WEC
Supporting PreK-12
Wisconsin Evaluation Collaborative

July 2020

# Evaluation of the Achievement Gap Reduction Program

## 2015-16 through 2018-19

*for the* Wisconsin Department of Public Instruction

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Wisconsin Center for Education Research
UNIVERSITY OF WISCONSIN–MADISON

# About the Authors

## Jed Richardson

Jed Richardson is an Assistant Scientist with the Wisconsin Evaluation Collaborative. He holds a Ph.D. in Economics from the University of California, Davis. Questions about this report can be directed to jed.richardson@wisc.edu.

## Grant Sim

Grant Sim is a Researcher with the Wisconsin Evaluation Collaborative. He holds a Master's Degree in Public Affairs from the University of Wisconsin–Madison.

## About the Wisconsin Evaluation Collaborative

The Wisconsin Evaluation Collaborative (WEC) is housed at the Wisconsin Center for Education Research at the University of Wisconsin-Madison. WEC's team of evaluators supports youth-serving organizations and initiatives through culturally responsive and rigorous program evaluation. Learn more at http://www.wec.wceruw.org.

# Executive Summary

## Evaluation of the AGR Program, 2015-16 through 2018-19

The Achievement Gap Reduction (AGR) program is an initiative of the Wisconsin Department of Public Instruction (DPI), as specified by 2015 Wisconsin Acts 53 and 71. AGR aims to improve the academic performance of students in schools with high concentrations of low-income students. AGR functions as a revision and continuation of the Student Achievement Guarantee in Education (SAGE) program. Similar to SAGE, AGR spans kindergarten to third grade and provides funds to participating Wisconsin schools based on their numbers of economically disadvantaged students. To receive AGR funding, schools must implement one or more strategies in each participating grade:

- Provide professional development related to small group instruction and reduce the class size to one of the following:
    - No more than 18.
    - No more than 30 in a combined classroom having at least 2 regular classroom teachers.
- Provide data-driven instructional coaching for one or more teachers of one or more participating grades. The instruction shall be provided by licensed teachers who possess appropriate content knowledge to assist classroom teachers in improving instruction in math or reading and possess expertise in reducing the achievement gap.
- Provide data-informed, one-to-one tutoring to pupils in the class who are struggling with reading or mathematics or both subjects. Tutoring shall be provided during regular school hours by a licensed teacher using an instructional program to be found effective by the What Works Clearinghouse of the Institute of Education Sciences. [1]

This report presents the results of an evaluation completed by the Wisconsin Evaluation Collaborative (WEC) within the Wisconsin Center for Education Research at the University of Wisconsin-Madison. The goal of this year's evaluation was to examine the following questions:

1. How are AGR schools implementing the AGR program as specified by 2015 Wisconsin Acts 53 and 71?
    a. What is the breakdown of strategy usage across the state?
    b. How does implementation of these three strategies differ across schools?
2. To what extent is AGR meeting intended outcomes, including impacts on standardized test scores, attendance, and disciplinary events?
    a. How does AGR impact achievement gaps between low-income students and their higher-income peers?
    b. How does AGR impact vary by student characteristics?
    c. How does AGR's impact on outcomes compare to impacts associated with the SAGE program?
3. Are there differences between the three AGR strategies' impacts on intended outcomes?
4. How are AGR schools implementing instructional coaching and one-to-one tutoring?

Because AGR targets higher poverty schools where outcomes are typically lower and demographic profiles differ from Wisconsin averages, simple comparisons of outcomes between AGR schools and other, unfunded Wisconsin schools would provide biased results. To address this selection bias, WEC uses a two-step statistical method in order to better understand how AGR impacts student achievement, attendance, and discipline outcomes, and to compare AGR's impact to those of its predecessor, SAGE. The first step of the analysis uses propensity score matching to identify non-AGR Wisconsin schools that are similar to those receiving AGR funding. These observationally similar schools function as a comparison group

1 2015 Wisconsin Act 53. Wisconsin Senate. Section 118.44.

for the second step of the analysis, estimating the impact of AGR through multivariate regression techniques.

## How are AGR schools implementing the program?

**In 2018-19, the most recent year of data, 412 schools implemented the AGR program, serving nearly 75,000 students in kindergarten through third grades.** As previously noted, to fulfill AGR obligations schools could implement any combination of three strategies: reduced class size, instructional coaching, and/or tutoring.

- Nearly 70 percent of schools utilized multiple strategies–38.1 percent of schools implemented reduced class size and instructional coaching together and 21.8 percent of schools implemented all three.
- Single strategies were employed less frequently, although 11.9 percent of schools implemented instructional coaching alone and 18.7 percent of schools implemented reduced class size alone.
- Comparatively few schools used only tutoring as a strategy or in combination with one of the other strategies.

## To what extent is AGR meeting intended outcomes?

The impact analysis examined how AGR students performed compared to non-AGR students in similar schools while controlling for student characteristics. Results from this analysis included:

- A positive and significant impact of the AGR program on statewide reading growth in kindergarten as measured by the PALS assessment. AGR is associated with a 0.11 standard deviation increase in PALS scores relative to similar, non-AGR schools. Increased growth on PALS is associated with a 0.05 standard deviation narrowing of the statewide kindergarten achievement gap by income in 2018-19.
- No estimated impact of the AGR program on statewide reading or math growth in Grades 1-3, as measured by the MAP and STAR assessments.
- No statistically significant impact of the AGR program on statewide attendance or out-of-school suspension rates.
- The evaluation also examined the impact of the program by various subgroup populations and found:
    - Large, positive, and significant impacts of the AGR program on kindergarten reading growth for low-income students.
    - Large, positive, and significant impacts of AGR on kindergarten reading for English learners, Hispanic students, Asian students, and students in urban settings.
    - Positive and significant impacts of the AGR program on behavior, as measured through a reduction in suspensions, for English learners and Hispanic students.
    - AGR is associated with an approximately one-day per year increase in absences for urban students.

The evaluation also examined AGR program impacts compared to the previously implemented SAGE program. Results showed that AGR likely has a larger, positive impact on kindergarten reading and decreases kindergarten absences by approximately one day per year.

## Are there differences in outcomes depending on the AGR strategies schools use?

The evaluation provided preliminary evidence of associations (not causal impacts) between outcomes and the AGR strategies schools choose. Results included:

- Increased reading growth, relative to class size, associated with coaching in Grade 1 and class size and coaching combined in Grade 3.
- Increased absences associated with tutoring only, relative to class size across Grades K-3.
- Reduced suspensions associated with reduced class size relative to other strategies.
- Analyses of coaching and tutoring frequency and intensity found few associated differences in outcomes.

# Contents

Section 1

# Introduction

# Introduction

The Achievement Gap Reduction (AGR) program is an initiative of the Wisconsin Department of Public Instruction (DPI) that provides funding to improve the academic performance of students in schools with high concentrations of low-income or economically disadvantaged students. AGR functions as a revision and continuation of the Student Achievement Guarantee in Education (SAGE) program, which the Wisconsin legislature and DPI initiated in 1995 to address the need for additional resources for economically disadvantaged students, particularly in urban areas. Starting in the 1996-97 school year, the SAGE program administered state aid to schools that implemented reduced class sizes in kindergarten through third grade. A school typically qualified for the SAGE program if at least 30 percent of the student population was economically disadvantaged and its school district included one or more schools with at least 50 percent of the student population qualifying as economically disadvantaged.

In 2015, Wisconsin recognized the need to add flexibility to SAGE, reorganizing and renaming the program with the enactment of Wisconsin Acts 53 and 71. Wisconsin began a gradual phase-in of AGR in 2015-16 by transitioning schools from SAGE to AGR, with the final phase out of previous SAGE programs by the end of the 2017-18 school year. Like SAGE, AGR targets funding to schools with economically disadvantaged students through contracts to implement the program in kindergarten through third grade. Each year, the state provides approximately $110,000,000 to be distributed to participating schools. In order to receive funding under AGR contracts, schools must implement at least one of three prescribed strategies in each participating grade. Each school, and each grade within a school, may implement different strategies. The three strategies include:

- Provide professional development related to small group instruction and reduce the class size to one of the following:
  - No more than 18.
  - No more than 30 in a combined classroom having at least 2 regular classroom teachers.
- Provide data-driven instructional coaching for one or more teachers of one or more participating grades. The instruction shall be provided by licensed teachers who possess appropriate content knowledge to assist classroom teachers in improving instruction in math or reading and possess expertise in reducing the achievement gap.
- Provide data-informed, one-to-one tutoring to pupils in the class who are struggling with reading or mathematics or both subjects. Tutoring shall be provided during regular school hours by a licensed teacher using an instructional program to be found effective by the What Works Clearinghouse of the Institute of Education Sciences. [2]

The AGR program prioritizes achievement gap reduction for economically disadvantaged students. As noted in the 2019 report, National Assessment of Educational Progress (NAEP) assessment scores show a persistent gap in performance between economically disadvantaged students and non-economically disadvantaged students in Wisconsin.[3] Neither Wisconsin nor the nation made substantial progress in reducing this achievement gap from 2003 through 2017.

## This Evaluation

2015 Wisconsin Acts 53 and 71 include a provision for an annual evaluation of the AGR program starting in the 2018-19 school year. DPI contracted with the Wisconsin Evaluation Collaborative (WEC) within the Wisconsin Center for Education Research at the University of Wisconsin–Madison for these evaluation services. This report provides results from the evaluation of the AGR program from 2015-16 through 2018-19.

To serve as a foundation for the evaluation, WEC worked in collaboration with DPI to develop the following overarching evaluation questions:

1. How are AGR schools implementing the AGR program as specified by 2015 Wisconsin Acts 53 and 71?

2 2015 Wisconsin Act 53. Wisconsin Senate. Section 118.44.
3 Richardson, J., Sim, G., & Chapa, B. (2019). *Evaluation of Wisconsin's Achievement Gap Reduction Program*. Wisconsin Evaluation Collaborative. https://dpi.wi.gov/sites/default/files/imce/title-i/doc/Final_AGR_Evaluation_2019_Report.pdf

   a. What is the breakdown of implementation look like across the state with regard to the three strategies?
   b. How does implementation of these three strategies differ across schools?
2. To what extent is AGR meeting intended outcomes, including impacts on standardized test scores, attendance, and disciplinary events?
   a. How does AGR impact achievement gaps between low-income students and their higher-income peers?
   b. How does AGR impact vary by student characteristics?
   c. How does AGR's impact on outcomes compare to impacts associated with the SAGE program?
3. Are there differences between the three AGR strategies' impacts on intended outcomes?
4. How are AGR schools implementing instructional coaching and one-to-one tutoring?
   a. How much time are schools committing to coaching and tutoring?
   b. What are coaches' and tutors' qualifications?
   c. How do impacts differ between schools that commit substantial time to coaching/tutoring versus schools that commit less time?

This report has nine main sections including the introduction. The evaluation data and methodology section includes details on data, analysis designs, and statistical models used to evaluate program impacts, as well as the limitations of this evaluation. The AGR demographics section contains information on the characteristics of AGR students and schools compared to the state overall to provide context for later findings. This section also contains testing patterns and growth analysis samples which describes coverage of common assessments in Grades K-3 and the samples chosen for estimating AGR Impacts on growth. The AGR implementation section contains information on the AGR strategies used by schools in 2018-19. The AGR impacts section provides the results of analyses of AGR impact on math growth, reading growth, attendance, and discipline. This section is further divided to provide overall impacts, impacts of AGR compared to SAGE, impacts by student subgroups, and differences in outcomes by AGR strategy. A section on statewide achievement gaps explores AGR impacts on Wisconsin gaps in math and reading in Grades K-3. The section on school board report findings includes results from an examination of 2018-19 reports by AGR districts and schools. The End-of-Year Report findings provide the results from the 2018-19 survey of AGR schools. The final section of the report includes a summary of findings and thoughts on future evaluations. This report also contains two appendices, a technical appendix which provides further details on statistical methodology, and an appendix including the instrument for the 2018-19 End-of-Year Report survey.

Section 2

# Evaluation Data & Methodology

# Evaluation Data and Methodology

In order to understand how AGR impacts student achievement, attendance, and discipline outcomes, and to compare AGR's impacts to those of its predecessor, SAGE, we must identify a plausible comparison group of schools and students. Because AGR targets higher poverty schools where outcomes are lower on average, naïve comparisons of AGR schools' outcomes to those of other Wisconsin schools would show biased, negative program impacts. To address this selection bias, the evaluation uses Propensity Score Matching (PSM) to identify non-AGR-funded, Wisconsin schools that are similar to those receiving AGR funding. These observationally similar schools act as a comparison group for analyses of AGR impacts.

The analysis includes students in Grades K-3 at all schools that received SAGE and AGR funding during the 2012-13 through 2018-19 academic years. In addition, for purposes of comparison, the evaluation includes K-3 students at subsets of non-AGR, non-SAGE schools.

## Data

In order to identify plausibly equivalent, non-AGR schools for a comparison group and to estimate impacts, the evaluation combines several sources of student- and school-level data for the academic years 2012-13 through 2018-19. Student-level achievement test data, student demographics, and enrollment records came from DPI administrative data. DPI also provided school-level data on AGR and SAGE funding by year. School-level teacher average salaries were sourced from DPI Public Staff Reports, and school location information came from school report card files.[4]

- *Demographic characteristics* include gender, race/ethnicity, English learner status, special education status, and low-income or economic status as measured by free or reduced-price lunch eligibility. School- and grade-level measures of demographic characteristics were calculated from student-level data.
- *Achievement test data* include fall and spring administrations of the Phonological Awareness Literacy Screening (PALS), MAP, and STAR. For Grades 1-3, MAP and STAR scores were equated and combined into a single test measure in order to attaina sufficient student sample.
- *Attendance data* consist of total days absent and total possible attendance days. The associated outcome variable is the absence rate or the total days absent divided by the total possible attendance days.
- *Discipline data* consist of the number of out-of-school suspensions. The associated outcome variable, the suspension rate, is an indicator that is one for students with at least one out-of-school suspension during the school year and zero for those who had not been suspended. We use this discipline outcome as a proxy for student behavior throughout the evaluation.
- *Enrollment data* include school attended and grade.
- *School-level data* include SAGE and AGR funding by year, teacher average salaries, and school location (city, suburb, town, rural).

## Identifying Comparison Schools

Using the data described above, we aggregated each school's K-3 data to find a comparison group of non-AGR schools. Matching followed two separate strategies. For attendance and discipline outcomes, we matched schools based on 2012-13 data. For math and reading testing outcomes, however, wide variation in schools' testing coverage both across time and across grades prevented matching at the school level (see Table 5 - Table 7). Instead, we chose to match at the school-grade-year level using each school's fall data.

4 Public Staff Reports are available at https://publicstaffreports.dpi.wi.gov/PubStaffReport/Public/PublicReport. School report cards can be found at https://apps2.dpi.wi.gov/reportcards/.

During the 2019 evaluation, we tested multiple variations of PSM in order to, (1) achieve the best match between AGR and comparison schools, and (2) retain as many AGR observations as possible. To do so, we tested combinations of demographic and academic variables and several matching algorithms. This testing process resulted in a kernel matching procedure, which we continue to use during the 2020 evaluation. Kernels place higher weights on untreated observations nearest to a treatment observation and assign successively lower weights to untreated observations as their distance from a treatment observation increases. Table 1 lists the covariates in the matching model that provide the best balance and sample retention.

# Table 1

## Propensity Score Matching Controls by Analysis Type

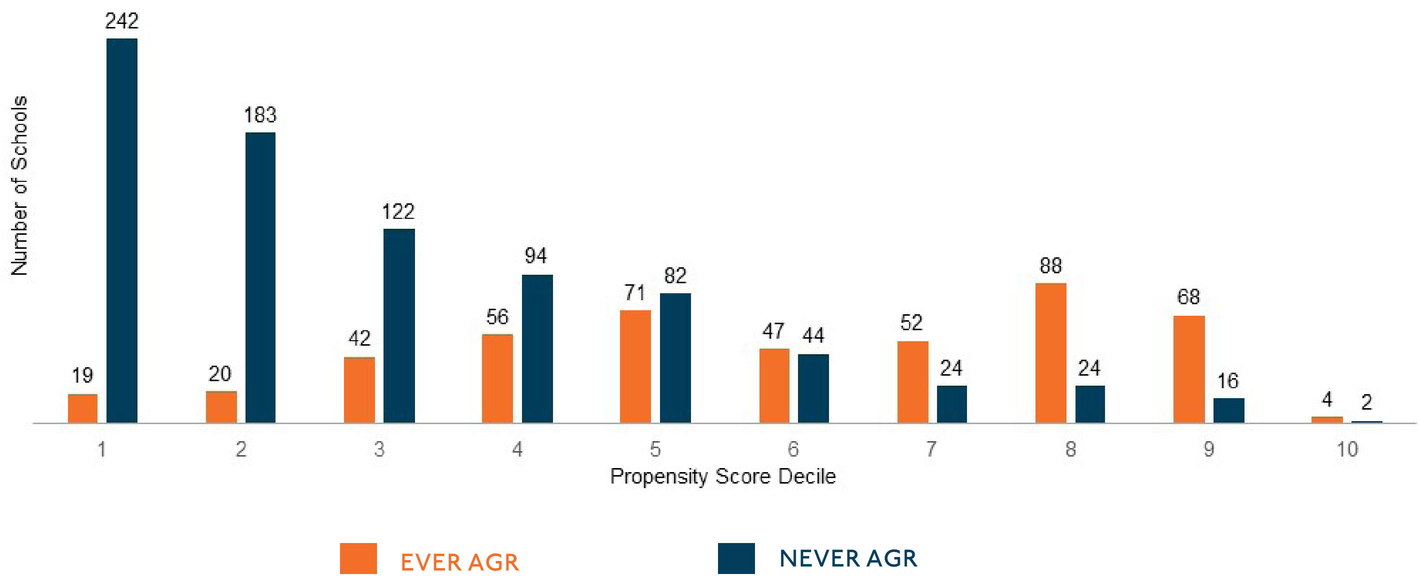
| CONTROL VARIABLE | GROWTH ANALYSIS | ATTENDANCE/ DISCIPLINE ANALYSIS |
|---|---|---|
| Student Population | ✓ | ✓ |
| % Black, Hispanic, White, Other Race/Ethnicity* | ✓ | ✓ |
| % Free/Reduced-price Lunch | ✓ | ✓ |
| % English Learner | ✓ | ✓ |
| % Special Education | ✓ | ✓ |
| Average Teacher Salary | ✓ | ✓ |
| Local Description (City, Suburb, Town, Rural)* | ✓ | ✓ |
| Average Standardized Fall Math Score** | ✓ | |
| Average Standardized Fall Reading Score | ✓ | |
| Grade Indicators*** | | ✓ |
| Attendance Rate in 2012-13 | | ✓ |
| Suspension Rate in 2012-13 | | ✓ |

Note: * Due to collinearity, we omitted one Race/Ethnicity category and one Local Description category from the model ** For PALS, only the PALS reading pretest is included, due to low participation in the MAP/STAR math exam in kindergarten. *** Indicators equal one if schools include that grade.

When matching is successful, there should be sufficient overlap in propensity scores of treated (AGR) and untreated (non-AGR) schools to ensure that there is a plausible comparison group for analysis. Figure 1 below shows the overlap between AGR and non-AGR schools for MAP/STAR math in 2018-19. In each decile of the propensity score distribution, there is at least one comparison (untreated) school. Most deciles have more than 10 comparison schools, showing sufficient overlap for the analysis. Overlap is similar across all models.

