VARC

VALUE-ADDED RESEARCH CENTER

UNIVERSITY OF WISCONSIN–MADISON

Education analytics to support students and educators

SAGE Program Evaluation Final Report

August 2014

Evaluation Team

Dr. Robert Meyer (Principal Investigator)

Penny Clark (Project Manager)

Dr. Emin Dokumaci (Economist)

Grant Sim (Project Lead)

Clarissa Steele (Researcher)

Jenn Vadas (Researcher)

Report Authorship

Dr. Emin Dokumaci

Grant Sim

Jenn Vadas

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

Executive Summary

The Student Achievement Guarantee in Education (SAGE) program is an initiative of the Wisconsin Department of Public Instruction (DPI) that provides funds to Wisconsin schools based on the number of free or reduced price lunch students for a strict guarantee to maintain small class sizes. The program specifies 18 students or fewer in single-teacher classrooms and 15 students or fewer per instructor in multi-teacher classrooms in kindergarten through third grade.

This report presents the evaluation work completed by the Value-Added Research Center (VARC) to measure the impact of the SAGE program. 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 within different sub groups, namely differential effects analyses; and
- Analyses of student academic growth comparing students in SAGE schools to students in non-SAGE schools over the course of multiple years, which serve as a prelude to the future analyses of the long-term effects of the SAGE program.

When comparing characteristics of students in SAGE versus non-SAGE schools, the Value-Added Research Center (VARC) observes 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 utilized a statistical method 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 yield:

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

Since the goal of the SAGE program is to improve the academic performance of economically disadvantaged students, VARC also statistically analyzed the differential impact of the SAGE program on students receiving a free or reduced price lunch. Results from the statistical analyses include:

- An estimated positive differential effect of the SAGE program on mathematical academic growth in schools with large proportions of economically disadvantaged students in first grade and third grade, and
- An estimated positive differential effect of the SAGE program on reading academic growth in schools with large proportions of economically disadvantaged students in kindergarten through third grade.

New to this year's evaluation, VARC added analysis of the effect on academic performance in reading and mathematics skills of students who were in the SAGE program over the course of multiple years. The first set of results from the statistical analyses shows:

- An undetectable effect of the SAGE program on mathematical academic growth from kindergarten through second grade, and
- A positive effect of the SAGE program on reading academic growth from kindergarten through second grade.

In order to improve the reliability of the estimates, VARC and DPI will continue to work collaboratively to address the issues related to the presence of small class sizes in non-SAGE schools[1] as well as the non-random selection of schools into the SAGE program.

[1] Lack of classroom size data in non-SAGE schools prevented VARC from attempting to control for it.

Table of Contents

I. Introduction

This report presents the ninth 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 1996-97 school year. The program allocates funds to 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 individualized instruction necessary to promote the educational development of students in early grades.

Historically, evaluations of SAGE lacked valid and reliable early-grade achievement data. However, districts have been increasingly recognizing the importance of testing students in early grades and are finding benchmark student testing useful when done at both the beginning and end of the school year. Such recent developments and trends in Wisconsin assessment systems have made it possible for VARC to evaluate the impact of SAGE on student achievement. While several benchmark assessments are available, by far the most prevalent assessment in Wisconsin is the Measures of Academic Progress (MAP) provided by the Northwest Evaluation Association (NWEA). The MAP generates scale scores called RIT scores to estimate achievement based on the specific Item Response Theory approach called Rasch Modeling.[2] RIT scores range from 100 to 320. This evaluation conducted by VARC uses MAP test scores for several main reasons:

- Districts widely use the MAP assessment across the state and this allows VARC to expand the evaluation geographically beyond the Madison and Milwaukee school districts. This, in turn, allows the examination of the impacts of the program across different types of districts.
- Districts typically administer the MAP assessment at least twice per year, once in the fall and once in the spring. This allows for examination of the impact of SAGE without including the academic loss or gain happening during the summer, out of the schools' control. This is a major improvement when VARC suspect that students in SAGE schools have different learning opportunities on average than non-SAGE students.
- Districts administer the MAP assessment in lower grades starting as early as kindergarten. This is a major benefit since the SAGE program runs from kindergarten through the third grade, while the Wisconsin Knowledge and Concepts Examination (WKCE), the state assessment, only starts in third grade.

The evaluation of the 2009-10 SAGE program was the first to make use of MAP data. The analysis used two-level (student and school) hierarchical linear models (HLM) to examine the impact of the SAGE program on student performance. The available student-level control variables were race,

[2] http://www.nwea.org/support/article/532/rit-scale

gender, and fall MAP scores; the availability of individualized education program (IEP) status, eligibility for free or reduced price lunch, and English Language Learner (ELL) status came in later years. The best methodology before VARC could include a broader set of student controls was to include school controls (percent free or reduced price lunch, percent minority, percent ELL, and percent IEP). For both reading and mathematics, the earlier analyses found positive but not statistically significant effects of the SAGE program. The magnitudes of the effect sizes were larger in earlier grades (approximately 0.1 to 0.3 standard deviations in kindergarten and first grade) for both reading and mathematics compared to later grades (approximately 0.05 standard deviations in second and third grade). These results suggest that the SAGE program may have a positive effect on both mathematics and reading achievement, but the analysis lacked sufficient power and controls to conclude if this effect was reliable.

The evaluation of the 2010-11 SAGE program improved in reliability by including a larger set of schools and through the feasibility of matching MAP data with state data. By matching MAP data with state data, this evaluation could include a more complete array of controls for isolating the SAGE effect at the student level. Furthermore, Milwaukee Public Schools began using the MAP as their benchmark assessment in 2010-11, which provided a larger analysis set. This analysis utilized two models to predict the impact of the SAGE program. The first model compared all SAGE schools to all non-SAGE schools within SAGE districts and controlled for student gender, race, free or reduced price lunch status, ELL status, IEP status, and both reading and mathematics baseline MAP scores. The first model estimated positive and significant effects of the SAGE program in first grade on both reading and mathematical academic growth. The second model compared SAGE schools to non-SAGE schools within the same school districts. This model used the same controls as the first model but also added in a district control. The second model estimated positive and nearly significant results for reading in kindergarten. Results from both models suggest that the SAGE program may have a positive effect on both mathematical and reading performance in the earlier grades.

The evaluation of the 2011-12 SAGE program differed in two main aspects. First, the analysis of MAP achievement in mathematics and reading included differential effects to determine the impact of the SAGE program on certain populations of students. Second, the year 8 evaluation also included a qualitative study to ascertain properties and impacts of the SAGE program beyond those that a quantitative evaluation can typically provide.

This year's evaluation allowed for a more comprehensive approach to the analysis through the increased availability of data and by building on previous results and methodology. 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. As a result, the reliability of the estimates continuously improves for the use of policymakers and educators. New to the year 9 evaluation, VARC:

- Obtained assessment data for students outside of SAGE districts;
- Utilized school matching to identify a control group; and
- Started a longitudinal analysis of the impact of SAGE over the course of multiple years.

This report has five main sections. The first section includes the main research questions. The second section provides background on SAGE demographics, the characteristics of schools utilizing the MAP assessment, and raw assessment results. The third section details the methodology of the evaluation including the proper selection of a control group and the analysis design models. The fourth section examines the results of the evaluation. Finally, in order to provide decision makers with a full understanding of the meaning of the estimation results, the fifth section includes a summary of findings, a list of data restrictions, and future directions to continue to improve the statistical analyses.

II. Research Questions

The year 9 analysis of the SAGE program evaluated the 2012-13 school year and a combination of the 2010-11, 2011-12, and 2012-13 school years to determine the impact of the SAGE program on student performance in mathematics and reading. To provide a foundation for this quantitative analysis, this evaluation had seven research questions:

1. What are the characteristics of SAGE students and schools?
2. What is the take-up rate of the MAP in SAGE districts and schools?
3. How does the set of schools using the MAP compare to the whole population of SAGE students?
4. How does the set of SAGE schools compare to the set of non-SAGE control schools?
5. What is the difference between MAP growth in SAGE schools and non-SAGE schools?
6. What is the difference between MAP growth in SAGE schools and non-SAGE schools for particular populations of students?
7. What is the difference between MAP growth in SAGE schools and non-SAGE schools over multiple years?

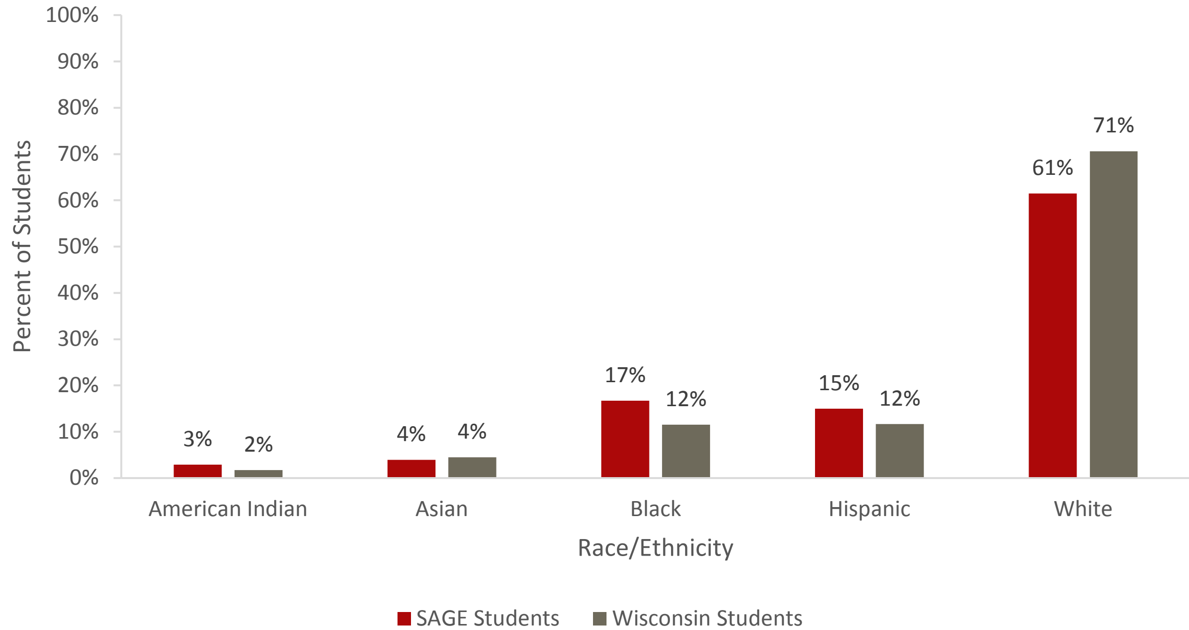
Research questions 1-4 provide the background necessary to design an appropriate method for analysis. For this evaluation's results to hold generalizability, the set of SAGE schools VARC analyzes should appear similar to the entire population of SAGE schools. Additionally, the analysis will have more power if a large percentage of both SAGE and non-SAGE schools utilize 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 the final three questions related to the impact of the SAGE program.

III. Overview of Demographics and Assessment

Characteristics of SAGE Schools and Students

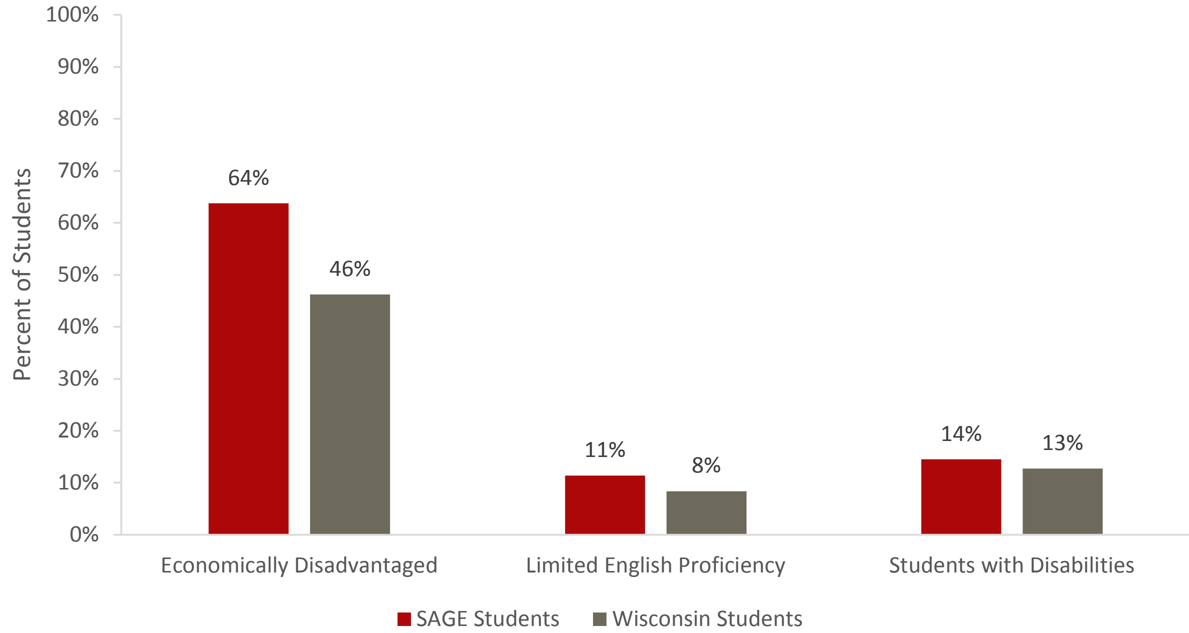
During the 2012-13 school year, 427 schools in 206 districts participated in the SAGE program. The number of students in SAGE schools in kindergarten through third grade was 81,677, with roughly equal proportions throughout each grade. Figures 1 – 3 show the demographic breakdown for students in SAGE 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 2012-13



