Evaluation of the 2010-2011 SAGE Initiative:

An analysis of MAP achievement growth

August, 2012

Evaluation Team

Dr. Robert Meyer (Principal Investigator)

Dr. Curtis Jones (Project Director)

Penny Clark (Project Manager)

Grant Sim (Researcher)

Report Authorship

Dr. Curtis Jones

For questions about this report or the evaluation, please contact Dr. Curtis Jones at cjjones5@wisc.edu or Penny Clark at plclark@wisc.edu.

Evaluation of the 2010-2011 Wisconsin SAGE Initiative:

An analysis of MAP achievement growth

This report presents the ongoing work of the Value-Added Research Center (VARC) of the University of Wisconsin's Wisconsin Center for Education Research (WCER) evaluating the Student Achievement Guarantee in Education (SAGE) program. An initiative of the Wisconsin Department of Public Instruction (DPI), SAGE provides funds to schools across Wisconsin to provide small class sizes to students in Kindergarten through 3rd grade. SAGE schools are required to have student to teacher ratios of either 18:1 or 30:2. These ratios ensure that students in early grades are given individualized instruction necessary to promote their educational development.

Historically, evaluations of SAGE have been hampered by a lack of valid and reliable early-grade achievement data. However, new developments and trends in the assessment systems of Wisconsin districts have opened up possibilities for evaluating the impact of SAGE on student achievement. First, districts are increasingly recognizing the importance of testing students in early grades. Second, districts are increasingly recognizing the utility of benchmark testing students at the beginning and end of the school year.

Although several benchmark assessments are available, by far, the most popular assessment in Wisconsin is the Measures of Academic Progress (MAP) published by the Northwest Evaluation Association (NWEA). The MAP is an adaptive test of student achievement. Adaptive tests represent a technological advancement from typical achievement tests in that they adjust the difficultly of questions according to the difficulty of previous questions and whether they were answered correctly. Thus, student achievement is estimated based on a larger number of questions calibrated to their actual achievement level and fewer questions that are too easy or too difficult. It is typically administered on a computer and can be used with students as early as kindergarten.

The evaluation of the 2009-10 SAGE program was the first to make use of MAP data. This preliminary analysis used two-level (student and school) hierarchical linear models (HLM) to examine the impact of the SAGE program on schools. Due to a lack of matching between MAP data and state data, student level control variables were limited to race, gender, and Fall MAP scores; it was not possible to identify which students had an IEP, were eligible for F/R lunch, or were English Language Learners (ELL). To partially address the lack of adequate student controls, school controls (%F/R lunch, %minority, %ELL, and %IEP), which are publicly available, were included. For both reading and mathematics, the analysis 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 math and reading achievement, but the analysis lacked sufficient power and controls to conclude if this effect was reliable. The current evaluation of the 2010-2011 SAGE program addresses these deficiencies

by including a larger sample of schools and by matching MAP data with state data. Each year more districts both begin to use the MAP as their benchmark test and start to use it in early grades. This was particularly evident in Wisconsin in the 2010-2011 school year, with the Milwaukee Public Schools beginning to use the MAP as their benchmark. Further, by matching MAP data with state data, the evaluation is able to include a more complete array of student covariates for statistically isolating the SAGE effect.

Evaluation Questions

Question 1: What is the take-up rate of the MAP in SAGE districts and schools?

Question 2: How closely do SAGE schools using the MAP represent all SAGE schools?

Question 3: What is the difference between MAP growth in SAGE schools and non-SAGE schools?

Both questions one and two are designed to provide the foundation for questions three. To trust the results of statistical analyses exploring the impact of SAGE on student achievement, it is important that an increasingly large percentage of both SAGE and non-SAGE school use the MAP. This speaks to the power of the evaluation for reliably estimating the difference between SAGE and non-SAGE achievement. Further, given that not all SAGE schools use the MAP, it is important to understand how well the results of our evaluation generalize to the entire population of SAGE schools.

Methods

During the spring of 2012, DPI contacted each SAGE district to collect consent for data use. Once DPI had written permission from each district, they provided VARC with a data file including all MAP RIT scores for students in SAGE districts from the 2010-2011 school year. RIT scores are the scale used by the MAP to estimate achievement and are calculated through a specific Item-Response Theory (IRT) called Rasch Modeling (http://www.nwea.org/support/article/532/rit-scale). RIT scores range from 100 to 320. 53,778 student 2010-2011 MAP records were provided to the evaluation team. First, matching with state records was done by birthday, grade, school, and last name. Then, manual matching was done. This process resulted in 52,616 matches, which represents an excellent match rate of 98%.

There are several benefits to using the MAP that are worth noting:

1. The MAP is used in districts across the entire state of Wisconsin; thus allowing the SAGE evaluation to examine the impact of SAGE across the entire state. Previous SAGE evaluations were typically confined to the Madison and Milwaukee School Districts.
2. MAP is typically administered three times each year; at the beginning, middle, and end of a school year. Thus, annual growth measures can be calculated from the MAP that better approximate the annual achievement growth of students than is possible using the WKCE, which is administered only once a year, each fall. MAP is inclusive of early-grade students (K

to 2). The WKCE starts in 3rd grade and thus cannot be used to compare the annual achievement growth of early-grade students in SAGE to non-SAGE schools.

3. The psychometric properties of the MAP allow it to be more validly and reliably used for evaluation than the WKCE. An analysis of Wisconsin reading and math MAP scores suggests that it is both vertically and horizontally aligned. Vertical alignment refers to the comparability of student scores across grades. Horizontal alignment refers to the comparability of student test scores over time. The results of the 2009-2010 SAGE evaluation suggest that the MAP has acceptable vertical and horizontal alignment.

The evaluation team used statistical modeling to isolate the impact of SAGE on spring MAP scores. Two model types were used to test this impact. In model one, all SAGE schools were compared to all non-SAGE schools in SAGE districts, and thus answers the question "How much better or worse do students in SAGE schools perform on the MAP compared to students in non-SAGE schools across the state?" This model controls for student gender, race, F/R lunch eligibility, ELL, IEP, and both reading and math baseline MAP scores. It used fixed school effects with a clustered error term to identify the SAGE effect within each grade level. Thus, different models were fit for K, 1st, 2nd, and 3rd grade students.

$$SPRING\ MAP_{ij} = \gamma_0 + \gamma_1 * SAGE_j + \gamma_2 * FEMALE_{ij} + \gamma_3 * WHITE_{ij} + \gamma_4 * F/R\ LUNCH_{ij} + \gamma_5 * IEP_{ij} + \gamma_6 * ELL_{ij} + \gamma_7 * MATHFALL\ MAP_{ij} + \gamma_8 * READFALL\ MAP_{ij} + \gamma_9 * SCHOOL_j + u_j + r_{ij}$$

Model two included the same student covariates but instead treated school as a random effect, and included the fixed effect of district. Thus, the results of this model control for differences in school districts and answer the question, "How much better or worse do SAGE schools do compared to non-SAGE schools within their school districts?"

$$SPRINGMAP_{ij} = \gamma_0 + \gamma_1 * SAGE_j + \gamma_2 * FEMALE_{ij} + \gamma_3 * WHITE_{ij} + \gamma_4 * F/R\ LUNCH_{ij} + \gamma_5 * IEP_{ij} + \gamma_6 * ELL_{ij} + \gamma_7 * MATHFALL\ MAP_{ij} + \gamma_8 * READFALL\ MAP_{ij} + \gamma_9 * DISTRICT_j + u_j + r_{ij}$$

Results – MAP usage in SAGE schools

214 districts received SAGE funds during the 2010-2011 school year. Within these districts, 458 schools received SAGE funds while 279 did not. There were a total of 86,041students in SAGE schools. These were broken down by 21,954 kindergarten students, 21,971 first graders, 21,481second graders, and 20,635 third graders.

