Year 6 SAGE Evaluation Final Report:

A Follow-up Analysis of SAGE Cohorts in Madison, and a Preliminary Study of SAGE across the State of Wisconsin

June, 2011

A report prepared by the Value-Added Research Center

Wisconsin Center of Education Research

University of Wisconsin, Madison

Research by: Curtis Jones, Ph.D.
Dan Bolt, Ph.D.
Mikhail Pyatigorsky, Ph.D.

For additional information about this report, please contact:

Curtis Jones at cjjones5@wisce.edu

Year 6 SAGE Evaluation – Analysis in Madison, and the State of Wisconsin

Executive Summary

In this report, we compare SAGE versus non-SAGE participating public schools with respect to student growth in Madison and statewide during the 09-10 school year.

Madison Results

In Madison, the current analyses follow two cohorts of students that have now been observed from kindergarten into fourth and fifth grades. The purpose of this was to examine if reading and math WKCE scores differ for the 04-05 and 05-06 SAGE cohorts, as compared to students in non-SAGE schools, as students move beyond their early-grade SAGE experiences, into 4th and 5th grade. The results suggest the following:

- After controlling for student and school differences, there were no statistically significant ($p < .5$) relationships found between SAGE participation and student achievement through 3rd and 4th grade.
- We found a clear trend that students attending SAGE schools, on average, outperformed what would be predicted of them had they attended non-SAGE schools.
- SAGE students in kindergarten during the 04-05 school year scored 12 scale points higher in reading and 7.7 points higher in math on the 09-10 WKCE (fall of 5th grade) than would be predicted had they not attended SAGE schools.
- SAGE students in kindergarten during the 05-06 school year scored 6.7 scale points higher in reading and 11 points higher in math on the 09-10 WKCE (fall of 4th grade) than would be predicted had they not attended SAGE schools.
- Although statistical significance criteria were not met, the magnitude of SAGE student effects ranged from .13 standard deviations in reading for the 05-06 SAGE cohort to .23 standard deviations for the 04-05 cohort in reading.
- Applied to school Value-Added, the SAGE effect magnitude ranged from .37 standard deviations in math for the 04-05 cohort to.74 in math for the 05-06 cohort.

Statewide Results

A statewide analysis of SAGE was made possible through a partnership between VARC and the Northwest Evaluation Association (NWEA). As part of this partnership NWEA obtained permission from districts across Wisconsin to share with VARC their scores on the Measures of Academic Progress (MAP) benchmark assessments. These assessments have several advantages for the evaluation including that it is a validated instrument for measuring early grade achievement in reading and math. Although the preliminary statewide results suggest positive trends, especially in Kindergarten, of SAGE participation in both reading and math, these analysis were inconclusive due to smaller MAP participation in grades K and first. However, the trend in Wisconsin is that more schools and districts have adopted the MAP each year. Thus, as the evaluation moves to its 7th year and beyond, more conclusive analyses will likely be possible.

Table of Contents

Madison Analyses:

Impact of SAGE on WKCE Scores of Students in Later Grades

Our analyses of the SAGE program in Madison were designed to answer the following two questions:

1. What is the impact of the SAGE for the 2004-2005 kindergarten cohort?
2. What is the impact of SAGE for the 2005-2006 kindergarten cohort?

To answer these questions, we analyzed 2009-2010 WKCE reading and math scale scores, comparing the performance of students who had attended SAGE schools to those who did not.

Methods

There are a total of 23 SAGE schools and 9 non-SAGE (comparison) schools in the Madison analyses. Only students who had complete test data (both a Kindergarten screener and WKCE scores) and had not switched schools since kindergarten were included in the analyses. This resulted in a total of 458 students from the 05-06 cohort and 525 from the 04-05 cohort being included in the analyses.

Hierarchical linear modeling (HLM) was used to estimate the impact the SAGE program had on WKCE reading and math scores. In previous SAGE reports, several differences between SAGE and non-SAGE schools were presented, many of which contribute to test score performance. Thus, controlling for these differences is necessary to disentangle the impact of SAGE from the impact of other factors related to SAGE. To this end, the current set of analysis controls for the following student level variables:

- Race/ethnicity
- Gender
- F/R lunch eligibility
- English Language Learner (ELL) Status
- Disability status
- Kindergarten screener scores

And the following school-level variables, which were obtained from the National Center of Education Statistics:

- Percent minority
- Percent eligible F/R lunch
- Percent male
- Average Kindergarten Screener score

Taken together, these controls were included in the following statistical model used test the impact of SAGE:

$$\textit{09-10 WKCE Scale Score}_{ij} = \gamma_{00} + \gamma_{01}*SAGE_j + \gamma_{02}*\%MALE_j + \gamma_{03}*\%WHITE_j + \gamma_{04}*\%FREE\_LUN_j + \gamma_{05}*K\_SCREENER_j + \gamma_{10}*K\_SCREENER_{ij} + \gamma_{20}*MALE_{ij} + \gamma_{30}*ELL_{ij} + \gamma_{40}*IEP_{ij} + \gamma_{50}*WHITE_{ij} + \gamma_{60}*LUNCH_{ij} + \gamma_{70}*HISPANIC_{ij} + \gamma_{80}*BLACK_{ij} + u_{0j} + r_{ij}$$

Results

Table 1 presents the summarized results of the two statistical models used to predict 09-10 WKCE reading scale scores for the 04-05 and 05-06 SAGE cohorts. The "non-SAGE" column presents the predicted scale score for students in non-SAGE schools. The "Coeff" column presents the "Value-Added" or impact of SAGE for each cohort. Thus, the coefficient of 12 suggests that SAGE students scored 12 scale points higher than would be predicted in a non-SAGE school. The "Standard Error" column is used to show how confident we can be that the coefficient is not due to chance. The coefficient and standard error together are used to determine the p-value. The student and school effect sizes represent the standardized SAGE effect on student and school achievement in terms of standard deviations. According to the modeling results, although neither model detected a statistically significant SAGE effect, the magnitude of both coefficients and their corresponding effect sizes suggest this was at least in part due to having an inadequate sample size. Twelve and seven scale point impacts on the WKCE are certainly of practical significance even if not statistically significant.