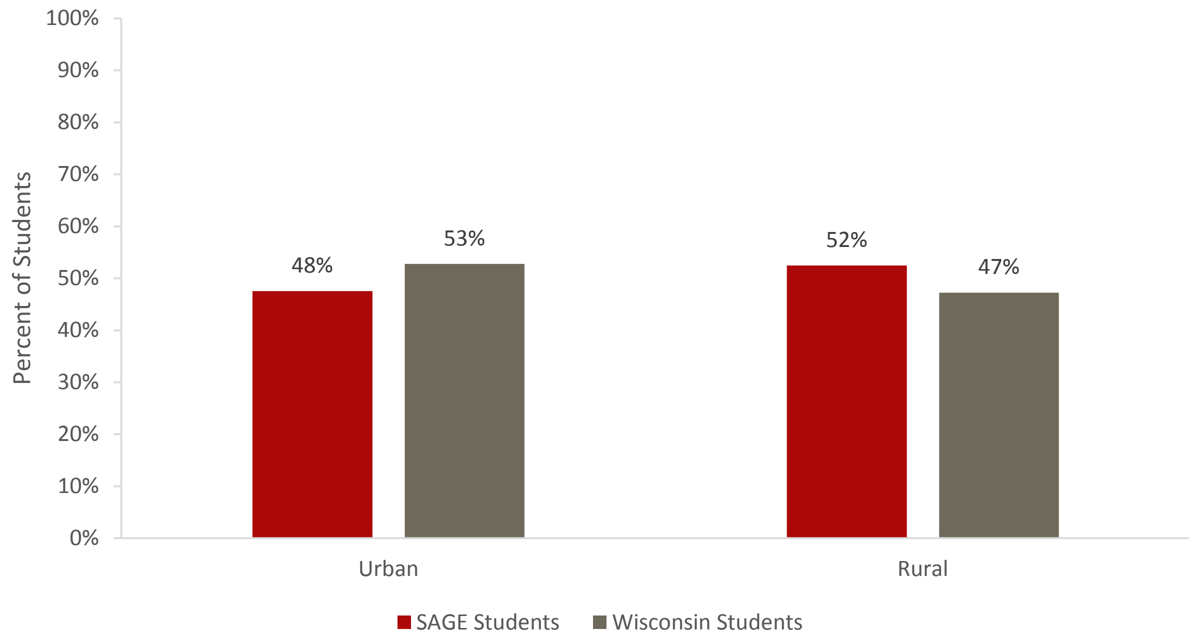
Source: WDPI LDS

Figure 2: Demographic breakdown of SAGE students by economic status, English proficiency, and disability status in 2012-13



Source: WDPI LDS

Figure 3: Demographic breakdown of SAGE students by urban and rural setting in 2012-13



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 17 percent of all SAGE students. When subtotaling these five districts, they comprise 32 percent of all SAGE students.

Table 1: Number and percent of SAGE students by district in kindergarten through third grade in 2012-13

District	SAGE Students	Percent of ALL SAGE Students
Beloit	2125	2.6%
Green Bay	2306	2.8%
Madison	5263	6.4%
Milwaukee	13648	16.7%
West Allis – West Milwaukee	2609	3.2%
Top 5 Subtotal	***25951***	***31.8%***
All Other SAGE Districts	*55726*	*68.2%*
Total	**81677**	**100.0%**

Source: LDS

Another important aspect to 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 2012-13. 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 2012-13

Configuration	Number of Classrooms	Percent of Classrooms
18:1	4995	96.1%
30:2	201	3.9%
45:3	3	0.1%
Total	5199	100.0%

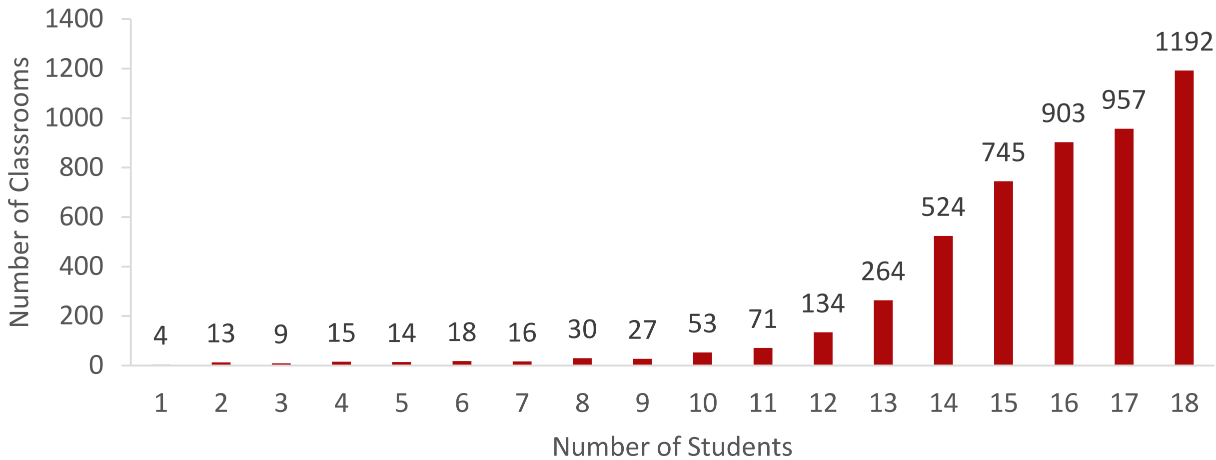
Source: SAGE Submission Report for 2012-13

Table 3: Mean number and standard deviation of students in SAGE classrooms by grade and classroom configuration in 2012-13

Grade	18:1 Configuration		30:2 Configuration		45:3 Configuration	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
K5	15.8	2.1	13.3	1.8	5.0	0.5
1st	15.8	2.0	12.9	1.9	N/A	N/A
2nd	15.8	2.0	13.2	1.6	6.0	N/A
3rd	15.8	2.0	12.1	1.8	N/A	N/A
Mixed	14.0	4.9	12.3	3.0	N/A	N/A

Source: SAGE Submission Report for 2012-13
Note: N/A indicates insufficient data

Figure 4: Distribution of students in 18:1 SAGE classrooms in 2012-13



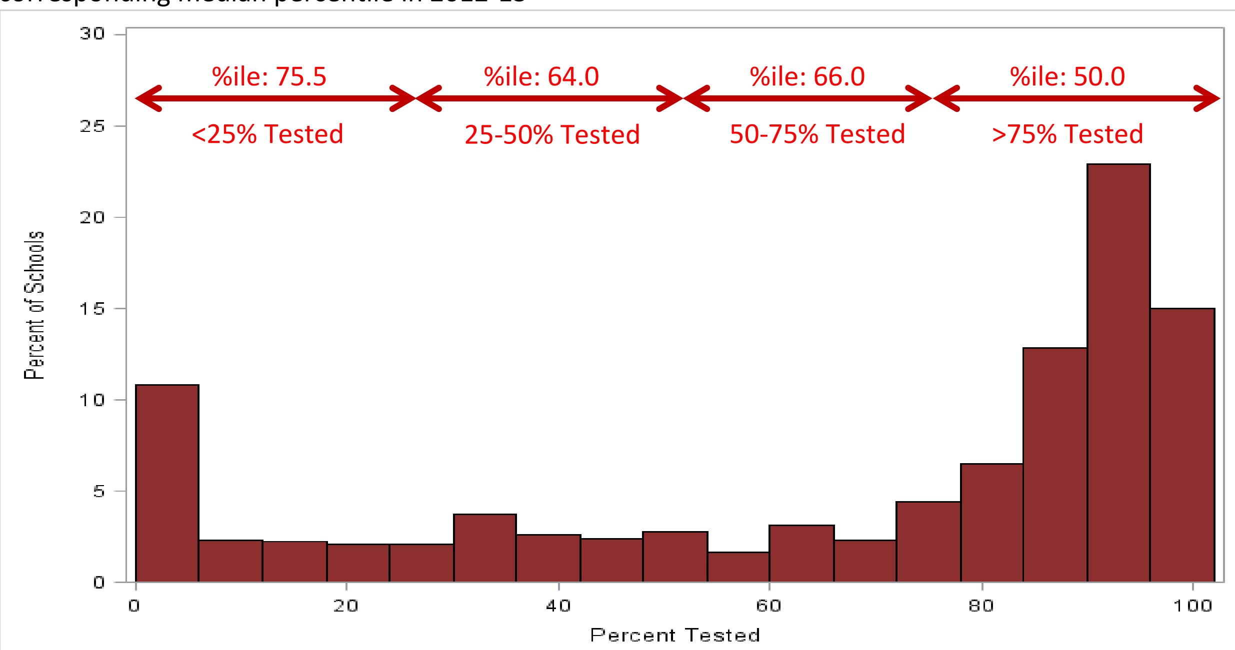
Source: SAGE Submission Report for 2012-13

MAP Assessment Utilization and Characteristics of MAP Students

Since this evaluation uses the MAP assessment as a measure of student achievement, and the state does not universally utilize the assessment across the entire population of all districts, the number of students who took the MAP was the largest factor limiting the size of the analysis set. The first step in the process of determining the set for this analysis was to define what it meant for a student to take the MAP assessment for purposes of the evaluation. VARC essentially placed two restrictions limiting what it meant for a student to take the MAP. The first criteria was that students needed both a pretest and a posttest score. This meant that each student in the set needed to take the MAP in both the fall and the spring. The second criteria was that students needed to take the assessment for benchmarking purposes. VARC assumes that when a school uses MAP testing for benchmarking purposes nearly all students within a grade and subject should have a test score. Hence, when a school has a large proportion of students without a test score, it indicates that the school might have used the assessment for purposes other than general benchmarking. Schools could use the MAP assessment to confirm the need of a specific intervention such as enrollment in a gifted program, needing special education, or justifying grade promotion/retention. Inclusion of schools using the MAP assessment for such purposes could skew the results.

As the data did not include a clear indicator for which students took the MAP for benchmarking purposes, this evaluation made an assumption from looking at trends in the results. Figure 5 shows the distribution of the percentage of students in each grade who took the MAP assessment. As seen in this figure, the majority of students who took the MAP were in grades where a large proportion of the students took the MAP. Not all grades followed this pattern, however, as just over ten 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. Figure 5 also shows the average MAP achievement median percentile across the fall and spring administrations in both reading and mathematics. As this figure demonstrates, the 0-25 percent tested range of students had a different average percentile range than the 25-50 percent tested, which was different from the 75-100 percent tested. Using this information, this evaluation assumed that schools used the MAP assessment for benchmarking purposes in a given grade if at least 75 percent of the students in that grade took the MAP assessment. Thus, VARC only included students in these grades as a part of the analysis set.

Figure 5: Distribution of the percent of students in each grade within schools taking the MAP and the corresponding median percentile in 2012-13



Source: 2012-13 MAP Files

VARC was then able to examine the proportion of students who took the MAP assessment and the demographics of those students. Table 4 shows the number and percent of students who took the MAP assessment in each grade for the three types of schools in the population: SAGE schools, non-SAGE schools in SAGE districts, and the new group for year 9, non-SAGE districts. Both SAGE students and non-SAGE students follow a similar trend of having a lower representation of the whole population in kindergarten and an increasing share up to third grade. The new group of schools for this year of the analysis had relatively low participation rates with only 13 percent of kindergarten through third grade students in non-SAGE districts participating in the MAP assessment.

Table 4: Number of students participating in the MAP assessment by grade for SAGE schools, non-SAGE schools in SAGE districts, and non-SAGE districts in 2012-13

Grade	SAGE Schools		Non-SAGE Schools, SAGE Districts		Non-SAGE Districts	
	MAP N	**% of Population**	**MAP N**	**% of Population**	**MAP N**	**% of Population**
K5	3821	18%	2921	22%	1852	7%
1st	4664	22%	4409	34%	3009	11%
2nd	5858	29%	5603	44%	4469	16%
3rd	7444	39%	7142	53%	5421	19%
K-3	21787	27%	20075	38%	14751	13%

Table 5 shows a breakdown of the number of SAGE students in the five most populous districts who took the MAP assessment and the corresponding percent of SAGE students who took the MAP assessment overall. As this table shows, districts with higher populations represented a larger proportion of the MAP-taking population and represented a larger proportion of the analysis set than the general population.

Table 5: Number of SAGE students participating in the MAP assessment in kindergarten through third grade in 2012-13

District	MAP SAGE Students	Percent of MAP SAGE Students
Appleton	925	4.3%
Beloit	1981	9.1%
Madison	890	4.1%
Milwaukee	9412	43.2%
West Allis – West MKE	1878	8.6%
Top 5 Subtotal	***15086***	***69.2%***
All Other SAGE Districts	*6701*	*30.8%*
Total	**21787**	**100.0%**

Since less than one-fourth 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 urban or rural setting for MAP and non-MAP students within the general population of both SAGE and non-SAGE schools. Figures 6 – 8 show these comparisons for all students in both SAGE and non-SAGE schools in kindergarten through third grade.[3] As seen, more students who are African-American, economically disadvantaged, and in urban settings took the MAP

[3] The Appendix shows the number of students and schools used in the analysis as a basis for comparison.

and fewer who have limited English proficiency and in 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 6: Race/ethnicity of all students in kindergarten through third grade by MAP utilization in 2012-13

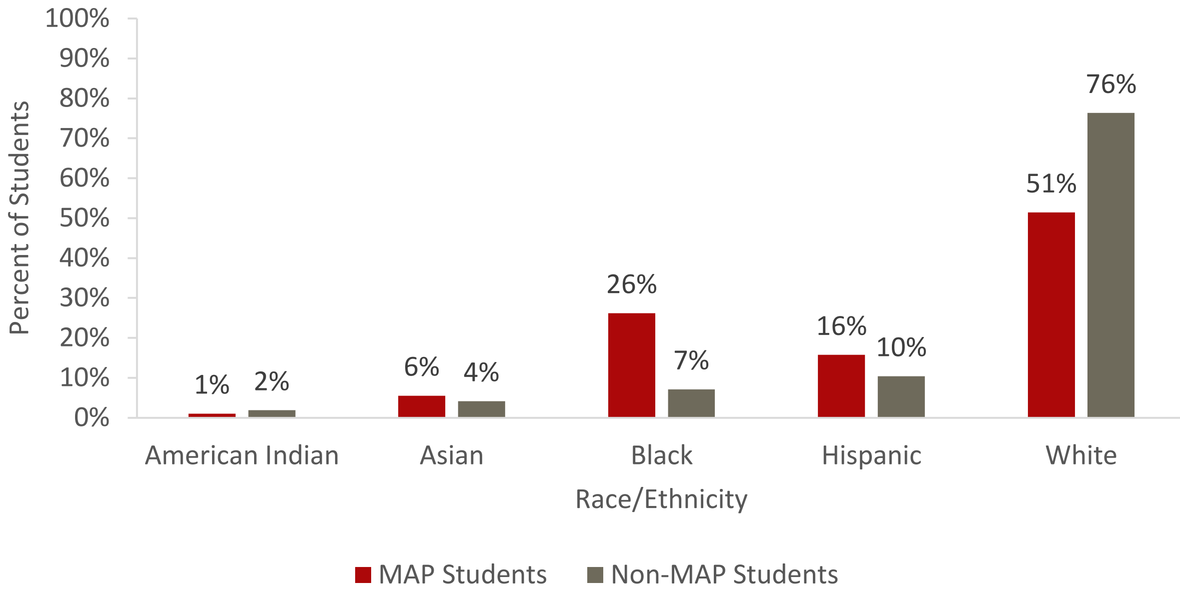


Figure 7: Percent of all kindergarten through third grade students who were economically disadvantaged, students with disabilities, and limited English proficient by MAP utilization in 2012-13