Of these schools, about half used the MAP as their benchmark test, but a much smaller percentage used it for kindergarten and first grade.[1] Overall, 87 out of 214 districts used the MAP for at least

[1] Only grade levels within schools that tested more than half of their students were included in the analyses. This was done to insure that the school was using the MAP as a benchmark test and not for another purpose.

one SAGE grade (41%), while 226 out of 458 SAGE schools (49%) and 156 out of 279 non-SAGE schools in SAGE districts used the MAP (56%) (Table 1, Figure 1).

Table 1: SAGE school MAP utilization

	SAGE Schools		Non-SAGE Schools	
K	104	23%	46	16%
1st	142	31%	87	31%
2nd	188	41%	126	45%
3rd	209	46%	135	48%
Overall	226	49%	156	56%

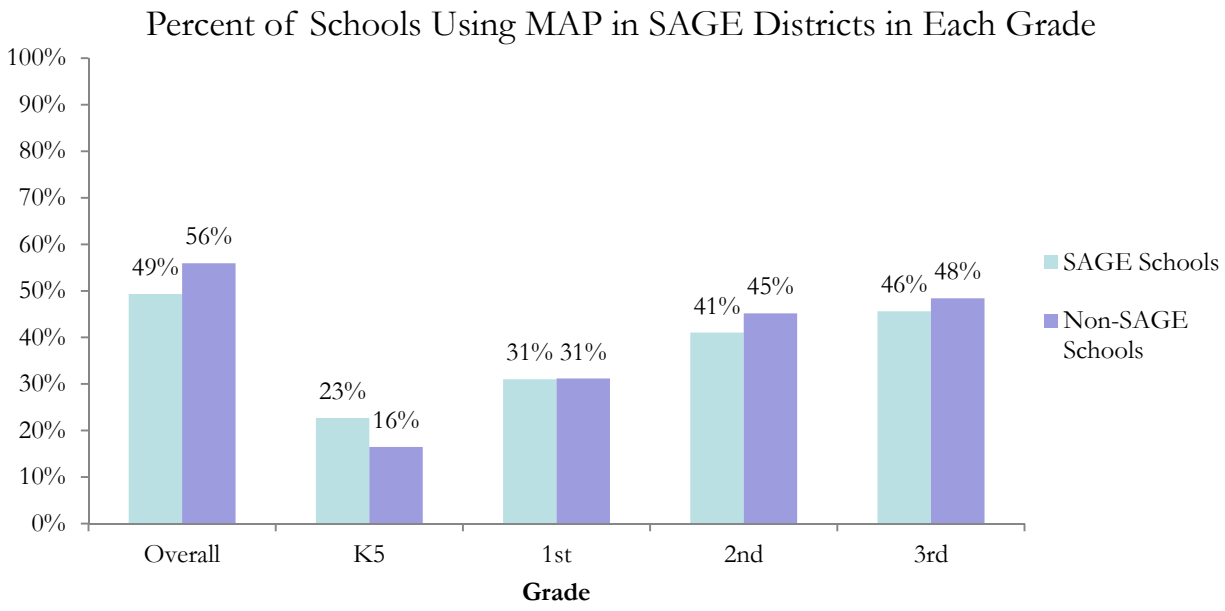

Figure 1: SAGE school MAP utilization

Fewer than half of all SAGE and non-SAGE students in SAGE districts take the MAP (Table 2, Figure 2). Again, a much smaller percentage of students took the MAP in kindergarten and first grade.

Table 2: SAGE student MAP utilization

	SAGE Students		Non-SAGE Students	
K	5,962	23%	2,158	16%
1st	7,412	31%	4,547	34%
2nd	9,459	41%	6,216	44%
3rd	10,585	46%	6,798	50%
Overall	33,148	39%	19,719	36%

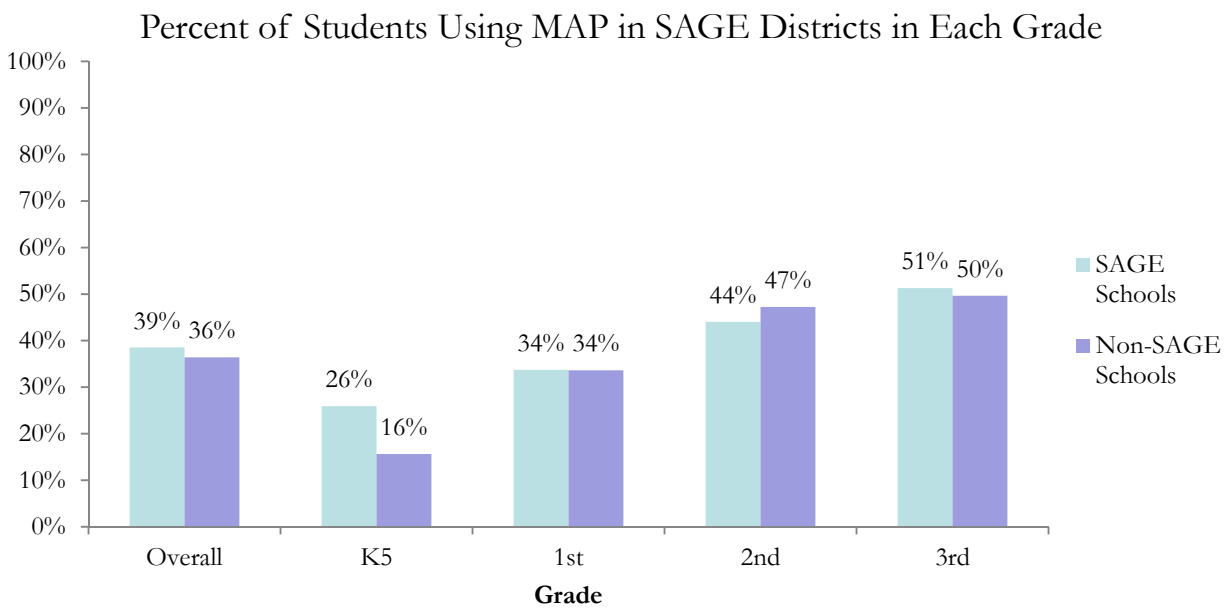

Figure 2: SAGE student MAP utilization

Taken together, these results show that although a larger number of SAGE schools and districts are using the MAP, the adoption of the MAP in early grades continues to lag. Thus, the power to detect an impact of SAGE, especially in kindergarten, is somewhat diminished.

Results – Characteristics of schools using and not using the MAP

Since fewer than half of SAGE schools and students take the MAP, it is important to consider the generalizability of any MAP analysis to the overall population of SAGE schools and districts. Thus, we compared the racial, ELL, F/R Lunch, and IEP characteristics of students taking the MAP to the total population of students in SAGE schools, and also to the overall population of SAGE districts. These comparisons were done across all grades and within each SAGE grade.

Figures three and four present the characteristics of SAGE students taking the MAP across all SAGE grade levels (K – 3rd). Based on these, it is clear that more SAGE students taking the MAP are African American, eligible for F/R lunch, and fewer are ELL than is represented in the population of all SAGE schools and even more so when compared to the entire population of students in SAGE districts.