# Figure 1

## Common Support for Matching - MAP/STAR Math (2018-19)



After matching, we estimate AGR impacts via multivariate regression models. These models include all school-level matching covariates listed in Table 1: Propensity Score Matching Controls by Analysis Type above, as well as student-level demographic variables, student-level pretest scores, and grade-by-year fixed effects. A full listing of analysis variables can be found in Table 2 below. All models include weights generated by the kernel PSM procedure.

## Table 2

### Analysis Model Controls

| CONTROL VARIABLE | GROWTH | ATTENDANCE | DISCIPLINE |
|---|---|---|---|
| **Student Demographics** Gender, Race/Ethnicity*, Free/Reduced-price Lunch, English Learner, Special Education | ✓ | ✓ | ✓ |
| **School Demographic Percentages** Gender, Black, Hispanic, White, Other Race/Ethnicity*, Free/Reduced-price Lunch, English Learner, Special Education | ✓ | ✓ | ✓ |
| **School Population** | ✓ | ✓ | ✓ |
| **Local Description (City, Suburb, Town, Rural)*** | ✓ | ✓ | ✓ |
| **Student Fall Test Scores**** | ✓ | | |
| **Student Fall Test Scores Squared***** | ✓ | | |
| **School Average Fall Test Scores** | ✓ | | |
| **School Attendance Rate in 2012-13** | | ✓ | |
| **School Suspension Rate in 2012-13** | | | ✓ |

Note: * Due to collinearity, we omitted one Race/Ethnicity category and one Local Description category from the model. ** For math and reading models, both subject pretests are included. For PALS, only the PALS reading pretest is included, due to low participation in the MAP/STAR math exam in kindergarten. *** PALS only.

## Limitations

The methodology outlined above provides the most rigorous possible evaluation given the rollout of AGR and available data. There are several limitations, however, that could impact this report's results and conclusions.

The most significant limitation stems from PSM's primary assumption that schools matched on observable characteristics such as test scores and demographics are also matched on unobserved characteristics, such as schools' ability to properly implement AGR strategies or instructor quality in the local hiring market. If unobserved characteristics are not balanced between AGR and comparison schools and are related to both outcomes and AGR participation, estimates of AGR impacts will be biased.

The second limitation occurs because all AGR schools previously participated in SAGE, which had been in operation for over 15 years at the beginning of this study's sample period. As a consequence, AGR schools are matched to non-AGR schools based on post-SAGE outcomes. Matching schools on post-program data risks biasing the results toward zero (toward estimating smaller impacts), because schools would be matched on previous-period outcomes that already include the treatment impact (in this case, the SAGE program is similar enough to AGR to raise similar concerns). Omitting these outcomes from the matching model, however, resulted in poor matches and would have caused significant bias.

To the extent that the first two limitations bias impact estimates, the results should not be considered causal. In particular, if AGR schools are systematically more (less) effective than schools in the matched comparison group, impact estimates will be biased upward (downward).

The final limitation occurs due to inconsistent testing patterns (described in detail in Testing Patterns and Growth Analysis Samples below). In general, during the sample period Wisconsin did not require schools to use specific assessments in Grades K-3, which creates difficulties for identifying a consistent, sufficiently sized sample for estimating growth impacts. Although the tested population of AGR schools used for the evaluation is observationally similar to the untested sample of AGR schools (see Figure 5 -Figure 7), available data cannot support analysis of whether schools' choices of tests are related to outcomes and participation in AGR.

Section 3

# AGR Demographics

# AGR Demographics

This section and the sections that follow present the evaluation results aligned to the evaluation questions listed above. We begin with information on the characteristics of AGR students and schools. Table 3 shows the number of AGR schools for each of the first four years of the program. The first AGR cohort started in 2015-16 with 96 schools, followed by the second cohort in 2016-17, which brought the total to 408 schools. The final cohorts added a small number of schools in 2017-18 and 2018-19.

## Table 3

### Number of AGR Schools by Grade and Year

| GRADE | 2015-16 | 2016-17 | 2017-18 | 2018-19 |
|---|---|---|---|---|
| Kindergarten | 88 | 393 | 392 | 394 |
| First | 91 | 398 | 398 | 402 |
| Second | 91 | 398 | 399 | 402 |
| Third | 88 | 391 | 392 | 394 |
| Any (K-3) | 96 | 408 | 409 | 412 |

The numbers of students in AGR schools from 2015-16 to 2018-19, overall and by grade, are presented in Table 4. The first cohort of AGR schools included approximately 18,000 students, while the addition of the second cohort in 2016-17 brought the total to over 77,000 students. Since then, student participation has plateaued around 75,000.

## Table 4

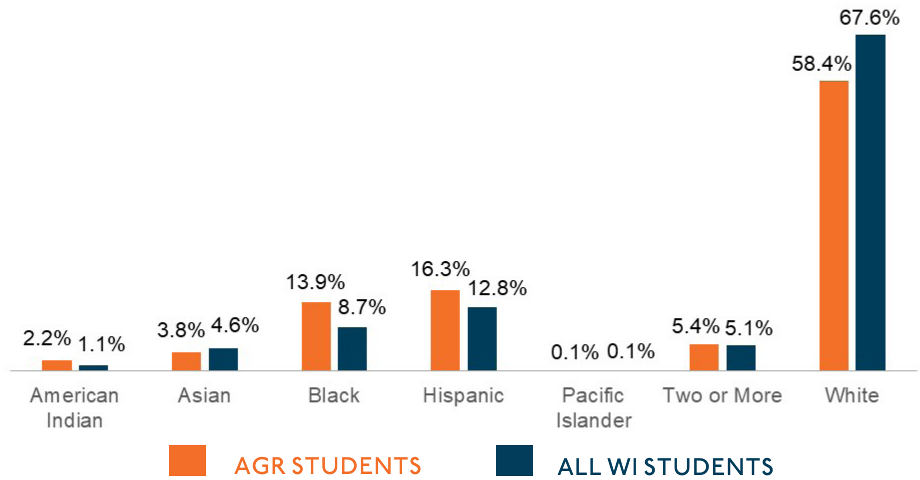
### Number of AGR Students by Grade and Year

| GRADE | 2015-16 | 2016-17 | 2017-18 | 2018-19 |
|---|---|---|---|---|
| Kindergarten | 4,139 | 18,384 | 18,797 | 18,340 |
| First | 4,571 | 19,288 | 18,855 | 18,741 |
| Second | 4,682 | 20,054 | 19,200 | 18,911 |
| Third | 4,544 | 19,508 | 19,257 | 18,328 |
| Overall (K-3) | 17,936 | 77,234 | 75,586 | 74,320 |

Figure 2 and Figure 3 compare the demographic characteristics of AGR students to all K-3 Wisconsin students (including AGR students) in 2018-19. Relative to Wisconsin as a whole, a higher proportion of AGR students were black, Hispanic, English learners, and eligible for free/reduced-price lunch. Students in AGR schools were less likely white.

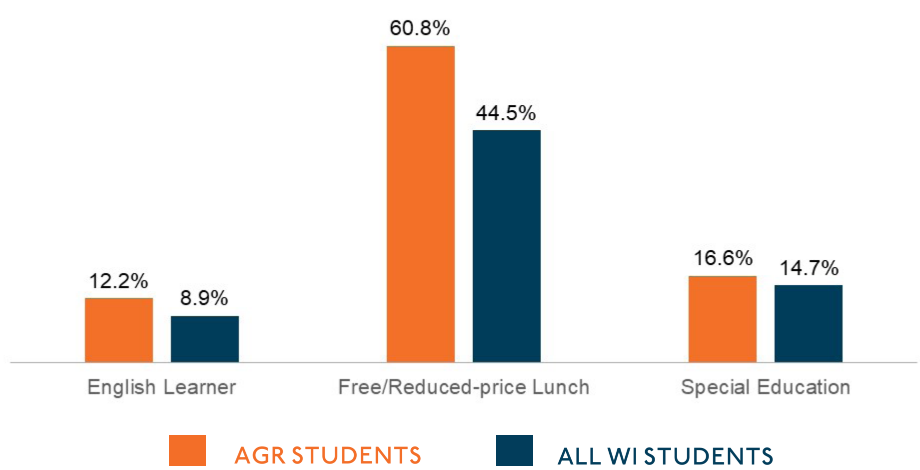
# Figure 2

## Race/Ethnicity of AGR and WI Students, 2018-19




# Figure 3

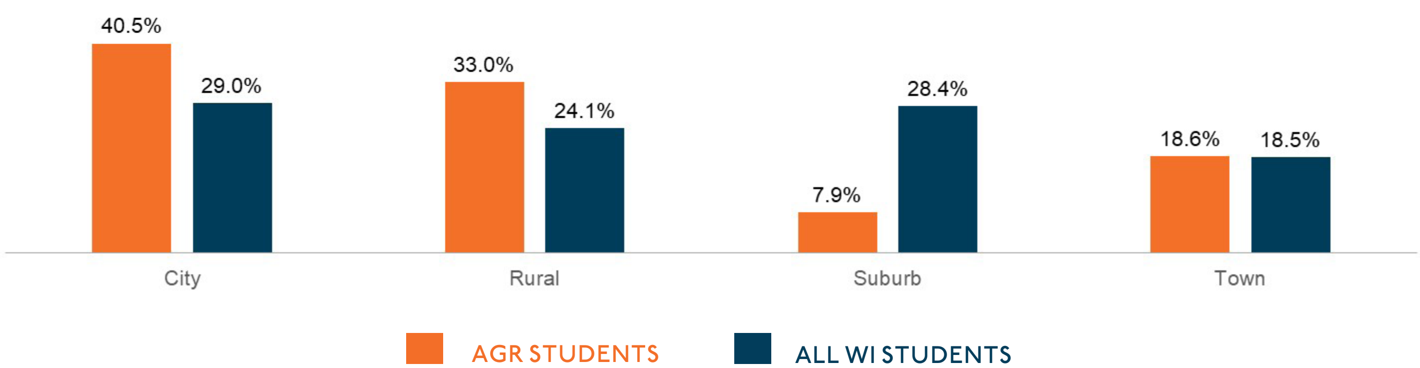
## Percentage of AGR and WI Students That Were English Learners, Eligible for Free/Reduced-price Lunch, and in Special Education, 2018-19



AGR schools were more likely to be located in urban or rural settings and less likely to be in suburban areas, as shown in Figure 4. This corresponds to the higher proportion of students eligible for free/reduced-price lunch students, seen previously, as city and rural areas of the state have larger populations with poverty.[5]

# Figure 4

## Locale Description of AGR and WI Students, 2018-19




5 See https://dpi.wi.gov/news/maps/free-reduced-lunch

## Testing Patterns and Growth Analysis Samples

Shifting testing patterns in Grades K-3 throughout the sample period complicate efforts to estimate AGR's impacts on test score growth. Under Wisconsin's current testing policy, the first common, state-mandated accountability test occurs during the spring of third grade. Although schools tested students throughout Grades K-3, schools or districts are allowed to choose their own assessments. This policy results in substantial variation in testing patterns both across and within schools. In addition to variation between schools regarding the assessments they select, many schools began administering a new test and/or quit using a test in the middle of the sample period. Other schools tested some of Grades K-3 but not others, and yet others changed which grades they tested during the sample period. As a result, less than half of the overall population of AGR schools and students would be appropriate for use  in growth analysis. Given testing patterns, we used two strategies to build sufficient samples. First, we split the growth analysis sample into two samples: Grade K and Grades 1-3. Table 5 shows that in 2018-19, less than half of all Wisconsin kindergarteners took the PALS, which had been a statemandated reading assessment for the grade from 2012-13 through 2015-16. For the purposes of this evaluation, PALS provided sufficient coverage for kindergarten reading, although no combination of assessments resulted in adequate coverage for kindergarten math.

# Table 5

Percentage of Wisconsin Schools Using PALS

| GRADE | 2015-16 | 2016-17 | 2017-18 | 2018-19 |
|---|---|---|---|---|
| Kindergarten | 96% | 54% | 47% | 40% |
| First | 96% | 48% | 40% | 33% |
| Second | 92% | 43% | 35% | 30% |

The second strategy we used to build sufficient growth analysis samples was to use both the MAP and STAR assessments for Grades 1-3 math and reading.[6] As shown in Table 6 and Table 7, in 2018-19 between 26-50 percent of Wisconsin schools used either the MAP or STAR in first, second, or third grade, with usage rates above 40 percent for second and third graders in both subjects. To combine MAP and STAR into a single measure, we equated assessment scores using national norms.[7]

## Table 6

Percentage of Wisconsin Schools Using MAP or STAR Math Tests

| GRADE | 2015-16 | 2016-17 | 2017-18 | 2018-19 |
|---|---|---|---|---|
| Kindergarten | 13% | 11% | 9% | 8% |
| First | 32% | 32% | 34% | 35% |
| Second | 44% | 43% | 43% | 45% |
| Third | 53% | 53% | 52% | 50% |

## Table 7

Percentage of Wisconsin Schools Using MAP or STAR Reading Tests

| GRADE | 2015-16 | 2016-17 | 2017-18 | 2018-19 |
|---|---|---|---|---|
| Kindergarten | 9% | 7% | 16% | 21% |
| First | 18% | 16% | 24% | 26% |
| Second | 42% | 42% | 41% | 43% |
| Third | 53% | 52% | 50% | 50% |

6 In addition to STAR Reading, the analysis also uses STAR Early Literacy results for Grade 1.

7 See Thum, Y. M. & Hauser, C.H. (2015). NWEA 2015 MAP norms for student and school achievement status and growth. NWEA Research Report. Portland, OR: NWEA. Retrieved from https://www.nwea.org/content/uploads/2018/01/2015-MAP-Norms-for-Student-and-School-Achievement-Status-and-Growth.pdf

Table 8 shows the number of schools overall and by grade that had testing information and were used in the analyses of academic growth. As a reference, the table also shows the percentage of all AGR schools in the tested population. Table 9 displays similar information but for students instead of schools. As seen in Table 8 and Table 9, testing patterns restrict the first grade sample most. In first grade, the growth analysis only includes approximately one-quarter of the entire sample of students in 2018-19. This restriction lessens as grade level increases.

# Table 8

## Number of Growth Analysis AGR Schools and Percentage of All AGR Schools by Grade and Year

| GRADE | 2015-16 | | 2016-17 | | 2017-18 | | 2018-19 | |
|---|---|---|---|---|---|---|---|---|
| | NUMBER | % | NUMBER | % | NUMBER | % | NUMBER | % |
| Kindergarten | 87 | 98.9 | 203 | 51.7 | 166 | 42.3 | 145 | 36.8 |
| First | 12 | 13.2 | 44 | 11.1 | 85 | 21.4 | 107 | 26.6 |
| Second | 28 | 30.8 | 160 | 40.2 | 156 | 39.1 | 158 | 39.3 |
| Third | 29 | 33.0 | 209 | 53.5 | 202 | 51.5 | 195 | 49.5 |
| Overall (K-3) | 93 | 96.9 | 309 | 75.7 | 296 | 72.4 | 282 | 68.4 |

# Table 9

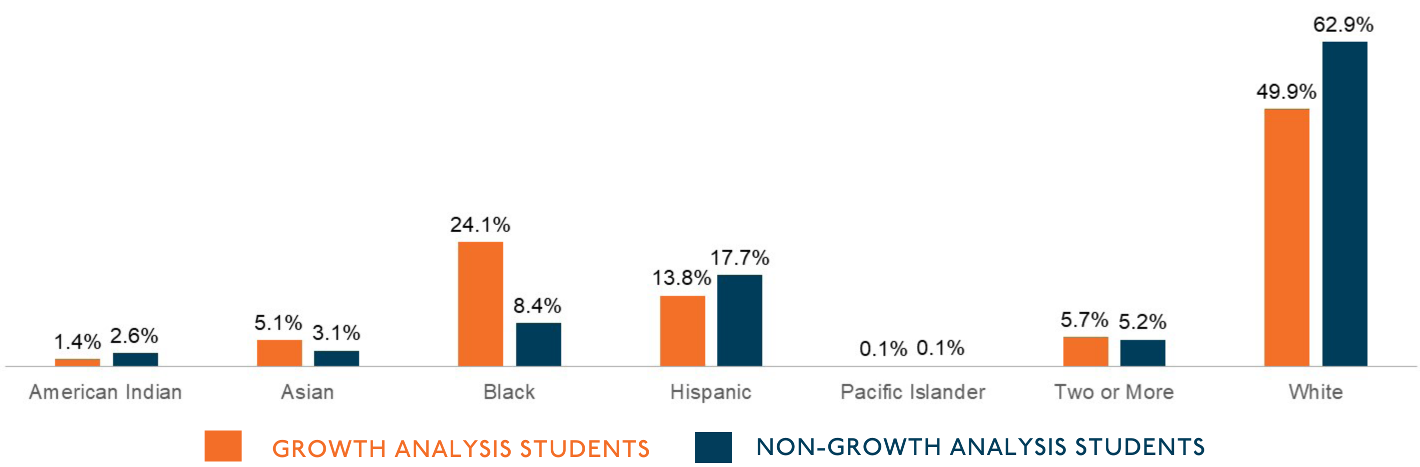
## Number of Growth Analysis AGR Students and Percentage of All AGR Schools by Grade and Year

| GRADE | 2015-16 | | 2016-17 | | 2017-18 | | 2018-19 | |
|---|---|---|---|---|---|---|---|---|
| | NUMBER | % | NUMBER | % | NUMBER | % | NUMBER | % |
| Kindergarten | 3,927 | 94.9 | 8,989 | 48.9 | 7,382 | 40.4 | 6,209 | 33.9 |
| First | 710 | 15.5 | 1,920 | 10.0 | 3,326 | 17.6 | 4,352 | 23.2 |
| Second | 1,475 | 31.5 | 7,328 | 36.5 | 6,449 | 33.6 | 6,749 | 35.7 |
| Third | 1,549 | 34.1 | 10,062 | 51.6 | 9,371 | 48.7 | 8,475 | 46.2 |
| Overall (K-3) | 7,661 | 42.7 | 28,299 | 36.6 | 26,528 | 35.1 | 25,785 | 34.7 |

Due to the growth analysis sample of students being smaller than the entire population, as mentioned in the limitations section, the growth analysis results may not apply to all AGR students. To observe whether the tested and untested samples differ, Figure 5 - Figure 7 compare demographic characteristics between AGR students used in the growth analysis and AGR students not used in the growth analysis due to lack of assessment information. The growth analysis AGR sample has higher proportions of black students and those eligible for free/reduced-price lunch and lower proportions of white students, although the differences are small. The schools included in the growth analysis are more likely urban and less likely rural.

