods to alleviate some of these concerns, and finds positive effects or small differences showing the success of the initiative on improving test scores, ultimately VARC could see different differences in achievement if VARC had better information from both SAGE and control schools. Thus, VARC will continue to collaborate with DPI to improve data sources for future evaluations.

Appendix: Technical Specifications

This appendix describes the statistical methods and data used by VARC to provide DPI with an estimate of how the SAGE program affected schools’ contribution to students’ academic performance and high school outcomes.

Single-Year MAP Assessment Analysis

For each district, DPI provided VARC with data describing students, schools, and, when available, student MAP test scores. Conceptually, the analysis used statistical techniques to isolate the component of measured student knowledge that is attributable to the schools’ productivity from other factors such as prior knowledge and student characteristics, and provide estimates on SAGE program effectiveness on productivity of schools by only comparing comparable schools, with comparability defined by statistical methods and by available data.

Analysis Data Set

In a given school year, VARC created an analysis data set for each grade and subject. Since the SAGE program is for kindergarten through third-grade students, in any given year, eight different data sets were constructed. For each student, each analysis data set included a grade level, two pretests and a posttest, a value for each of the control variables used in the model (described below), the ID of the school attended, and whether the school received SAGE funding for that grade. For each pretest, a measure of the standard error of measurement was also required.

Table A1 shows the number of students in the analysis set in both SAGE and non-SAGE schools by grade and Table A2 shows the number of schools in the analysis set by SAGE status and grade.

Table A1: Number of SAGE and non-SAGE students used in the main analysis and proportion of total students by grade in 2013-14

Grade	SAGE Students in Analysis	% of All SAGE Students	Non-SAGE Students in Analysis	% of All Non-SAGE Students	Total Students in Analysis	% of All Students
K5	3991	19%	5291	13%	9282	15%
1 st	4937	23%	8857	21%	13794	22%
2 nd	7461	35%	16724	41%	24185	39%
3 rd	9897	49%	19395	47%	29292	47%

Table A2: Number of SAGE and non-SAGE schools used in the main analysis and proportion of total schools by grade in 2013-14

Grade	SAGE Schools in Analysis	% of all SAGE Schools	Non-SAGE Schools in Analysis	% of All Non-SAGE Schools	Total Schools in Analysis	% of All Schools
K5	82	20%	99	14%	181	16%
1 st	104	25%	168	23%	272	24%
2 nd	160	39%	302	42%	462	41%
3 rd	207	50%	343	47%	550	48%

Pretest and Posttest Scores

For each grade in both reading and mathematics, the fall MAP scores in RIT scale points were the student's pretest scores, or the measures of the student's level of academic knowledge before receiving instruction from their teacher and school that year. Likewise, for each grade in both reading and mathematics, the spring MAP score in RIT scale points was the student's posttest, or the measure of the student's level of academic knowledge after receiving instruction from their teacher and school that year. The test scores dataset was the most restrictive because the MAP assessment is not mandatory in Wisconsin; thus, only schools choosing to administer the MAP test could be included in the analysis. Moreover, in order to avoid problems due to schools using the assessment for purposes other than benchmarking, VARC used MAP data only when in given grade in a given school, at least 80 percent of the students took the test. Lastly, because value-added estimation requires a pretest and posttest score for each student, VARC kept students with values for both.

Standard Errors of Measurement of Pretest Scores

VARC included standard errors of measurement (SEMs) provided by the test vendor NWEA in order to correct for measurement error associated with each pretest. Each score in the same grade and subject had an associated SEM. Estimating the measured student achievement without controlling for pretest measurement error yields biased estimates of all parameters, including the SAGE effect coefficient. Estimating the desired parameters can be consistent if external information is available on the variance of measurement error for prior achievement; Fuller's (1987) *Measurement Error Models*³ describes approaches for consistent estimation in the presence of measurement error.

Biographical Student Variables

Gender, race, free or reduced price lunch status, and disability status came from the biographical dataset. Gender categories were male and female. Race categories were Asian, Black, Hispanic, American Indian, and white. If a student had a disability, VARC assigned the disability status dummy variable a value of one; it is zero otherwise. Likewise, if a student was eligible to receive free or reduced price lunch, VARC assigned the poverty dummy variable a value of one, zero otherwise. Finally, if a student was an English language learner, VARC assigned the ELL dummy variable a value of one, zero otherwise. Students with no missing value for each variable were kept in the analysis.

³ Fuller, W. A. (1987). *Measurement Error Models*. *Wiley Series in Probability and Statistics*.