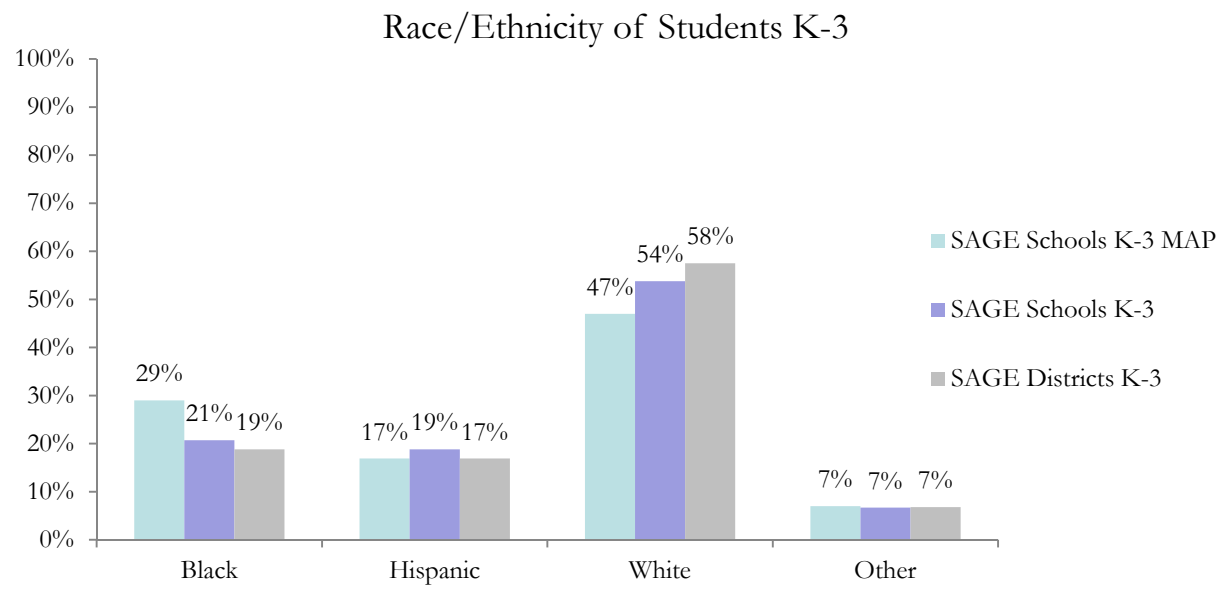

Figure 3: Overall comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – Race/Ethnicity

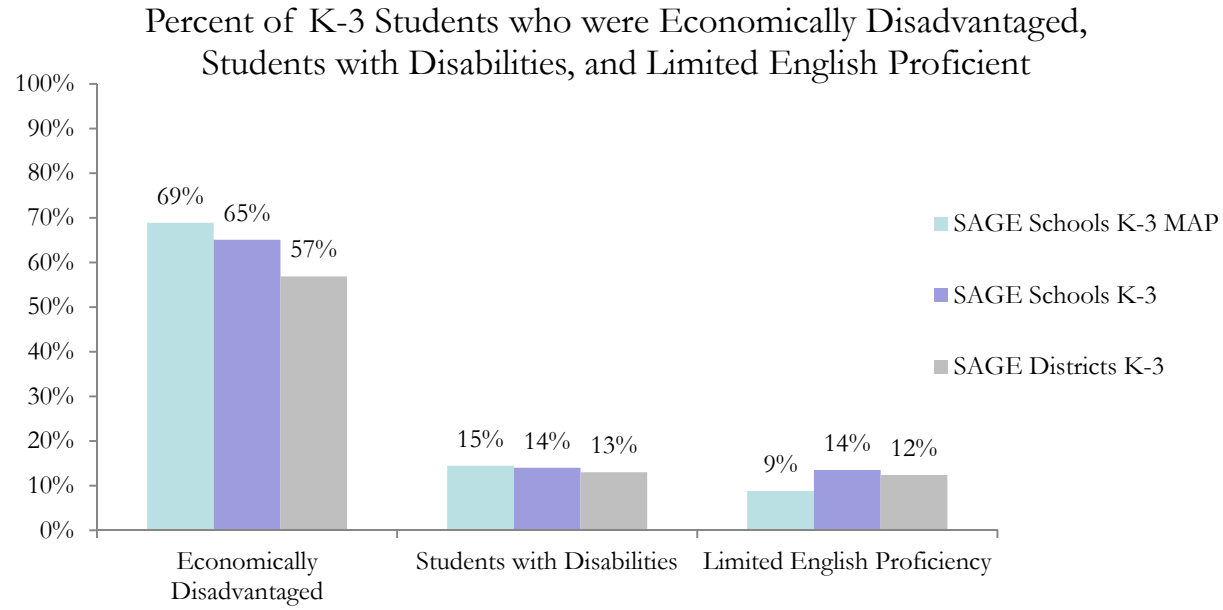

Figure 4: Overall comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – F/R lunch, IEP, ELL

Figures five and six present the characteristics of SAGE students taking the MAP specific to kindergarten. It is clear that the differences found overall are even more pronounced in kindergarten; more SAGE students taking the MAP are African American, eligible for F/R lunch, and fewer are ELL.

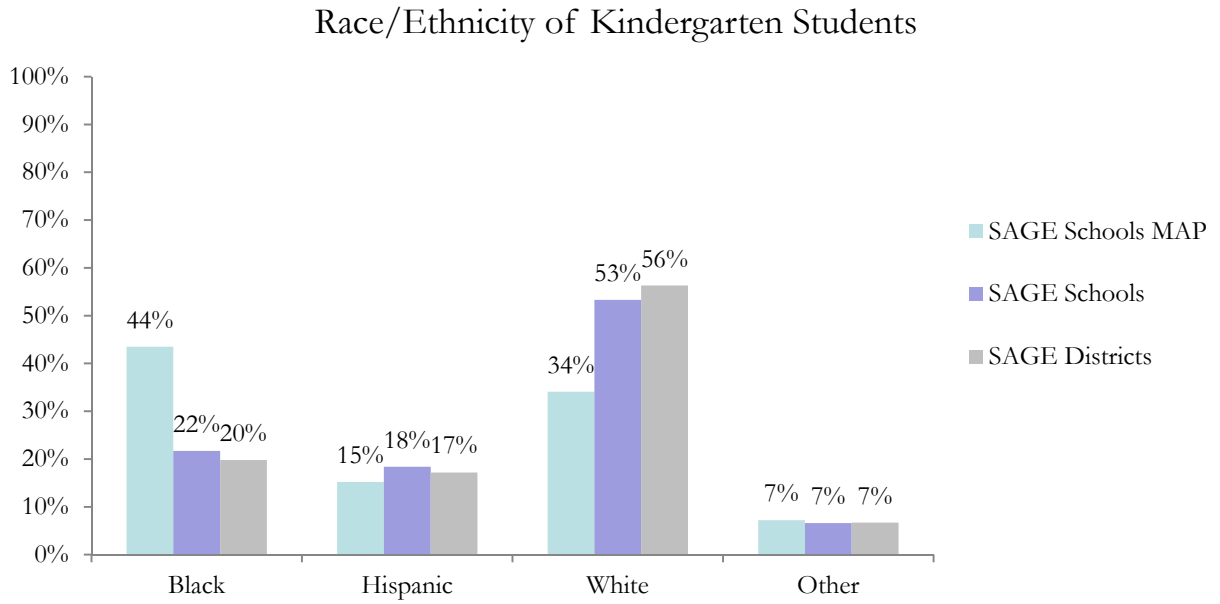

Figure 5: Kindergarten comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – Race/Ethnicity