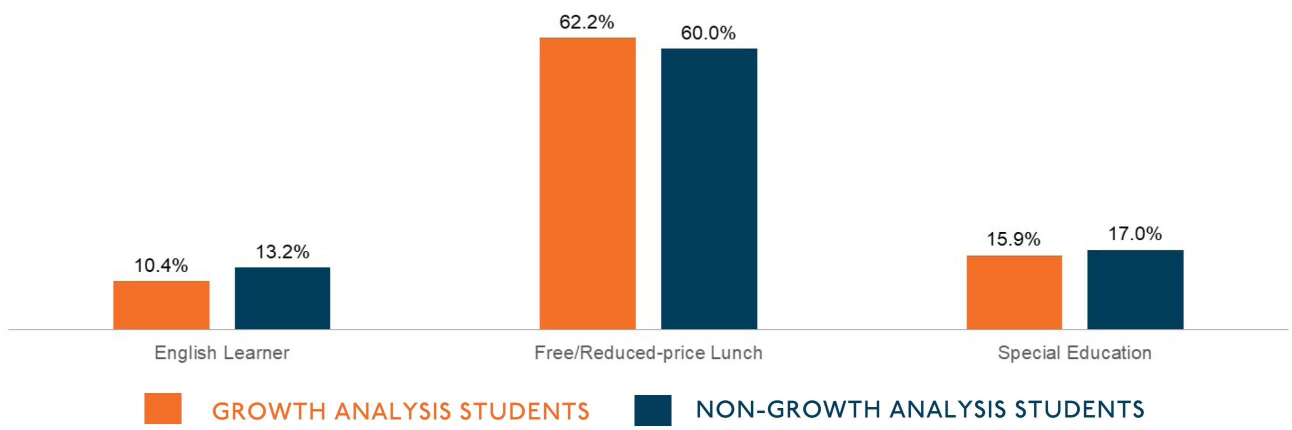
# Figure 5

Race/Ethnicity of AGR Growth Analysis and Non-Growth Analysis Students, 2018-19



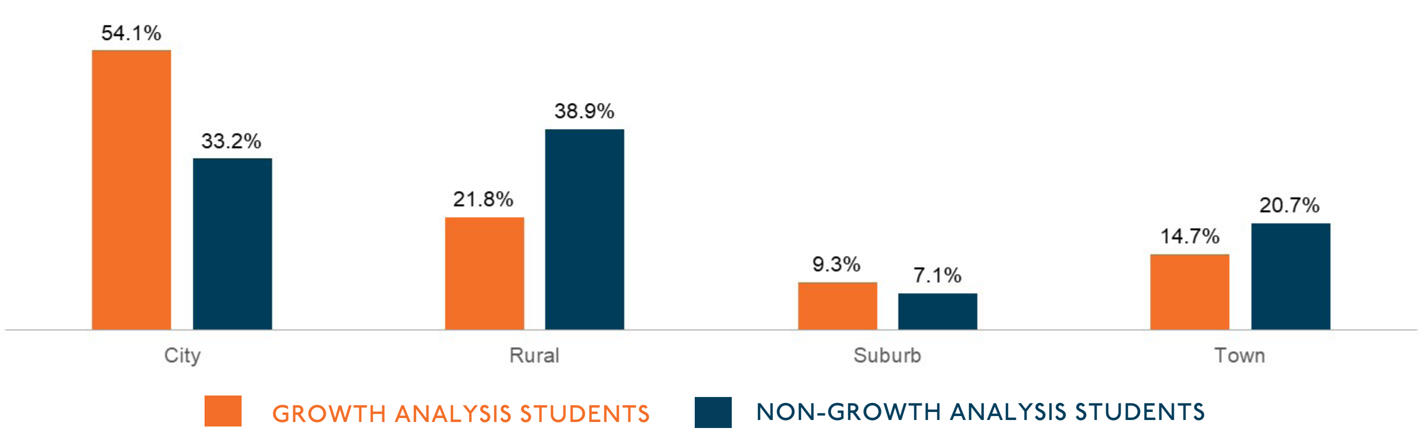
# Figure 6

Percentage of AGR Growth Analysis and Non-Growth Analysis Students that were English Learners, Free/Reduced-price Lunch Eligible, and in Special Education, 2018-19




# Figure 7

Locale Description of AGR Growth Analysis and Non-Growth Analysis Students, 2018-19



Section 4

# AGR Implementation

# AGR Implementation

This section of the report examines the usage of the three possible AGR strategies that schools could use as part of AGR. As noted previously, the three strategies include:

- Provide professional development related to small group instruction and reduce the class size to one of the following:
  - No more than 18.
  - No more than 30 in a combined classroom having at least 2 regular classroom teachers.
- Provide data-driven instructional coaching for the class teachers.
- Provide data-informed, one-to-one tutoring to pupils in the class who are struggling with reading or mathematics or both subjects.

As the program allowed schools to use more than one strategy within a school, there are seven possible combinations schools could implement: class size reduction only, coaching only, tutoring only, class size reduction and coaching, class size reduction and tutoring, coaching and tutoring, and all three strategies. Table 10 provides information on the strategy combinations AGR schools implemented during 2018-19. This table also provides information on the number and percentage of students affected by each strategy combination. The most frequently used strategies were class size reduction and coaching, all three strategies, coaching only, and class size reduction only. Very few schools used only tutoring as a strategy.

## Table 10

Distribution of AGR Strategies, 2018-19

| STRATEGY | NUMBER OF SCHOOLS | PERCENTAGE OF SCHOOLS | NUMBER OF STUDENTS | PERCENTAGE OF STUDENTS |
|---|---|---|---|---|
| Class Size Only | 77 | 18.7% | 15,448 | 20.8% |
| Coaching Only | 49 | 11.9% | 18,155 | 24.4% |
| Tutoring Only | 1 | 0.2% | 412 | 0.6% |
| Class Size and Coaching | 157 | 38.1% | 21,259 | 28.6% |
| Class Size and Tutoring | 27 | 6.6% | 2,941 | 4.0% |
| Coaching and Tutoring | 11 | 2.7% | 4,373 | 5.9% |
| All Three | 90 | 21.8% | 11,732 | 15.8% |

Section 5

# AGR Impacts

# AGR Impacts

This section of the report examines the results from statistical analyses undertaken to determine the impact of the AGR program. The intent is to answer the second and third guiding evaluation question:

2. To what extent is AGR meeting intended outcomes, including impacts on standardized test scores, attendance, and disciplinary events?
   a. How does AGR impact achievement gaps between low-income students and their higher-income peers?
   b. How does AGR impact vary by student characteristics?
   c. How does AGR's impact on outcomes compare to impacts associated with the SAGE program?
3. Are there differences between the three AGR strategies' impacts on intended outcomes?

To answer these questions, we begin by providing results on the program's overall and by-grade impacts. We then examine the impact of AGR compared to previous SAGE implementation, followed by impacts of AGR by student subgroup populations. Finally, we provide the results of analyses for different AGR strategy combinations.

All AGR impact analyses examine how students performed on four different outcome measures including math growth, reading growth, absences, and discipline. For each of these outcomes, this report provides a table of results at each applicable grade level and overall (across all applicable grades). These tables show a measure or measures of the program impact and a p-value that indicates the likelihood of observing the reported impact or more extreme assuming that there is no actual impact of the program. Larger p-values indicate weaker evidence of an impact, while smaller p-values indicate stronger evidence of an impact. Throughout the report, the evaluation uses a threshold of 0.05 to determine if a result was statistically significant from zero. All p-values presented in this report are corrected to account for multiple estimates (see the Technical Appendix for details).

## Overall Impacts

The impact analysis examines how AGR students performed compared to non-AGR students in similar schools, while controlling for student characteristics. Impacts are shown for each grade and for all grades combined.

Table 11 presents the impacts of AGR on math growth using two different measures. The first measure of impact is on a standardized scale representing the number of standard deviations from zero while the second measure of impact is in approximate MAP scale scores. Both show the difference between average AGR student growth and non-AGR student growth for students in similar schools. Results across all grades reveal little difference in math growth between AGR students and non-AGR students.

Table 12 shows impacts of AGR on reading growth. As with math growth, this table shows average differences in growth on two different scales, a standardized scale and points on the assessment scale (PALS in kindergarten and MAP in Grades 1-3). There are statistically significant impacts of AGR on kindergarten reading growth. On average, AGR students grew 0.11 standard deviations (or 1.4 PALS score points) more than their non-AGR counterparts at similar schools. The size of this impact is substantive, approximately equal to the impact of a one standard deviation increase in teacher effectiveness.[8] In contrast, results from MAP/STAR reading assessments indicate little difference in average growth between AGR and non-AGR students in Grades 1-3.

# Table 11

## Overall Impact of AGR on Math Growth

| OUTCOME | GRADE | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| MAP/STAR Math | First | -0.005 | -0.07 | 0.941 |
| | Second | 0.025 | 0.33 | 0.639 |
| | Third | -0.039 | -0.54 | 0.196 |
| | Overall (1-3) | -0.011 | -0.15 | 0.827 |

Note: P-values corrected to account for multiple estimates. * Statistically signicant at the 0.05 level.

# Table 12

## Overall Impact of AGR on Reading Growth

| OUTCOME | GRADE | IMPACT (STANDARDIZED) | IMPACT (APPROX. PALS/MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| PALS | Kindergarten | 0.105* | 1.44 | 0.039 |
| MAP/STAR Reading | First | -0.004 | -0.06 | 0.948 |
| | Second | 0.021 | 0.32 | 0.552 |
| | Third | -0.006 | -0.09 | 0.873 |
| | Overall (1-3) | 0.003 | 0.05 | 0.944 |

Note: P-values corrected to account for multiple estimates. * Statistically signicant at the 0.05 level.

8 Hanushek, E. A. & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4: 131-157.

The impacts of AGR on absence rates appear in Table 13. As indicated, while overall there was little impact of AGR on absence rates, there were higher, statistically significantly absence rates for third grade students in AGR compared to non-AGR students. On average, AGR students had an absence rate 0.4 percentage points higher than their third grade peers in matched non-AGR schools. This translates to approximately 0.6 more absence days, a relatively small impact that is measured with statistical precision due to the large sample size included in the analysis.

# Table 13

## Overall Impact of AGR on Absences

| OUTCOME | GRADE | IMPACT (PERCENTAGE POINTS) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|
| Absence Rate | Kindergarten | 0.27 | 0.5 | 0.339 |
| | First | 0.31 | 0.5 | 0.175 |
| | Second | 0.33 | 0.6 | 0.095 |
| | Third | 0.36* | 0.6 | 0.045 |
| | Overall (K-3) | 0.33 | 0.6 | 0.122 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 14 presents the impacts of the AGR program on student discipline as measured by out-of-school suspensions. While the overall impact of the program shows a decrease in the suspension rate, this result is not statistically significant.

# Table 14

## Overall Impact of AGR on Discipline

| OUTCOME | GRADE | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|
| Suspension Rate | Kindergarten | -0.5 | 0.081 |
| | First | -0.4 | 0.374 |
| | Second | -0.4 | 0.368 |
| | Third | -0.3 | 0.695 |
| | Overall (K-3) | -0.4 | 0.262 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

## Impacts Compared to SAGE

The following results provide information on the impact of the AGR program compared to previous SAGE implementation. This analysis estimates impacts using AGR schools both before and after their transition from SAGE to AGR. Table 15 shows the impact of AGR compared to SAGE on math growth using both a standardized measure and a measure on the MAP scale. Overall, differences in math growth between AGR and SAGE were small and non-statistically significant.

## Table 15

### Impact of AGR Compared to SAGE on Math Growth

| OUTCOME | GRADE | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| MAP/STAR Math | First | 0.033 | 0.45 | 0.484 |
| | Second | 0.040 | 0.54 | 0.184 |
| | Third | -0.050 | -0.69 | 0.060 |
| | Overall (1-3) | -0.004 | -0.06 | 0.886 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 16 includes differences in reading growth between AGR and SAGE. While the evaluation found few differences in reading growth in first through third grade, the estimated impact on kindergarten reading growth found that AGR students had higher average growth (by approximately 1.2 PALS score points) than previous SAGE students. The estimated difference between SAGE and AGR impacts has a p-value of 0.053, just above the level required for statistical significance.

## Table 16

### Impact of AGR Compared to SAGE on Reading Growth

| OUTCOME | GRADE | IMPACT (STANDARDIZED) | IMPACT (APPROX. PALS/MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| PALS | Kindergarten | 0.087 | 1.20 | 0.053 |
| MAP/STAR Reading | First | 0.029 | 0.42 | 0.616 |
| | Second | -0.009 | -0.13 | 0.857 |
| | Third | -0.003 | -0.50 | 0.114 |
| | Overall (1-3) | -0.014 | -0.21 | 0.685 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Absence rate comparisons between AGR and SAGE reveal few differences, as seen in Table 17. The estimated average AGR absence rates are slightly lower than SAGE rates. The kindergarten estimate is statistically significant, and the first and second grade estimates are nearly significant. As with the overall results, however, this is a relatively small impact that is measured with statistical precision due to the large sample size included in the analysis.

## Table 17

### Impact of AGR Compared to SAGE on Absences

| OUTCOME | GRADE | IMPACT (PERCENTAGE POINTS) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|
| Absence Rate | Kindergarten | -0.64* | -1.1 | 0.041 |
| | First | -0.46 | -0.8 | 0.053 |
| | Second | -0.40 | -0.7 | 0.055 |
| | Third | -0.37 | -0.6 | 0.089 |
| | Overall (K-3) | -0.46 | -0.8 | 0.059 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 18 shows the AGR impact compared to the SAGE impact for student discipline. The evaluation found little difference in the suspension rates between AGR students and SAGE students.

## Table 18

### Impact of AGR Compared to SAGE on Discipline

| OUTCOME | GRADE | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|
| Suspension Rate | Kindergarten | 0.0 | 0.898 |
| | First | -0.1 | 0.864 |
| | Second | 0.0 | 0.891 |
| | Third | 0.5 | 0.111 |
| | Overall (K-3) | 0.1 | 0.821 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

## Impacts by Demographics

In addition to examining the overall, statewide impact of the AGR program on student performance, this evaluation also examines whether the AGR program has different impacts for different subgroups of students. Given AGR's focus on closing the achievement gap, examining the program impact for specific subgroups, particularly economically disadvantaged students, is important for determining whether the program is meeting its goals. For each of the subgroups, results compare outcomes of subgroup students that attended AGR schools to outcomes of subgroup students attended observationally similar, non-AGR schools (e.g. economically disadvantaged students at AGR schools versus economically disadvantaged students at similar, non-AGR schools). Results are pooled across all applicable grade levels to measure the impact of AGR for the following subgroups of students: females, Asian students, black students, Hispanic students, white students, students of other race or ethnicity, economically disadvantaged students (as measured by free or reduced-price lunch, or FRL status), English learner (EL) students, special education students, and students in schools within cities.

## Impacts for Students Eligible for Free/ Reduced-price Lunch

Given AGR's focus on reducing gaps between economically disadvantaged students and their higher income peers, we begin by presenting AGR impacts for students who are eligible for free or reduced-price lunch. Outcomes for students eligible for free/reduced-price lunch at AGR schools are compared to outcomes of students eligible for free/reduced-price lunch at observationally similar, non-AGR schools. These results mirror pooled impact results shown above. AGR's impact is large for kindergarten reading scores of students eligible for free/reduced-price lunch, but there are no statistically significant impacts on other outcomes. Students eligible for free/reduced-price lunch had an average reading growth 0.13 standard deviations higher than students eligible for free/reduced-price lunch in similar non-AGR schools (approximately 2 points higher growth on the PALS assessment). Outside of kindergarten reading, AGR impacts on students eligible for free/reduced-price lunch are generally small and not statistically significant. A potential exception is the impact on suspension rates. A 0.5 percentage point reduction in suspension rates is substantial relative to the low (2.6 percent) incidence of suspensions in Grades K-3.

## Table 19

Impact of AGR on FRL Students - All Outcomes

| OUTCOME | GRADE | IMPACT (STANDARDIZED) | IMPACT (SCALED) | P-VALUE |
|---|---|---|---|---|
| MAP/STAR Math | Overall (1-3) | -0.025 | -.034 MAP points | 0.549 |
| PALS | Kindergarten | 0.134* | 1.84 PALS points | 0.038 |
| MAP/STAR Reading | Overall (1-3) | 0.003 | 0.05 MAP points | 0.934 |
| Absence Rate | Overall (K-3) | 0.23 percentage points | 0.4 days | 0.490 |
| Suspension Rate | Overall (K-3) | -0.5 percentage points | N/A | 0.174 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

## Impacts for Other Subgroups

Table 20 shows the impact of the AGR program on math growth for each subgroup, relative to the same subgroup of students in similar non-AGR schools. Across all subgroups, the evaluation found little difference in math growth between AGR students and non-AGR students in similar schools, regardless of subgroup.

# Table 20

### Impact of AGR on Math Growth by Student Subgroup

| OUTCOME | GRADE | SUBGROUP | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|---|
| MAP/STAR Math | Overall (1-3) | Female Students | -0.004 | -0.06 | 0.934 |
| | | Asian Students | -0.038 | -0.52 | 0.760 |
| | | Black Students | -0.058 | -0.79 | 0.386 |
| | | Hispanic Students | 0.027 | 0.37 | 0.539 |
| | | White Students | 0.001 | 0.01 | 0.959 |
| | | Other Race/Ethnicity | -0.009 | -0.12 | 0.883 |
| | | EL Students | 0.009 | 0.12 | 0.939 |
| | | Special Ed. Students | -0.065 | -0.89 | 0.069 |
| | | City Students | -0.017 | -0.23 | 0.840 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

As shown in Table 21, for kindergarten PALS the evaluation found significantly positive impacts of AGR for a variety of subgroups including females, Asian students, Hispanic students, English learners, and students in urban areas. The largest impacts found were for kindergarten English learners who had an average reading growth of 0.37 standard deviations, or 5.1 PALS score points, higher than English learners in similar non-AGR schools. The evaluation found only small, not statistically significant differences in reading growth for first through third grade overall between AGR students and non-AGR students in similar schools, regardless of the type of student.

## Table 21

### Impact of AGR on Reading Growth by Student Subgroup

| OUTCOME | GRADE | SUBGROUP | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|---|
| PALS | Kindergarten | Female Students | 0.103* | 1.41 | 0.037 |
| | | Asian Students | 0.240* | 3.31 | 0.006 |
| | | Black Students | 0.187 | 2.57 | 0.137 |
| | | Hispanic Students | 0.253* | 3.48 | 0.005 |
| | | White Students | 0.053 | 0.73 | 0.219 |
| | | Other Race/Ethnicity | 0.004 | 0.06 | 0.958 |
| | | EL Students | 0.370* | 5.08 | 0.000 |
| | | Special Ed. Students | 0.083 | 1.14 | 0.383 |
| | | City Students | 0.206* | 2.83 | 0.018 |
| MAP/STAR Reading | Overall (1-3) | Female Students | 0.005 | 0.07 | 0.913 |
| | | Asian Students | -0.008 | -0.13 | 0.947 |
| | | Black Students | -0.036 | -0.53 | 0.552 |
| | | Hispanic Students | 0.015 | 0.22 | 0.773 |
| | | White Students | 0.016 | 0.24 | 0.547 |
| | | Other Race/Ethnicity | 0.016 | 0.24 | 0.765 |
| | | EL Students | 0.012 | 0.18 | 0.904 |
| | | Special Ed. Students | -0.004 | -0.06 | 0.943 |
| | | City Students | -0.010 | -0.15 | 0.902 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

With the exception of student absences in schools located in cities, AGR impacts on absence rates, as shown in Table 22, show only small differences across all subgroups between AGR students and non-AGR students in similar schools. On average, city students at AGR schools had one more absence than their counterparts at non-AGR city schools.