Table 1: Results of models predicting 09-10 WKCE Reading Scale Scores for the 04-05 and 05-06 SAGE cohorts

	09-10 WKCE Reading Results					
	Non-SAGE	Coeff	Standard error	p-value	Student effect size	School Value-Added effect size
0405 cohort (5th graders)	480.8	12*	6	0.06	0.23	0.68
0506 cohort (4th graders)	476.4	6.7	6	0.286	0.13	0.47

Table 2 presents the summarized results of the two statistical models used to predict 09-10 WKCE math scale scores for the 04-05 and 05-06 SAGE cohorts. The results of these models are similar to the reading models presented in Table 1. Although neither model detected a statistically significant "Value-Added" or SAGE impact, the magnitude of the 8 and 11 scale point differences in SAGE student performance are certainly of practical significance. The large standard error terms suggest that our sample size is inadequate for testing the statistical significance of the SAGE effect on these cohorts.

Table 2: Results of models predicting 09-10 WKCE Math Scale Scores for the 04-05 and 05-06 SAGE cohorts

	09-10 WKCE Math Results					
	Non-SAGE	Coeff	Standard error	p-value	Student effect size	School Value-Added effect size
0405 cohort (5th graders)	498.7	7.7*	7.1	0.293	0.15	0.37
0506 cohort (4th graders)	473.8	11	5.3	0.055	0.22	0.74

Summary Conclusions

In summary, the results in Madison suggest a moderately-sized positive effect associated with the SAGE program. Even after moving beyond SAGE grades, SAGE students seem to achieve greater than would be predicted had they attended a non-SAGE school. A number of limitations with the current analyses should be noted. First, the analysis is descriptive, meaning alternative interpretations to the differences between SAGE and comparison schools (other than effects due to SAGE) are possible. Second, patterns of missing data (particularly for pre-test scores) may result in biased estimates of school effects. Finally, the sample of students and schools was not adequate for testing the significance of SAGE impacts. As more cohorts move beyond the SAGE grades, it will be possible to pool the cohorts together to more reliably test the impact of SAGE in latter grades.

A Statewide Analysis of the 2009-2010 SAGE Program:

Effects on MAP scores

This section presents the preliminary results of a state-wide analysis of the impact of the SAGE program on student and school reading and math achievement during the 2009-2010 school year. Although the results in the section are useful for understanding the impact of SAGE, these should be treated as preliminary due to data limitations explained later in the document. This section begins with an introduction to the Measures of Academic Progress (MAP) and the strengths and challenges for using to evaluate SAGE. It then presents MAP participation numbers and characteristics of students and schools taking the MAP across Wisconsin. Finally, the outcomes of statistical models built to test the differences between the academic growth of SAGE students and schools compared to non-SAGE students and schools are presented.

The MAP

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 most commonly administered on a computer, and can be used with students as early as kindergarten.

The publisher of the MAP, Northwest Evaluation Association (NWEA), partnered with VARC to collect consent from all school districts in the state of Wisconsin that use the MAP to allow VARC to use their MAP data for Value-Added modeling and evaluation purposes. During the fall of 2010 and winter of 2011, NWEA contacted each district and collected consent from willing districts. Districts were given the choice to share student and school identifiable data with VARC, to not share, or to share de-identified data. Once NWEA had permission from a district, they provided VARC with a data file. Data were provided from the 2005-2006 to 2009-2010 school years.

MAP RIT scores are what will be used in the SAGE evaluation. 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 120 to 270.

Benefits of using the MAP in the SAGE evaluation

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 test the impact of SAGE across the entire state. Previous SAGE evaluations were typically confined to the Madison and Milwaukee School Districts. Further, since the MAP is administered across the entire state it may provide the ability to explore the impact of SAGE across and within diverse locations. Thus, it may be possible to disentangle the impact of SAGE in rural, urban, and suburban 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. A vertically aligned test like the MAP suggests that an achievement scale score in 3rd grade is directly comparable to a 7th grade score. Figures 1 and 2 present the typical progression of student MAP scale scores from Kindergarten to 9th grade. Horizontal alignment refers to the comparability of student

test scores over time. Figures 3 and 4 present the historical math and reading scale scores in Wisconsin broken down by grade level. Although these figures do show some variability across years, much of that variability is concentrated in the early grades, where smaller sample sizes may result in the observed pattern. In later grades, where larger samples result in smaller standard error terms, average scores are more consistent from year to year. Taken together, these figures suggest that the MAP has acceptable vertical and horizontal alignment.