Value-Added Regressions

Once VARC constructed the analysis datasets, the evaluation estimated school value-added coefficients using an ordinary least squares (OLS) regression with measurement error correction for each grade and subject. Formally, we can write the equation as

$$Y_2 = \lambda_1 Y_1 + \lambda_2 Y_{alt1} + X\beta' + S\alpha' + e \quad (0.1)$$

Where,

- Y_2 is a vector of posttest scores. For each subject and grade, the spring MAP assessment is used.
- Y_1 is a vector of pretest scores in the same subject as the posttest. For each subject and grade, the fall MAP assessment is used.
- Y_{alt1} is a vector of pretest scores in the other subject than the posttest. For each subject and grade, the fall MAP assessment is used.
- X is a matrix of student characteristics. It includes gender, race, English language learners, free or reduced price lunch, and disability.
- S is a vector of school dummy variables. Each line represents a student observation and each column a school. If student i is in school k , then the dummy equals 1, and 0 otherwise.
- e is the error term.

VARC ran eight regressions, one for each grade and subject, taking into account the pretests measured with error. From each regression, VARC obtained a vector of $\hat{\alpha}$ providing an estimation of each school's performance measured in test score gains in reading and mathematics, at each grade. Hence, each school has eight scores.

The estimated coefficient $\hat{\alpha}_{kg}$ for each school k and each grade g was then centered so that the estimates had a dose-weighted mean of zero. Formally,

$$\hat{\alpha}_{kg}^{centered} = \hat{\alpha}_{kg}^c = \hat{\alpha}_{kg} - \frac{\sum n_{kg} \cdot \hat{\alpha}_{kg}}{\sum n_{kg}} \quad (0.2)$$

To simplify notation, we will further refer to the centered estimate as $\hat{\alpha}_{kg}$.

Propensity Score Matching

Propensity score matching (PSM) employs a predicted likelihood of a school participating in SAGE based on observed predictors. It uses a logistic regression to estimate these likelihoods.

Since VARC wanted comparability at the school level (for grades K through 3) where the SAGE program was relevant, rather than grade level, the unit of observation in the analysis data sets for propensity score estimation was a unique school (unlike the analysis data for value-added estimation where unit of observation was a unique student.) The steps to create the PSM analysis data sets were as follows:

- Drop observations that were not in kindergarten through third grade.

- Calculate school level (K-3) percentages of each demographic characteristic of the student body (percentage of students who were eligible for a free or reduced price lunch, English language learner, and special education/disability status and percentage of a given race or gender). Once VARC computed percentages for the free or reduced price lunch (FRL) variable, VARC created new variables by computing the corresponding squared and cubic values. This variable transformation was done in order to account for a possible non-linear relationship between the percentage of FRL in a school and the likelihood of that school being in SAGE.
- Import the urban/rural indicator for each school.
- Import the average teacher salary in \$1,000 intervals (including fringe benefits) for each district. Formally, average teacher salary = (average teacher salary + average fringe)/1000

Formally, the equation for the logistic regression

$$\Pr(SAGE = 1|X = x) = \frac{e^{\beta'x}}{1 + e^{\beta'x}} \quad (0.3)$$

where X is the school characteristics defined above. The results of the regression generate estimates of likelihood of a school being a SAGE school based on the school's characteristics.

Using estimated propensity scores, we used a one-to-one matching algorithm with replacement in order to estimate effectiveness of the SAGE program on the productivity of schools. In one-to-one matching, each SAGE school was matched with a non-SAGE school whose propensity score was closest to the propensity score of the SAGE school. By allowing replacement, multiple SAGE schools could be matched with the same non-SAGE school to provide a better group of matches. Specifically, VARC used the `psmatch2` algorithm in Stata to match the SAGE schools by grade.

Weighted Linear Regressions

Lastly, VARC used the matched sample of schools to perform a weighted OLS regression for each grade and subject to find the SAGE estimate. Formally, we can write this equation as

$$\hat{\alpha}_{kg} = \gamma SAGE + e \quad (0.4)$$

Where,

- $\hat{\alpha}_{kg}$ is vector of the estimates of school value added as calculated in equations 0.1 and 0.2.
- $SAGE$ is a SAGE dummy variable. Each line represents a school. If school k is in SAGE, then the dummy equals 1, and 0 otherwise.
- e is the error term.

VARC again ran eight regressions, one for each grade and subject, while weighting these regressions by a factor of both the number of students in the school as well as that school's propensity score. This weighting placed a higher value on those schools with more students and those schools with a higher likelihood of participation in SAGE given their characteristics. The coefficient γ was the SAGE impact estimate.