## Table 22

### Impact of AGR on Absences by Student Subgroup

| OUTCOME | GRADE | SUBGROUP | IMPACT (STANDARDIZED) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|---|
| Absence Rate | Overall (K-3) | Female Students | 0.30 | 0.5 | 0.184 |
| | | Asian Students | 0.32 | 0.6 | 0.680 |
| | | Black Students | 0.87 | 1.5 | 0.053 |
| | | Hispanic Students | 0.40 | 0.7 | 0.167 |
| | | White Students | 0.21 | 0.4 | 0.518 |
| | | Other Race/Ethnicity | 0.27 | 0.5 | 0.753 |
| | | EL Students | 0.39 | 0.7 | 0.145 |
| | | Special Ed. Students | 0.40 | 0.7 | 0.189 |
| | | City Students | 0.66* | 1.2 | 0.006 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 23 provides the results of the estimated impact of AGR on student discipline for the same subgroups of students. As seen from this table, although all subgroups of AGR students had lower suspension rates, only estimates for Hispanic and English learner students were statistically significant. Hispanic AGR students had a suspension rate roughly 0.6 percentage points lower than their non-AGR counterparts and English learner AGR students had a suspension rate roughly 0.5 percentage points lower. These are meaningfully large impacts considering the already low suspension rate (2.6 percent) at these grade levels.

## Table 23

Impact of AGR on Discipline by Student Subgroup

| OUTCOME | GRADE | SUBGROUP | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|---|
| Suspension Rate | Overall (K-3) | Female Students | -0.2 | 0.378 |
| | | Asian Students | 0.0 | 0.886 |
| | | Black Students | -0.7 | 0.307 |
| | | Hispanic Students | -0.6* | 0.048 |
| | | White Students | 0.0 | 0.963 |
| | | Other Race/Ethnicity | -0.9 | 0.387 |
| | | EL Students | -0.5* | 0.043 |
| | | Special Ed. Students | -0.9 | 0.190 |
| | | City Students | -0.5 | 0.361 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

## Differences in Outcomes by Strategy

This set of student performance results examines the difference in AGR impacts for each combination of strategies. For the four following tables, the impact is the difference in the outcome between AGR students in schools with the strategy combination listed and AGR students in schools with small class size only. The strategy combinations examined include coaching only, tutoring only, small class size and coaching, small class size and tutoring, coaching and tutoring, and all three strategies.

Differences in outcomes by strategy should be considered only limited evidence of how strategy usage might impact test score growth, absences, and discipline. The results in Table 24 - Table 27 should not be interpreted as causal in nature. The analysis includes only AGR schools with no comparison schools. Because AGR schools are allowed to select their strategies, differences in outcomes could be biased by omitted variables. For example, more effective schools may systematically choose certain strategies, in which case differences in outcomes could be caused by school effectiveness rather than the strategies themselves.

Table 24 shows the impact on math growth for each strategy combination compared to small class size. Strategy combinations associated with higher average math growth compared to small class size included coaching only in first grade and tutoring only in third grade.

# Table 24

## Differences in Math Growth by Strategy, Compared to Small Class Size Only

| OUTCOME | GRADE | STRATEGY | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|---|
| MAP/STAR Math | First | Coaching Only | 0.235* | 3.21 | 0.007 |
| | | Tutoring Only | N/A | N/A | N/A |
| | | Class Size and Coaching | 0.116 | 1.58 | 0.500 |
| | | Class Size and Tutoring | 0.080 | 1.09 | 0.692 |
| | | Coaching and Tutoring | N/A | N/A | N/A |
| | | All Three | 0.183 | 2.50 | 0.126 |
| | Second | Coaching Only | 0.071 | 0.96 | 0.641 |
| | | Tutoring Only | N/A | N/A | N/A |
| | | Class Size and Coaching | 0.014 | 0.19 | 0.962 |
| | | Class Size and Tutoring | 0.154 | 2.09 | 0.200 |
| | | Coaching and Tutoring | 0.059 | 0.81 | 0.738 |
| | | All Three | -0.062 | -0.83 | 0.745 |
| | Third | Coaching Only | 0.021 | 0.29 | 0.875 |
| | | Tutoring Only | 0.374* | 5.17 | 0.000 |
| | | Class Size and Coaching | 0.035 | 0.49 | 0.734 |
| | | Class Size and Tutoring | -0.001 | -0.02 | 1.003 |
| | | Coaching and Tutoring | 0.077 | 1.06 | 0.501 |
| | | All Three | 0.076 | 1.05 | 0.509 |
| | Overall (1-3) | Coaching Only | 0.075 | 1.03 | 0.473 |
| | | Tutoring Only | 0.292 | 3.99 | 0.107 |
| | | Class Size and Coaching | 0.042 | 0.58 | 0.680 |
| | | Class Size and Tutoring | 0.076 | 1.04 | 0.408 |
| | | Coaching and Tutoring | 0.106 | 1.45 | 0.178 |
| | | All Three | 0.051 | 0.69 | 0.646 |

Note: P-values corrected to account for multiple estimates. N/A indicates too few schools employing a strategy to accurately estimate results. * Statistically significant at the 0.05 level.

Looking at the differences in kindergarten PALS reading growth for each strategy combination compared to small class size only, the analysis found few differences across strategy combinations, as seen in Table 25. Results on differences in reading growth in Grades 1-3, found in the same table, indicate higher average reading growth for coaching only in first grade and class size and coaching across first through third grade when compared to small class size only.

# Table 25

## Differences in Reading Growth by Strategy, Compared to Small Class Size Only

| OUTCOME | GRADE | STRATEGY | IMPACT (STANDARIZED) | IMPACT (APPROX. PALS OR MAP SCALE) | P-VALUE |
|---|---|---|---|---|---|
| PALS | Kindergarten | Coaching Only | -0.132 | -1.82 | 0.369 |
| | | Tutoring Only | 0.001 | 0.01 | 0.996 |
| | | Class Size and Coaching | -0.104 | -1.43 | 0.252 |
| | | Class Size and Tutoring | -0.001 | -0.01 | 1.001 |
| | | Coaching and Tutoring | -0.109 | -1.50 | 0.518 |
| | | All Three | -0.030 | -0.42 | 0.842 |
| MAP/STAR Reading | First | Coaching Only | 0.264* | 3.60 | 0.002 |
| | | Tutoring Only | N/A | N/A | N/A |
| | | Class Size and Coaching | 0.211 | 2.87 | 0.102 |
| | | Class Size and Tutoring | -0.039 | -0.53 | 0.894 |
| | | Coaching and Tutoring | N/A | N/A | N/A |
| | | All Three | 0.101 | 1.38 | 0.497 |
| | Second | Coaching Only | 0.066 | 0.89 | 0.582 |
| | | Tutoring Only | N/A | N/A | N/A |
| | | Class Size and Coaching | 0.081 | 1.10 | 0.520 |
| | | Class Size and Tutoring | 0.038 | 0.51 | 0.817 |
| | | Coaching and Tutoring | 0.052 | 0.70 | 0.709 |
| | | All Three | -0.032 | -0.44 | 0.871 |
| | Third | Coaching Only | 0.016 | 0.22 | 0.881 |
| | | Tutoring Only | 0.133 | 1.84 | 0.235 |
| | | Class Size and Coaching | 0.095 | 1.32 | 0.053 |
| | | Class Size and Tutoring | -0.027 | -0.37 | 0.887 |
| | | Coaching and Tutoring | 0.049 | 0.68 | 0.689 |
| | | All Three | 0.014 | 0.19 | 0.919 |

# Table 25, continued

| OUTCOME | GRADE | STRATEGY | IMPACT (STANDARIZED) | IMPACT (APPROX. PALS OR MAP SCALE) | P-VALUE |
|---|---|---|---|---|---|
| MAP/STAR Reading | Overall (1-3) | Coaching Only | 0.075 | 1.120 | 0.273 |
| | | Tutoring Only | 0.104 | 1.56 | 0.578 |
| | | Class Size and Coaching | 0.109* | 1.63 | 0.046 |
| | | Class Size and Tutoring | -0.013 | -0.19 | 0.958 |
| | | Coaching and Tutoring | 0.092 | 1.37 | 0.201 |
| | | All Three | 0.015 | 0.22 | 0.884 |

Note: P-values corrected to account for multiple estimates. N/A indicates too few schools employing a strategy to accurately estimate results. * Statistically significant at the 0.05 level.

Table 26 displays differences in absence rates for each of the strategy combinations compared to class size reduction only. As this table illustrates, students in AGR schools using tutoring only, when compared to small class sizes only, had higher absence rates in kindergarten through third grade overall.

# Table 26

## Differences in Absences by Strategy, Compared to Small Class Size Only

| OUTCOME | GRADE | STRATEGY | IMPACT (PERCENTAGE POINTS) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|---|
| Absence Rate | Kindergarten | Coaching Only | -0.31 | -0.5 | 0.706 |
| | | Tutoring Only | 0.89 | 1.6 | 0.212 |
| | | Class Size and Coaching | -0.30 | -0.5 | 0.733 |
| | | Class Size and Tutoring | 0.91 | 1.6 | 0.117 |
| | | Coaching and Tutoring | -0.99 | -1.7 | 0.706 |
| | | All Three | -0.19 | -0.3 | 0.874 |
| | First | Coaching Only | 0.05 | 0.1 | 0.966 |
| | | Tutoring Only | 0.75 | 1.3 | 0.104 |
| | | Class Size and Coaching | 0.04 | 0.1 | 0.975 |
| | | Class Size and Tutoring | -0.15 | -0.3 | 0.954 |
| | | Coaching and Tutoring | -0.05 | -0.1 | 0.964 |
| | | All Three | 0.44 | 0.8 | 0.506 |

# Table 26, continued

| OUTCOME | GRADE | STRATEGY | IMPACT (PERCENTAGE POINTS) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|---|
| Absence Rate | Second | Coaching Only | 0.13 | 0.2 | 0.819 |
| | | Tutoring Only | 0.78 | 1.4 | 0.122 |
| | | Class Size and Coaching | 0.25 | 0.4 | 0.695 |
| | | Class Size and Tutoring | 0.10 | 0.2 | 0.973 |
| | | Coaching and Tutoring | 0.34 | 0.6 | 0.625 |
| | | All Three | 0.24 | 0.4 | 0.688 |
| | Third | Coaching Only | 0.09 | 0.2 | 0.890 |
| | | Tutoring Only | 0.68 | 1.2 | 0.097 |
| | | Class Size and Coaching | 0.15 | 0.3 | 0.872 |
| | | Class Size and Tutoring | 0.87 | 1.5 | 0.103 |
| | | Coaching and Tutoring | -0.19 | -0.3 | 0.911 |
| | | All Three | -0.14 | -0.3 | 0.881 |
| | Overall (K-3) | Coaching Only | 0.02 | 0.0 | 0.984 |
| | | Tutoring Only | 0.75* | 1.3 | 0.037 |
| | | Class Size and Coaching | 0.04 | 0.1 | 0.976 |
| | | Class Size and Tutoring | 0.44 | 0.8 | 0.685 |
| | | Coaching and Tutoring | -0.19 | -0.3 | 0.885 |
| | | All Three | 0.10 | 0.2 | 0.889 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 27 shows differences in student discipline, as measured by the out-of-school suspension rate, for each combination of AGR strategy as compared to small class size only. Students in schools and grades implementing the tutoring only strategy in third grade and coaching only strategy in all grades had suspension rates significantly higher than students in schools and grades using small class sizes only.

# Table 27

## Differences in Discipline by Strategy, Compared to Small Class Size Only

| OUTCOME | GRADE | STRATEGY | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|---|
| Suspension Rate | Kindergarten | Coaching Only | 1.9 | 0.474 |
| | | Tutoring Only | 0.9 | 0.164 |
| | | Class Size and Coaching | 0.3 | 0.877 |
| | | Class Size and Tutoring | -0.5 | 0.330 |
| | | Coaching and Tutoring | 0.6 | 0.693 |
| | | All Three | 0.3 | 0.710 |
| | First | Coaching Only | 0.7 | 0.788 |
| | | Tutoring Only | 0.0 | 0.998 |
| | | Class Size and Coaching | 1.2 | 0.620 |
| | | Class Size and Tutoring | 0.4 | 0.687 |
| | | Coaching and Tutoring | 0.1 | 0.957 |
| | | All Three | 0.3 | 0.818 |
| | Second | Coaching Only | 3.1 | 0.164 |
| | | Tutoring Only | 0.8 | 0.222 |
| | | Class Size and Coaching | 0.9 | 0.704 |
| | | Class Size and Tutoring | 1.1 | 0.115 |
| | | Coaching and Tutoring | 0.8 | 0.614 |
| | | All Three | 0.3 | 0.801 |

# Table 27, continued

| OUTCOME | GRADE | STRATEGY | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|---|
| Suspension Rate | Third | Coaching Only | 2.1 | 0.593 |
| | | Tutoring Only | 2.0* | 0.001 |
| | | Class Size and Coaching | 1.5 | 0.501 |
| | | Class Size and Tutoring | 0.7 | 0.517 |
| | | Coaching and Tutoring | 0.5 | 0.758 |
| | | All Three | 0.4 | 0.689 |
| | Overall (K-3) | Coaching Only | 1.0* | 0.018 |
| | | Tutoring Only | 0.3 | 0.607 |
| | | Class Size and Coaching | 1.8 | 0.200 |
| | | Class Size and Tutoring | 0.9 | 0.518 |
| | | Coaching and Tutoring | 0.4 | 0.658 |
| | | All Three | 0.4 | 0.589 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

## Differences in Outcomes by Intensity and Frequency of Strategy Use

New to this year's evaluation is an examination of the frequency and intensity of strategies on the four outcomes of interest. Results from the End-of-Year Report (see below) provide information on the frequency of coaching per semester, the frequency of tutoring per month, the characteristics of instructional coaches, the usage of high-quality instructional coaching practices, and the usage of high-quality tutoring practices. Impacts analyzed by frequency show the effect associated with one additional instructional coaching session or one additional tutoring session. Impacts analyzed by qualification or practice, show the effect associated with that qualification or practice regardless of the other qualifications or practices employed. As with the previous section, results shown below should not be considered causal due to selection concerns associated with schools choosing which strategies to use.

Table 28 shows the impact of the frequency and intensity of coaching and tutoring on math growth across all grades 1-3. There were no statistically significant impacts of frequency, coach characteristics, coach practices, or tutoring practices.

# Table 28

## Differences in Math Growth by Frequency and Intensity of Coaching and Tutoring

| OUTCOME | FREQUENCY, CHARACTERISTIC, OR PRACTICE | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| MAP/STAR Math | **Frequency** | | | |
| | Coaching frequency per semester | -0.004 | -0.05 | 0.655 |
| | Tutoring frequency per month | -0.019 | -0.25 | 0.489 |
| | **Coach Characteristics** | | | |
| | Coach training | -0.022 | -0.29 | 0.886 |
| | Previous coaching experience | -0.002 | -0.03 | 0.993 |
| | Content specialist in their subject | -0.055 | -0.75 | 0.573 |
| | **Coaching Practices** | | | |
| | One-to-one teacher coaching | 0.072 | 0.99 | 0.597 |
| | Team teacher coaching | 0.069 | 0.95 | 0.568 |
| | Keeps a coaching log | -0.011 | -0.16 | 0.936 |
| | Advises teachers to set goals | 0.010 | 0.13 | 0.953 |
| | Coaching focuses on teacher goals | -0.007 | -0.09 | 0.960 |
| | Maintains a focus on equity | -0.024 | -0.33 | 0.812 |
| | Encourages reflective practices | 0.044 | 0.60 | 0.730 |
| | Discusses data with teachers | -0.028 | -0.38 | 0.819 |
| | Observes teacher practices | -0.063 | -0.86 | 0.528 |
| | **Tutoring Practices** | | | |
| | Reviews student data | -0.059 | -0.80 | 0.796 |
| | Models appropriate learning behavior | 0.142 | 1.94 | 0.116 |
| | Adapts to student learning styles | 0.063 | 0.87 | 0.490 |
| | Maintains a focus on equity | -0.038 | -0.52 | 0.682 |
| | Provides scaffolding | -0.120 | -1.64 | 0.265 |
| | Communicates regularly with teacher | -0.168 | -2.30 | 0.167 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

As with the math outcome, the impact of frequency coaching and tutoring, coach characteristics, coach practices, and tutoring practices were not statistically significant on PALS growth, as seen from Table 29.

# Table 29

## Differences in PALS Growth by Frequency and Intensity of Coaching and Tutoring

| OUTCOME | FREQUENCY, CHARACTERISTIC, OR PRACTICE | IMPACT (STANDARDIZED) | IMPACT (APPROX. PALS SCALE) | P-VALUE |
|---|---|---|---|---|
| PALS | **Frequency** | | | |
| | Coaching frequency per semester | -0.006 | -0.08 | 0.617 |
| | Tutoring frequency per month | -0.008 | -0.10 | 0.829 |
| | **Coach Characteristics** | | | |
| | Coach training | 0.049 | 0.68 | 0.750 |
| | Previous coaching experience | 0.022 | 0.31 | 0.893 |
| | Content specialist in their subject | -0.150 | -2.06 | 0.100 |
| | **Coaching Practices** | | | |
| | One-to-one teacher coaching | -0.006 | -0.08 | 0.984 |
| | Team teacher coaching | -0.042 | -0.58 | 0.820 |
| | Keeps a coaching log | -0.007 | -0.10 | 0.970 |
| | Advises teachers to set goals | 0.045 | 0.61 | 0.789 |
| | Coaching focuses on teacher goals | 0.015 | 0.20 | 0.958 |
| | Maintains a focus on equity | -0.081 | -1.12 | 0.583 |
| | Encourages reflective practices | -0.030 | -0.42 | 0.899 |
| | Discusses data with teachers | 0.103 | 1.42 | 0.515 |
| | Observes teacher practices | -0.004 | -0.06 | 0.994 |
| | **Tutoring Practices** | | | |
| | Reviews student data | -0.102 | -1.40 | 0.532 |
| | Models appropriate learning behavior | -0.170 | -2.34 | 0.454 |
| | Adapts to student learning styles | -0.027 | -0.38 | 0.887 |
| | Maintains a focus on equity | 0.061 | 0.84 | 0.702 |
| | Provides scaffolding | 0.127 | 1.75 | 0.500 |
| | Communicates regularly with teacher | -0.127 | -1.75 | 0.509 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 30 shows the impact of coaching and tutoring frequency and intensity on reading growth across grades 1-3 overall. Significant results included a positive impact on reading growth for a tutor modeling appropriate learning behavior and for a tutor adapting to student learning styles and a negative impact on reading growth for a coach being a content specialist in their subject, a coach observing teacher practices, and a tutor communicating regularly with the classroom teacher.