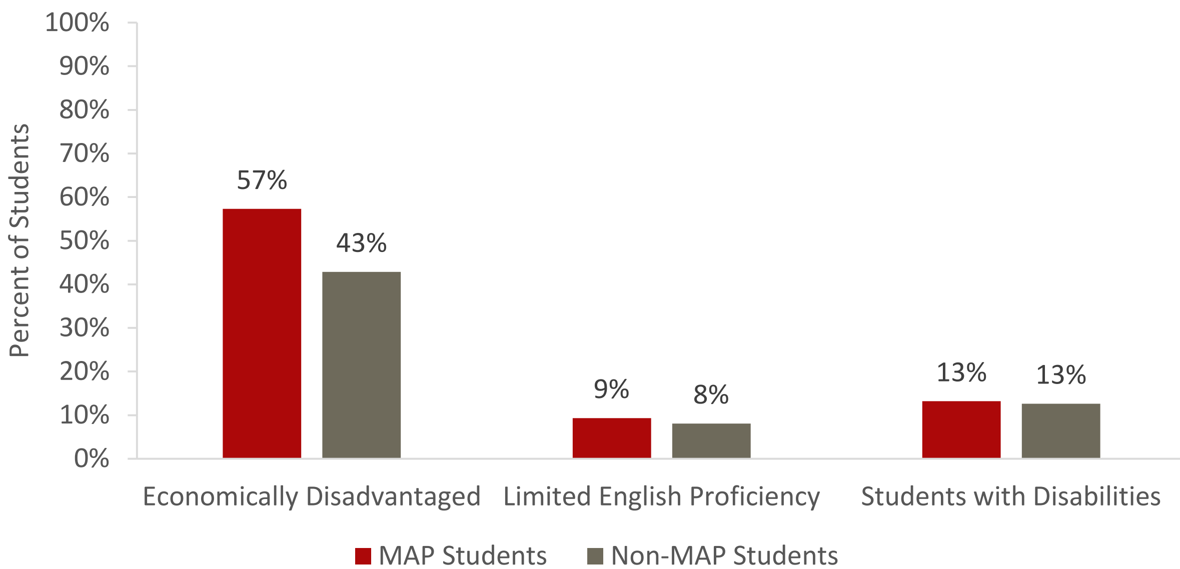
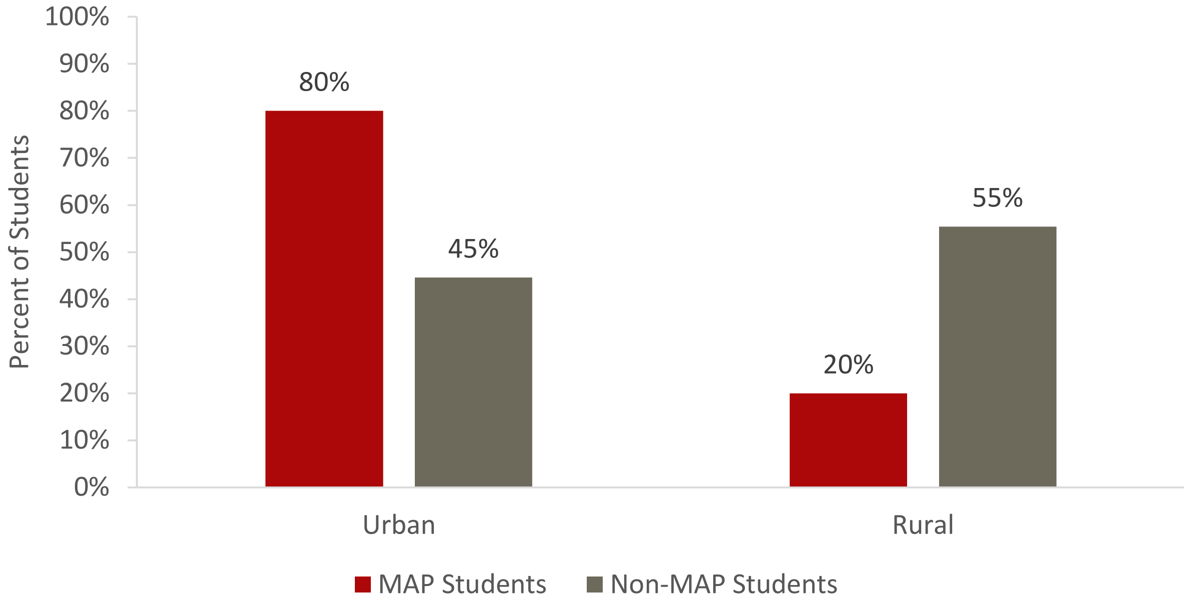


Figure 8: Percent of all kindergarten through third grade students who were in urban or rural settings by MAP utilization in 2012-13



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 9 – 11 show the breakdown of the percent of students by race/ethnicity, economically disadvantaged status, students with disabilities, limited-English proficiency, and urban 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.[4] 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 new group of analysis this year, non-SAGE districts, had a higher proportion of rural students than the other two categories.

[4] The Appendix shows the number of students and schools used in the analysis as a basis for comparison.

Figure 9: Race/ethnicity of tested students in kindergarten through third grade by SAGE participation in 2012-13

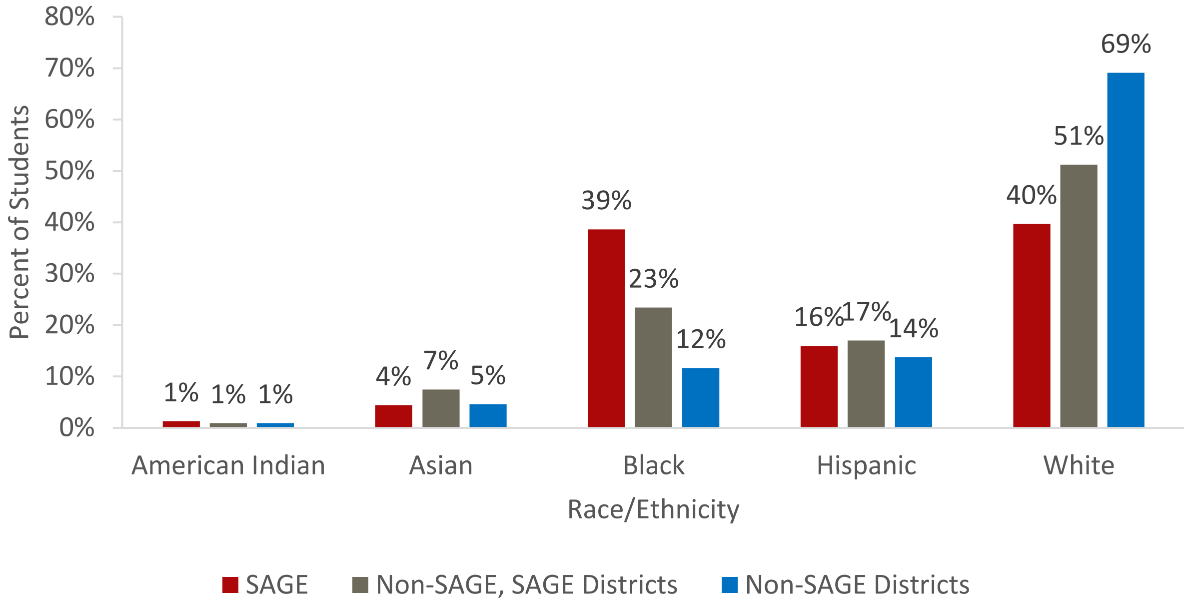


Figure 10: Percent of kindergarten through third grade tested students who were economically disadvantaged, students with disabilities, and limited English proficient by SAGE participation in 2012-13

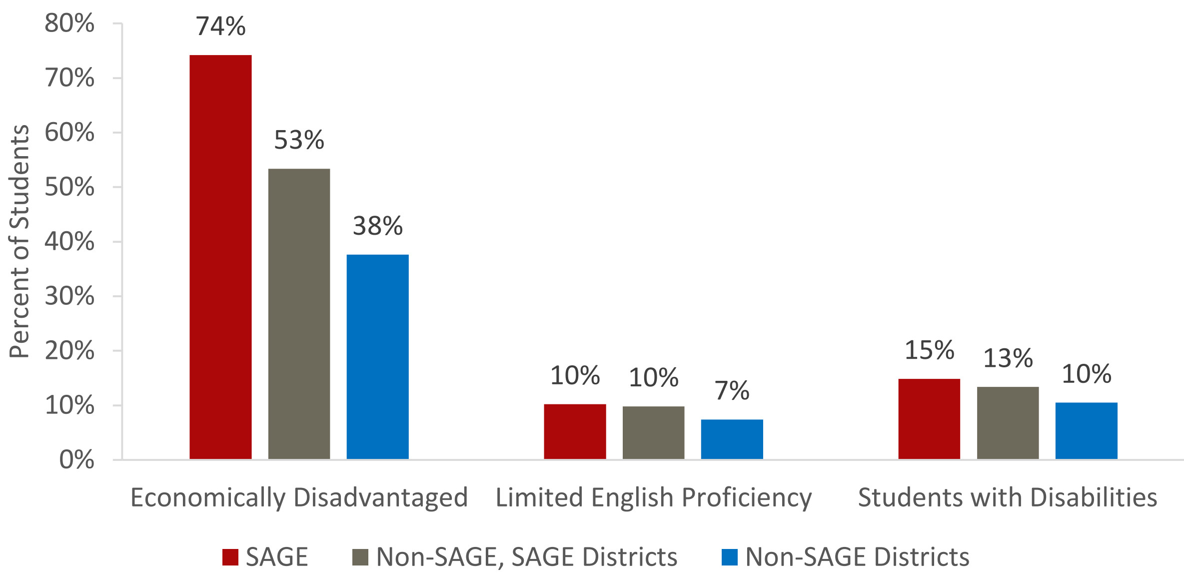
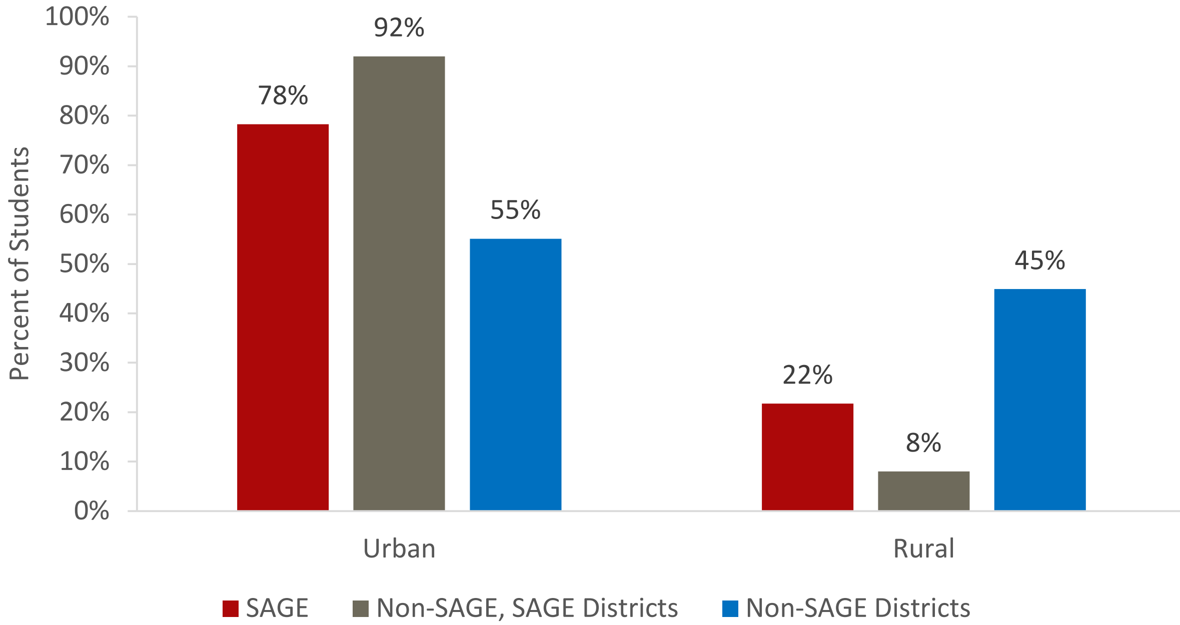


Figure 11: Percent of kindergarten through third grade tested students who were in urban or rural settings by SAGE participation in 2012-13



Overall, these differences between SAGE, non-SAGE within SAGE district, and non-SAGE district students not only demonstrate the need for more statistically advanced forms of analysis that can account for these differences, but also the need for careful attention in control group selection. 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.

Unadjusted MAP Results for SAGE and Non-SAGE Schools

Prior to analyzing the difference in academic performance between students in SAGE schools and non-SAGE schools with a more sophisticated model, VARC examined the simple differences in MAP scores and growth. Results from these base results show that for both reading and mathematics, SAGE students started at a lower average RIT score in the fall than non-SAGE students in SAGE districts and students from non-SAGE districts. By spring, SAGE students had average RIT scores similar to non-SAGE students in SAGE districts in kindergarten, but lower scores in first through third grade. Students in non-SAGE districts typically had the highest performance in spring with the exception of reading in kindergarten. Tables 6 – 9 and Figures 12 – 15 show the unadjusted results in mathematics, and Tables 10 – 13 and Figures 16 – 19 show the unadjusted results in reading.[5]

[5] The Appendix shows the number of students and schools used in the analysis as a basis for comparison.