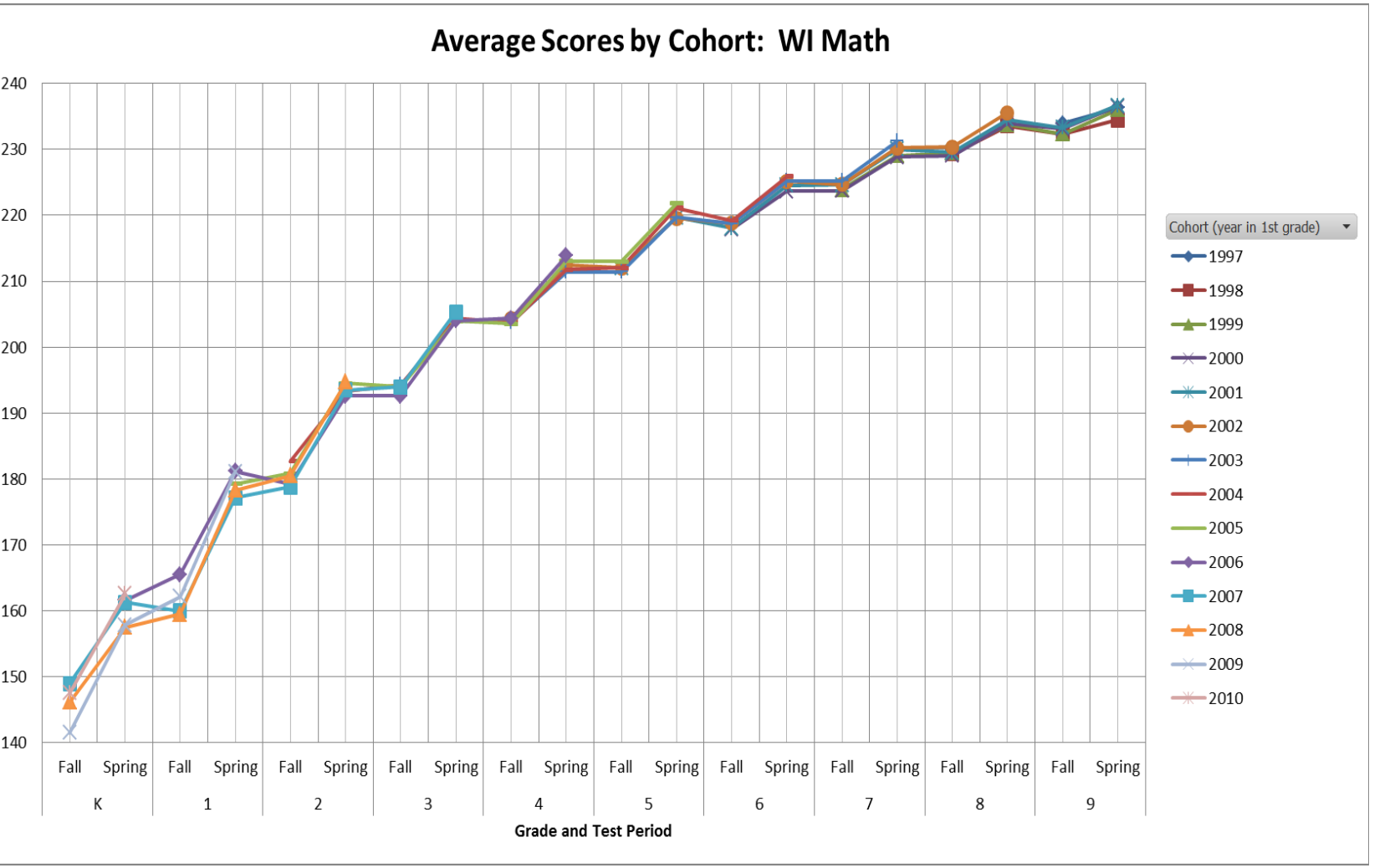

Figure 1: Vertical alignment of Math MAP Test in Wisconsin

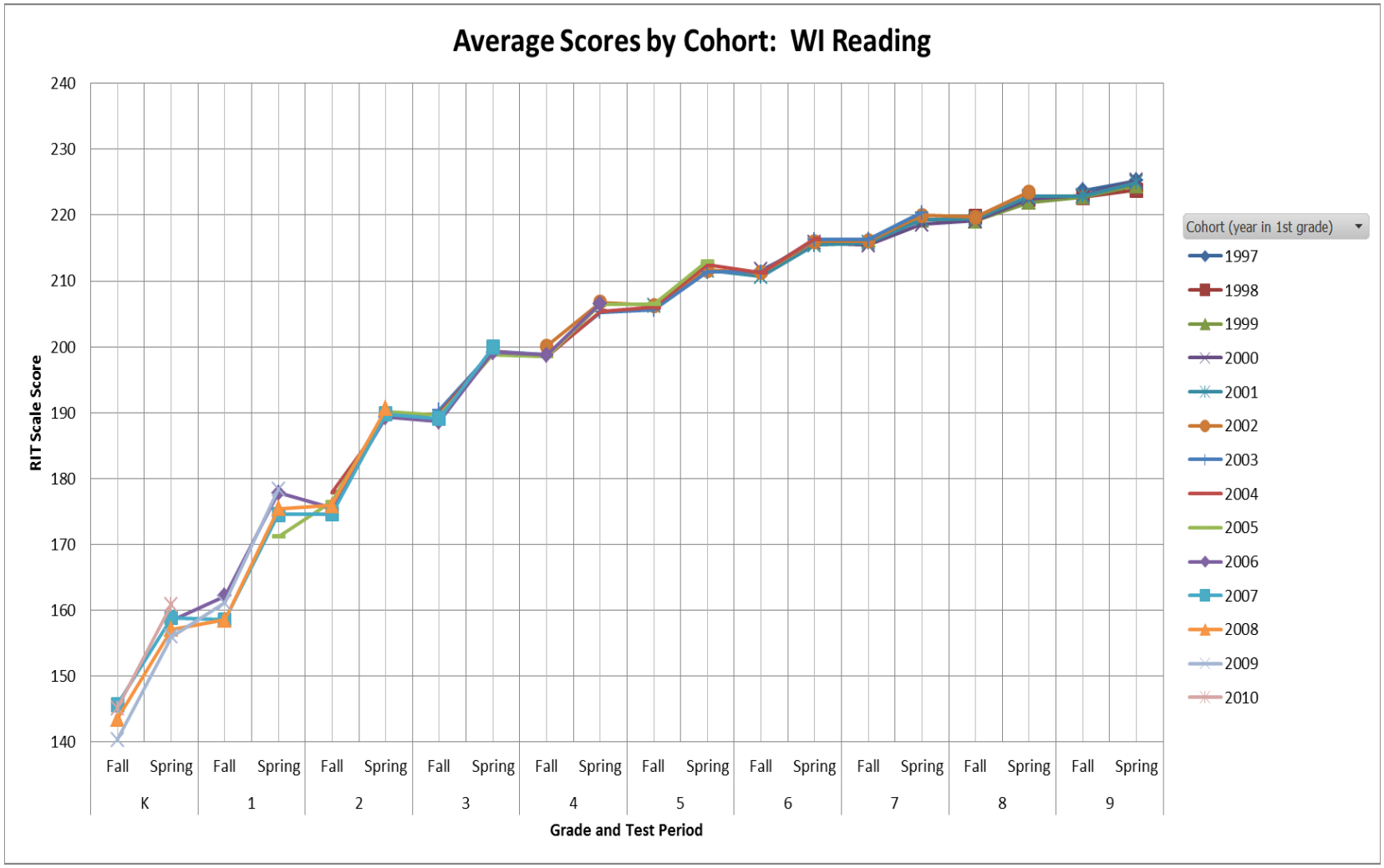

Figure 2: Vertical alignment of Reading MAP Test in Wisconsin

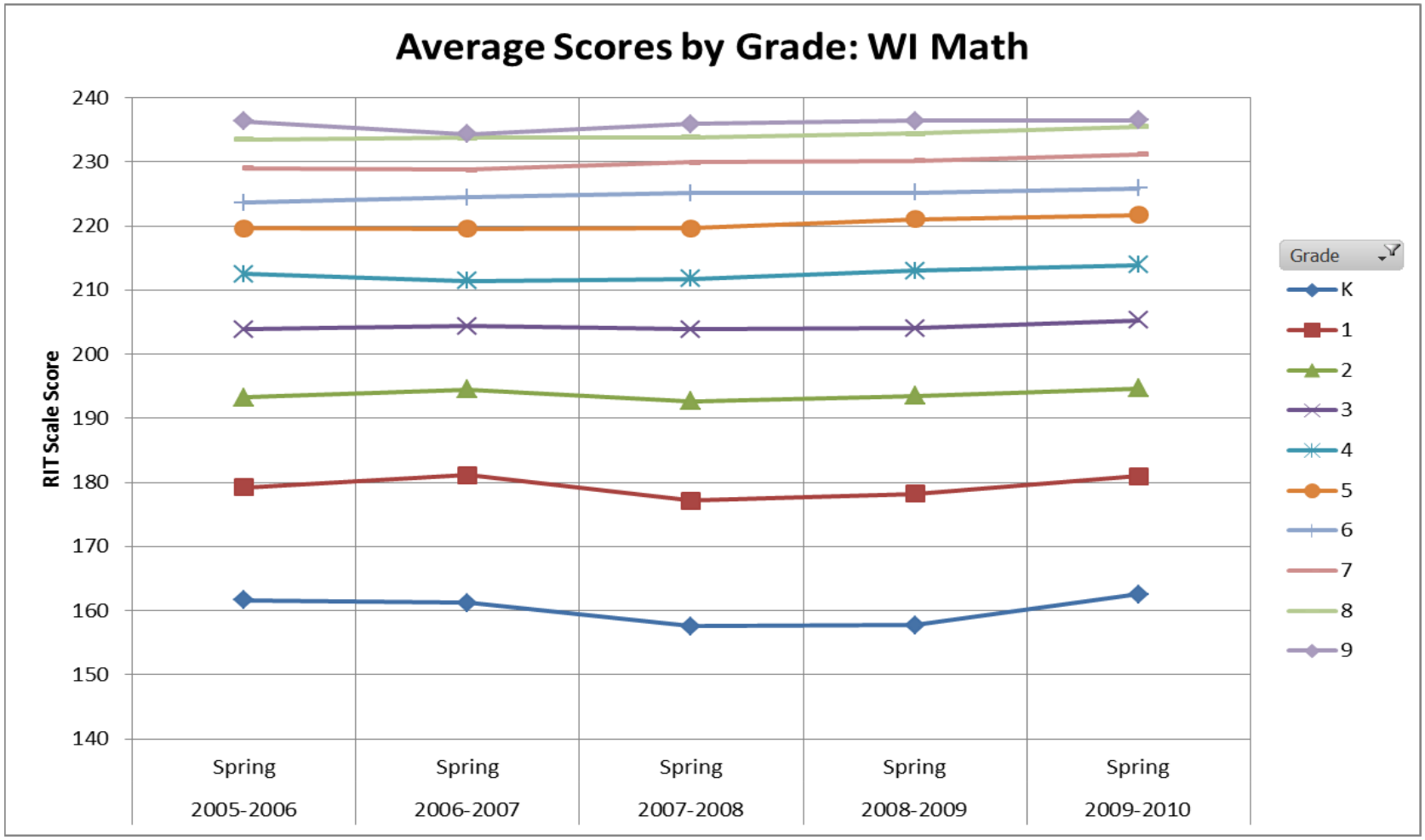

Figure 3: Horizontal alignment of Math MAP Test in Wisconsin

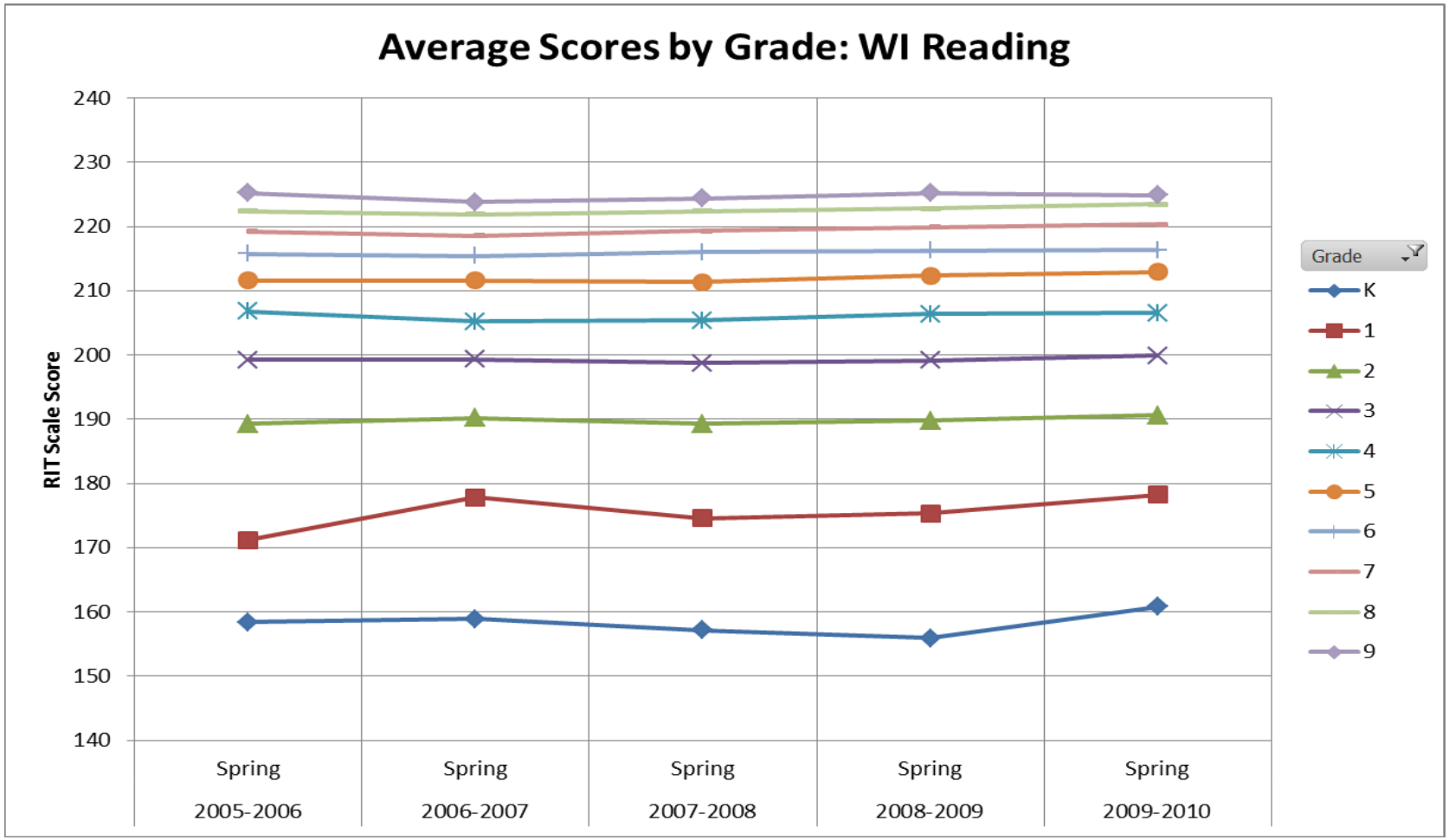

Figure 4: Horizontal alignment of Reading MAP Test in Wisconsin

Challenges for using the MAP in the SAGE evaluation

There are also some serious challenges to using the MAP for SAGE evaluation.

1 One challenge to using the MAP for the SAGE evaluation is that fewer schools have implemented the MAP in the grades the SAGE program directly affects (K to 3rd grade). The smaller samples both result in lower statistical power for measuring a SAGE effect and limit the generalizability of the results to the entire state of Wisconsin. It may be that schools using the MAP in early grades are qualitatively different that those only using it in later grades. However, it is important to note that there is a clear trend toward more schools and districts using the MAP in early grades and thus, over time, lessoning the importance of this challenge.

2 Another challenge is that the MAP is relatively new to Wisconsin. Since the MAP hasn't been used in Wisconsin for many years, the evaluation of SAGE cannot use it to longitudinally determine if student and school achievement growth has changed as a result of the SAGE program.

3 Another challenge is that MAP scores are not easily linked with state data systems, which limits the information available on students. Thus, statistical modeling procedures are less able to isolate the impact of SAGE. However, VARC is currently working with NWEA to address this deficit and improve the array of controls available to use in analyses.