Multiple-Year MAP Assessment Analysis

For each district, DPI provided VARC with data describing students, schools, and, when available, student MAP test scores. Conceptually, this analysis was similar to the single-year MAP analysis, using statistical techniques to isolate the component of measured student knowledge that is attributable to the schools' productivity, and providing estimates of SAGE program effectiveness on productivity of schools by only comparing comparable schools; however, this analysis looked across multiple years of data to identify a cumulative SAGE effect.

Analysis Data Set

For this analysis, VARC created one analysis set for each subject. Each analysis set included two pretests and a posttest, a value for each of the control variables used in the model, the ID of the school attended in the 2013-14 school year, and whether the school received SAGE funding for that grade. For each pretest, a measure of the standard error of measurement was also required. In addition, a student was only included in the analysis set if they made normal grade progression and either attended a SAGE school for all four years of the analysis period or attended a non-SAGE school for all four years. Students could change schools and districts, but only if they did not move between SAGE and non-SAGE schools. All students were attributed to their 2013-14 school for the purposes of estimating the SAGE effect. Students who attended a school that opted out of SAGE for second or third grade were excluded from the analysis.

Table A3 shows the number of students in the analysis set in both SAGE and non-SAGE schools and Table A4 shows the number of schools in the analysis set by SAGE status.

Table A3: Number of SAGE and non-SAGE students used in the multiple-year analysis and proportion of total students in 3rd grade in 2013-14

SAGE Students in Analysis	% of All SAGE Students	Non-SAGE Students in Analysis	% of All Non-SAGE Students	Total Students in Analysis	% of All Students
2024	10%	1490	4%	3514	6%

Table A4: Number of SAGE and non-SAGE schools used in the multiple-year analysis and proportion of total schools in 2013-14

SAGE Schools in Analysis	% of all SAGE Schools	Non-SAGE Schools in Analysis	% of All Non-SAGE Schools	Total Schools in Analysis	% of All Schools
91	22%	103	14%	194	17%

Pretest and Posttest Scores

As with the single-year analysis, the 2014 spring MAP score in RIT scale points was the student's posttest. Instead of using same-year fall pretests, however, this analysis used fall 2010 MAP tests in RIT scale points so that the test data spanned four full years of instruction. VARC dropped students with missing pretest or posttest data and excluded students who attended schools where less than 80 percent of third-grade students took the MAP assessment in the 2013-14 school year.

Biographical Student Variables

Gender, race, free or reduced price lunch status, and disability status came from the biographical dataset and were defined the same way in the single- and multiple-year analyses. VARC dropped students missing a value for any of these variables from the analysis.

Value-Added Regressions

Once VARC constructed the analysis datasets, the evaluation estimated school value-added coefficients using an ordinary least squares (OLS) regression with measurement error correction for each subject, as outlined in the single-year analysis section above.

Propensity Score Matching

For comparability across schools, VARC also used a PSM design to match schools with SAGE students to schools without SAGE students. These matches were made on the third-grade schools and used a similar design to the PSM technique used in the single-year analysis.

Weighted Linear Regressions

Finally, VARC used the matched sample of schools with students with four years of SAGE and students with no SAGE in a weighted OLS regression for each subject to estimate the SAGE impact over four years. This regression design was similar to the one used in the single-year analysis.

Long-Term Outcomes Analysis

The long-term outcomes analysis differs significantly from both MAP assessment analyses. For this analysis, DPI provided VARC with historical Wisconsin Reading Comprehension Test (WRCT) data from 1998-2004; historical data from Wisconsin's longitudinal student data system (ISES) beginning in the 2003-04 school year; and a crosswalk of student names, birth dates, and ID numbers.

Analysis Sets

For each student, the analysis set contained

- Student biographical variables
- WRCT cohort year
- First 9th-grade observation year
- Final ISES observation year
- School in which they were enrolled for 3rd grade
- An indicator of whether the school received SAGE funding at the time of their 3rd-grade enrollment
- Exit code and, if applicable, completion credential type from their final ISES observation
- A set of indicators identifying their completion status based on their exit code

Records in the WRCT datasets from 1998 through 2004 were matched with all available ISES records from 2004 to 2014 using names, birth dates, and student IDs from the crosswalk. No restriction was placed on the final ISES observation year beyond the requirement that the student have a 9th-grade observation. Students with missing WRCT, crosswalk, and/or 9th-grade ISES data were excluded from the analysis, as were students who remained enrolled and neither eventually completed high school nor dropped out. This analysis included students who completed high school regardless of how long it took

the student to complete. The analysis also excluded students from schools that joined the SAGE program prior to 2000-01 to maintain an unbiased control group.