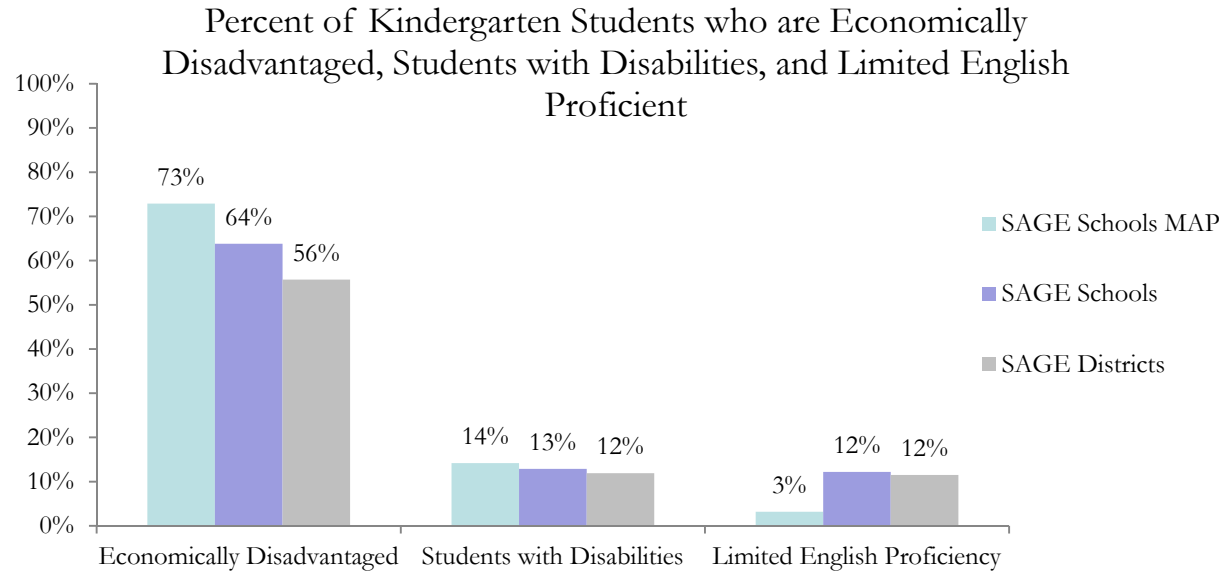

Figure 6: Kindergarten comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – F/R lunch, IEP, ELL

Figures seven and eight present the characteristics of SAGE students taking the MAP specific to first grade. Although still clearly present, the differences found in kindergarten are somewhat diminished.

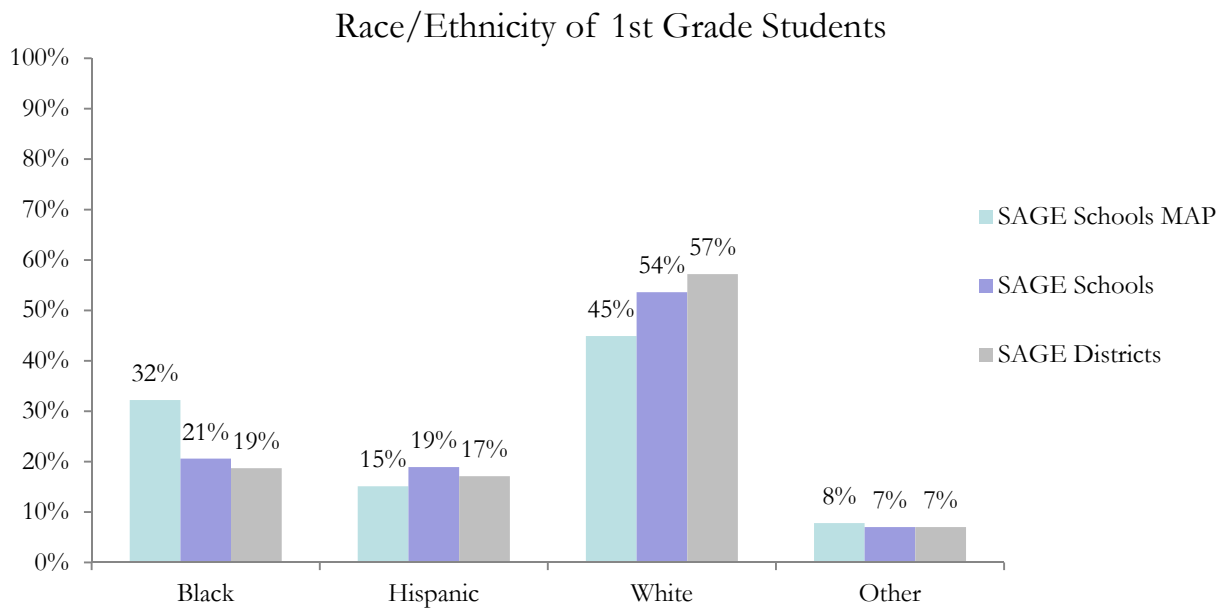

Figure 7: First grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – Race/Ethnicity

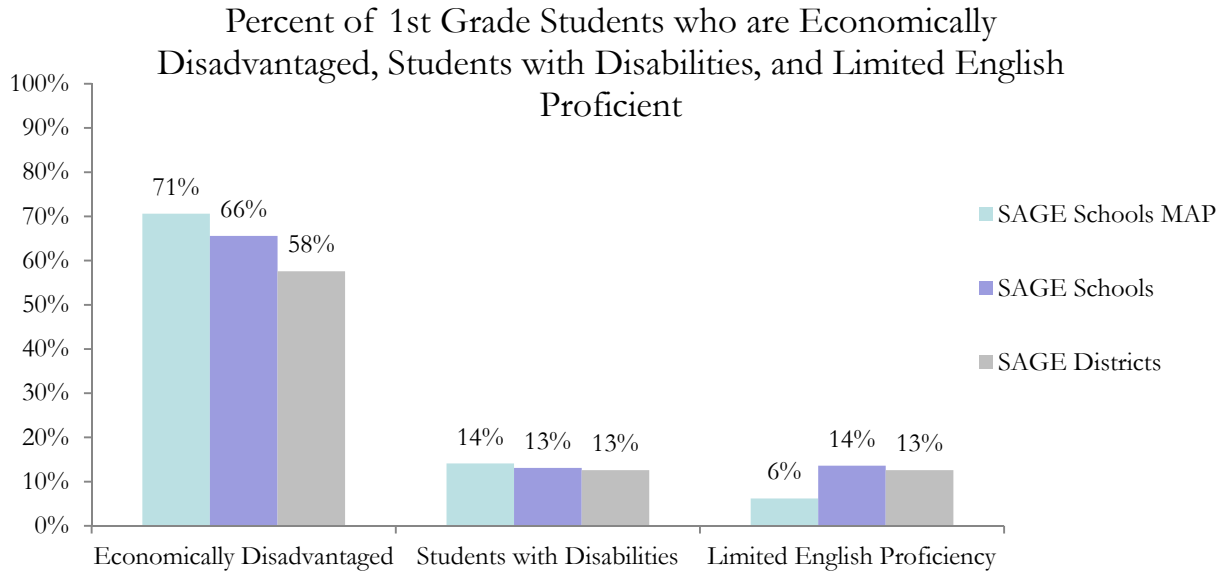

Figure 8: First grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – F/R lunch, IEP, ELL

Figures nine and ten present the characteristics of SAGE students taking the MAP specific to second grade. Again, the racial and economic differences are still clearly present. However, SAGE students taking the MAP in second grade are beginning to look more like the overall populations of SAGE schools and districts.