## Table 30

### Differences in Reading Growth by Frequency and Intensity of Coaching and Tutoring

| OUTCOME | FREQUENCY, CHARACTERISTIC, OR PRACTICE | IMPACT (STANDARDIZED) | IMPACT (APPROX. MAP SCALE) | P-VALUE |
|---|---|---|---|---|
| MAP/STAR Reading | **Frequency** | | | |
| | Coaching frequency per semester | -0.002 | -0.03 | 0.814 |
| | Tutoring frequency per month | -0.003 | -0.05 | 0.916 |
| | **Coach Characteristics** | | | |
| | Coach training | -0.015 | -0.23 | 0.887 |
| | Previous coaching experience | 0.004 | 0.06 | 0.974 |
| | Content specialist in their subject | -0.113* | -1.69 | 0.037 |
| | **Coaching Practices** | | | |
| | One-to-one teacher coaching | 0.043 | 0.64 | 0.654 |
| | Team teacher coaching | 0.019 | 0.28 | 0.858 |
| | Keeps a coaching log | -0.081 | -1.21 | 0.203 |
| | Advises teachers to set goals | -0.008 | -0.12 | 0.944 |
| | Coaching focuses on teacher goals | 0.045 | 0.67 | 0.524 |
| | Maintains a focus on equity | 0.022 | 0.33 | 0.787 |
| | Encourages reflective practices | 0.034 | 0.51 | 0.800 |
| | Discusses data with teachers | -0.010 | -0.16 | 0.942 |
| | Observes teacher practices | -0.122* | -1.82 | 0.027 |
| | **Tutoring Practices** | | | |
| | Reviews student data | -0.027 | -0.40 | 0.885 |
| | Models appropriate learning behavior | 0.133* | 1.99 | 0.001 |
| | Adapts to student learning styles | 0.174* | 2.60 | 0.000 |
| | Maintains a focus on equity | -0.048 | -0.72 | 0.479 |
| | Provides scaffolding | 0.012 | 0.17 | 0.876 |
| | Communicates regularly with teacher | -0.286* | -4.28 | 0.000 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Table 31 and Table 32 show the impacts of coaching and tutoring frequency, characteristics, and practices on absence rate and out-of-school suspension rates respectively. As seen from these two tables, there were no statistically significant impacts.

# Table 31

## Differences in Absences by Frequency and Intensity of Coaching and Tutoring

| OUTCOME | FREQUENCY, CHARACTERISTIC, OR PRACTICE | IMPACT (PERCENTAGE POINTS) | IMPACT (APPROX. DAYS) | P-VALUE |
|---|---|---|---|---|
| Absence Rate | **Frequency** | | | |
| | Coaching frequency per semester | -0.03 | -0.1 | 0.601 |
| | Tutoring frequency per month | -0.13 | -0.2 | 0.497 |
| | **Coach Characteristics** | | | |
| | Coach training | 0.29 | 0.5 | 0.872 |
| | Previous coaching experience | -0.43 | -0.8 | 0.516 |
| | Content specialist in their subject | 0.50 | 0.9 | 0.584 |
| | **Coaching Practices** | | | |
| | One-to-one teacher coaching | -0.02 | 0.0 | 0.996 |
| | Team teacher coaching | -0.07 | -0.1 | 0.966 |
| | Keeps a coaching log | 0.11 | 0.2 | 0.852 |
| | Advises teachers to set goals | 0.26 | 0.4 | 0.815 |
| | Coaching focuses on teacher goals | 0.25 | 0.4 | 0.831 |
| | Maintains a focus on equity | -0.05 | -0.1 | 0.968 |
| | Encourages reflective practices | -0.38 | -0.7 | 0.658 |
| | Discusses data with teachers | -0.56 | -1.0 | 0.413 |
| | Observes teacher practices | 0.56 | 1.0 | 0.616 |
| | **Tutoring Practices** | | | |
| | Reviews student data | -1.51 | -2.6 | 0.475 |
| | Models appropriate learning behavior | 1.39 | 2.4 | 0.498 |
| | Adapts to student learning styles | -2.38 | -4.2 | 0.342 |
| | Maintains a focus on equity | 0.29 | 0.5 | 0.748 |
| | Provides scaffolding | 0.63 | 1.1 | 0.817 |
| | Communicates regularly with teacher | -1.96 | -3.4 | 0.472 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

# Table 32

## Differences in Discipline by Frequency and Intensity of Coaching and Tutoring

| OUTCOME | FREQUENCY, CHARACTERISTIC, OR PRACTICE | IMPACT (PERCENTAGE POINTS) | P-VALUE |
|---|---|---|---|
| Suspension Rate | **Frequency** | | |
| | Coaching frequency per semester | 0.0 | 0.650 |
| | Tutoring frequency per month | 0.0 | 0.944 |
| | **Coach Characteristics** | | |
| | Coach training | 0.6 | 0.494 |
| | Previous coaching experience | -0.2 | 0.852 |
| | Content specialist in their subject | 0.3 | 0.691 |
| | **Coaching Practices** | | |
| | One-to-one teacher coaching | -0.9 | 0.198 |
| | Team teacher coaching | -0.4 | 0.647 |
| | Keeps a coaching log | 0.1 | 0.940 |
| | Advises teachers to set goals | 0.1 | 0.954 |
| | Coaching focuses on teacher goals | 0.1 | 0.964 |
| | Maintains a focus on equity | -1.2 | 0.026 |
| | Encourages reflective practices | 0.5 | 0.577 |
| | Discusses data with teachers | 0.0 | 1.000 |
| | Observes teacher practices | 0.3 | 0.744 |
| | **Tutoring Practices** | | |
| | Reviews student data | -0.1 | 0.979 |
| | Models appropriate learning behavior | -0.3 | 0.833 |
| | Adapts to student learning styles | -0.1 | 0.968 |
| | Maintains a focus on equity | 0.0 | 0.994 |
| | Provides scaffolding | 0.4 | 0.743 |
| | Communicates regularly with teacher | 0.4 | 0.596 |

Note: P-values corrected to account for multiple estimates. * Statistically significant at the 0.05 level.

Section 6

# Impacts on Statewide Achievement Gaps

# Impacts on Statewide Achievement Gaps

The Wisconsin Legislature designed AGR to reduce achievement gaps between low-income students and their higher-income peers. The results presented above show that, for students who are eligible for free or reduced-price lunch, AGR is associated with large increases on Kindergarten PALS but only small changes in grades 1-3 math and reading. In this section, for each of Kindergarten PALS and Grades 1-3 MAP and STAR math and reading, we calculated how AGR might have impacted statewide achievement gaps in 2019.

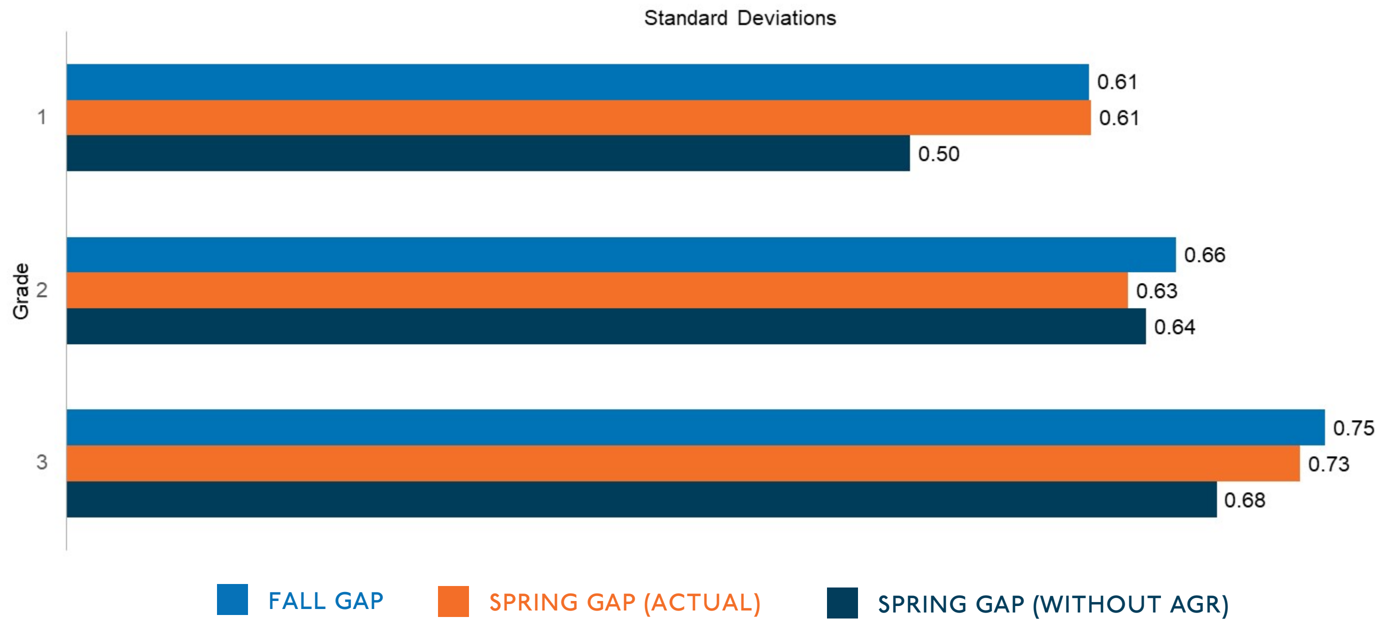
Statewide gaps are measured using data for all Wisconsin students with both fall and spring scores on the appropriate assessment (either PALS Reading for Kindergarteners or STAR/MAP for Grades 1-3 math and reading), not just those who appear in the analysis samples described above. For each grade and subject, we calculate the fall test score gap, prior to AGR impacts, and the spring test score gap, which includes AGR impacts. We also calculate the spring gap under the hypothetical that AGR did not exist. These hypothetical spring gaps are estimated under the assumption that AGR impacts are constant across all schools, not just those that appear in the analysis samples. For more details on the methodology for the gap analysis, please see the Technical Appendix.

Figure 8 and Figure 9 describe AGR's impacts on gaps in math and reading, respectively. In each figure, the top bar is the actual fall test score gap between students eligible for free or reduced-price lunches in 2018-19 and those who were not. A positive gap indicates that higher-income students scored higher on average than students who received free or reduced-price lunch. The middle bar shows the same gap from the spring of 2018-19. If test score gaps are shrinking, the middle bar (spring gap) would be shorter than the top bar (fall gap). The lowest bar in each grade represents what the spring gap would have been had AGR not been in effect. If the lowest bar is longer than the middle bar, AGR reduced the test score gap from what it would have been without the program.

In both figures, actual spring gaps are generally wider than fall gaps (the only exception is Kindergarten reading as measured by PALS). This does not mean, however, that AGR widened the gap. AGR affects only a subset of students who were eligible for free or reduced-price lunch and also benefits many students who are not eligible for free or reduced-price lunch but attend a school that receives AGR funds. To understand how AGR impacted statewide gaps, we need to compare actual spring gaps to the gaps that would have occurred had AGR not existed. If AGR reduces statewide gaps, then the lower bars in each grade will be longer than the middle bars, meaning that the gap would have been wider had AGR not existed. For the most part, actual spring gaps and hypothetical spring gaps are approximately equal, reflecting the lack of AGR impacts that we estimate in Grades 1-3. The exception is kindergarten reading, where the statewide spring gap would have been 0.05 standard deviations wider without AGR.

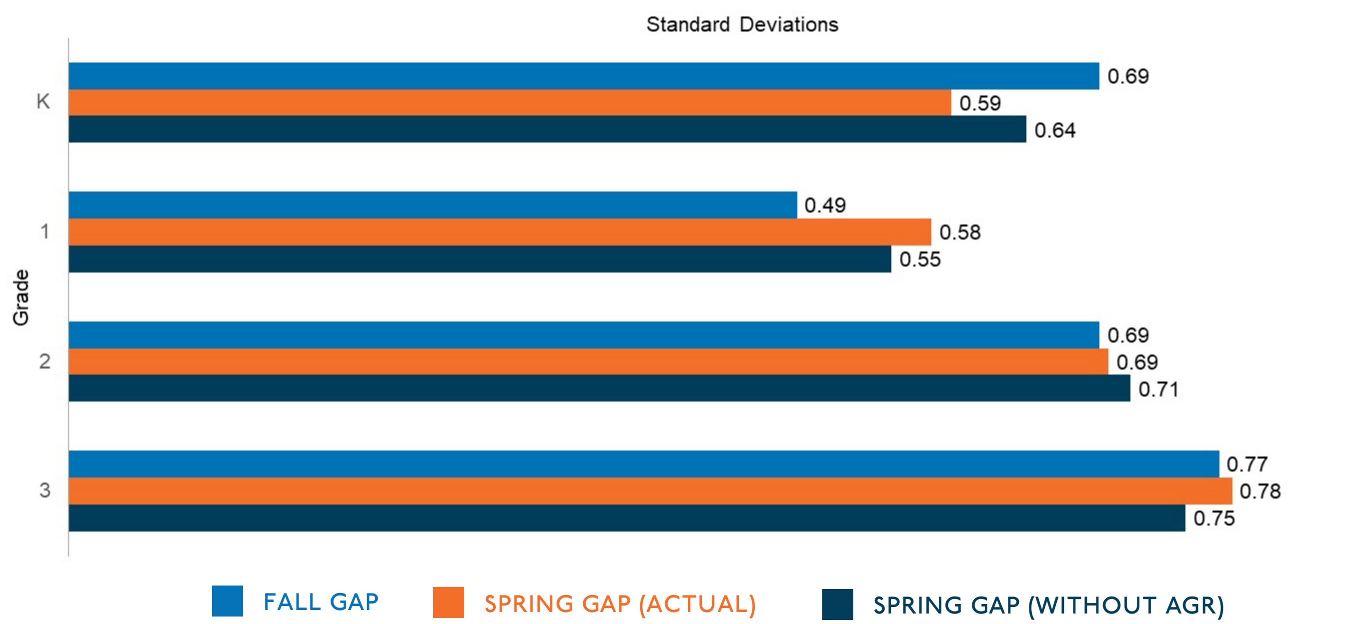
# Figure 8

2019 Statewide Math Gaps with and without AGR




# Figure 9

2019 Statewide Reading Gaps with and without AGR



Section 7

# School Board Report Findings

# School Board Report Findings

As part of participation in AGR, schools and districts agree to report to their boards on the strategies they implemented and their success in meeting the performance objectives listed in their AGR contracts. DPI provides a suggested reporting template that the majority of schools use.[9] The impact evaluation uses data from these school board reports, in conjunction with data from the End-of-Year Report, to determine the strategies that schools use in each grade and year. Due to reporting inconsistencies between schools, however, data from school board reports is less reliable and covers fewer schools than the End-of-Year Report. Below, we describe strategies and performance objectives data from the school board reports.

Table 33 below lists all possible combinations of the three strategy types–reduced class sizes, instructional coaching, and one-to-one tutoring–similar to strategies data from the End-of-Year Report (see Table 35). The breakdown of strategies is very similar to those provided from the End-of-Year Report. Schools most commonly reduce class sizes, although instructional coaching was used by over half the reporting schools. Schools are more likely to use combinations of strategies, with 17 percent using all three.

## Table 33

### 2018-19 School-level AGR Strategies, School Board Report Data

| STRATEGIES | % |
|---|---|
| Coaching Only | 16% |
| Class Size Only | 26% |
| Tutoring Only | 2% |
| Coaching and Class Size | 33% |
| Coaching and Tutoring | 3% |
| Class Size and Tutoring | 6% |
| All Three | 14% |

9 DPI's school board report template can be found at https://dpi.wi.gov/sites/default/files/imce/sage/doc/agr_performance_objectives_and_school_board_report_template.docx

Data on the types of performance objectives schools use appears in Table 34. We classify performance objectives into three primary types–achievement (e.g. bringing all students up to proficiency), growth (e.g. improving student scores by 10 points), and gap closing (e.g. improving scores for students receiving free and reduced-price lunch relative to other students' scores). Approximately 90 percent of schools set achievement objectives, and 30 percent set growth objectives. Very few set performance objectives to close school achievement gaps, although it should be noted that the AGR program offers supplemental school-based funding to close gaps across schools, not necessarily within schools.

# Table 34

## 2018-19 School-level AGR Performance Objectives, School Board Report Data

| TYPES OF PERFORMANCE OBJECTIVES | % |
|---|---|
| Achievement only | 64% |
| Growth only | 5% |
| Gap closing only | 2% |
| Achievement and growth | 24% |
| Achievement and gap closing | 2% |
| Growth and gap closing | 1% |
| All 3 goals | 3% |

Section 8

# End-of-Year Report Findings

# End-of-Year Report Findings

This section of the report provides the results from the 2018-19 End-of-Year Report survey of AGR schools. As exemplified earlier in the report, by 2018-19 schools that had transitioned from SAGE to AGR had taken advantage of AGR's increased flexibility. As shown in Table 35, only 13 percent of responding schools employed only the reduced class size strategy. A majority of schools opted for multiple strategies–65 percent used more than one strategy, including 16 percent that used all three strategies. Table 36 shows that reduced class size (83 percent of schools) and instructional coaching (72 percent of schools) were more common than one-to-one tutoring, which was used by only 30 percent of sample schools.

## Distribution of AGR Strategies Across Schools

### Table 35

2018-19 School-level AGR Strategies from End-of-Year Report

| STRATEGIES | % |
|---|---|
| Coaching Only | 13% |
| Class Size Only | 22% |
| Tutoring Only | <1% |
| Coaching and Class Size | 39% |
| Coaching and Tutoring | 4% |
| Class Size and Tutoring | 6% |
| All Three | 16% |

Note: 413 respondents.

# Table 36

## Percentage of AGR Schools Using Each Strategy

| STRATEGIES | % |
|---|---|
| Instructional Coaching | 83% |
| Reduced Class Size | 30% |
| One-to-one Tutoring | 72% |

Note: 413 respondents.

In general, schools chose to use the same strategies across classrooms within each grade (Table 37, Table 38, and Table 39). This trend was strongest for grades with reduced size classrooms. Within each grade, over 90 percent of schools chose to use reduced class sizes in at least three-quarters of classrooms or not at all (Table 37). For one-to-one tutoring (Table 38) and instructional coaching (Table 39), approximately 65 to 70 percent of schools used a strategy in at least three-quarters of the classrooms or not at all.

# Table 37

## Schools' Distributions of Classrooms with Reduced Class Size, by Grade

| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
|---|---|---|---|---|---|
| Kindergarten | 9% | 1% | 6% | 2% | 83% |
| First | 25% | 1% | 5% | 3% | 67% |
| Second | 27% | 1% | 3% | 4% | 65% |
| Third | 31% | 1% | 4% | 3% | 62% |

Note: 342 respondents. Categories are mutually exclusive.

# Table 38

## Schools' Distribution of Classrooms with One-to-One Tutoring, by Grade

| | PERCENT OF CLASSROOMS | | | | |
|---|---|---|---|---|---|
| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
| Kindergarten | 10% | 17% | 6% | 7% | 59% |
| First | 10% | 16% | 6% | 7% | 61% |
| Second | 12% | 15% | 8% | 8% | 57% |
| Third | 10% | 18% | 6% | 8% | 57% |

Note: 109 respondents. Categories are mutually exclusive.

# Table 39

## Schools' Distribution of Classrooms with Instructional Coaching, by Grade

| | PERCENT OF CLASSROOMS | | | | |
|---|---|---|---|---|---|
| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
| Kindergarten | 5% | 13% | 10% | 10% | 62% |
| First | 5% | 6% | 10% | 16% | 63% |
| Second | 5% | 7% | 8% | 17% | 62% |
| Third | 6% | 6% | 8% | 15% | 65% |

Note: 297 respondents. Categories are mutually exclusive.

## Reduced Class Size Instructional Strategies and Benefits

Schools using reduced class sizes reported using a variety of instructional strategies (Table 40), including small group instruction (97 percent), one-on-one time with the teacher (79 percent), differentiation of instruction (89 percent), strategic placement of students in groups (75 percent), and, to a lesser extent, strategic placement of students in classrooms (52 percent). Only two percent of responding schools reported that they use no additional instructional strategies due to AGR class size reductions.