Table 6: Average MAP *mathematics* RIT scores and standard deviations for SAGE and non-SAGE kindergarten students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	140.9	*11.6*	159.7	*15.0*
Non-SAGE Schools, SAGE Districts	143.2	*12.2*	159.7	*14.8*
Non-SAGE Districts	146.7	*15.7*	164.0	*15.1*

Figure 12: Average MAP *mathematics* RIT scores for SAGE and Non-SAGE kindergarten students in fall and spring in 2012-13

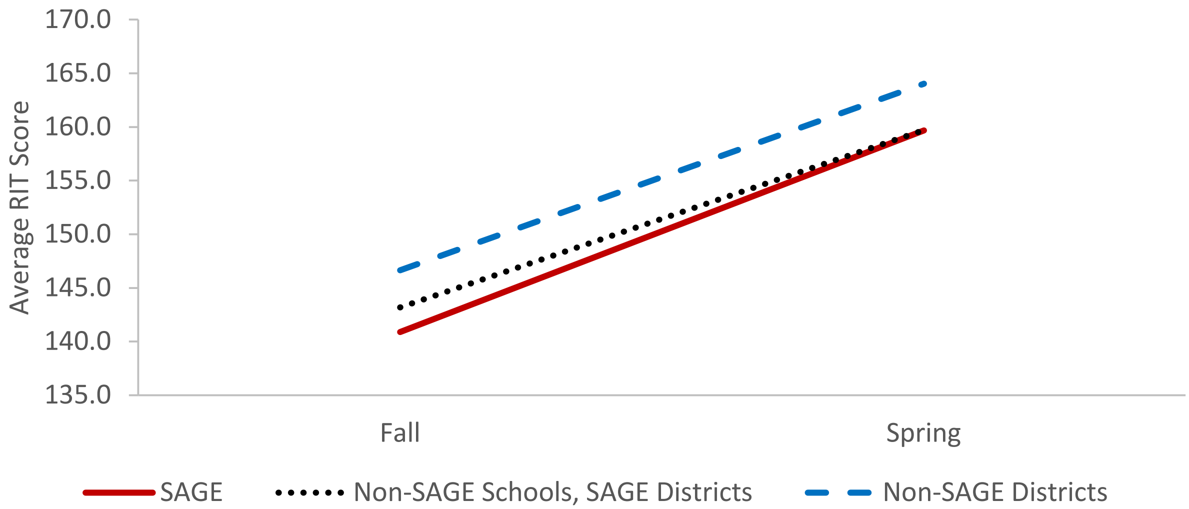


Table 7: Average MAP *mathematics* RIT scores and standard deviations for SAGE and non-SAGE first-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	160.1	*14.4*	177.3	*15.2*
Non-SAGE Schools, SAGE Districts	166.0	*19.1*	180.7	*16.5*
Non-SAGE Districts	166.3	*16.4*	186.0	*16.4*

Figure 13: Average MAP *mathematics* RIT scores for SAGE and Non-SAGE first-grade students in fall and spring in 2012-13

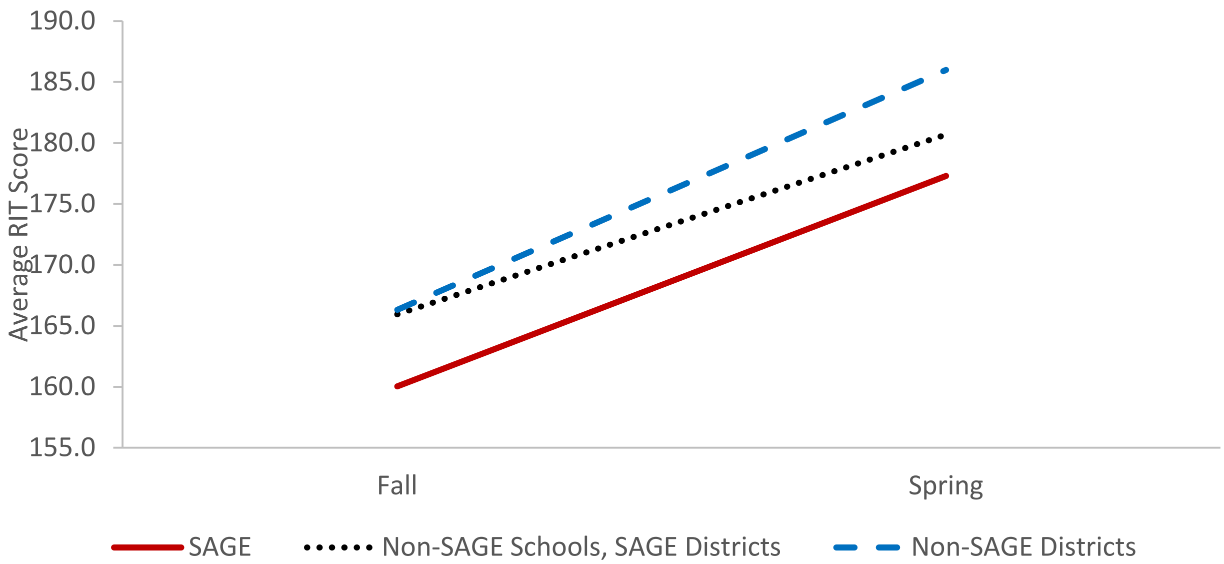


Table 8: Average MAP *mathematics* RIT scores and standard deviations for SAGE and non-SAGE second-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	176.2	*14.7*	190.0	*14.2*
Non-SAGE Schools, SAGE Districts	179.3	*16.8*	192.3	*14.1*
Non-SAGE Districts	182.3	*15.3*	197.9	*14.7*

Figure 14: Average MAP *mathematics* RIT scores for SAGE and Non-SAGE second-grade students in fall and spring in 2012-13

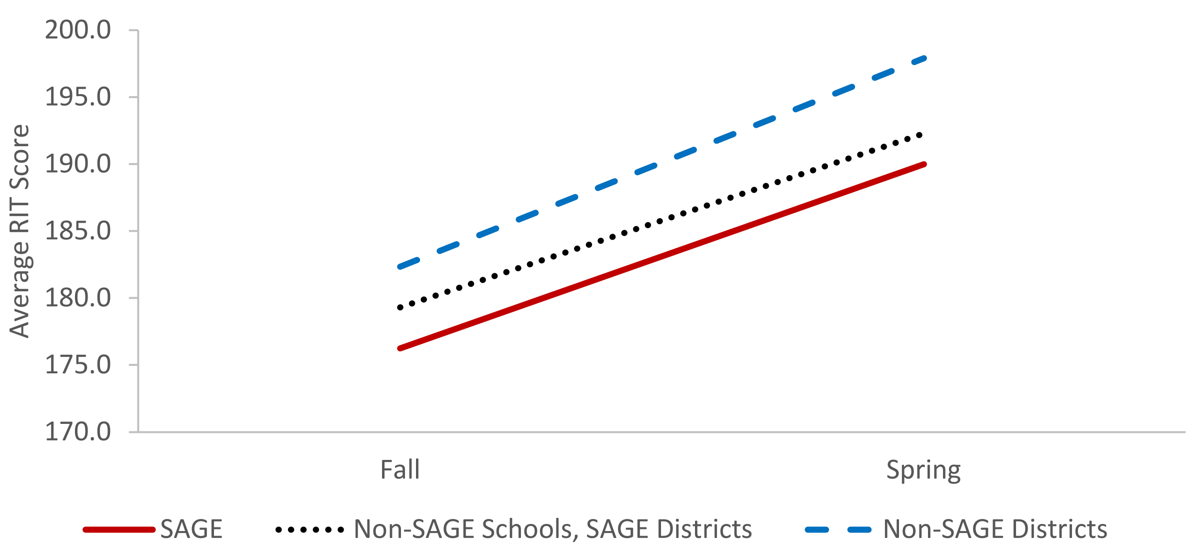


Table 9: Average MAP *mathematics* RIT scores and standard deviations for SAGE and non-SAGE third-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	197.7	*14.2*	199.6	*14.2*
Non-SAGE Schools, SAGE Districts	192.6	*16.4*	203.6	*15.1*
Non-SAGE Districts	194.5	*13.8*	206.9	*13.6*

Figure 15: Average MAP *mathematics* RIT scores for SAGE and Non-SAGE third-grade students in fall and spring in 2012-13

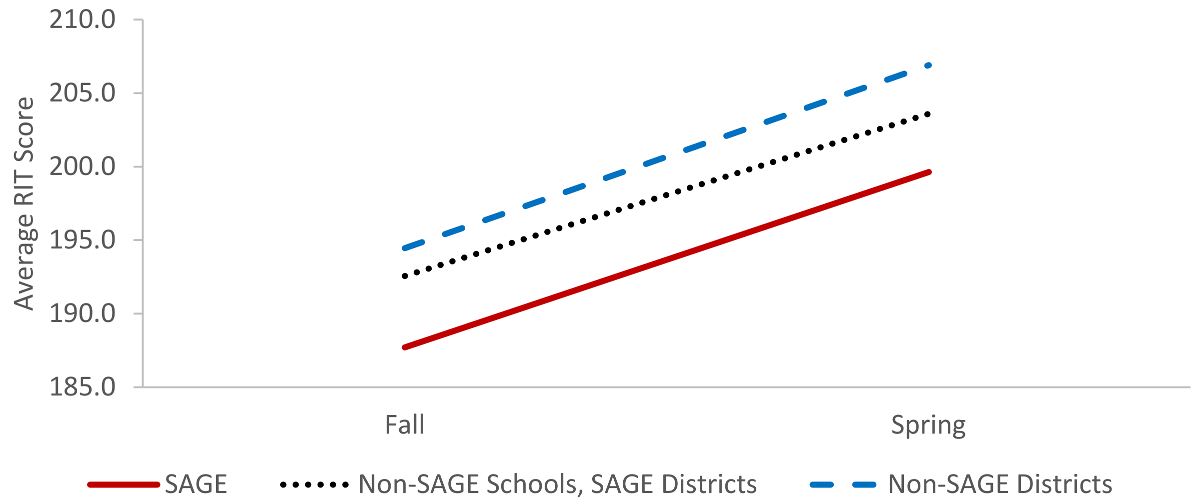


Table 10: Average MAP *reading* RIT scores and standard deviations for SAGE and non-SAGE kindergarten students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	144.7	*10.5*	163.6	*12.7*
Non-SAGE Schools, SAGE Districts	147.0	*10.3*	164.4	*13.1*
Non-SAGE Districts	146.1	*14.0*	161.3	*13.7*

Figure 16: Average MAP *reading* RIT scores for SAGE and Non-SAGE kindergarten students in fall and spring in 2012-13

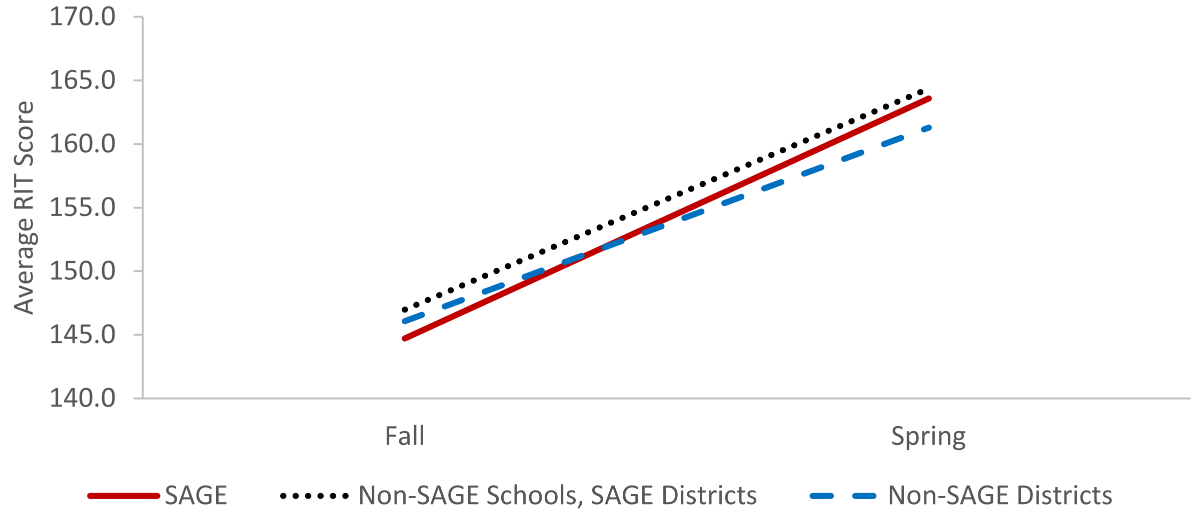


Table 11: Average MAP *reading* RIT scores and standard deviations for SAGE and non-SAGE first-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	158.4	*13.4*	174.0	*15.1*
Non-SAGE Schools, SAGE Districts	163.9	*18.3*	177.0	*16.5*
Non-SAGE Districts	164.8	*15.6*	181.4	*15.5*

Figure 17: Average MAP *reading* RIT scores for SAGE and Non-SAGE first-grade students in fall and spring in 2012-13