Outcome Measures

Based on their exit codes and completion credential types, students could be classified into several categories – including high school completers, continuing students, dropouts, and ineligible students – based on their final ISES observations. A student was classified as a high school completer if

- The student had an exit code of “HSC” with any credential type recorded within four years of their first 9th grade ISES observation

A student was classified as continuing if

- The student had an exit code of “HSC” with any credential type recorded more than four years after their first 9th grade ISES observation
- The student had an exit code of “TC” or “NA”
- The student had an exit code of “MA”

A student was classified as a dropout if

- The student had an exit code of “ETC”, “ODO”, “OOS”, or “ISM”

A student was classified as ineligible and removed from the cohort if

- The student had an exit code of “TNC”, “TOS”, “INM”, “BCA”, or “DE”

VARC focused on the dropout classification for this analysis. VARC classified the student as a dropout if the student’s final ISES observation showed that

- The student was supposed to transfer to another WSLs school but did not re-enroll at the destination school,
- The student moved to another state but was not known to re-enroll in the destination state, or
- The student was assigned an “other dropout or possible dropout” code in ISES.

Biographical Student Variables

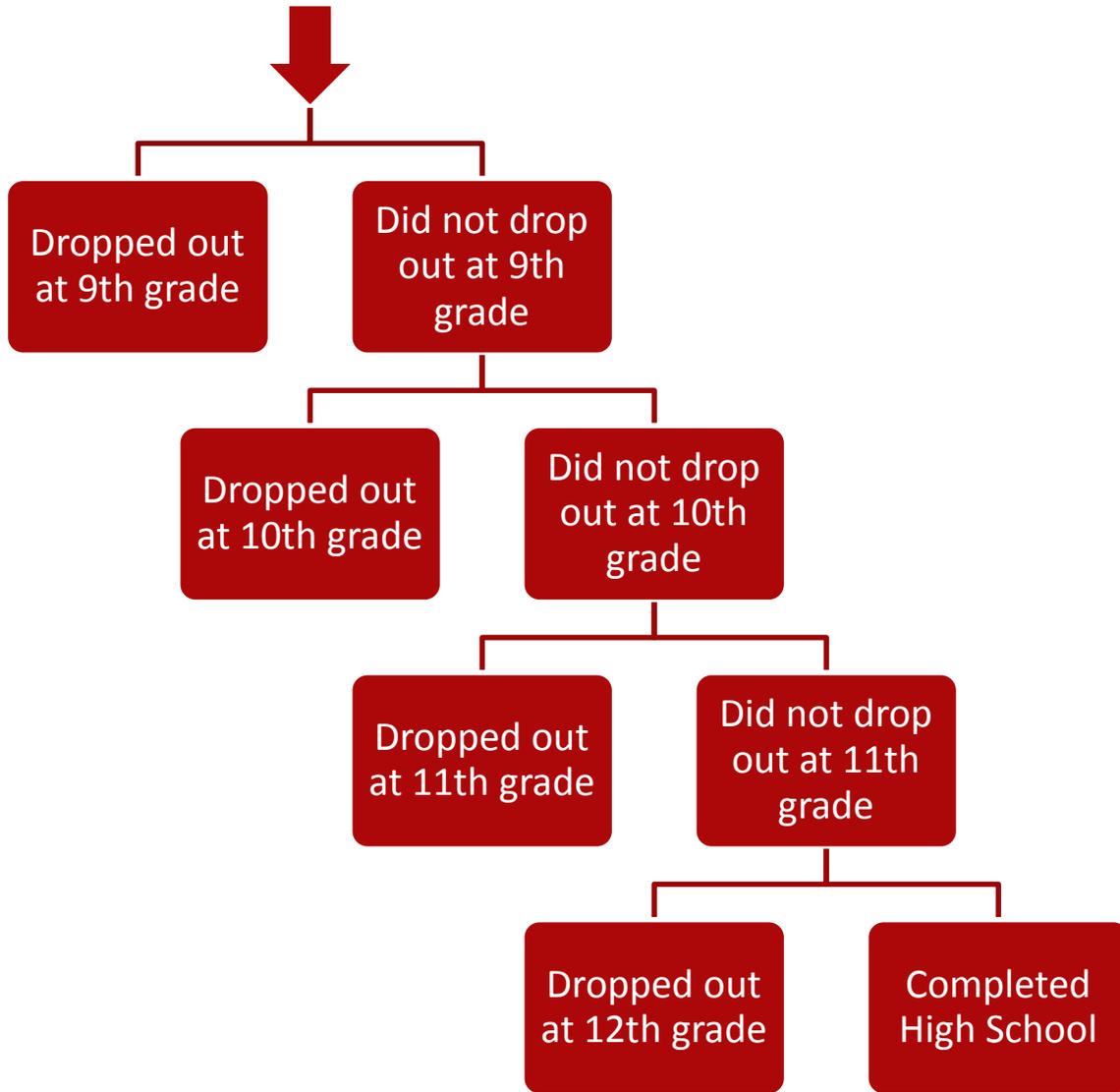
Gender, race, free or reduced price lunch status, and disability status came from the biographical dataset and were defined similarly in the long-term outcomes analysis and in the single-year and multiple-year analyses. VARC dropped students missing a value for any of these variables from the analysis.