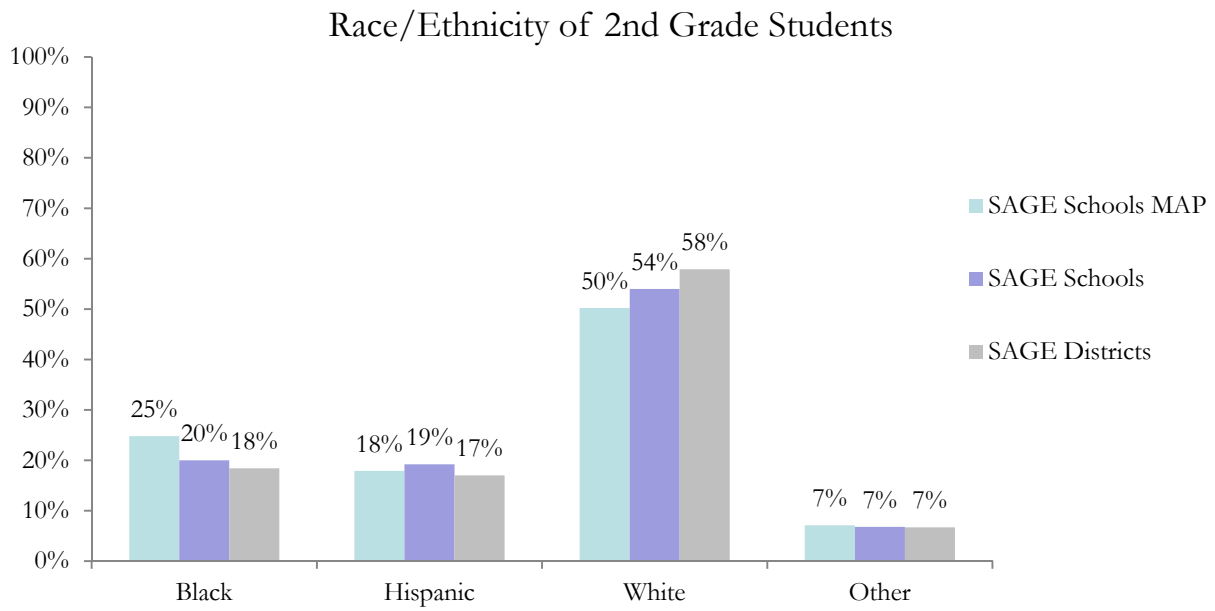

Figure 9: Second grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – Race/Ethnicity

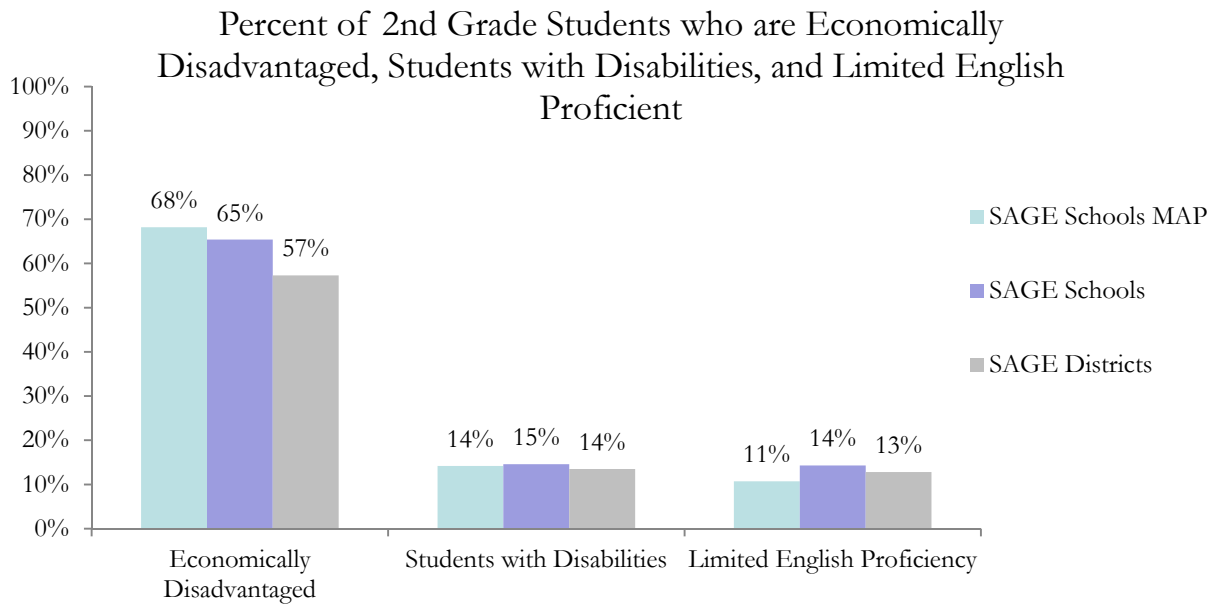

Figure 10: Second grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – F/R lunch, IEP, ELL

Figures eleven and twelve present the characteristics of SAGE students taking the MAP specific to third grade. In this grade, the racial and economic differences between SAGE students taking the MAP and SAGE schools are nearly gone. However, differences still remain between SAGE schools and non-SAGE schools more generally.

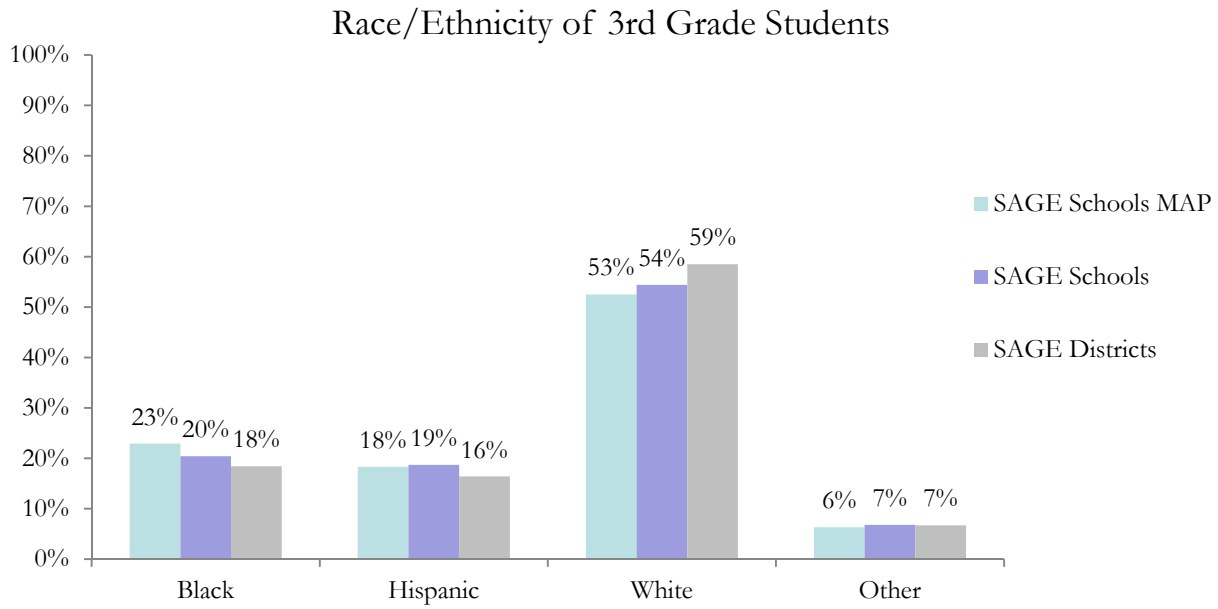

Figure 11: Third grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – Race/Ethnicity