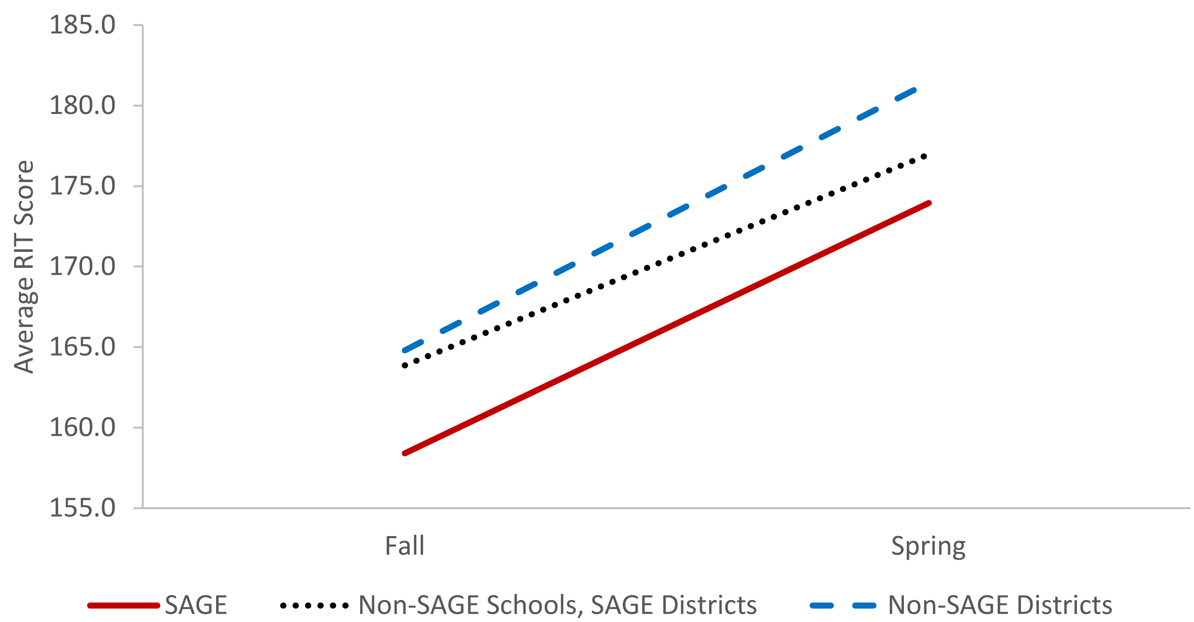


Table 12: Average MAP *reading* RIT scores and standard deviations for SAGE and non-SAGE second-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	172.2	*17.0*	186.0	*15.7*
Non-SAGE Schools, SAGE Districts	175.5	*19.0*	188.0	*16.2*
Non-SAGE Districts	178.0	*16.5*	191.8	*14.8*

Figure 18: Average MAP *reading* RIT scores for SAGE and Non-SAGE second-grade students in fall and spring in 2012-13

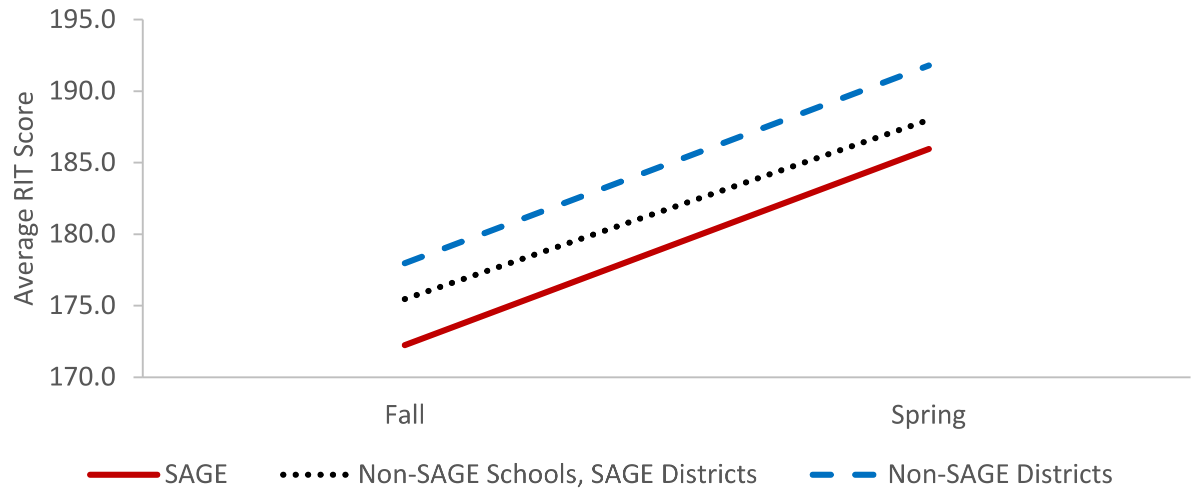


Table 13: Average MAP *reading* RIT scores and standard deviations for SAGE and non-SAGE third-grade students in fall and spring in 2012-13

School Type	Fall RIT	*Std. Dev.*	Spring RIT	*Std. Dev.*
SAGE Schools	183.8	*17.6*	194.5	*16.4*
Non-SAGE Schools, SAGE Districts	189.4	*18.5*	198.8	*16.3*
Non-SAGE Districts	190.6	*16.4*	200.9	*14.7*

Figure 19: Average MAP *reading* RIT scores for SAGE and Non-SAGE third-grade students in fall and spring in 2012-13

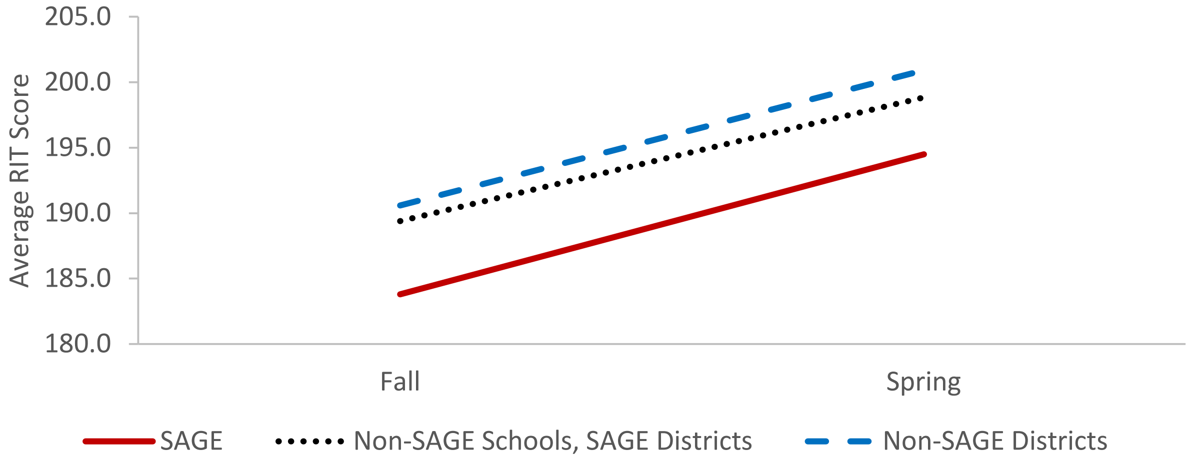


Figure 20 and Figure 21 show the same information converted to MAP growth for mathematics and reading, respectively. These unadjusted results suggest a differing impact of the SAGE program on student performance depending on the control group. To account for these differences, and the differences in student characteristics for these groups, the next section of this report describes the issue of control group selection and the models for analysis.

Figure 20: Average unadjusted MAP *mathematics* RIT score growth for students in SAGE schools, students in non-SAGE schools in SAGE districts, and students in non-SAGE districts by grade in 2012-13

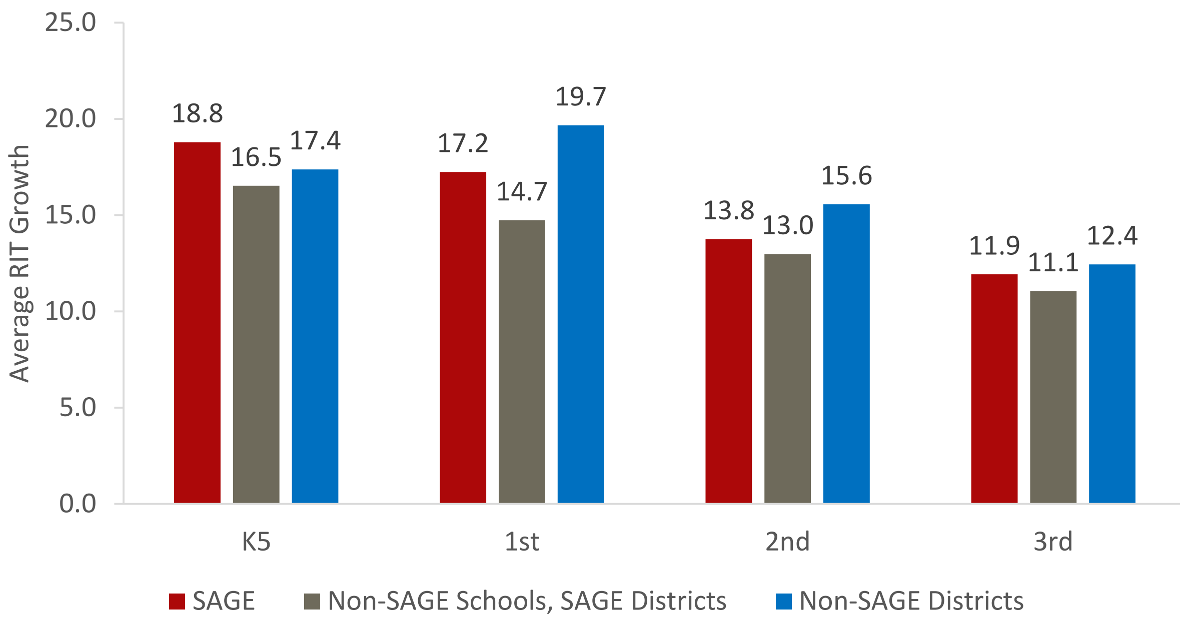
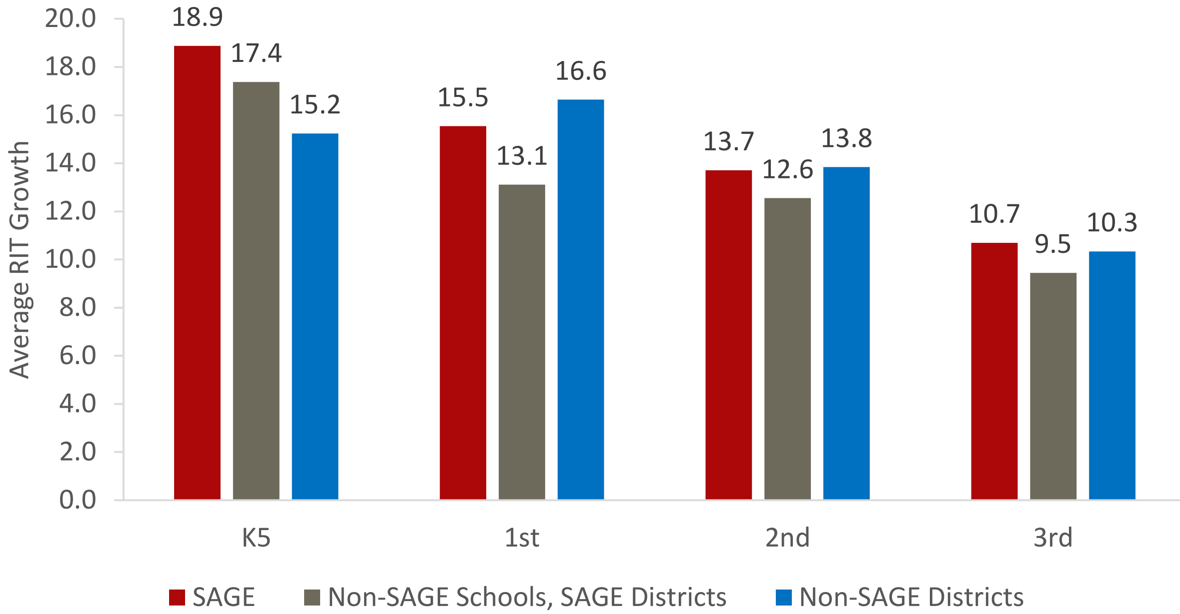


Figure 21: Average unadjusted MAP *reading* RIT score growth for students in SAGE schools, students in non-SAGE schools in SAGE districts, and students in non-SAGE districts by grade in 2012-13



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 of the SAGE program re-evaluated the selection of the control group to better account for selection bias. Additionally, this evaluation utilized statistical modeling in an attempt to detect the impact of the SAGE program on MAP growth in mathematics and reading. [6] This section of the report examines the selection of a control group and the analysis design models.

Control Group Selection

The introduction of an additional pool of comparison schools in non-SAGE districts provided yet another choice for how to model a control group for SAGE schools. In prior year's evaluations, the control group consisted of non-SAGE schools within SAGE districts. With the entire state as a potential pool, and with large differences between the types of schools to use as a control, this year's evaluation re-considered the approach to control group selection.

Ideal features of a SAGE control school include similar demographic characteristics, similar funding beyond the SAGE contribution, similar curriculum and interventions, similar school culture, a class size larger than 18:1, and participation for non-biasing reasons. While it is unlikely that this evaluation could control for curriculum, interventions, school culture, and non-SAGE class size at this

[6] For a detailed and technical description of the design model, refer to the Appendix.

time, this evaluation can control for demographic characteristics and funding to some extent. This left SAGE participation as a factor for which to control.

To determine how schools make participation decisions, VARC contacted several districts including Milwaukee, Beloit, and Waukesha, to gather information on this topic. All contacts responded that the decision was an economic one. This had the effect of rephrasing this factor to the question: can a school afford to participate in SAGE? One question from the 2012-13 End-of-Year Report asked SAGE schools, "What proportion of SAGE expenses are covered by program funding?"[7] Just over 60 percent of respondents reported that SAGE funding covered 51-100 percent of SAGE expenses, which left just under 40 percent of SAGE schools with funding not covering more than 50 percent of the additional expense of SAGE. The results from this question exemplify that SAGE schools need to supplement the program funding to afford participation in some instances.

If the decision to participate in SAGE is an economic one, the program funding model, the ability for a district to raise additional funding, and the cost associated with participation all directly contribute to this decision. This evaluation identified five specific measures to account for these factors.

The first measure is the number of free or reduced price lunch students within the school. The SAGE funding formula provided schools with approximately $2050 for each free or reduced price lunch student in kindergarten through third grade in 2012-13. This funding formula directly relies on the number of free or reduced price lunch students. Additionally, Title I funding, which is a common supplement to SAGE, also relies on the number of free or reduced price lunch students for its funding formula.