## Table 40

Instructional Strategies Associated with AGR Reduced Class Size

| STRATEGIES | % |
|---|---|
| Small-group instruction | 97% |
| Differentiation of instruction | 89% |
| One-on-one time with the teacher | 79% |
| Strategic placement of students in groups | 75% |
| Strategic placement of students in classrooms | 52% |
| We don't use any specific instructional strategies because of smaller class sizes | 2% |
| Other | 5% |
| Not sure/don't know | 0% |

Note: 342 respondents.

The survey included an open-ended response item for perceived benefits reduced class size provided. Over 300 schools responded to the item, and all responses were positive. Responses to this item included the following categories, in order of prevalence:

1. Changes in teacher behavior
2. Changes in student behavior and performance
3. Changes in classroom organization
4. Changes in family involvement

Respondents described changes in teacher behavior more than the other three categories combined. These changes most commonly included individually-targeted instruction, student assessment/performance monitoring, increased quality of teacher-student relationships, and increased interactions with coaches. One representative response reads,

> "Teachers have less data to analyze and can get students into smaller groups for differentiated instruction. Teachers are able to more frequently progress monitor students and then adjust their instruction to meet the needs of the students. Teachers are able to more frequently hold one-on-one conferences with students to provide feedback and lift their thinking."

## One-to-one Tutoring Frequency, Practices, and Benefits

Schools using one-to-one tutoring did so frequently, as seen in Table 41. Eighty-four percent offered tutoring at least weekly, and 62 percent engaged in tutoring 3 times a week or more.

# Table 41

### Frequency of AGR One-to-One Tutoring

| FREQUENCY | % |
|---|---|
| 3 times a week or more | 62% |
| 2 times a week | 11% |
| Weekly | 11% |
| Biweekly | 2% |
| Monthly | 0% |
| Other | 0% |
| As needed | 14% |

Note: 109 respondents to survey item.

Table 42 shows that schools reported using almost all of the tutoring practices listed on the survey, although an exception was maintaining a focus on equity (50 percent).

# Table 42

## AGR One-to-One Tutoring Practices

| PRACTICE | % |
|---|---|
| Reviews student data | 84% |
| Communicates regularly with classroom teacher | 83% |
| Provides scaffolding | 82% |
| Adapts to student learning styles | 79% |
| Models appropriate learning behavior | 78% |
| Maintains a focus on equity | 50% |
| Other | 6% |
| Not sure/don't know | 0% |

Note: 109 respondents to survey item.

The survey incorporated an open-response item for benefits one-to-one tutoring provides. Approximately one-third of the 100 responses to this item included both general descriptions of benefits, (e.g. "Instruction to support students' skill deficits") while two-thirds of respondents answered with more specific descriptions of benefits. Most commonly, respondents described using data to identify student learning needs that tutors attempt to address. For example,

> "In the fall, winter, and spring students take district assessments. After each assessment a group will sit down using data ... and then target students for interventions on individual skills they may be lacking. The one to one tutoring gives us the ability to then have a teacher focus on what the individual student is lacking in skills."

## Instructional Coaching Frequency, Practices, and Coach Characteristics

Table 43 shows that at 72 percent of schools, instructional coaches met with teachers at least monthly, and 52 percent met weekly.

# Table 43

### Frequency of AGR Instructional Coaches Meeting with Teachers

| FREQUENCY | % |
|---|---|
| Weekly | 52% |
| Monthly | 20% |
| Quarterly | 4% |
| Each semester | 1% |
| Other | 8% |
| As needed | 13% |
| Not sure/don't know | 1% |

Note: 297 respondents to survey item.

Schools responded affirmatively to almost all listed coaching practices, although less than half indicated that their coaches kept coaching logs or maintained a focus on equity (Table 44).

# Table 44

## Instructional Coaching Practices

| PRACTICES | % |
|---|---|
| Encourages reflective practices | 85% |
| Discusses data with teachers | 85% |
| One-to-one teacher coaching | 81% |
| Observes teacher practices | 71% |
| Team teacher coaching | 65% |
| Advises teachers to set goals | 61% |
| Coaching focuses on teacher goals | 61% |
| Maintains a focus on equity | 46% |
| Keeps a coaching log | 45% |
| Other | 3% |
| Not sure/don't know | 1% |

Note: 297 respondents to survey item.

Schools were able to successfully find trained, experienced coaches with content specialization, as seen from Table 45.

# Table 45

## Characteristics of AGR Instructional Coaches

| FREQUENCY | % |
|---|---|
| Coach training | 84% |
| Content specialist in their subject of coaching | 66% |
| Previous instructional coaching experience | 61% |
| Other | 5% |
| Not sure/don't know | 1% |

Note: 297 respondents to survey item.

The survey also asked an open-ended question regarding what benefits instructional coaching provides. The 277 responses included the following categories:

- Teacher professional development and growth (N = 85)
- Descriptions of coaches' duties (e.g. meetings with teachers, classroom observations, analyzing data) (N = 85)
- Improved communication/planning, collaboration, data use among teachers (N=55)
- Description of impacts of instructional coaching on student performance and behavior (N=41)
- Other (N=11)

Section 9

# Summary/ Conclusions

# Summary/Conclusions

This report provides evidence regarding program impacts on math and reading growth, student attendance, and out-of-school suspensions. These results are presented at the state level and disaggregated by grade and student demographic characteristics. The report also contains data on the AGR strategies schools have implemented, the intensity of strategy use, and preliminary evidence on the relative effectiveness of combinations of strategies.

From 2015-16 to 2018-19, AGR impacts on achievement are limited to strong kindergarten reading growth statewide. Attending an AGR school is associated with a moderate (0.11 standard deviation) increase in PALS growth from fall to spring, relative to a comparison group of students from observably similar schools. Students eligible for free/reduced-price lunch in AGR schools experienced moderate growth (0.13 standard deviations) greater than that of similar students in non-AGR schools. Impacts on PALS growth was large for Hispanic students (0.25 standard deviations), English learner students (0.37 standard deviations), urban students (0.21 standard deviations), and Asian students (0.24 standard deviations). Math and reading MAP and STAR growth in Grades 1-3 was near zero and insignificant. Although estimates of impacts on math and reading ranged across subgroups, none were significantly different from zero.

The report also estimated impacts for non-testing outcomes. We found some evidence that AGR is associated with fewer out-of-school suspensions. Although suspensions are rare events in Grades K-3, Hispanic students in AGR schools were 0.6 percentage points less likely to receive a suspension relative to similar students in comparable non-AGR schools. Similarly, English learner students at AGR schools were 0.5 percentage points less likely to be suspended. We found very few statistically significant impacts on attendance. Most point estimates of attendance impacts showed decreases in attendance at AGR schools, although these decreases were too small to be significant to state policy.

Looking at the AGR strategies that schools chose to implement, we found that most AGR schools took advantage of the program's flexibility and chose to implement instructional coaching and one-to-one tutoring strategies that were not included in the state's previous SAGE policy. Over 60 percent of schools combined 2 or more strategies.

As in any observational study, this evaluation has several limitations. The PSM methodology matches schools on observable characteristics, but comparison schools may not match AGR schools on unobserved characteristics such as schools' ability to properly implement AGR or instructor quality in the local hiring market. The long history of SAGE, AGR's precursor program that provided funding for reduced class sizes only, also limits the study. Previous school outcomes used for matching likely include SAGE impacts as well, which would bias AGR impacts toward zero. Finally, inconsistent testing patterns in Grades K-3 restricted the sample of AGR and non-AGR schools included in the growth analysis samples, potentially limiting how growth impact estimates can be generalized to schools not in the sample.

## Section 10

# Technical Appendix

# Technical Appendix

In order to credibly estimate AGR impacts, we must address two primary challenges to identification. First, a plausible comparison or control group must be identified. Schools that receive AGR funding are different from schools statewide (see AGR Demographics above) because those selected for SAGE, and subsequently eligible for AGR, were required to meet certain thresholds of students receiving free and reduced-price lunches. Second, because all AGR schools previously participated in SAGE, total AGR impacts cannot be determined solely through changes over time in AGR schools' outcomes. In most evaluations, schools participating in a program (the treatment) are previously untreated, meaning that, under certain conditions, comparing pre-treatment and post-treatment outcomes results in plausible estimates of the treatment impact. For AGR, however, comparing pre- and post-treatment outcomes only provides estimates of the difference between the AGR and SAGE treatment impacts, not the AGR impact itself.

To find a plausible control group and identify the AGR impact, we use Propensity Score Matching (PSM). PSM addresses selection bias by choosing a control group with observable characteristics similar to those of the treatment group. As described above (see AGR Demographics), schools that receive AGR funding are observably different than other Wisconsin schools. This is because AGR targets funding to schools with higher percentages of students eligible for free or reduced price lunch. Coincident with being located in higher poverty environments, relative to their non-AGR counterparts AGR schools have lower pre-program (2013) average test scores and attendance, and higher numbers of suspensions. As a result, naive comparisons of outcomes across non-AGR and AGR schools would find negative program impacts based only on program selection. To address this selection bias, PSM identifies Wisconsin schools that are observably similar to AGR schools in order to create an apples-to-apples comparison when estimating program impacts. Successful matching relies on both the quality of matches and overlap (or common support) of propensity scores between AGR and non-AGR schools.

Comparison schools with high percentages of students eligible for free or reduced-price lunch are not AGR participants for two primary reasons. First, poverty in those schools may have increased since the last SAGE eligibility period. Those schools would be eligible for AGR based on poverty thresholds but are ineligible because they did not participate in SAGE. To test this potential source of bias, we include school-specific time trends in robustness checks below. Impact estimates from these analyses are similar to those from our preferred models. Second, schools may have opted out of SAGE. Opt-out schools would be systematically different from AGR schools due to characteristics of the district or school. Although we cannot test for bias resulting from selection bias associated with opting into or out of SAGE, the final round of SAGE enrollment occurred in 2011-12, and many school and district characteristics, particularly those associated with administration, have since changed.

Despite limitations of the PSM regarding unobserved characteristics, it represents the best available methodology given program rollout and available data. Below, we describe the choices of variables to include in the matching model, the overlap in propensity scores between AGR and non-AGR schools, and the covariate balance among the matched sample. In addition, we present multiple robustness checks to provide evidence of whether unobserved school characteristics might bias AGR impact estimates. The primary limitation of PSM is that it rests on the strong assumption that balancing AGR and non-AGR schools on observed characteristics also balances those schools on unobserved characteristics. The most typical method of addressing bias from fixed, unobserved characteristics would be to include school fixed effects in the estimation. For the AGR analysis, however, including school fixed effects would only allow comparisons of AGR to SAGE because all AGR schools previously participated in SAGE. The included robustness checks compare the reports main results to the results of various impact models with partial controls for unobserved school characteristics.

Finally, we present results from the Benjamini-Hochberg correction for multiple comparisons. These corrections adjust the p-values from impact estimates to account for the increased probability of finding statistically significant results due to the large number of models included in the report.

## Propensity Score Matching

We estimated the probability of a school receiving AGR with the logit model of treatment shown below. The probability that a school participates in AGR, Pr(EverAGRs), is a function of an intercept term α, a vector of school-level covariates Xs, and a school-specific error term εs.

$$ln\left[\frac{Pr(EverAGR_s)}{1-Pr(EverAGR_s)}\right] = \alpha + \beta X_s + \varepsilon_s \qquad (1)$$

In the equation above, matching occurs at the school-level (defined by the grades included in the model, not necessarily all of the grades that a school contains) because AGR is a school-level treatment.[10] We use this matching strategy for both the attendance and discipline models. For the models of test score outcomes, however, we match at the school-grade-year level due to inconsistent testing coverage both across and within schools. As described in Table 5 through Table 7, during the 2015-16 through 2018-19 sample period, only a minority of schools used the PALS, MAP, and STAR tests.[11] Underlying Table 5 through Table 7 is even greater variation both across and within schools. Many schools began a new test and/or quit using a test in the middle of the sample period. Other schools tested some of Grades K-3 but not others, and yet others changed which grades they tested during the sample period. Due to this variation, it is not possible to build a sufficiently sized, consistent sample while matching at the school level. To provide DPI with the most complete and generalizable evaluation of AGR impacts, we prioritized the inclusion of as many AGR schools as possible. As a result, we chose to match all models of test outcomes at the school-grade-year level.[12] For these matches, we use school-year averages of demographic and academic characteristics due to instability in school-grade-year level averages, particularly in small schools, but match within school-grade-year to ensure that matches only occur between schools and grades that were tested in the same year.

### Specifying the Propensity Score Model

To determine which variables to include in the propensity score matching model above, we tested the influence of many demographic and academic variables. The final list of covariates appears in Table 1.

For each of the models, the most important matching variables measure the average outcome in a previous time period (pretests), such as the school's average test scores from the previous time period. The choice of pretest was complicated by both the level of matching (school or school-grade-year) and by the fact that AGR schools previously participated in SAGE. To the greatest extent possible, we aimed to remove previous program impacts from the matching model. Matching schools on post-program data risks biasing the results toward zero, because schools would be matched on previous-period outcomes that already include the treatment impact. However, at the beginning of our sample period, SAGE had been in operation for over 15 years, so it was not possible to include pre-program data. We used two strategies to address matching on post-program outcomes. For the attendance and discipline models, we matched once using school average attendance rate and suspension data from 2012-13, limiting the effect of including a post program outcome to just one year. For the PALS and MAP/STAR testing models, we focused on growth instead of achievement. Focusing on growth lessens the impact of previous test scores, because, with appropriate pretest controls in the analysis model, the potential for growth is roughly equal regardless of initial pretest score.

In order to find the best PSM model to balance covariates across AGR and comparison schools, retaining as many school observations as possible, and stability of matches, we tested different matching algorithms, including caliper matching with various bandwidths, kernel matching, and Mahalanobis. For the analysis in the report, we used a kernel matching procedure that places higher weights on control observations nearest to a treatment observation and places successively lower weights on control observations as their distance from a treatment observation

10 Stuart, E. (2007). Estimating causal effects using school-level datasets. *Educational Researcher*, 36, 187-198.

11 Schools administered dozens of different types of tests across all grades. PALS, MAP, and STAR were the most common.

12 We tested models that limit the sample to schools that tested throughout 2012-13 to 2017-18, but these models omitted most AGR schools.

increases.[13]

Prior to matching we limited the sample using two additional rules. First, we removed any schools that had participated in SAGE but never participated in AGR, including those that declined to participate in AGR. Second, we limited the testing models to schools that tested at least 75 percent of the relevant population in Grades K-3, following previous SAGE evaluations.[14] Table 46 illustrates the matching and subsequent analysis strategies for each outcome.

# Table 46

## Matching and Analysis Strategies

| OUTCOME | GRADES | MATCHING LEVEL | MATCHING DATA | ANALYSIS YEARS |
|---|---|---|---|---|
| PALS Growth | K | School-grade-year | Fall 2012-13 through Fall 2018-19 | 2012-13 through 2018-19 |
| MAP/STAR Reading Growth | 1-3 | School-grade-year | Fall 2012-13 through Fall 2018-19 | 2012-13 through 2018-19 |
| MAP/STAR Math Growth | 1-3 | School-grade-year | Fall 2012-13 through Fall 2018-19 | 2012-13 through 2018-19 |
| Absence Rate | K-3 | School | 2012-13 | 2013-14 through 2018-19 |
| Suspension Rate | K-3 | School | 2012-13 | 2013-14 through 2018-19 |

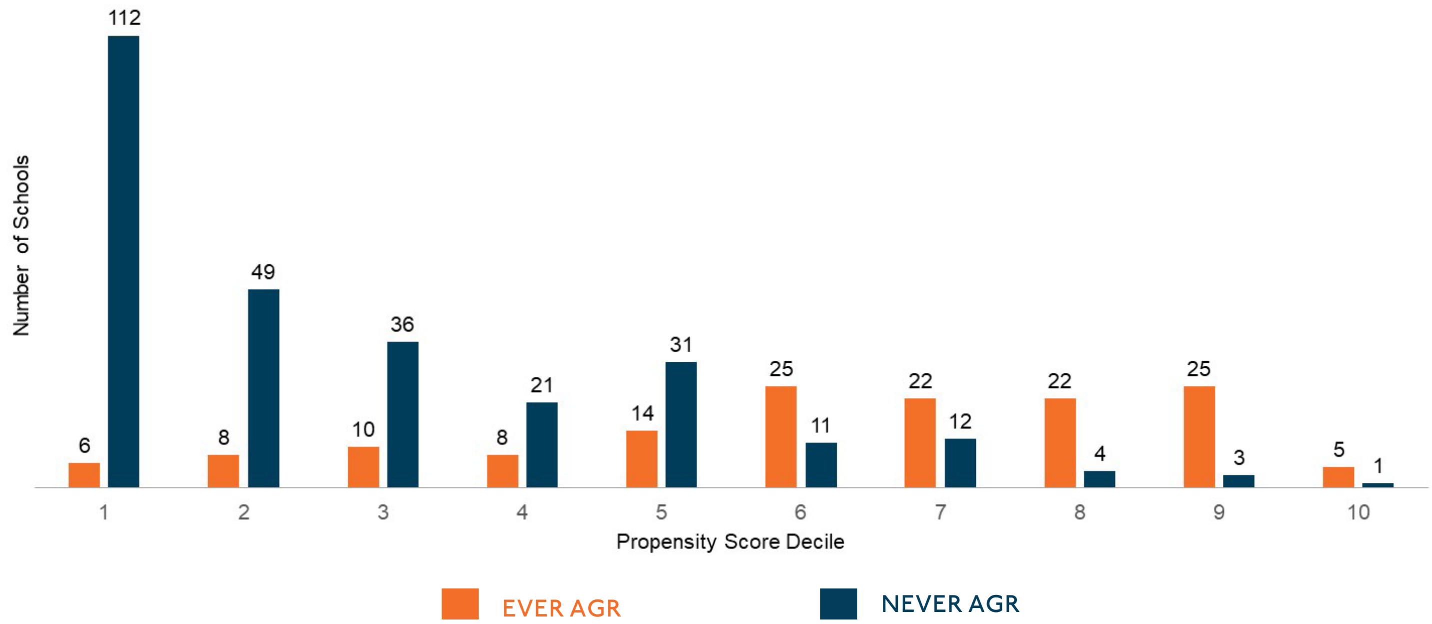
When matching is successful, there is sufficient overlap in the propensity scores of treated (AGR) and comparison (non-AGR) schools to ensure that there is a plausible control group for the analysis. Figure 10 through Figure 13 display common support for PALS, Math, Reading, and attendance and discipline (which were matched together), respectively. Each figure shows the number of AGR and non-AGR schools by deciles of the propensity score distributions. For each of the outcomes, there are substantial numbers of non-AGR schools in most propensity score deciles and at least one control school in every decile.

13 Specifically, we used Stata's kmatch package with an Epanechnikov kernel and allowed Stata to select the optimal bandwidth.

14 The 75 percent threshold helps to ensure that students were tested for benchmarking purposes and not because they had been singled out for testing or had tested at another school before moving. See Meyer, R., Dokumaci, E., Sim, G., Steele, C., Suchor, K., & Vadas, J. (2015). *SAGE program evaluation final report*. University of Wisconsin-Madison, Value-Added Research Center.