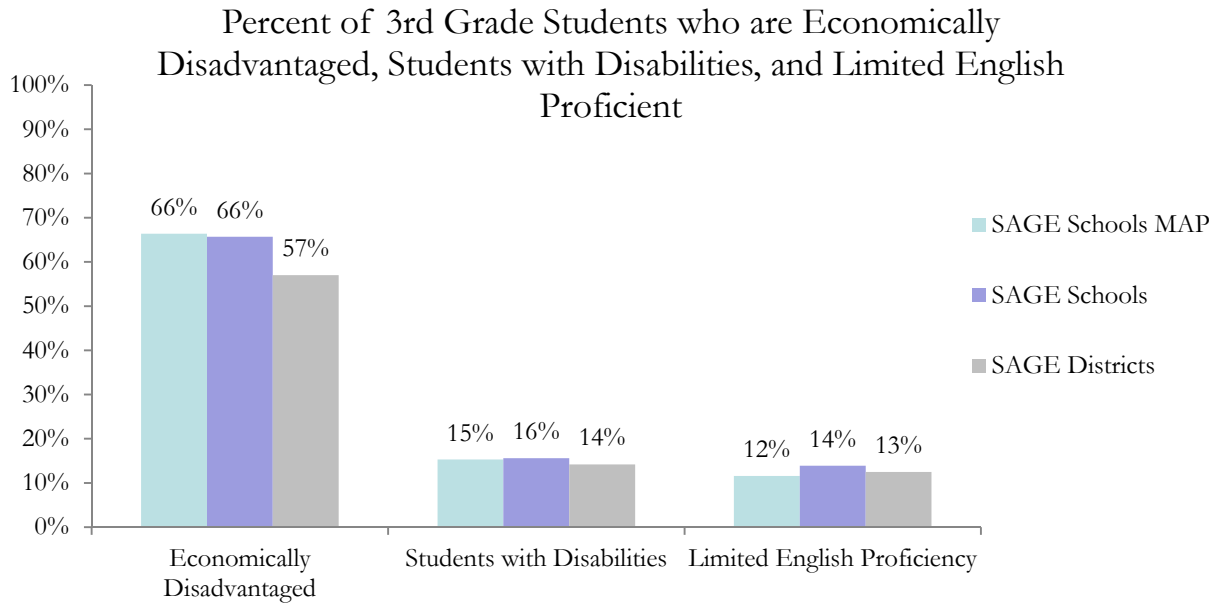

Figure 12: Third grade comparison of SAGE schools using the MAP to SAGE schools and SAGE Districts – F/R lunch, IEP, ELL

Based on these results, the evaluation of SAGE needs to be particularly cognizant of differences in student characteristics in kindergarten and first grade, and how these differences complicate the analysis of outcomes and the generalizability of findings. The less common the usage of the MAP within a grade level, the less similar students taking the MAP are as compared to both SAGE schools and SAGE districts. Although the statistical analysis the evaluation employs will include some statistical controls for these differences, it may be the case that SAGE schools using the MAP are qualitatively different than non-SAGE schools, and that statistical controls may not be adequate to fully control for these differences.

Results – MAP growth in SAGE and non-SAGE schools

Before statistically analyzing the differences between students in SAGE and non-SAGE schools, we explored the unadjusted differences in MAP scores and growth within each grade level analyzed. Based on the unadjusted numbers, students in SAGE schools start each grade level behind students in non-SAGE schools (Tables 3 & 4, Figure 13 & 14). For instance, SAGE students started kindergarten 1.54 RIT points behind non-SAGE students in reading. However, these differences are smaller at the end of the year, and in the case of Kindergarten and first grade, they actually disappear entirely.

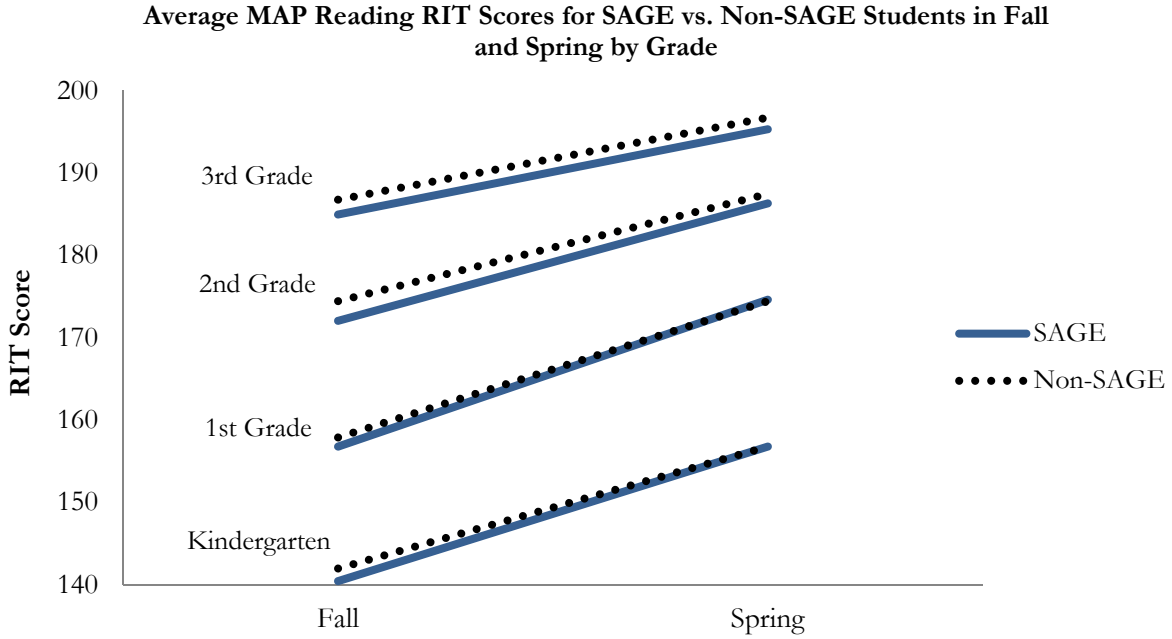

Figure 13: Average MAP reading scores of SAGE and non-SAGE students.

Table 3: Average reading RIT scale scores and standard deviations on the 2010-11 MAP for students in SAGE and non-SAGE schools

Grade	SAGE Fall	SD	SAGE Spring	SD	Non-SAGE Fall	SD	Non-SAGE Spring	SD
K5	140.44	10.47	156.78	13.02	141.98	10.64	156.78	13.08
1st	156.79	12.37	174.62	14.67	157.89	12.81	174.45	14.41
2nd	172.06	14.91	186.32	14.38	174.46	15.53	187.48	14.77
3rd	184.96	15.79	195.32	14.57	186.77	16.36	196.75	15.14

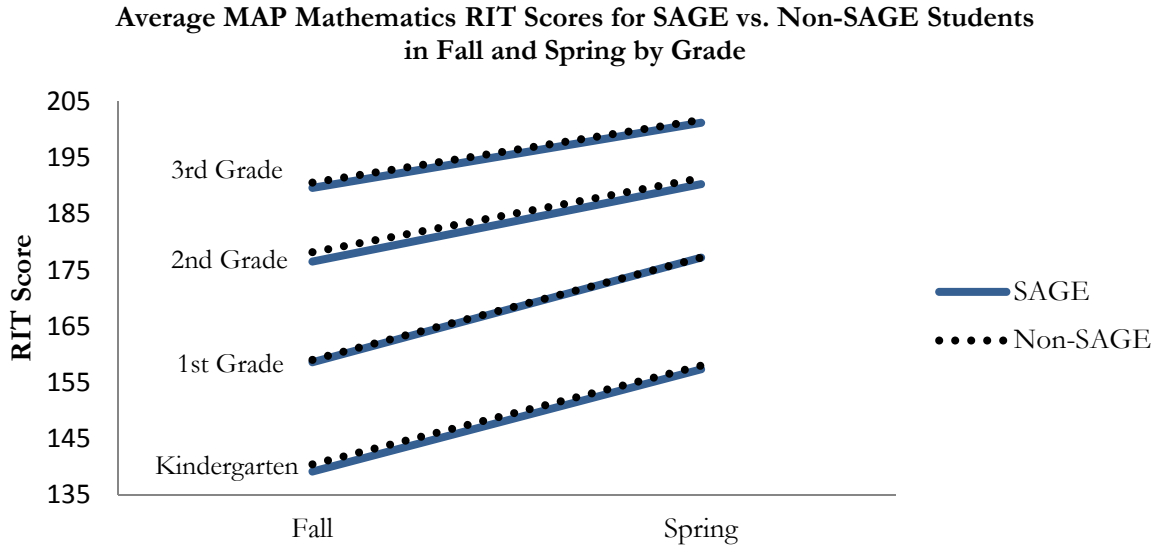

Figure 14: Average MAP math scores of SAGE and non-SAGE students

Table 4: Average math RIT scale scores and standard deviations on the 2010-11 MAP for students in SAGE and non-SAGE schools