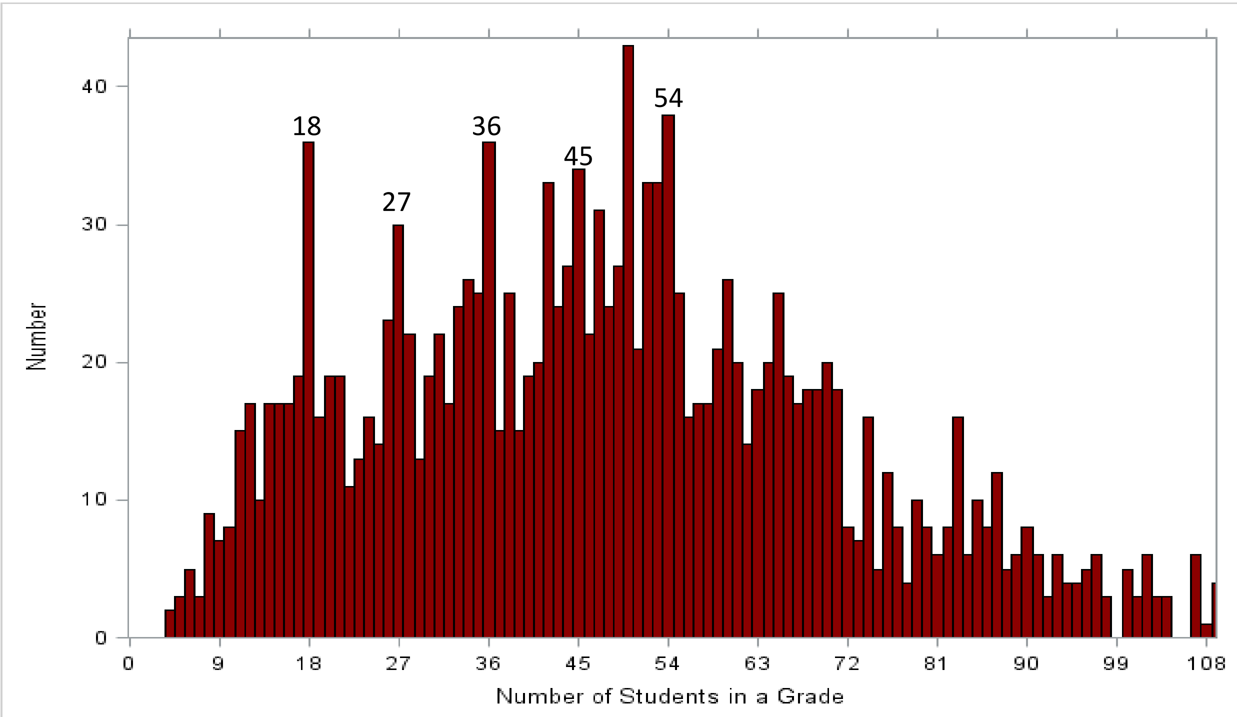
The second measure in SAGE participation is the average teacher salary. While not always, often schools that participate in SAGE require a higher number of teachers than non-SAGE schools, all else being equal. When an additional teacher is required, average teacher salary directly contributes to the cost associated with the SAGE program. This occurs when the number of students in a grade is larger than a multiple of 18 but smaller than a multiple of the non-SAGE class size cap. In school districts that the evaluation staff spoke to, this non-SAGE class size cap varied from 20 to 28 depending on funding and school decisions. In all cases, however, the cap in non-SAGE schools was not a hard cap. If an additional student came in during the middle of the year, the school was not likely to create a new classroom. In contrast, if a SAGE school had 20 students in first grade, SAGE guidelines would require the school to meet the 18:1 cap and, thus, the school would split the 20 students into two 10-student classrooms. This requires the SAGE school to hire an additional teacher. The law does not constrain non-SAGE schools, and such a school could keep the 20-student classroom with no additional cost.

The third measure in SAGE participation is the number of students in each school. At certain intervals of student populations, SAGE schools do not require an additional teacher as compared to their non-SAGE counterparts. For example, at intervals of 17 or 18 students in a grade, it is less expensive for a school to participate in SAGE than if they have intervals of 22 to 25 students, as they would not require

[7] For full results from the 2012-13 End-of-Year Report, refer to the Year 8 SAGE Program Evaluation Final Report.

an additional teacher. SAGE schools with intervals of more than 18 students over time will suffer additional costs every year to participate without some way to alleviate or supplement this extra expense. Thus, SAGE schools often have more grades with intervals of 18 or fewer students or risk the extra cost every year. Figure 22 shows that indeed SAGE schools exhibit this behavior. This figure shows the number of SAGE schools that have each exact number of students within that grade. Clearly, spikes occur at intervals of 18 (18, 36, 54, etc.).

Figure 22: Number of students in SAGE grades by the number of total grades within schools in 2012-13



Another interesting phenomenon apparent in Figure 22 is the spikes seen around 27 and 45 students. These intervals show the prevalence of mixed-grade classrooms. Instead of a first grade with 27 students and a second grade with 27 students creating four classrooms of 13 or 14 students, SAGE schools can create one first-grade class with 18 students, one second-grade class with 18 students, and one mixed-grade class with 18 students. This reduces the cost of participation by one full teacher while still maintaining the 18:1 requirement. The use of mixed-grade classrooms is an example of the fourth measure in SAGE participation, flexibility of class size. SAGE schools can use these strategies to reduce the costs associated with participation. Another example of these strategies is open enrollment. In larger districts with more than one elementary school, an open enrollment policy allows them the flexibility to turn students away when a school is full, or in relation to SAGE, when all the classes reach 18 students in size. This saves funds for SAGE schools in larger districts compared to smaller districts that must create a new classroom when the class receives the nineteenth student.

A fifth and final measure of SAGE participation is the district revenue limit. Two of the largest portions of district revenue are local property taxes and state equalization aid. Wisconsin state law applies a revenue limit each year on the additional amount of funds that a district can raise from these two sources. Because of this revenue limit, districts have limited funding from the two sources that they could use to supplement SAGE funding. While the evaluation may control for each districts' funding, VARC did not see large variation in how close districts are to their revenue limits across the state. In 2012-13, 95 percent of districts were within 5 percent of the revenue limit and 90 percent of districts were at the revenue limit. With nearly all districts facing this issue, controlling for the revenue limit does not help this evaluation with regard to participation decision making.

With all five of these measures available to examine, this evaluation could then begin to model the likelihood of a school being in the SAGE program. The next section will examine the propensity model used in the evaluation as well as the analysis models.

Analysis Design Models

Since a direct comparison of SAGE to non-SAGE schools is not feasible due to differences in student and school characteristics, this evaluation of the SAGE program uses statistical modeling to detect the impact of the SAGE program on MAP growth in mathematics and reading. The main effect model compares students in SAGE schools to students in non-SAGE schools by utilizing matching methods. The outcome variable of this evaluation is value added; hence, first. VARC obtained the value-added estimates from student-level information for each school; see the Appendix for technical specifications. Then, using propensity score modeling, VARC estimated the propensity score of each school. Finally, using estimated propensity scores, VARC matched each SAGE school with non-SAGE schools to obtain the SAGE effect, or impact, from the school-level information obtained in the first stage.

The value-added model uses the following information as explanatory student variables for the analysis:

- Fall MAP RIT scores (both mathematics and reading),
- Gender,
- Race/ethnicity,
- IEP or disability status,
- Economically disadvantaged status (through the free or reduced price lunch indicator), and
- ELL level.

The propensity score model used 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:

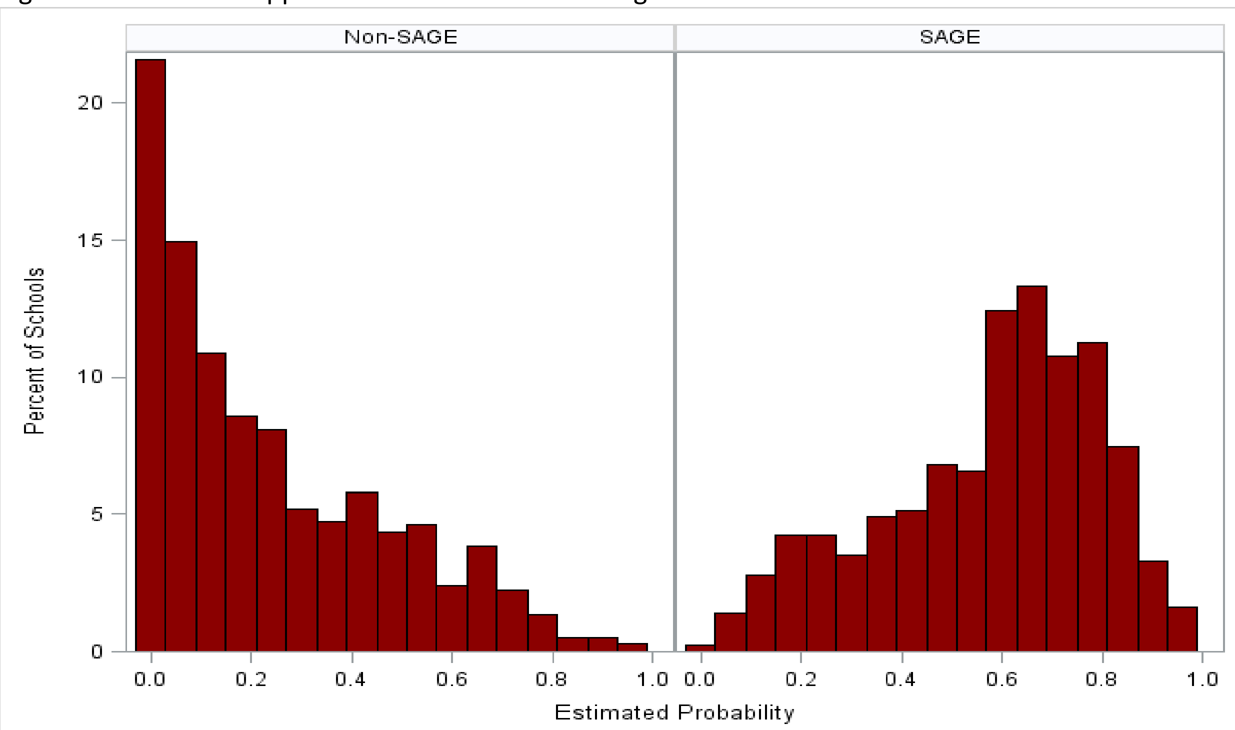
- Number of students,
- Percent of free/reduced price lunch students,
- Percent female,
- Percent race for each race,
- Percent special education,
- Percent ELL,

- Percent IEP or disability status,
- Average teacher compensation, and
- Urban/Rural indicator.

Finally, the analysis matches SAGE schools to a weighted average of non-SAGE schools using a kernel propensity score matching method. This method more heavily weights non-SAGE schools with propensity scores closer to the scores of a SAGE school and is often more efficient than one-to-one matching as it utilizes more information from the control group.

Examining the results from the propensity score model, the evaluation found common support for SAGE participation between SAGE schools and non-SAGE schools. Figure 23 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 23: Common support for the likelihood of being in SAGE in 2012-13



This evaluation also examined the differential impacts of SAGE on different types of students and schools. These differential effects compared the effect of the SAGE program on students within schools with certain characteristics to non-SAGE students within schools with the same characteristics. The areas of differential impact evaluated included:

- Urban/rural district setting,
- Proportion of economically disadvantaged students,
- Proportion of limited-English proficient students,
- Proportion of students with disabilities,

- Proportion of African-American students,
- Proportion of white students, and
- SAGE school funding.

The analysis estimated the differential SAGE effect model for urban/rural setting by restricting the matching to occur within the same subgroup. The analysis estimated the differential effect model for schools with high or low percentages of economically disadvantaged students similarly, with three subgroups defined by whether the school's percentage of free or reduced price lunch students is 0-33 percent, 33-66 percent, or 66-100 percent. The analysis estimated the differential SAGE effect model for English proficiency with two subgroups defined by whether the school's percentage of limited-English proficient students is less than five percent or greater than five percent. The analysis estimated the differential SAGE effect model for students with disabilities with two subgroups defined by whether the school's percentage of students with disabilities is less than ten percent or greater than ten percent. The analysis estimated the differential SAGE effect model for race/ethnicity for both African-American students and white students. The evaluation did not include differential results for other sub-populations of students by race/ethnicity due to relatively low proportions of these students within schools. The analysis examines the differential SAGE effect model for African-American students with two subgroups defined by whether the school's percentage of African-American students is less than ten percent or greater than ten percent. The analysis also examines the differential SAGE effect model for white students with two subgroups defined by whether the school's percentage of white students is less than 50 percent or greater than 50 percent. Finally, the evaluation included differential results based on the qualitative information from the 2012-13 End-of-Year Report. One question from the 2012-13 End-of-Year Report asked SAGE schools, "What proportion of SAGE expenses are covered by program funding?" This analysis examines the differential SAGE effect model for SAGE funding with two subgroups defined by whether a school responded with 0-50 percent of SAGE expenses covered by program funding or a school responded with 51-100 percent of SAGE expenses covered by program funding.

New to this year's analysis of the SAGE program was a longitudinal analysis of SAGE outcomes. Since this was the first time the evaluation had access to three years of comprehensive MAP assessment data (2010-11, 2011-12, and 2012-13), VARC started to work on examining the longitudinal 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 2012-13. The analysis included students who took the MAP examination every year, remained in either a SAGE school for all three years or a non-SAGE school for all three years, and progressed from kindergarten in 2010-11 to first grade in 2011-12 to second grade in 2012-13. The analysis did not restrict students to be at the same school or same district for all three years. With an established cohort set, the analysis matched SAGE students to non-SAGE students with a propensity score model similar to that used in prior analyses only matching for students instead of schools. Finally, to calculate a longitudinal effect, the analysis compared the growth of SAGE students to the growth of their matched non-SAGE students on average. In future years, VARC will work to improve the longitudinal analysis with more years of MAP assessment data.

V. Evaluation Results

Main SAGE Effect

Table 14 shows the number of SAGE schools and non-SAGE schools used in the analysis. In first grade, for example, the model matched each of the 92 SAGE schools with the weighted average of 133 non-SAGE schools. As the previous section mentioned, non-SAGE schools with a closer propensity score to that of the matched SAGE school received a higher weight.