Wisconsin MAP participation – SAGE and Non-SAGE Schools

Based on the low early grade MAP participation numbers in past years, only MAP data from the most recent complete school year (2009-2010) were used to conduct preliminary analyses of the SAGE effect. Two additional selection criteria were used to select the study sample. First, only students who had both fall and spring scores were included. Without both tests, it is not possible to measure or model achievement growth for a specific student. Second, only schools with 50% or greater MAP take

up rates in a particular grade were considered to be using the MAP for benchmarking achievement purposes. To clarify, it is possible that schools might use the test to track the progress of a select group of students being provided an intervention. In this case, MAP scores would not be representative of the entire grade in the school. Based on these criteria, the evaluation has MAP data for a total of 33,536 students across 280 schools. Of these, 8873 students and 82 schools received SAGE funding.

Characteristics of SAGE Students and Schools

Some limited student data (test scores, race, gender, grade level, school, district) were provided as part of the MAP file by NWEA. These were used to explore the characteristics of students taking the MAP in SAGE and non-SAGE schools and to explore the possibility of using the MAP to test the SAGE effect for specific subgroups of students. Table 3 presents a breakdown of early grade student demographics (K to 4th) across SAGE and non-SAGE schools. From this table, it is clear how the MAP was not utilized nearly as frequently in Kindergarten as in first grade, which in turn was not utilized as often as it was in later grades. Although, roughly equal numbers of male and female students took the MAP, more students in SAGE schools taking the MAP were Hispanic.

Table 3: Demographic breakdown of Wisconsin students taking the 2009-2010 MAP

	Not SAGE Students	SAGE Students	Total
Grade			
K	1515	481	1996
1st	4503	1795	6298
2nd	8456	3056	11512
3rd	10229	3541	13770
4th	10663	3150	13813
Gender			
Female	17375	5891	23266
Male	17991	6132	24123
Race/ethnicity			
African American	2646	911	3557
Asian/Pacific Islander	1120	297	1417
Hispanic	3049	1667	4716
Other	702	542	1244
White	27849	8606	36455
Total	35366	12023	47389

Table 4 and 5 compares the fall and spring reading MAP RIT score descriptive statistics between SAGE and non-SAGE students broken down by grade level. Fourth grade scores were included for reference purposes. SAGE students on average started the 2009-2010 school year behind students in non-SAGE schools. Upon entering school in the fall, kindergarten SAGE students scored 3.2 RIT points (.3 standard deviations) lower than students not in SAGE schools (Table 4). The

difference between SAGE and non-SAGE students in later grades was somewhat less, averaging between .18 and .22 standard deviations. Student fall reading scores show that, at the end of the school year, SAGE kindergarten students scored 1.7 RIT points (.14 standard deviations) higher than non-SAGE students (Table 5). SAGE students in later grades continued to score lower than non-SAGE students.

Table 4: Fall MAP reading descriptive statistics

	Fall 2009 Reading MAP RIT Scores					
	Not SAGE Students			SAGE Students		
Grade	Mean	SD	N	Mean	SD	N
K	146.1	10.1	1515	142.9	9.8	481
1st	161.9	12.1	4503	159.8	11.5	1795
2nd	176.9	15.7	8456	174.1	15	3056
3rd	190.1	15	10229	187.4	15.1	3541
4th	199.8	14.3	10663	196.2	14.7	3150

Table 5: Spring MAP reading descriptive statistics

	Spring 2010 Reading MAP RIT Scores					
	Not SAGE Students			SAGE Students		
Grade	Mean	SD	N	Mean	SD	N
K	160.7	11.8	1515	162.4	11.9	481
1st	178.5	13.4	4503	177.7	12.9	1795
2nd	191.4	13.6	8456	189.3	13.6	3056
3rd	200.7	13.2	10229	198.2	13	3541
4th	207.5	13	10663	203.8	13.4	3150

Table 6 and 7 compares the fall and spring math MAP RIT score descriptive statistics between SAGE and non-SAGE students. Again, SAGE students on average started the 2009-2010 school year behind students in non-SAGE schools. Upon entering school in the fall, kindergarten SAGE students scored 3.5 RIT points (.32 standard deviations) lower than students not in SAGE schools (Table 6). The difference between SAGE and non-SAGE students in later grades was somewhat less, averaging between .11 and .18 standard deviations. Student fall math scores show that, at the end of the school year, SAGE kindergarten students scored 1.1 RIT points (.08 standard deviations) higher than non-SAGE students (Table 7). SAGE students in later grades continued to score lower than non-SAGE students.

Table 6: Fall MAP math descriptive statistics

	Fall 2009 Math MAP RIT Scores					
	Not SAGE Students			SAGE Students		
Grade	Mean	SD	N	Mean	SD	N
K	148.5	10.9	1574	145	11.2	633
1st	162.7	13.6	4618	161.2	13.7	1957
2nd	181.3	12.5	8679	179.2	12.3	3231
3rd	194.6	12.2	10414	192.4	12.1	3608
4th	205	12.4	9972	202.3	12.4	3236

Table 7: Spring MAP math descriptive statistics

	Spring 2010 Math MAP RIT Scores					
	Not SAGE Students			SAGE Students		
Grade	Mean	SD	N	Mean	SD	N
K	162.2	13.7	1574	163.3	13.8	633
1st	181.4	13.3	4618	180.3	13.2	1957
2nd	195.1	12.1	8679	193.7	12	3231
3rd	205.8	12.2	10414	203.8	11.6	3608
4th	214.6	13.4	9972	211.5	13	3236

Characteristics of schools were obtained from the National Center of Education Statistics (NCES). These included % free/reduced price lunch, %minority, %male/female, and school community types (rural/urban/town/suburban). Characteristics of SAGE and non-SAGE schools using the MAP in Wisconsin are presented in Table 8. School-level characteristics presented here reflect characteristics of the whole school, not just grades benefitting from small class sizes due to the SAGE program. Several differences between SAGE and non-SAGE schools are noteworthy. First,

SAGE schools have a much higher free/reduced lunch population (51%) than non-SAGE schools (28%). This makes sense considering that percentage of school free/reduced lunch status is a qualifying characteristic for the SAGE program. It is also noteworthy that SAGE schools were less white (75%) than non-SAGE schools (81%). Finally, SAGE schools using the MAP were located in different community types than non-SAGE schools. Nearly half of all SAGE schools (43%) were located in rural areas as compared to 28% of non-SAGE schools. More SAGE schools were located in towns (24%) than non-SAGE schools (17%). Fewer SAGE schools were located in suburban (16%) areas and urban (19%) areas than non- SAGE schools, where 31% and 25% were located respectively. This distribution will in for the 2010-2011 school year, since the entirety of the Milwaukee Public Schools began using the MAP this year.