Sequential Logit

To analyze the dropout rates, VARC utilized a sequential logit approach taking into account the fact that dropping out at a particular grade requires a student not to drop out prior. Figure A1 shows this sequence. While the analysis did have access to information related to whether a student proceeded to the next grade if he/she did not drop out, this factor was not included in the analysis due to the additional complexity. If the analysis included information on this, VARC would have split the “did not drop out” branches into two pieces in the tree shown Figure A1; one for proceeding to the next grade and another for repeating grade. Notice that in this case, the tree becomes quite complex since the “repeating grade” branch is essentially the copy of the main tree itself. Because of this complexity, VARC preferred the simplified tree presented in Figure A1, since a repeating student would eventually either

drop out at that grade or continue to the next grade. Hence, for example if a student repeated grade 9, and dropped out at 10th grade, the data shows him as he "dropped out at 10th grade." In the analysis data, VARC only kept students who eventually dropped out or completed high school. In other words, we did not include students who were still in school in our analysis time span or students who reached the maximum age. Hence, not including information on whether a student who did not drop out is repeating the grade or proceeded to the next grade was not overly detrimental to the analysis. However, it does affect the interpretation of results, since the analysis did not include whether a student completed high school in four years or more, whether they dropped out in 10th grade after a year of schooling after 9th grade or more, and so on.

Figure A1: Sequential Logit Diagram



In order to estimate the SAGE program effect on dropout rates, we ran a conditionally independent logistic regression at each node of the tree in Figure A1. Starting from the last node, in the subset of data that included only students who completed high school or dropped out at 12th grade, VARC ran a logistic regression. In this regression, the independent variable was whether they completed high school or dropped out at 12th grade and explanatory variables were

- whether the student took the WRCT in 3rd grade in a SAGE school after 2000 (SAGE program effect coefficient);

- whether the student took the WRCT in 3rd grade between 1998 and 2004 in a school that participated in the SAGE program after 2000;
- gender, race, English language learner status, free or reduced price lunch eligibility, and disability status;, and
- 1998-2004 year binary variables.

VARC repeated this logistic regression for the other three nodes. For example, the 11th-grade dropout and not dropping out analysis was done on a subset of data consisting of students who dropped out at 11th grade, dropped out at 12th grade, and completed high school.

To calculate the probability of not dropping out at any grade, i.e. not dropping out at 9th grade and 10th grade and 11th grade and 12th grade, using the independence assumption. Thus, we multiplied probability of not dropping out in each node for given characteristics of students for each case of SAGE school participation and year. In Figure 17, students are white, non-special education, non-ELP, free lunch status males. The figure shows the predicted probabilities of not dropping out in any grade, hence the probability of graduating eventually. The marginal effect of attending in a SAGE school on probability of eventually graduating then can be calculated by differentiating the probability with respect to the SAGE indicator. One then calculate the standard error on this marginal effect by using delta method. Table 13 presents these marginal effects and their standard error for a few different combination of student characteristics.

This analysis required conditional independence between consecutive nodes. That is, conditional on not dropping out at 9th grade, the likelihood of dropping out 10th grade or not does not depend on the characteristics of the students (and of the schools that they attended) who dropped out at the 9th grade. This assumption might be violated if explanatory variables were not rich enough. To see this, consider the null hypothesis that the students who attended the SAGE schools were less likely to drop out than their equivalent peers in non-SAGE schools in each grade. Under this null hypothesis, the students who dropped out at the 9th grade were more likely to be from non-SAGE schools if all the students were equivalent other than SAGE and non-SAGE school participation. Hence, if the dropout rate of non-SAGE students was less than observationally equivalent to SAGE students at 9th grade, then they must have had some unobservable (or unobservable to the researcher) characteristics that reduced their dropout rate. This kind of violation of the conditional independence would cause the model to underestimate the effect of SAGE program.