Grade	SAGE Fall	SD	SAGE Spring	SD	Non-SAGE Fall	SD	Non-SAGE Spring	SD
K5	139.12	11.86	157.33	14.01	140.41	12.61	157.98	14.2
1st	158.58	13.88	177.22	14.05	159.01	14.4	177.22	14.26
2nd	176.51	13.03	190.3	13.57	178.18	13.64	191.41	13.84
3rd	189.65	12.86	201.22	13.17	190.57	13.49	201.75	13.99

We also compared the unadjusted MAP scores of SAGE and non-SAGE students, only including students who took both the fall and spring MAP tests (Tables 5 & 6, Figure 15 & 16). These numbers were consistent with the results found in the previous set of figures; kindergarten and first grade SAGE students made up ground during the course of the year, while second and third grade SAGE students continued to lag behind.

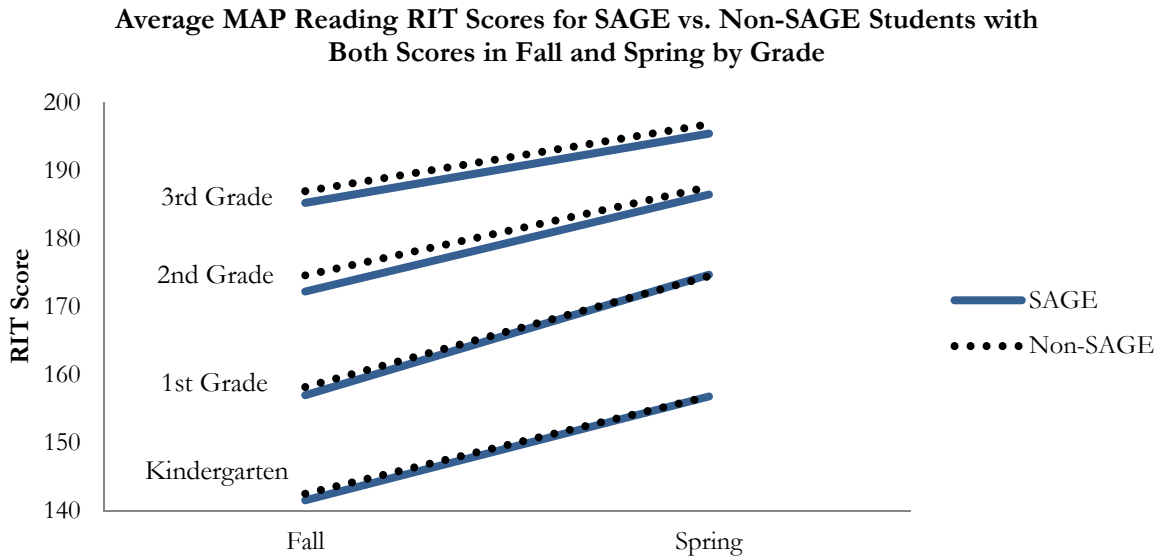

Figure 15: Average MAP reading scores of SAGE and non-SAGE students: Students with both tests

Table 5: Average reading RIT scale scores and standard deviations on the 2010-11 MAP for students in SAGE and non-SAGE schools: Students with both tests.

Grade	SAGE Fall	SD	SAGE Spring	SD	Non-SAGE Fall	SD	Non-SAGE Spring	SD
K5	141.51	10.03	156.79	13.03	142.49	10.44	156.8	13.08
1st	156.97	12.32	174.68	14.69	158.17	12.79	174.47	14.42
2nd	172.21	14.8	186.45	14.35	174.58	15.43	187.58	14.8
3rd	185.24	15.64	195.4	14.52	186.94	16.26	196.81	15.09

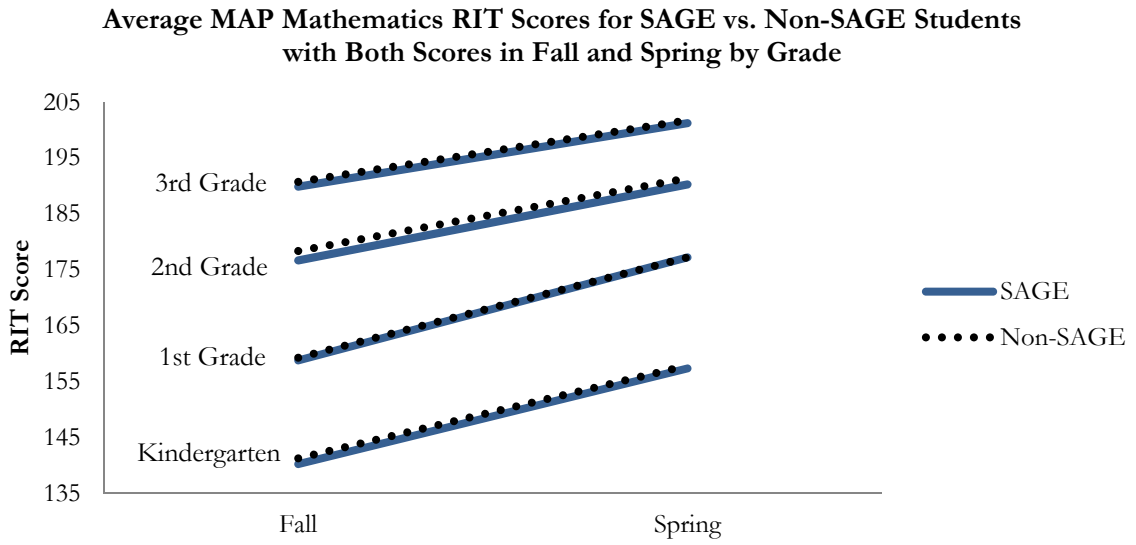

Figure 16: Average MAP math scores of SAGE and non-SAGE students: Students with both tests

Table 6 Average mathematics RIT scale scores and standard deviations on the 2010-11 MAP for students in SAGE and Non-SAGE Schools: Students with both tests.

Grade	SAGE Fall	SD	SAGE Spring	SD	Non-SAGE Fall	SD	Non-SAGE Spring	SD
K5	140.23	11.47	157.38	13.96	141.27	12.17	157.87	13.99
1st	158.79	13.78	177.25	14.02	159.29	14.35	177.25	14.25
2nd	176.7	12.9	190.31	13.55	178.39	13.51	191.43	13.84
3rd	189.9	12.73	201.26	13.15	190.75	13.42	201.78	13.97

Figures 17 and 18 present the numbers converted into MAP growth. Again, these numbers suggest that SAGE students may have benefitted from SAGE, especially in the early grades and in reading. However, as mentioned earlier, great differences exist between the characteristics of SAGE and non-SAGE schools that may limit the accuracy of these findings. The following set of analyses is designed to statistically control for these differences and obtain a more valid estimate of the SAGE effect.