Table 14: Number of SAGE and matched non-SAGE schools used in main analysis in 2012-13

Grade	SAGE Schools	Non-SAGE Schools
K5	77	88
1st	92	133
2nd	126	186
3rd	151	223

As a useful comparison, Table 15 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 15: Total number of SAGE and non-SAGE schools in Wisconsin in 2012-13

Grade	SAGE Schools	Non-SAGE Schools
K5	413	688
1st	416	698
2nd	411	704
3rd	398	714

Table 16 and Table 17 show the statistically adjusted MAP RIT score growth differences between the students in SAGE and non-SAGE schools from the main analysis design model presented in the previous section for mathematics and reading, respectively. In each of these tables, and the differential effect results that follow, VARC provides both the SAGE coefficient in RIT scale scores and the SAGE coefficient in standard deviations of the post-test results. The coefficient in RIT 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 RIT scores divided by the standard deviation of the whole set's spring test results.

As seen in Table 16, VARC estimated positive and significant effects of the SAGE program on mathematics growth in kindergarten and first grade as compared to students in non-SAGE schools. The evaluation found no statistically significant results in second or third grade for mathematics. This means that, on average, students in SAGE schools grew at a higher rate than students in non-SAGE schools in kindergarten and first grade in mathematics.

Table 16: Results of statistical analysis of the SAGE program effect on *mathematics* MAP growth by grade in 2012-13

Grade	Effect Size (Scale Scores)	Effect Size (Std. Dev.)	Standard Error	T-stat
K5	**2.41**	**0.16**	**1.130**	**2.14**
1st	**3.15**	**0.20**	**0.895**	**3.52**
2nd	0.65	0.04	0.636	1.02
3rd	0.46	0.03	0.457	1.02

Note: Results in bold indicate a statistically significant effect.

As seen in Table 17, VARC also estimated positive and significant effects of the SAGE program on reading growth in kindergarten through second grade as compared to students in non-SAGE schools. The evaluation found no statistically significant results in 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, first grade, and second grade in reading.

Table 17: Results of statistical analysis of the SAGE program effect on *reading* MAP growth by grade in 2012-13

Grade	Effect Size (Scale Scores)	Effect Size (Std. Dev.)	Standard Error	T-stat
K5	**3.67**	**0.28**	**1.350**	**2.72**
1st	**3.61**	**0.23**	**0.904**	**3.99**
2nd	**1.56**	**0.10**	**0.596**	**2.62**
3rd	0.76	0.05	0.441	1.71

Note: Results in bold indicate a statistically significant effect.

Differential SAGE Effects

In addition to the main effects of the SAGE program across all students, this report also presents information on the impact of SAGE upon specific subpopulations of students. Results from these analyses follow.

Differential Effects by Economic Status

The first differential impact this evaluation examined was the effect of the SAGE program upon schools that had higher or lower percentages of economically disadvantaged students based on the students' free or reduced price lunch status. Table 18 shows the number of SAGE and non-SAGE schools used in the analysis for each of the three categories identified in the previous section. Since the state bases SAGE funding on the number of free or reduced price lunch students, there were too few schools in the lowest group (i.e. schools where student population consist of 0-33 percent of free or reduced price lunch students) to perform statistical analyses with these schools.

Table 18: Number of SAGE and matched non-SAGE schools used in the economically disadvantaged differential analysis in 2012-13

Grade	Schools with 0-33% FRPL		Schools with 33-66% FRPL		Schools with 66-100% FRPL	
	SAGE	Non-SAGE	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	1	21	16	29	58	34
1st	2	40	25	44	62	45
2nd	3	53	48	73	72	54
3rd	3	69	56	90	89	57

Table 19 and Table 20 show the results of the analyses for the differential effects of SAGE on economically disadvantaged students in both mathematics and reading. As the SAGE program's initial stated goal was to improve the academic achievement of students from economically disadvantaged backgrounds, this differential impact can shed light on the SAGE program's progress in attaining this goal. Table 19 indicates positive and significant impacts of the SAGE program on mathematics growth for schools with a high proportion of economically disadvantaged students in first and third grade. Table 20 indicates positive and significant impacts on reading growth for schools with a high percentage of economically disadvantaged students in all SAGE grades.

Table 19: Results of the SAGE program effect on *mathematics* MAP growth for economically disadvantaged students in 2012-13

Grade	Schools with 33-66% FRPL				Schools with 66-100% FRPL			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	**5.51**	**0.37**	**1.766**	**3.12**	2.02	0.13	1.391	1.45
1st	**3.57**	**0.22**	**1.644**	**2.17**	**3.37**	**0.21**	**1.156**	**2.91**
2nd	-0.57	-0.04	0.862	-0.66	1.04	0.07	0.832	1.24
3rd	-0.55	-0.04	0.574	-0.96	**1.68**	**0.11**	**0.629**	**2.67**

Note: Results in bold indicate a statistically significant effect.

Table 20: Results of the SAGE program effect on *reading* MAP growth for economically disadvantaged students in 2012-13

Grade	Schools with 33-66% FRPL				Schools with 66-100% FRPL			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	2.46	0.19	1.878	1.31	**4.21**	**0.32**	**2.084**	**2.02**
1st	**3.41**	**0.21**	**1.586**	**2.15**	**4.08**	**0.26**	**1.332**	**3.06**
2nd	-0.15	-0.01	0.749	-0.20	**2.40**	**0.15**	**0.827**	**2.90**
3rd	-0.32	-0.02	0.537	-0.59	**2.18**	**0.14**	**0.763**	**2.86**

Note: Results in bold indicate a statistically significant effect.

Differential Effects by District Setting

Another type of differential impact this evaluation examined was the difference between SAGE and non-SAGE schools by urban and rural setting. Table 21 shows the number of schools used in the analysis for both settings. The number of schools in rural settings for both SAGE and non-SAGE was smaller than that of urban settings due to limited numbers of these schools utilizing the MAP assessment. This limited the evaluation's ability to detect any effect of the program on rural students.

Table 21: Number of SAGE and matched non-SAGE schools used in the differential analysis for urban and rural schools in 2012-13

Grade	Urban Schools		Rural Schools	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	66	80	11	8
1st	77	115	15	18
2nd	88	142	38	44
3rd	116	165	35	58

Table 22 and Table 23 show the effects of the SAGE program for urban and rural districts. For mathematics, the analysis estimated a positive and significant effect of the SAGE program for students in first grade in urban areas. For reading, the analysis estimated positive and significant effects of the SAGE program for students in all of the SAGE grades in urban areas, with the largest effect in the earlier grades. The evaluation found no statistically significant impacts of the SAGE program for rural students in either mathematics or reading.

Table 22: Results of the SAGE program effect on *mathematics* MAP growth for urban and rural students in 2012-13

Grade	Urban Schools				Rural Schools			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	2.14	0.14	1.169	1.83	2.67	0.18	3.256	0.82
1st	**2.98**	**0.19**	**0.936**	**3.19**	3.11	0.19	2.930	1.06
2nd	0.97	0.07	0.745	1.29	-0.87	-0.06	1.162	-0.75
3rd	1.01	0.07	0.545	1.84	-0.73	-0.05	0.825	-0.88

Note: Results in bold indicate a statistically significant effect.

Table 23: Results of the SAGE program effect on *reading* MAP growth for urban and rural students in 2012-13

Grade	Urban Schools				Rural Schools			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	**5.00**	**0.38**	**1.425**	**3.51**	0.03	0.00	3.251	0.01
1st	**3.28**	**0.21**	**0.939**	**3.49**	3.17	0.20	2.558	1.24
2nd	**2.05**	**0.13**	**0.702**	**2.93**	-0.28	-0.02	1.028	-0.28
3rd	**1.34**	**0.08**	**0.543**	**2.47**	-0.82	-0.05	0.667	-1.23

Note: Results in bold indicate a statistically significant effect.

Differential Effects by English Proficiency

Table 24 shows the number of schools used in the analysis divided by the proportion of English proficient students. Table 25 and Table 26 show the results of the analyses for differential effects of SAGE on students with limited-English proficiency for both mathematics and reading. In mathematics, VARC estimated a positive and significant impact of the SAGE program in first and second grade in both schools with a high proportion of limited-English proficient students and schools with a low proportion of limited-English proficient students. In reading, VARC estimated positive and significant effects in first grade and second grade for schools with low proportions of limited-English proficient students. VARC estimated positive and significant effects in all SAGE grades in reading in schools with high proportions of limited-English proficient students.

Table 24: Number of SAGE and matched non-SAGE schools used in the English proficiency differential analysis in 2012-13

Grade	Schools with <5% ELL		Schools with >5% ELL	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	59	58	19	29
1st	62	78	30	55
2nd	77	92	49	94
3rd	77	109	74	114

Table 25: Results of the SAGE program effect on *mathematics* MAP growth by English proficiency in 2012-13

Grade	Schools with <5% ELL				Schools with >5% ELL			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	2.60	0.17	1.468	1.77	2.31	0.17	1.680	1.37
1st	**2.30**	**0.14**	**1.154**	**1.99**	**4.01**	**0.25**	**1.422**	**2.82**
2nd	1.55	0.11	1.005	1.55	0.40	0.03	0.850	0.47
3rd	0.67	0.05	0.760	0.88	0.61	0.04	0.549	1.10

Note: Results in bold indicate a statistically significant effect.

Table 26: Results of the SAGE program effect on *reading* MAP growth by English proficiency in 2012-13

Grade	Schools with <5% ELL				Schools with >5% ELL			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	2.24	0.15	1.948	1.15	**4.23**	**0.32**	**1.668**	**2.53**
1st	**2.91**	**0.18**	**1.261**	**2.31**	**4.37**	**0.27**	**1.292**	**3.38**
2nd	**1.87**	**0.13**	**0.935**	**2.00**	**1.88**	**0.12**	**0.765**	**2.46**
3rd	0.46	0.03	0.654	0.70	**1.27**	**0.08**	**0.563**	**2.25**

Note: Results in bold indicate a statistically significant effect.

Differential Effects by Disability Status

Table 27 shows the number of schools used in the analysis broken up by the proportion of students with disabilities. Table 28 and Table 29 show the results of the analyses for differential effects of SAGE on students with disabilities for both mathematics and reading. In both mathematics and reading, VARC estimated a positive and significant impact of the SAGE program in kindergarten, first grade, and second grade in schools with low proportions of students with disabilities. In schools with high proportions of students with disabilities, VARC estimated positive and significant effects in first grade in both mathematics and reading.

Table 27: Number of SAGE and matched non-SAGE schools used in the disability status differential analysis in 2012-13

Grade	Schools with <10% Students with Disabilities		Schools with >10% Students with Disabilities	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	7	22	70	66
1st	10	31	82	102
2nd	15	54	111	132
3rd	26	63	125	160

Table 28: Results of the SAGE program effect on *mathematics* MAP growth by disability status in 2012-13

Grade	Schools with <10% Students with Disabilities				Schools with >10% Students with Disabilities			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	**7.96**	**0.53**	**2.203**	**3.61**	2.07	0.16	1.311	1.58
1st	**4.40**	**0.27**	**1.808**	**2.43**	**3.27**	**0.20**	**1.058**	**3.09**
2nd	**5.39**	**0.37**	**1.300**	**4.16**	0.06	0.00	0.733	0.09
3rd	1.51	0.10	0.833	1.81	0.38	0.02	0.530	0.72

Note: Results in bold indicate a statistically significant effect.

Table 29: Results of the SAGE program effect on *reading* MAP growth by disability status in 2012-13

Grade	Schools with <10% Students with Disabilities				Schools with >10% Students with Disabilities			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	**8.42**	**0.56**	**1.564**	**5.38**	1.86	0.14	1.972	0.95
1st	**6.23**	**0.39**	**2.077**	**3.00**	**3.53**	**0.22**	**1.043**	**3.38**
2nd	**6.49**	**0.44**	**1.449**	**4.48**	1.07	0.07	0.663	1.61
3rd	1.33	0.09	0.896	1.48	0.63	0.04	0.512	1.23

Note: Results in bold indicate a statistically significant effect.

Differential Effects by Race/Ethnicity

Table 30 shows the number of schools used in the analysis divided by the proportion of African-American students. Table 31 and Table 32 show the results of the analyses for differential effects of SAGE on African-American students for both mathematics and reading. In both mathematics and reading, VARC estimated positive and significant impacts of the SAGE program in all SAGE grades in schools with high proportions of African-American students. VARC found no statistically significant impact of the SAGE program on students in schools with low proportions of African-American students in either mathematics or reading.

Table 30: Number of SAGE and matched non-SAGE schools used in the African-American differential analysis in 2012-13

Grade	Schools with <10% African-American Students		Schools with >10% African-American Students	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	13	38	64	50
1st	22	66	70	67
2nd	53	111	73	75
3rd	53	133	98	90

Table 31: Results of the SAGE program effect on *mathematics* MAP growth by proportion African-American in 2012-13

Grade	Schools with <10% African-American Students				Schools with >10% African-American Students			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	2.46	0.16	2.100	1.17	**2.88**	**0.22**	**1.200**	**2.40**
1st	2.70	0.17	1.620	1.66	**3.26**	**0.20**	**1.078**	**3.03**
2nd	-0.66	-0.04	0.843	-0.79	**1.86**	**0.12**	**0.935**	**1.98**
3rd	-0.47	-0.03	0.615	-0.76	**1.63**	**0.10**	**0.606**	**2.68**

Note: Results in bold indicate a statistically significant effect.

Table 32: Results of the SAGE program effect on *reading* MAP growth by proportion African-American in 2012-13

Grade	Schools with <10% African-American Students				Schools with >10% African-American Students			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	0.74	0.05	2.28	0.32	**6.56**	**0.50**	**1.811**	**3.62**
1st	2.01	0.13	1.41	1.43	**3.95**	**0.25**	**1.166**	**3.39**
2nd	0.34	0.02	0.715	0.48	**2.80**	**0.18**	**0.848**	**3.30**
3rd	-0.56	-0.04	0.526	-1.07	**1.93**	**0.12**	**0.687**	**2.81**

Note: Results in bold indicate a statistically significant effect.

Table 33 shows the number of schools used in the analysis divided by the proportion of white students. Table 34 and Table 35 show the results of the analyses for differential effects of SAGE on white students for both mathematics and reading. In both mathematics and reading, VARC estimated a positive and significant impact of the SAGE program in first grade in schools with low proportions of white students and in kindergarten in schools with high proportions of white students.