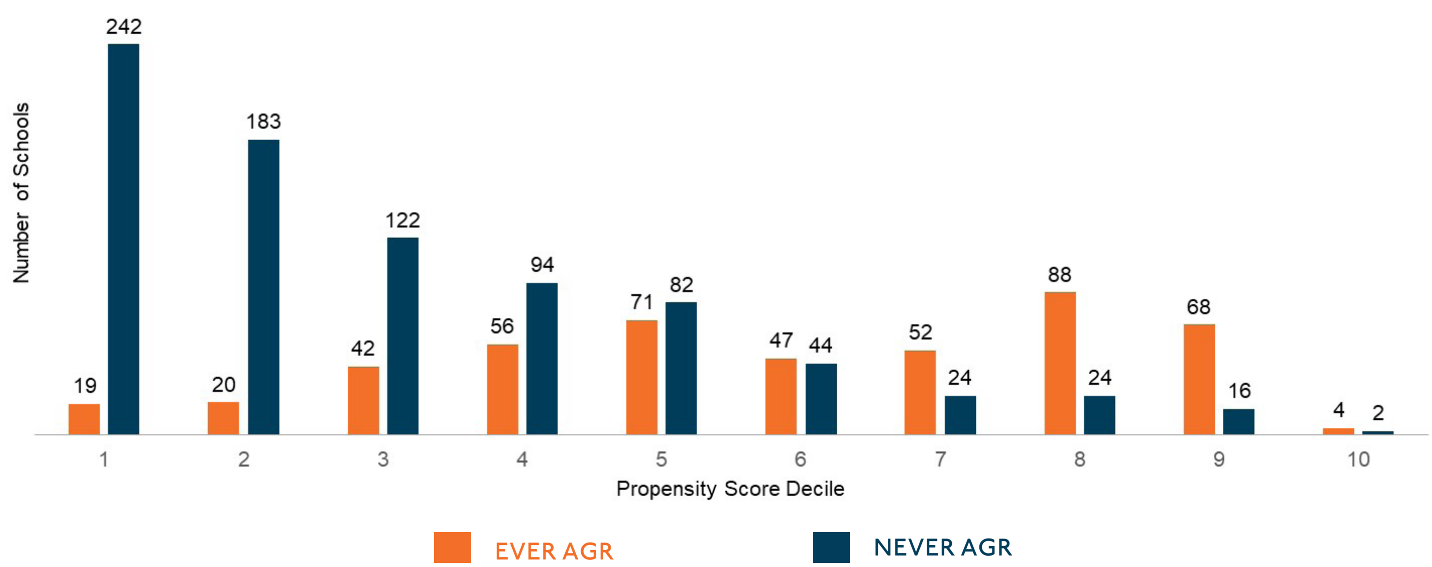
# Figure 10

Common Support for Matching - PALS Reading (2018-19)



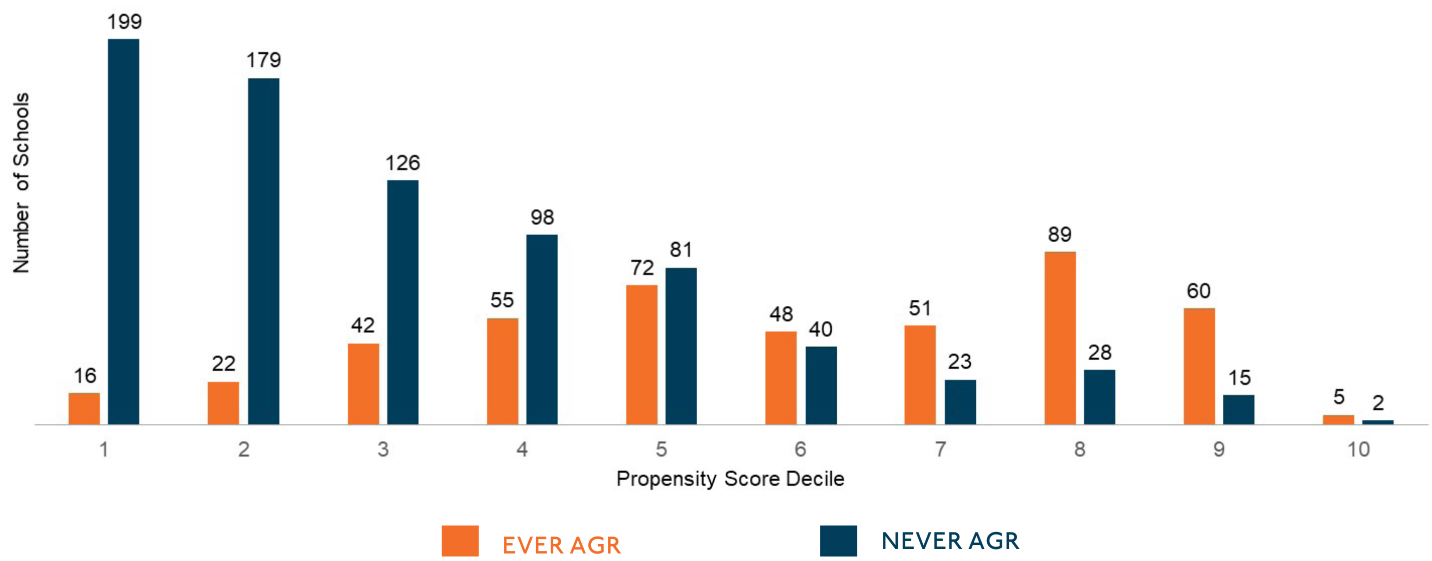

# Figure 11

Common Support for Matching - MAP/STAR Math (2018-19)



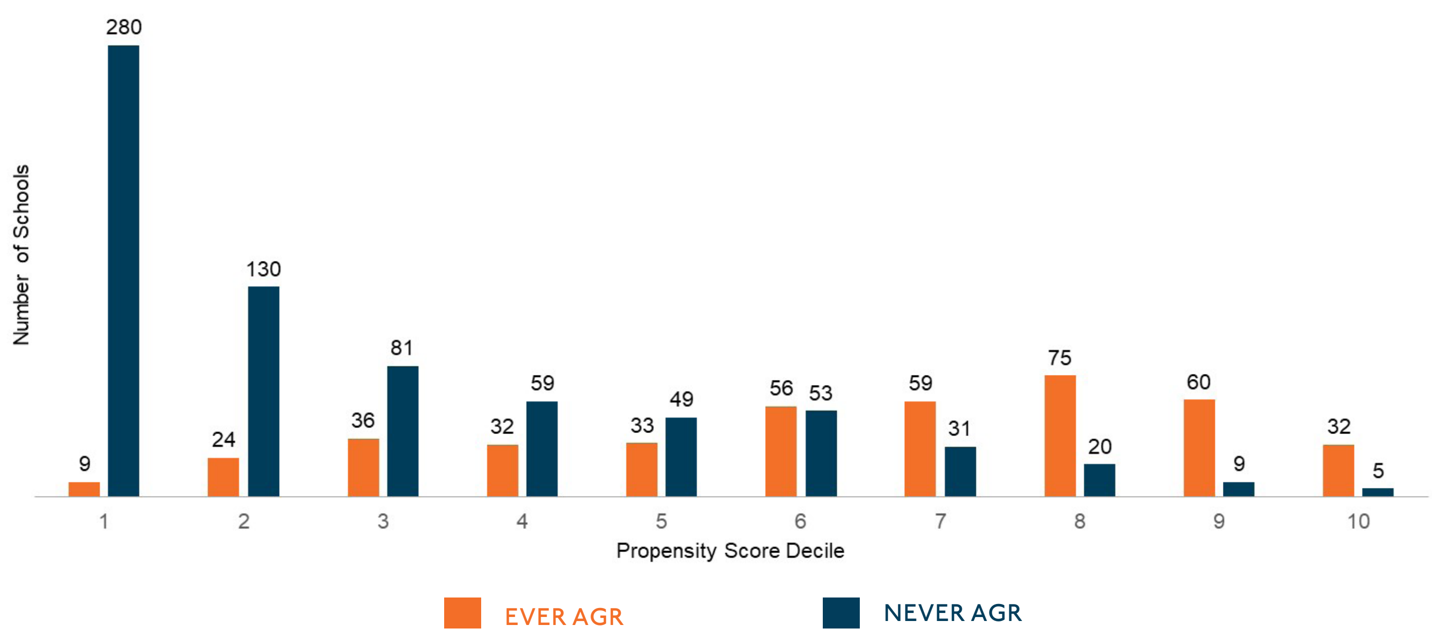
# Figure 12

Common Support for Matching - MAP/STAR Reading (2018-19)




# Figure 13

Common Support for Matching - Attendance and Discipline (2013-14)



Successful matching should also result in balanced covariates across the treatment and control groups. Table 47 through Table 50 describe student-level balance for each of the matched samples. In keeping with the recommendations of the What Works Clearinghouse (WWC), we assess equivalence using both the p-values from t-tests of differences in means, and with standardized differences. The WWC specifies that standardized differences[15] over 0.25 are signals of imbalance, and those between 0.05 and 0.25 require that the covariates be included as covariates in the impact analysis. In Table 47 through Table 50, no standardized differences reach the 0.25 threshold, and we include all covariates in all impact analyses for double robustness.

# Table 47

## Balance of Matched Sample - PALS

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| N | 98,384 | 99,454 | | |
| Fall PALS Score | -0.17 | -0.14 | 0.00 | 0.03 |
| Std. Dev. | 1.03 | 1.04 | | |
| School Fall PALS Score | -0.18 | -0.15 | 0.00 | 0.08 |
| Std. Dev. | 0.39 | 0.46 | | |
| Female | 0.49 | 0.48 | 0.26 | 0.01 |
| Std. Dev. | 0.50 | 0.50 | | |
| Black | 0.14 | 0.16 | 0.00 | 0.06 |
| Std. Dev. | 0.35 | 0.37 | | |
| Hispanic | 0.15 | 0.14 | 0.00 | 0.02 |
| Std. Dev. | 0.35 | 0.35 | | |
| Other Race | 0.10 | 0.11 | 0.00 | 0.03 |
| Std. Dev. | 0.30 | 0.31 | | |
| FRL | 0.62 | 0.62 | 0.59 | 0.00 |
| Std. Dev. | 0.49 | 0.49 | | |
| SPED | 0.14 | 0.14 | 0.57 | 0.00 |
| Std. Dev. | 0.35 | 0.35 | | |
| ELL | 0.11 | 0.11 | 0.77 | 0.00 |
| Std. Dev. | 0.31 | 0.31 | | |
| Urban | 0.39 | 0.44 | 0.00 | 0.10 |
| Std. Dev. | 0.49 | 0.50 | | |

15 What Works Clearinghouse. (2020). Standards Handbook (Version 4.1). Retrieved from https://ies.ed.gov/ncee/wwc/handbooks

# Table 47, continued

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| Suburb | 0.09 | 0.11 | 0.00 | 0.05 |
| Std. Dev. | 0.29 | 0.31 | | |
| Town | 0.18 | 0.18 | 0.04 | 0.01 |
| Std. Dev. | 0.39 | 0.38 | | |
| School Population | 244.95 | 244.87 | 0.89 | 0.00 |
| Std. Dev. | 102.49 | 101.69 | | |
| School Avg Teacher Salary | 46,878.27 | 45,627.94 | 0.00 | 0.12 |
| Std. Dev. | 8,178.04 | 13,047.58 | | |
| School % Female | 0.48 | 0.48 | 0.00 | 0.02 |
| Std. Dev. | 0.04 | 0.04 | | |
| School % Black | 0.15 | 0.17 | 0.00 | 0.09 |
| Std. Dev. | 0.26 | 0.28 | | |
| School % Hispanic | 0.15 | 0.14 | 0.00 | 0.04 |
| Std. Dev. | 0.20 | 0.17 | | |
| School % Other Race | 0.10 | 0.11 | 0.00 | 0.07 |
| Std. Dev. | 0.12 | 0.16 | | |
| School % FRL | 0.63 | 0.62 | 0.21 | 0.01 |
| Std. Dev. | 0.20 | 0.23 | | |
| School % SPED | 0.15 | 0.15 | 0.00 | 0.04 |
| Std. Dev. | 0.05 | 0.05 | | |
| School % ELL | 0.11 | 0.11 | 0.00 | 0.03 |
| Std. Dev. | 0.15 | 0.15 | | |

# Table 48

## Balance of Matched Sample - MAP/STAR Math

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| N | 151,916 | 153,886 | | |
| Fall Math Score | -0.04 | -0.05 | 0.00 | 0.01 |
| Std. Dev. | 1.02 | 1.03 | | |
| Fall Reading Score | -0.19 | -0.21 | 0.00 | 0.02 |
| Std. Dev. | 1.06 | 1.07 | | |
| School Fall Math Score | -0.05 | -0.06 | 0.00 | 0.03 |
| Std. Dev. | 0.39 | 0.43 | | |
| School Fall Read Score | -0.19 | -0.20 | 0.00 | 0.04 |
| Std. Dev. | 0.38 | 0.42 | | |
| Female | 0.49 | 0.49 | 0.87 | 0.00 |
| Std. Dev. | 0.50 | 0.50 | | |
| Black | 0.23 | 0.22 | 0.00 | 0.04 |
| Std. Dev. | 0.42 | 0.41 | | |
| Hispanic | 0.14 | 0.14 | 0.06 | 0.01 |
| Std. Dev. | 0.35 | 0.35 | | |
| Other Race | 0.10 | 0.10 | 0.00 | 0.03 |
| Std. Dev. | 0.29 | 0.30 | | |
| FRL | 0.65 | 0.64 | 0.00 | 0.03 |
| Std. Dev. | 0.48 | 0.48 | | |
| SPED | 0.15 | 0.15 | 0.65 | 0.00 |
| Std. Dev. | 0.35 | 0.36 | | |
| ELL | 0.10 | 0.11 | 0.00 | 0.03 |
| Std. Dev. | 0.30 | 0.31 | | |
| Urban | 0.52 | 0.54 | 0.00 | 0.03 |
| Std. Dev. | 0.50 | 0.50 | | |
| Suburb | 0.11 | 0.10 | 0.00 | 0.02 |
| Std. Dev. | 0.31 | 0.30 | | |
| Town | 0.17 | 0.15 | 0.00 | 0.06 |
| Std. Dev. | 0.37 | 0.35 | | |
| School Population | 235.66 | 233.09 | 0.00 | 0.03 |
| Std. Dev. | 93.84 | 92.81 | | |

# Table 48, continued

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| School Avg Teacher Salary | 48,833.65 | 48,461.55 | 0.00 | 0.04 |
| Std. Dev. | 8,436.91 | 10,914.85 | | |
| School % Female | 0.48 | 0.48 | 0.00 | 0.03 |
| Std. Dev. | 0.04 | 0.04 | | |
| School % Black | 0.23 | 0.22 | 0.00 | 0.05 |
| Std. Dev. | 0.33 | 0.31 | | |
| School % Hispanic | 0.15 | 0.14 | 0.00 | 0.02 |
| Std. Dev. | 0.18 | 0.16 | | |
| School % Other Race | 0.10 | 0.11 | 0.00 | 0.07 |
| Std. Dev. | 0.11 | 0.14 | | |
| School % FRL | 0.66 | 0.64 | 0.00 | 0.07 |
| Std. Dev. | 0.22 | 0.23 | | |
| School % SPED | 0.16 | 0.15 | 0.00 | 0.04 |
| Std. Dev. | 0.05 | 0.05 | | |
| School % ELL | 0.10 | 0.11 | 0.00 | 0.07 |
| Std. Dev. | 0.14 | 0.14 | | |

# Table 49

## Balance of Matched Sample - MAP/STAR Reading

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| N | 151,873 | 153,508 | | |
| Fall Math Score | -0.04 | -0.05 | 0.00 | 0.01 |
| Std. Dev. | 1.02 | 1.03 | | |
| Fall Reading Score | -0.19 | -0.21 | 0.00 | 0.02 |
| Std. Dev. | 1.06 | 1.07 | | |
| School Fall Math Score | -0.05 | -0.06 | 0.00 | 0.02 |
| Std. Dev. | 0.39 | 0.43 | | |
| School Fall Read Score | -0.19 | -0.20 | 0.00 | 0.04 |
| Std. Dev. | 0.38 | 0.43 | | |
| Female | 0.49 | 0.49 | 0.58 | 0.00 |
| Std. Dev. | 0.50 | 0.50 | | |
| Black | 0.23 | 0.22 | 0.00 | 0.04 |
| Std. Dev. | 0.42 | 0.41 | | |
| Hispanic | 0.14 | 0.14 | 0.07 | 0.01 |
| Std. Dev. | 0.35 | 0.35 | | |
| Other Race | 0.10 | 0.10 | 0.00 | 0.02 |
| Std. Dev. | 0.29 | 0.30 | | |
| FRL | 0.65 | 0.64 | 0.00 | 0.03 |
| Std. Dev. | 0.48 | 0.48 | | |
| SPED | 0.15 | 0.15 | 0.82 | 0.00 |
| Std. Dev. | 0.35 | 0.35 | | |
| ELL | 0.10 | 0.11 | 0.00 | 0.03 |
| Std. Dev. | 0.30 | 0.31 | | |
| Urban | 0.52 | 0.54 | 0.00 | 0.02 |
| Std. Dev. | 0.50 | 0.50 | | |
| Suburb | 0.11 | 0.10 | 0.00 | 0.02 |
| Std. Dev. | 0.31 | 0.30 | | |
| Town | 0.17 | 0.15 | 0.00 | 0.05 |
| Std. Dev. | 0.37 | 0.35 | | |
| School Population | 236.01 | 233.12 | 0.00 | 0.03 |
| Std. Dev. | 94.01 | 92.50 | | |

# Table 49, continued

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| School Avg Teacher Salary | 48,798.32 | 48,449.39 | 0.00 | 0.04 |
| Std. Dev. | 8,483.33 | 10,841.03 | | |
| School % Female | 0.48 | 0.48 | 0.00 | 0.03 |
| Std. Dev. | 0.04 | 0.04 | | |
| School % Black | 0.23 | 0.22 | 0.00 | 0.05 |
| Std. Dev. | 0.33 | 0.31 | | |
| School % Hispanic | 0.15 | 0.14 | 0.00 | 0.02 |
| Std. Dev. | 0.18 | 0.16 | | |
| School % Other Race | 0.10 | 0.11 | 0.00 | 0.06 |
| Std. Dev. | 0.11 | 0.14 | | |
| School % FRL | 0.66 | 0.64 | 0.00 | 0.08 |
| Std. Dev. | 0.22 | 0.22 | | |
| School % SPED | 0.16 | 0.15 | 0.00 | 0.04 |
| Std. Dev. | 0.05 | 0.05 | | |
| School % ELL | 0.10 | 0.11 | 0.00 | 0.06 |
| Std. Dev. | 0.14 | 0.14 | | |