Table 8: Characteristics of Wisconsin Schools Using the MAP

	Not SAGE Schools		SAGE schools	
	Mean	N	Mean	N
Percent Male	51.3%	198	52.0%	82
Percent White	80.9%	198	74.9%	82
Percent Free/Reduced Lunch	28.0%	198	50.7%	82
Percent Rural	27.3%	54	43.2%	35
Percent Town	16.7%	33	23.5%	19
Percent Suburban	31.3%	62	16.0%	13
Percent Urban	24.7%	49	18.5%	15

Estimating the SAGE effect

Methods

The evaluation used hierarchical linear modeling to disentangle the impact of SAGE on spring MAP scores from the other differences in SAGE and non-SAGE schools likely to result in different patterns of MAP growth. In each model, the following variables were used as controls:

Student variables:

- Race (white/non-white)
- Gender
- Fall MAP scores

School variables:

- SAGE/Non-SAGE
- Percent free/reduced lunch
- Percent male
- Percent minority
- Location (Urban, Rural, Suburban, Town)

Different models were estimated for each grade and subject, all following the basic model presented below with error terms for both school (u_{0j}) and student (r_{ij}). Interaction terms of SAGE participation with community types (urban, suburban, etc.) were are explored in the models but due to inadequate power were not able to be reliably estimated and are therefore not included in this report.

$$Spring\ MAP_{ij} = \gamma_{00} + \gamma_{01}{*}SAGE_j + \gamma_{02}{*}MALE\%_j + \gamma_{03}{*}WHITE\%_j + \gamma_{04}F/R\ LUNCH\%_j + \gamma_{05}{*}URBAN_j + \gamma_{06}{*}SUBURBAN_j + \gamma_{07}{*}TOWN_j + \gamma_{10}{*}FALL\ MAP_{ij} + \gamma_{20}{*}WHITE_{ij} + \gamma_{30}{*}MALE_{ij} + u_{0j} + r_{ij}$$

Math Results

Table 9 presents the results of grade specific models estimating the difference in fall MAP scores between SAGE and non-SAGE schools. The SAGE effect reflects the difference in spring MAP RIT scores between SAGE and non-SAGE schools. Although there were no statistically significant findings, the overall pattern of results suggest a potential positive SAGE effect. First, even though none of the SAGE effects were statistically significant ($p > .05$), all the SAGE effects were in the positive direction. Also, the SAGE effect magnitude in kindergarten was considerably higher than in later grades. The four RIT point effect is of a considerable magnitude, with a student effect size of .3 and a school value-added effect size of .7. Finally, the finding that the 4th and 5th grade SAGE effects are considerably lower than the K, 1st, or 2nd grade effects suggests that the SAGE effects are not likely the result of endogenous, unmeasured school differences.

Table 9: Estimation of the SAGE effect on MAP math scores

Grade	Coef (SAGE Effect)	standard error	t-ratio	d.f.	p-value	schools	students	Student Effect Size	School VA Effect Size
K	4	2.3	1.768	35	0.086	43	2207	0.32	0.73
1st	0.8	0.6	1.193	115	0.235	123	6575	0.06	0.15
2nd	1.1	0.6	1.744	201	0.083	210	11910	0.1	0.22
3rd	0.5	0.4	1.366	235	0.173	244	14022	0.05	0.11
4th	-0.3	0.4	-0.821	224	0.412	232	13208		
5th	0.3	0.4	0.696	205	0.487	213	10809		

Given that the larger SAGE effects in the early grades are based on a much smaller sample of schools and students, it is not clear if the Kindergarten, 1st, and 2nd grade SAGE effects are generalizable to the total population of SAGE schools. To partially address the possibility that the sample of schools using the MAP in early grades are different than those not using the MAP in early grades, additional models were developed that maintained the sample of early grade students and schools to test the SAGE effect in later grades. If the pattern of results in these analyses are consistent with those found in the overall analyses, it provides a limited degree of confidence that the early grade effects were not driven by differences between SAGE schools utilizing the early grade MAP and those not. If the results show the magnitude of the SAGE effects remains elevated in latter grades, it is likely that the early grade effect is not representative of the total population of SAGE schools.

Table 10 presents the results of these analyses. Keeping the kindergarten sample of schools constants across the other grade analyses resulted in a similar trend of SAGE effects than the overall analysis. The same pattern was found keeping the 1st and 2nd grade school samples constant as well. Taken together, these results support the idea that early grade SAGE effects are generalizable to the entire sample of SAGE schools included. However, the evidence is far from conclusive and will require a larger sample of schools utilizing the early grade MAP and additional years of data to adequately test the robustness of the effects found in these analyses.