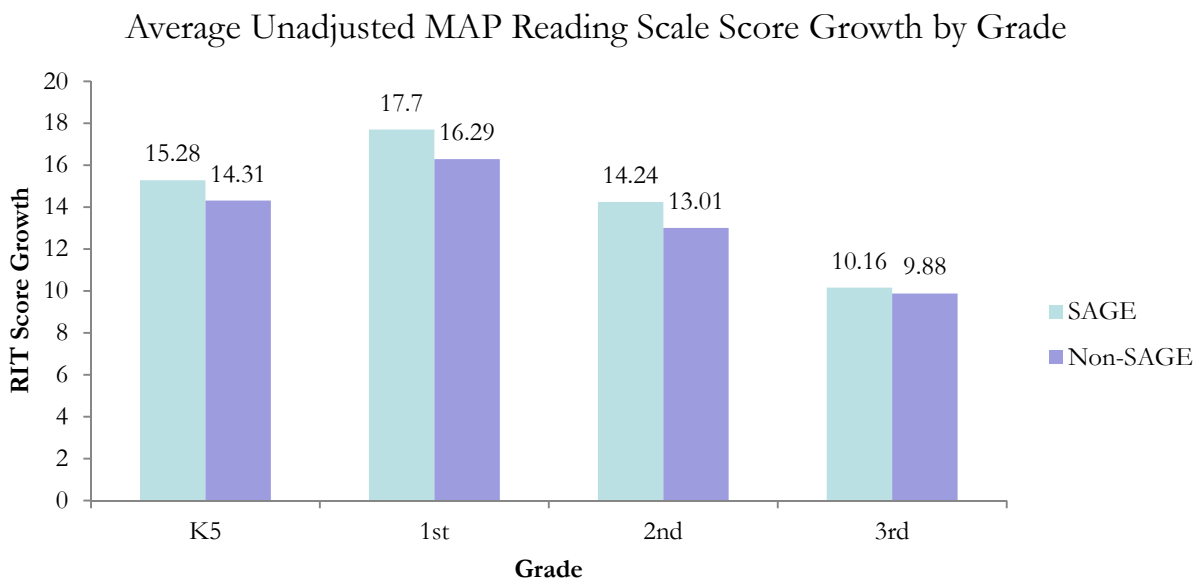

Figure 17: Average RIT reading growth of SAGE and non-SAGE students.

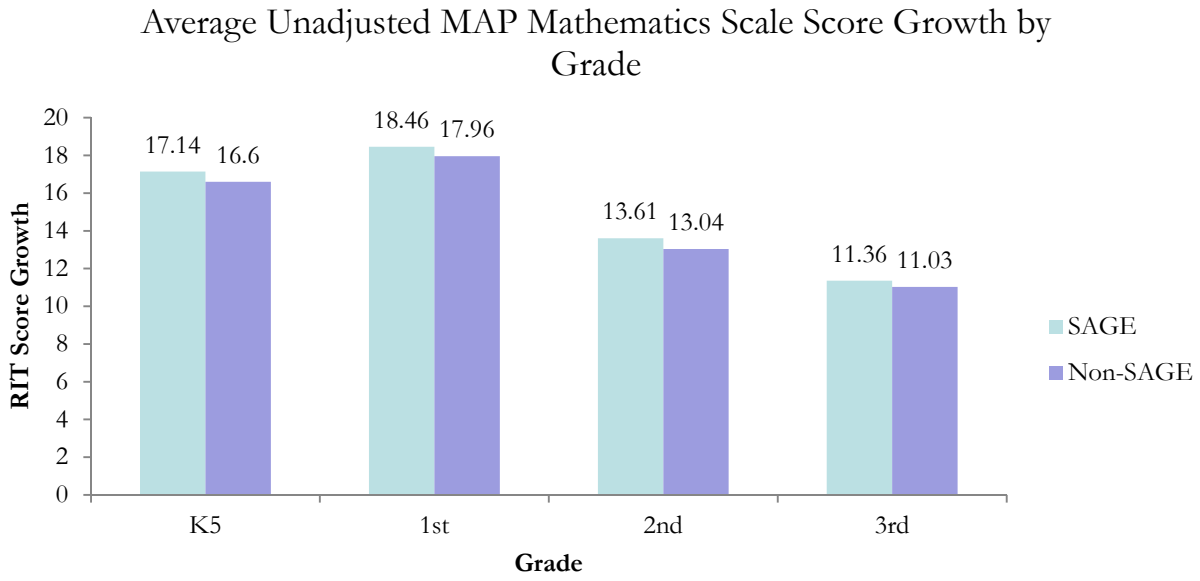

Figure 18: Average RIT math growth of SAGE and non-SAGE students.

Tables seven and eight present the RIT score differences between students in SAGE and non-SAGE schools, adjusted for student baseline achievement, and differences in racial and social characteristics (see methods section for the specific models). In model one, students in SAGE schools were compared to students in non-SAGE schools across all SAGE districts. According to this model, first-grade SAGE students gained significantly more in reading and math than students in non-SAGE schools. Kindergarten SAGE students also gained more than non-SAGE students in reading, but this difference only approached significance ($p = .08$). In model two, comparing students in SAGE schools to students in non-SAGE schools within the same districts, resulted in no first grade differences. However, the kindergarten reading difference still approached significant ($p = .07$). No differences were found, using either model, in the reading or math achievement of second or third grade students. The differences in the results of each model suggests a significant school district effect on MAP scores.

Table 7: Results of statistical analyses of the SAGE program effect upon mathematics MAP scale scores

Model 1			**Model 2**		
Grade	SAGE Coefficient	P-value	Grade	SAGE Coefficient	P-value
K5	1.129	0.476	K5	0.331	0.608
1st	4.149	0.000	1st	0.727	0.451
2nd	-1.119	0.293	2nd	-0.839	0.064
3rd	1.042	0.384	3rd	0.464	0.429

Table 8: Results of statistical analyses of the SAGE program effect upon reading MAP scale scores

Model 1			**Model 2**		
Grade	SAGE Coefficient	P-value	Grade	SAGE Coefficient	P-value
K5	1.468	0.080	K5	0.997	0.074
1st	5.433	0.010	1st	0.629	0.410
2nd	-0.078	0.957	2nd	-0.059	0.911
3rd	0.814	0.451	3rd	0.050	0.917

Conclusions

The results of the evaluation of the 2010-2011 SAGE program are consistent with the results from the 2009-2010 evaluation. Specifically, there is evidence that the SAGE program is having an impact on the reading and math achievement of early grade (K and 1st) but not later grade students. Again though, the ability to detect an impact on kindergarten students is inhibited by smaller numbers of schools participating in the MAP at these grade levels.

Further, large differences in the characteristics of SAGE schools and non-SAGE schools, especially in the early grades, represents an analytic challenge. Generally, SAGE schools participating in the MAP have many more African American and F/R lunch eligible students than non-SAGE schools. Since the numbers of F/R lunch eligible students in a school is a factor for their deciding to become a SAGE school, this difference of course makes sense. It is likely that there are no schools in Wisconsin that look like SAGE schools that are not receiving SAGE funding. Controlling for student characteristics may not adequately control for this selection bias. Still, if a selection bias exists, one would expect that the achievement of schools with greater numbers of low-income students would actually be negatively affected by this. Thus, the unobserved bias that may affect the results of the current evaluation may actually be suppressing the SAGE effect.

Another interesting result found in the current evaluation was that, although kindergarten and first grade students in SAGE schools seem to make up the achievement gap that existed at the beginning of the school year, second and third grade students again present with this achievement gap at the beginning of the year. While this may be due to differences in cohorts, it may also be due to environmental factors, like the summer slide, that disproportionally affect students in SAGE schools. If this is true, most of the gains made by SAGE students are lost over the summer. Future evaluation work following students across multiple years will provide a more clear opportunity for testing this hypothesis.