Table 33: Number of SAGE and matched non-SAGE schools used in the white differential analysis in 2012-13

Grade	Schools with <50% White Students		Schools with >50% White Students	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	42	54	23	46
1st	58	53	34	80
2nd	60	63	65	123
3rd	86	69	64	154

Table 34: Results of the SAGE program effect on *mathematics* MAP growth by proportion white in 2012-13

Grade	Schools with <50% White Students				Schools with >50% White Students			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	0.84	0.06	1.297	0.64	**5.46**	**0.41**	**1.862**	**2.93**
1st	**2.67**	**0.17**	**1.096**	**2.44**	1.92	0.12	1.604	1.20
2nd	0.53	0.04	0.842	0.62	-0.47	-0.03	0.855	-0.55
3rd	0.95	0.06	0.602	1.57	-0.13	-0.01	0.601	-0.22

Note: Results in bold indicate a statistically significant effect.

Table 35: Results of the SAGE program effect on *reading* MAP growth by proportion white in 2012-13

Grade	Schools with <50% White Students				Schools with >50% White Students			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	3.26	0.22	3.057	1.07	**3.46**	**0.26**	**1.763**	**1.96**
1st	**3.42**	**0.21**	**1.270**	**2.69**	1.19	0.07	1.367	0.87
2nd	1.42	0.10	0.843	1.68	0.54	0.03	0.699	0.77
3rd	1.21	0.08	0.705	1.72	-0.03	0.00	0.527	-0.06

Note: Results in bold indicate a statistically significant effect.

Differential Effects by SAGE Funding

The final differential impact this evaluation examined was the effect of the SAGE program given schools reporting that SAGE funding may not cover all SAGE expenses. Table 36 shows the number of SAGE and non-SAGE schools used in the analysis based on the proportion of expenses covered by funding. In each of these cases, the analysis matched the two sub-groups of SAGE schools to all of the potential control schools.

Table 36: Number of SAGE and matched non-SAGE schools used in the SAGE funding differential analysis in 2012-13

Grade	0-50% Expenses Covered by SAGE Funding		51-100% Expenses Covered by SAGE Funding	
	SAGE	Non-SAGE	SAGE	Non-SAGE
K5	56	87	14	87
1st	63	133	19	133
2nd	69	186	42	186
3rd	73	223	58	223

Table 37 and Table 38 show the results of the analyses for the differential effects of SAGE given the proportion of program expenses covered through state SAGE program funding. Results from these analyses indicate that, with the exception of reading in kindergarten, SAGE schools with a higher proportion of their expenses covered by SAGE funding have larger positive effects upon student growth in comparison to the control group in mathematics and reading than schools with a lower proportion of their expenses covered by SAGE funding.

Table 37: Results of the SAGE program effect on *mathematics* MAP growth by SAGE funding in 2012-13

Grade	0-50% Expenses Covered by SAGE Funding				51-100% Expenses Covered by SAGE Funding			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	1.65	0.11	1.169	1.41	**4.88**	**0.37**	**1.748**	**2.79**
1st	**2.73**	**0.17**	**0.899**	**3.03**	**4.54**	**0.28**	**1.245**	**3.64**
2nd	-0.18	-0.01	0.689	-0.26	**1.92**	**0.12**	**0.818**	**2.35**
3rd	-0.18	-0.01	0.524	-0.35	**1.24**	**0.08**	**0.552**	**2.24**

Note: Results in bold indicate a statistically significant effect.

Table 38: Results of the SAGE program effect on *reading* MAP growth by SAGE funding in 2012-13

Grade	0-50% Expenses Covered by SAGE Funding				51-100% Expenses Covered by SAGE Funding			
	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
K5	**3.72**	**0.25**	**1.385**	**2.69**	3.07	0.23	2.292	1.34
1st	**2.99**	**0.19**	**0.916**	**3.26**	**5.59**	**0.35**	**1.414**	**3.95**
2nd	1.00	0.07	0.672	1.48	**3.17**	**0.20**	**0.803**	**3.95**
3rd	0.23	0.02	0.536	0.43	**1.37**	**0.08**	**0.533**	**2.58**

Note: Results in bold indicate a statistically significant effect.

Longitudinal SAGE Effects

As stated previously, this year VARC added an exploration of the effect of the SAGE program over multiple years. With three years of assessment data, the evaluation could examine growth from kindergarten through second grade. Table 39 shows the breakdown of the number and proportion of students by their status over the three years. The numbers on this table represent SAGE status in each of the three years. For instance, “000” represents a student being in a non-SAGE school for all three years, while “111” represents a student being in a SAGE school for all three years. As this table shows, of the 7,633 students tested in kindergarten of the first year, approximately 57 percent remained in a tested school for each of the three years. Of those who remained in tested schools, approximately 29 percent remained in non-SAGE schools for three years, and 45 percent remained in SAGE schools for three years.

Table 39: Number and proportion of tested students in kindergarten of 2010-11, SAGE status, and attrition

Specifications	Number of Students	Percent of Tested Three Years	Percent of Tested in Kindergarten
Tested in Kindergarten	7633	N/A	100.0%
000	1264	29.0%	16.6%
001	38	0.9%	0.5%
010	6	0.1%	0.0%
011	335	7.7%	4.4%
100	678	15.6%	8.9%
101	27	0.6%	0.4%
110	56	1.3%	0.7%
111	1948	44.8%	25.5%
Tested Three Years	*4352*	*100.0%*	*57.0%*
Untested or Mobile	*3281*	*N/A*	*43.0%*

Note: The first numeral in the Specifications refers to kindergarten, the second numeral to first grade, and the third numeral to second grade. Thus, 000 represents a student in a non-SAGE school for three years, 111 represents a student in a SAGE school for three years, 001 represents a student in a non-SAGE school for kindergarten and first grade who completed second grade in a SAGE school, and so forth.

Table 40 shows the results of the analysis that compared the difference in average growth between SAGE and non-SAGE students over three years. As these preliminary results indicate, the SAGE program had a positive and significant impact upon reading growth from kindergarten through second grade. The analysis could not detect a longitudinal effect of the SAGE program on mathematics growth.

Table 40: Results of the longitudinal SAGE program effect on *mathematics and reading* MAP growth from 2010-11 to 2012-13

Subject	Effect Size (S.S.)	Effect Size (Std. Dev.)	Std. Err.	T-stat
Mathematics	0.67	0.04	0.485	1.39
Reading	**1.94**	**0.12**	**0.530**	**3.65**

Note: Results in bold indicate a statistically significant effect.

VI. 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 9 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 improvements for future evaluations of the SAGE program.

Summary of Results

This evaluation utilized results from the 2012-13 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 students throughout the state take 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. These differences were larger in the earlier grades and diminished by third grade. This indicates a possible limited generalization of the evaluation's results to the larger population of all SAGE students, especially in the earlier grades. VARC did not dismiss the evaluation's results outright, though, as the tested population represents over 50,000 students or approximately one-quarter of the entire population of SAGE and non-SAGE students in kindergarten through third grade.

By examining the variance in characteristics between SAGE and non-SAGE schools and students within the set, other differences existed. SAGE schools in the set had a higher proportion of African-American students, a higher proportion of economically disadvantaged students, a higher proportion of students residing in rural districts, and a lower proportion of white students. Given these differences, the analysis first calculated a propensity score for each school, then used a two-stage statistical model to estimate value added while controlling for many of these characteristics, and finally matched schools by their propensity scores and compared the differences in average value-added results.

General results from the analysis found a trend of positive and significant effects of the SAGE program on mathematics and reading growth in kindergarten and first grade with a smaller but significant effect of the SAGE program on reading growth in second grade.

To determine if any difference in impact existed for a particular subset of the set, the evaluation also examined differential effects. The results from differential estimates for urban schools, schools with high proportions of economically disadvantaged students, and high proportions of limited-English proficient students indicated positive and significant effects of the SAGE program in first grade in mathematics and in all SAGE grades in reading. Differential effects for schools with low proportions of students with disabilities found positive and significant effects of the SAGE program in kindergarten through second grade in both mathematics and reading. Differential effects for schools with high proportions of African-American students found positive and significant effects of the SAGE program in all the SAGE grades in both mathematics and reading. Too few schools administered the MAP assessment in rural districts to estimate a significant effect of the SAGE program upon rural populations. Finally, differential effects based on SAGE funding found that schools with a higher proportion of their program expenses covered by SAGE funding had positive and significant effects of the program in both mathematics and reading in nearly all of the SAGE grades. Overall, these results suggest that while the SAGE program may not have a universal impact on all types of students, in general, the SAGE program may have a positive impact on student growth in the earlier grades and seems to have large impacts on targeted populations.

Preliminary results from the longitudinal analysis found a positive and significant effect of multiple years of the SAGE program on reading growth and found no significant effect on mathematics growth.

Current Context and Future Analyses

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 of this data and provide cautions around the use of the estimates in any decision-making processes.

Propensity Score Matching in Further Analyses

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.

While VARC constantly strives to improve the estimation method, VARC can further account for the selection of schools into SAGE by improving the understanding of how schools handle student enrollment processes and classroom configurations and by investigating district-level decision making with regard to funding.

Small Class Sizes in 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 allows for a comparison in performance between schools receiving SAGE funding and the ones 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. The possibility exists for the Year 10 evaluation to use a new set of data from DPI's Coursework Completion System data collection. This set of data contains information related to class size for all schools throughout the state. VARC will examine these data sources in the coming year to evaluate their usability as new control variables.

Non-Academic Outcomes

At this point of the SAGE program, VARC can only measure the effects of class size reduction on test scores. However, the small class size literature shows that greater quality classrooms and, in particular, small class sizes have positive impacts on non-cognitive untested skills, which in turn have large impacts on life outcomes. Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011)[8] found that students in classrooms with small class sizes were more likely to enroll in college by the age of 20. Additionally, students in classrooms with small classes exhibit statistically significant improvements on a summary index of home ownership, 401(k) savings, mobility rates, percent college graduates within a ZIP code, and marital status. Finn Gerber and Boyd-Zaharias (2005)[9] showed that students in classrooms with small class sizes are more likely to complete high school, and Krueger and Whitmore (2001)[10] showed that authorities are less likely to arrest these students for crime. In general, Chetty *et. al.* (2011) show that a better classroom environment from ages 5 to 8 has substantial long-term benefits even without intervention at earlier ages. The authors also document the fade-out and re-emergence effects and the potential role of non-cognitive skills in explaining this pattern. VARC's evaluation work next year will include a study of long-term outcomes for SAGE students which, given data limitations, may include an examination of high school completion.

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.

[8] Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics, 126* (4), 1593-1660.

[9] Finn, J. D., Gerber, S. B., & Boyd-Zaharais, J. (2005). Small Classes in the Early Grades, Academic Achievement, and Graduating from High School. *Journal of Educational Psychology, 97* (2), 214-223.

[10] Krueger, A. B. & Whitmore, D. M. (2001). The Effects of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *The Economic Journal, 111*, 1-28.

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.

For each district, DPI provided VARC with data describing students, schools, and, when available, student MAP test scores. Conceptually, the analysis uses 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 are constructed. Each analysis data set must include for each student: a grade level, a pretest and a posttest, a value for each of the control variables used in the model, the ID of the school attended, and whether the school received SAGE funding. For each pretest, a measure of the standard error of measurement is 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 2012-13

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	3821	18%	4773	12%	8594	14%
1st	4664	22%	7418	18%	12082	20%
2nd	5858	29%	10072	25%	15930	26%
3rd	7444	39%	12563	30%	20007	33%

Table A2: Number of SAGE and non-SAGE schools used in the main analysis and proportion of total schools by grade in 2012-13

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	77	19%	88	13%	165	15%
1st	92	22%	133	19%	225	20%
2nd	126	31%	186	26%	312	28%
3rd	151	38%	223	31%	374	34%

Pretest and Posttest Scores

For each grade in both reading and mathematics, the fall MAP score in RIT scale points is the student's pretest score, or the measure of the student's level of academic knowledge before they receive 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 is the student's posttest, or the measure of the student's level of academic knowledge after they received instruction from their teacher and school that year. The data on test scores are the most restrictive because the MAP assessment is not mandatory in Wisconsin; thus, only a subset of schools are included in the analysis. Moreover, VARC can only utilize MAP data when schools administer the test to 75 percent or more of their students to avoid problems due to schools using the assessment for purposes other than benchmarking. Because value-added estimation requires a pretest and posttest score from each student, VARC dropped students missing a pretest or a posttest from the analysis.

Standard Errors of Measurement of Pretest Scores

VARC included standard errors of measurement (SEMs) to correct for measurement error associated with each pretest. NWEA provides these SEMs; each score in the same grade and subject has 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*[11] 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 come from the biographical dataset. Gender categories are male and female. Race categories are Asian, African-American, Hispanic, American Indian, and white. If a student has 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 is qualified as being an English language learner, VARC assigned the ELL

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

dummy variable a value of one, zero otherwise. VARC dropped students missing a value for one 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 grade and subject. Formally, we can write the equation as

$$Y_2 = \lambda Y_1 + \mathrm{X}\beta' + S\alpha' + e \qquad (0.1)$$

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. 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 pretest 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 were then centered so that the estimates had a dose-weighted mean of zero. Formally,

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

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

- Drop observations that are not in kindergarten through third grade.
- Calculate school level (K-3) percentages of each demographic characteristic of the student body (percentage of students who are 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 we compute percentages for the free or reduced price lunch variable, we created new variables by computing the corresponding squared and cubic values. This variable transformation is 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

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

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 schools' characteristics.

Using estimated propensity scores, we use the Gaussian kernel matching algorithm in order to estimate effectiveness of the SAGE program on the productivity of schools. Kernel matching is a generalization of one-to-one matching. In one-to-one matching, each SAGE school is matched with a non-SAGE school whose propensity score is closest to the propensity score of the SAGE school. In the kernel matching algorithm, a SAGE school is matched with all non-SAGE schools, but non-SAGE schools whose propensity scores are closer to that of the SAGE school receive higher weights than the ones with further away propensity scores.

Specifically, we use the psmatch2 algorithm with Gaussian kernel matching algorithm option in Stata to estimate the SAGE effects by grade, where the left hand variable is the SAGE school indicator, and the outcome variable is value-added estimates and the propensity scores are the ones estimated previously.