Table 10: Math results robustness checks

		coefficient	standard error	t-ratio	d.f.	p-value	schools	students
Kinder sample								
	K	4	2.3	1.768	35	0.086	43	2207
	1st	0.1	1.4	0.077	35	0.939	43	2173
	2nd	1.5	1.5	1.009	35	0.32	43	2275
	3rd	1.2	1.1	1.032	32	0.31	40	2178
	4th	0.2	1.3	0.17	30	0.866	38	1993
	5th	-0.1	1.5	-0.072	29	0.943	37	1900
1st grade sample								
	1st	0.8	0.6	1.193	115	0.235	123	6575
	2nd	1	0.8	1.249	113	0.214	121	6619
	3rd	0.5	0.5	1.056	99	0.294	107	5714
	4th	-0.4	0.6	-0.606	93	0.546	101	5424
	5th	-0.3	0.6	-0.564	93	0.574	101	4863
2nd grade sample								
	2nd	1.1	0.6	1.744	201	0.083	210	11910
	3rd	0.6	0.4	1.362	179	0.175	187	10512
	4th	-0.3	0.5	-0.64	164	0.523	172	9563
	5th	0.3	0.4	0.726	141	0.469	149	7338

Reading Results

Table 11 presents the results of grade specific models estimating the difference in fall MAP reading scores between SAGE and non-SAGE schools. As was the case in math, there were no statistically significant findings ($p > .05$). However, the overall pattern of results again suggests a potential positive SAGE effect. Again, all the SAGE effects were in the positive direction and their magnitudes in early grades were considerably higher than in later grades.

Table 11: SAGE reading results

Grade	coefficient	standard error	t-ratio	d.f.	p-value	schools	students	Student Effect Size	School VA Effect Size
K	2.2	1.5	1.443	29	0.16	37	1929	0.2	0.51
1st	1.4	0.7	1.978	107	0.051	115	6204	0.11	0.28
2nd	0.6	0.6	0.985	190	0.326	198	11324	0.05	0.14
3rd	0.3	0.4	0.676	225	0.499	233	13536	0.02	0.06
4th	-0.1	0.4	-0.297	210	0.767	218	12567		
5th	0.6	0.4	1.583	202	0.115	210	10196		

As was the case in the math analyses, the larger SAGE effects in the early grades are based on a much smaller sample of schools that may or may not be representative of the entire sample of SAGE schools using the MAP. To partially address the possibility that the sample of schools using the MAP in early grades are different than those not using the MAP in early grades, additional models were tested that maintained the sample of early grade students and schools to test the SAGE effect in later grades. Table 12 presents the results of these analyses. Keeping the kindergarten sample of schools constants across the other grades resulted in a similar trend of SAGE effects than the overall analysis, although the results do show somewhat inflated 1st and 2nd grade SAGE effects. The results maintaining the 1st

grade sample constant more closely represent the pattern of results including the entire sample. Taken together, these results again seem to support the idea that early grade SAGE reading effects are generalizable to the entire sample of SAGE schools included. Again, the evidence is far from conclusive and will require a larger sample of schools utilizing the MAP in early grades and additional years of data to be adequately tested.

Table 12: Reading results robustness checks

		coefficient	standard error	t-ratio	d.f.	p-value	schools	students
Kinder sample								
	K	2.2	1.5	1.443	29	0.16	37	1929
	1st	2.1	1.3	1.711	29	0.098	37	1913
	2nd	1.5	1.7	0.864	29	0.395	37	1968
	3rd	0.9	0.9	0.994	26	0.33	34	1894
	4th	-0.4	1.2	-0.293	24	0.772	32	1685
	5th	0.5	0.8	0.664	23	0.514	31	1605
1st grade sample								
	1st	1.4	0.7	1.978	107	0.051	115	6204
	2nd	0.2	0.9	0.217	104	0.829	112	6114
	3rd	0.7	0.5	1.502	90	0.137	98	5287
	4th	0.5	0.6	0.831	84	0.408	92	4945
	5th	0.2	0.5	0.456	88	0.65	96	4576

Summary Conclusions

Utilizing the MAP for the SAGE evaluation will provide for the opportunity to test the SAGE effect across the entire state of Wisconsin. Several characteristics of the MAP make it especially appealing for this use. First, the MAP is a validated instrument for use in the early grades targeted by SAGE. The MAP is typically administered at least twice, at the beginning and end of the school year, thus providing an annual measure of academic growth. A growing number of schools and districts across the entire state use the MAP as an early grade benchmark test. The biggest challenge in using the MAP for the SAGE evaluation is that relatively few schools use the MAP in Kindergarten and first grade. But as more schools adopt the MAP, it will be possible to conduct more formative analyses of SAGE, testing the impact of SAGE across diverse student groups and school settings.

An analysis of student and school characteristics of SAGE and non-SAGE schools using the MAP during the 2009-2010 school year shows that SAGE schools, compared to non-SAGE schools, have a higher free/reduced lunch population, have more Hispanic students, are more rural, and are less urban and suburban. On average, students in SAGE schools demonstrated lower achievement, starting kindergarten .3 standard deviations behind students in non-SAGE schools in both reading and math. The difference was somewhat reduced in later grades. Interestingly, by the end of Kindergarten, students in SAGE schools scored higher on both the reading and math MAP than did students in non-SAGE schools. However, in first, second, and third grade, students in SAGE schools remained behind students in non-SAGE schools.

The results of statistical modeling suggest that SAGE may be having a significant impact on both math and reading achievement. However, low statistical power, especially in early grades, prevent more conclusive statements about the impact of SAGE. Still, the results are promising and hold up to robustness checks of the generalizability of these findings. As additional schools adopt the MAP, it will

be possible to estimate with more confidence the impact that SAGE is having on students and schools.