# Table 50

## Balance of Matched Sample - Attendance & Discipline

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| N | 476,753 | 446,445 | | |
| School Attendance Rate 2012-13 | 0.95 | 0.95 | 0.00 | 0.07 |
| Std. Dev. | 0.02 | 0.02 | | |
| School Suspension Rate 2012-13 | 0.03 | 0.03 | 0.00 | 0.02 |
| Std. Dev. | 0.05 | 0.05 | | |
| Female | 0.48 | 0.48 | 0.43 | 0.00 |
| Std. Dev. | 0.50 | 0.50 | | |
| Black | 0.15 | 0.16 | 0.00 | 0.03 |
| Std. Dev. | 0.35 | 0.36 | | |
| Hispanic | 0.16 | 0.15 | 0.00 | 0.03 |
| Std. Dev. | 0.36 | 0.35 | | |
| Other Race | 0.10 | 0.10 | 0.00 | 0.02 |
| Std. Dev. | 0.31 | 0.30 | | |
| FRL | 0.62 | 0.59 | 0.00 | 0.07 |
| Std. Dev. | 0.48 | 0.49 | | |
| SPED | 0.15 | 0.15 | 0.00 | 0.01 |
| Std. Dev. | 0.36 | 0.36 | | |
| ELL | 0.12 | 0.12 | 0.00 | 0.01 |
| Std. Dev. | 0.32 | 0.32 | | |
| Urban | 0.41 | 0.41 | 0.00 | 0.02 |
| Std. Dev. | 0.49 | 0.49 | | |
| Suburb | 0.08 | 0.13 | 0.00 | 0.16 |
| Std. Dev. | 0.28 | 0.34 | | |
| Town | 0.18 | 0.21 | 0.00 | 0.06 |
| Std. Dev. | 0.39 | 0.40 | | |
| School Population | 246.21 | 247.70 | 0.00 | 0.02 |
| Std. Dev. | 98.23 | 99.05 | | |
| School Avg Teacher Salary | 47,261.52 | 47,399.40 | 0.00 | 0.02 |
| Std. Dev. | 8,039.82 | 10,044.05 | | |
| School % Female | 0.48 | 0.48 | 0.00 | 0.01 |
| Std. Dev. | 0.04 | 0.04 | | |

# Table 50, continued

| | AGR | NON-AGR | P-VALUE (T/C DIFFERENCE) | EFFECT SIZE |
|---|---|---|---|---|
| School % Black | 0.15 | 0.16 | 0.00 | 0.05 |
| Std. Dev. | 0.26 | 0.27 | | |
| School % Hispanic | 0.15 | 0.14 | 0.00 | 0.06 |
| Std. Dev. | 0.20 | 0.17 | | |
| School % Other Race | 0.09 | 0.08 | 0.00 | 0.10 |
| Std. Dev. | 0.12 | 0.08 | | |
| School % FRL | 0.64 | 0.61 | 0.00 | 0.12 |
| Std. Dev. | 0.20 | 0.22 | | |
| School % SPED | 0.15 | 0.15 | 0.00 | 0.10 |
| Std. Dev. | 0.05 | 0.05 | | |
| School % ELL | 0.11 | 0.11 | 0.04 | 0.01 |
| Std. Dev. | 0.16 | 0.14 | | |

## Impact Analysis

After matching, we model impact estimates for both SAGE and AGR using the following, student-level specification:

(2) $$Y_{isgy} = \alpha + \gamma_0 SAGE_{sy} + \gamma_1 AGR_{sy} + \beta X_{iy} + \pi Z_{sy} + \partial_{gy} + \varepsilon_{isgy}$$

where $Y_{isgy}$ is an outcome for student *i* in grade *g*, school *s*, and year *y*. $SAGE_{sy}$ and $AGR_{sy}$ are indicators for whether a school received SAGE or AGR funding, respectively, in each year. $X_{iy}$ represents a vector of student-level covariates, including lagged values of the outcome Y, and $Z_{sy}$ represents a vector of school-level covariates. Grade-by year fixed effects, $\delta_{gy}$, are included to control for any unobserved, statewide effects that vary by grade and/or time.[16] All analysis variables are described in Table 2 above. As described above, the models include all school-level variables from the PSM procedure as well as individual-level controls. For PALS, due to nonlinearity in the pre-post relationship, we include variables for both the fall pretest and a squared measure of the fall pretest.

16 PALS models, which only include kindergarten, and models that estimate differential effects by grade, contain only year fixed effects.

All models include weights generated by the kernel PSM procedure. Standard errors are clustered at the school-level. Models for PALS, MAP/STAR math and reading, and absence rate use Weighted Least Squares, and the suspension rate model, where the outcome is an indicator of whether a student received at least one suspension during the year, uses a logit specification. To account for the non-linearity of absence rate as an outcome, we first converted absence rates onto the standard normal distribution using a probit transformation. To provide meaningful results, we then use an inverse transformation of the raw impact estimates before reporting.

For reference, Table 51 provides information on the average and standard deviation of each of the outcomes of interest in 2018-19 using the appropriate and weighted analysis sample.

## Table 51

### Outcomes Summary Statistics, 2018-19

| OUTCOME | MEAN | S.D. |
|---|---|---|
| PALS | -0.329 | 1.346 |
| MAP/STAR Math | 0.080 | 1.121 |
| MAP/STAR Reading | -0.134 | 1.109 |
| Absence Rate | 0.060 | 0.069 |
| Suspension Rate | 0.026 | 0.159 |

Note: Statistics are weighted and sample-specific.

## Robustness to Alternative Estimation Strategies

To assess the robustness of our findings, we tested several alternative estimation strategies that attempt to address limitations of the matching and estimation strategies described above. These strategies include school fixed effects, school random effects, and school-specific time trends, shown in Table 52 through Table 56 below. In each of the tables, Column (1) displays the results from the preferred specification used in the main analysis above. Columns (2)-(4) of the tables display separate robustness checks.

The matching and estimation strategies described above rely on the assumption that schools matched on observable characteristics (e.g. test scores, demographics) are also matched on unobservable characteristics (e.g. schools' ability to implement AGR, teacher quality available in the local hiring market) that might be related to both outcomes and SAGE/AGR participation, and therefore bias impact estimates. However, there is no way to test this assumption. Including school fixed effects in the estimation would control for differences in unobservable characteristics between schools by comparing outcomes before and after AGR implementation within the same school. For the AGR analysis, however, including school fixed effects only allows comparisons of AGR to SAGE because all AGR schools previously participated in SAGE. With school fixed effects, comparisons to non-SAGE, non-AGR comparison schools would be impossible, because comparison schools, whose program participation does not change over the sample period, would not contribute to the AGR impact estimate. Nevertheless, comparing the AGR-SAGE difference from the preferred specification to a specification with school fixed effects provides useful information about the extent that unobservable school characteristics may bias estimations. To that end, in Table 52 - Table 56, Column (2) contains results of school fixed effects regressions. These results are qualitatively similar to the preferred specification in Column (1), although for PALS the difference between AGR and SAGE is less than half that of the preferred specification.

As an alternative to fixed effects, we also include school-specific random effects, which produce a weighted average of between-school and within-school effects.[17] Random effects, however, do not allow for variation in weights within schools, which occurs when matching testing outcomes within year and grade. As discussed above, our preferred matching strategy enables us to significantly increase the sample and improve generalizability by including schools that did not consistently test throughout the sample period or across Grades 1-3. Conversely, both attendance and discipline outcomes are available statewide in every year, allowing a less restrictive matching strategy that gives control schools the same weights in every year. Column (3) of Table 52 - Table 56 shows results from regressions that include school random effects. For both attendance and discipline, key coefficients are qualitatively similar to those from the preferred specification in Column (1).

17 Cameron, A. C. & Trivedi, P. K. (2005). *Microeconometrics methods and applications*. Cambridge, UK: Cambridge University Press, p. 711.

Finally, we test for the presence of time trends in outcomes that may differ between AGR/SAGE and control schools and bias results. For example, If AGR/SAGE schools are more likely on positive trajectories unrelated to their participation in the program, estimates of AGR and SAGE impacts would be biased upward. We chose not to include school-specific time trends in our preferred specification because these school trends could be the result of SAGE and AGR, and there is no method to differentiate between unrelated trends and program impacts. However, in Table 52 - Table 56 we include estimates from regressions with school-specific linear time trends to provide readers with as much information as possible. In general, including trends has only small impacts on estimated impacts. For PALS (Table 52), the SAGE impact increases and the AGR impact decreases. For reading (Table 53), estimated AGR impacts increase substantially, while time trends have little effect on attendance or discipline impacts (Table 55 and Table 56).

# Table 52

## Robustness to Alternative Estimation Strategies - PALS

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| SAGE vs. non | 0.18 | | NA | 0.053 |
| p-value | 0.387[18] | | NA | 0.095 |
| AGR vs. non | 0.105 | | NA | 0.086 |
| p-value | 0.004 | | NA | 0.077 |
| AGR vs. SAGE | 0.087 | 0.03 | NA | 0.033 |
| p-value | 0.011 | 0.213 | NA | 0.217 |
| Individual controls | YES | YES | YES | YES |
| School-level controls | YES | YES | YES | YES |
| School fixed effects | NO | YES | NO | NO |
| School random effects | NO | NO | YES | NO |
| School-specific time trends | NO | NO | NO | YES |

18 We report p-values before the multiple comparisons correction.

# Table 53

## Robustness to Alternative Estimation Strategies - MAP/STAR Math

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| SAGE vs. non | -0.007 | | NA | 0.008 |
| p-value | 0.654 | | NA | 0.739 |
| AGR vs. non | -0.011 | | NA | 0.017 |
| p-value | 0.559 | | NA | 0.676 |
| AGR vs. SAGE | -0.004 | -0.023 | NA | 0.008 |
| p-value | 0.797 | 0.184 | NA | 0.718 |
| Individual controls | YES | YES | YES | YES |
| School-level controls | YES | YES | YES | YES |
| School fixed effects | NO | YES | NO | NO |
| School random effects | NO | NO | YES | NO |
| School-specific time trends | NO | NO | NO | YES |

# Table 54

## Robustness to Alternative Estimation Strategies - MAP/STAR Reading

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| SAGE vs. non | 0.017 | | NA | 0.059 |
| p-value | 0.185 | | NA | 0.025 |
| AGR vs. non | 0.003 | | NA | 0.109 |
| p-value | 0.826 | | NA | 0.004 |
| AGR vs. SAGE | -0.014 | -0.012 | NA | 0.050 |
| p-value | 0.359 | 0.434 | NA | 0.009 |
| Individual controls | YES | YES | YES | YES |
| School-level controls | YES | YES | YES | YES |
| School fixed effects | NO | YES | NO | NO |
| School random effects | NO | NO | YES | NO |
| School-specific time trends | NO | NO | NO | YES |

# Table 55

## Robustness to Alternative Estimation Strategies - Absences

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| SAGE vs. non | 0.113[19] | | 0.088 | 0.113 |
| p-value | 0.000 | | 0.007 | 0.003 |
| AGR vs. non | 0.050 | | 0.035 | 0.059 |
| p-value | 0.024 | | 0.279 | 0.265 |
| AGR vs. SAGE | -0.063 | -0.053 | -0.053 | -0.054 |
| p-value | 0.009 | 0.018 | 0.018 | 0.059 |
| Individual controls | YES | YES | YES | YES |
| School-level controls | YES | YES | YES | YES |
| School fixed effects | NO | YES | NO | NO |
| School random effects | NO | NO | YES | NO |
| School-specific time trends | NO | NO | NO | YES |

# Table 56

## Robustness to Alternative Estimation Strategies - Suspensions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| SAGE vs. non | -0.005 | | -.003 | -0.002 |
| p-value | 0.007 | | 0.141 | 0.381 |
| AGR vs. non | -0.005 | | -0.005 | -0.001 |
| p-value | 0.058 | | 0.083 | 0.877 |
| AGR vs. SAGE | 0.001 | -0.001 | -0.001 | 0.002 |
| p-value | 0.784 | 0.596 | 0.527 | 0.389 |
| Individual controls | YES | YES | YES | YES |
| School-level controls | YES | YES | YES | YES |
| School fixed effects | NO | YES | NO | NO |
| School random effects | NO | NO | YES | NO |
| School-specific time trends | NO | NO | NO | YES |

19 Reported coefficients are for the probit transformation of absences.

## Statewide Achievement Gap Impacts

To calculate how the evaluation's AGR findings would impact statewide achievement gaps, for each year we calculate the actual fall and spring test score gaps and an adjusted spring test score gap under the hypothetical that AGR did not exist. If AGR is reducing the statewide test score gap, then the actual spring gap should be less than the hypothetical spring gap.

We use the sample of students with both fall and spring scores on PALS, MAP, or STAR during the same academic year from 2013-14 through 2018-19. We begin by calculating mean fall test scores by year, subject, grade, and FRL status. The fall gap is simply the difference in means across FRL and non-FRL students by year, subject, and grade. The actual spring test score gap is calculated in the same way.

For calculating the hypothetical spring test gap, we begin by calculating spring test score means and sample sizes by year, subject, grade, FRL, and whether a student attended a school receiving AGR funding in the relevant year. These statistics allow us to decompose FRL and non-FRL test score means by whether a student attended an AGR school:

$$\frac{(\bar{Y}_{spring,cy} * N|FRL_{iy} = 1, AGR_{sy} = 1) + (\bar{Y}_{spring,cy} * N|FRL_{iy} = 1, AGR_{sy} = 0)}{(N|FRL_{iy} = 1, AGR_{sy} = 1) + (N|FRL_{iy} = 1, AGR_{sy} = 0)} \quad (3)$$

Similarly, the decomposed mean for non-FRL students is:

$$\frac{(\bar{Y}_{spring,cy} * N|FRL_{iy} = 0, AGR_{sy} = 1) + (\bar{Y}_{spring,cy} * N|FRL_{iy} = 0, AGR_{sy} = 0)}{(N|FRL_{iy} = 0, AGR_{sy} = 1) + (N|FRL_{iy} = 0, AGR_{sy} = 0)} \quad (4)$$

Calculating the hypothetical mean spring test score for FRL and non-FRL students requires that we replace the actual means for AGR students:

$$(\bar{Y}_{spring,cy} * N|FRL_{iy} = X, AGR_{sy} = 1) \quad (5)$$

with adjusted means under the hypothetical that AGR did not exist. To do so, we begin with regressions similar to our differential effects regression for FRL students, but instead of including only the simple interaction between FRL and the AGR effect, we also include full interactions between FRL, AGR, and indicators for grade and year. We estimate these regressions separately for each test – PALS, MAP/STAR math, and MAP/STAR reading. From these regressions, we calculate predicted spring test scores for AGR students assuming that AGR effects are zero (i.e. the coefficients on all regression terms that include AGR, both the average AGR effect and the interacted AGR effects, are all zero). These adjusted predicted test scores are estimates of what AGR students' scores would have been if their school had not received AGR funding.

Finally, we use the means of the adjusted predicted scores in place of actual spring test score means in the decomposed test score means in Equations 6 and 7:

$$\frac{(\bar{\hat{Y}}^{adj}_{spring,cy} * N|FRL_{iy} = 1, AGR_{sy} = 1) + (\bar{Y}_{spring,cy} * N|FRL_{iy} = 1, AGR_{sy} = 0)}{(N|FRL_{iy} = 1, AGR_{sy} = 1) + (N|FRL_{iy} = 1, AGR_{sy} = 0)} \quad (6)$$

$$\frac{(\bar{\hat{Y}}^{adj}_{spring,cy} * N|FRL_{iy} = 0, AGR_{sy} = 1) + (\bar{Y}_{spring,cy} * N|FRL_{iy} = 0, AGR_{sy} = 0)}{(N|FRL_{iy} = 0, AGR_{sy} = 1) + (N|FRL_{iy} = 0, AGR_{sy} = 0)} \quad (7)$$

The difference between Equation 6 and 7 is the hypothetical, statewide spring test score gap between FRL and non-FRL students.

We choose to report only 2019 impacts on statewide gaps to avoid confusion regarding gap behavior over time. Our preferred methodology, described throughout the report, re-matches schools in each year. This method is necessary because schools frequently change test vendors, and limiting the sample to schools with consistent test patterns would result in a small sample that is not plausibly representative of all AGR schools. By annually re-matching on fall test scores (which include AGR impacts), however, we are implicitly assuming that AGR impacts disappear at the end of each school year. Thus, for any cohort the AGR impact on the statewide gap would need to reset to zero at the beginning of each successive school year, discounting any previous AGR impacts (e.g. discounting the Kindergarten gap impact when estimating the first grade impact). This resetting gap does not reflect reality, only the necessities of the impact methodology. Therefore, we choose to report only one year of gap impacts to avoid understating AGR impacts on the gap over time.

## Multiple Comparisons Analysis

Estimating multiple impact models, as this report does, increases the likelihood for false positives–results that are statistically significant due to random chance rather than actual program impacts. For example, a 0.05 significance level implies that 5 percent of statistically significant estimates are produced by random chance. To adjust for potential false positives, we apply the Benjamini-Hochberg procedure, a common method of correcting for multiple comparisons by accounting for the total number of statistical tests as well as the strength of the estimates, as measured by p-values.[20]

Table 57 below shows results of the Benjamini-Hochberg procedure for AGR's main and subgroup impacts. According to the procedure, impact estimates are ranked in ascending order of p-values. We then calculate a critical value equal to the rank multiplied by a false discovery rate (chosen here to be 5 percent), divided by the total number of comparisons. For each estimate to be statistically significant, its p-value must be less than the critical value. In addition to the critical value, to aid in interpretation for readers accustomed to the 0.05 threshold for statistical significant, we calculate an adjusted p-value from the same formula used to produce the critical value.[21]

After correcting for multiple comparisons, approximately half of the AGR impact estimates with unadjusted p-values below 0.05 are no longer statistically significant. Only the strongest estimates, those with unadjusted p-values less than 0.007, remain significant.

# Table 57

### Results of the Benjamini-Hochberg Procedure for Multiple Comparisons–AGR Statewide and Subgroup Impacts

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| PALS K | Subgroup - EL | 0.370 | 0.000 | 1 | 0.001 | Yes | 0.000 |
| PALS K | Subgroup- Hispanic | 0.253 | 0.000 | 2 | 0.001 | Yes | 0.005 |
| Absences | Subgroup - Urban | 0.092 | 0.000 | 3 | 0.002 | Yes | 0.006 |
| PALS K | Subgroup - Asian | 0.240 | 0.000 | 4 | 0.003 | Yes | 0.006 |
| PALS K | Subgroup - Urban | 0.206 | 0.001 | 5 | 0.003 | Yes | 0.018 |
| PALS K | Subgroup - FRL | 0.134 | 0.003 | 6 | 0.004 | Yes | 0.038 |
| PALS K | Subgroup - Female | 0.103 | 0.003 | 7 | 0.004 | Yes | 0.037 |
| PALS K | Overall | 0.105 | 0.004 | 8 | 0.005 | Yes | 0.039 |
| OSS | Subgroup - EL | -0.005 | 0.005 | 9 | 0.006 | Yes | 0.043 |
| OSS | Subgroup - Hispanic | -0.006 | 0.006 | 10 | 0.006 | Yes | 0.048 |

20 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*: Series B (Methodological), 57(1), 289-300.

21 Specifically, the number of comparisons multiplied by the false discover rate (0.05), divided by the rank.

# Table 57, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Absences | Subgroup - Grade 3 | 0.062 | 0.006 | 11 | 0.007 | Yes | 0.045 |
| Absences | Subgroup - Black | 0.108 | 0.008 | 12 | 0.008 | No | 0.053 |
| Math | Subgroup - Special Ed. | -0.065 | 0.011 | 13 | 0.008 | No | 0.069 |
| OSS | Subgroup - Grade K | -0.005 | 0.014 | 14 | 0.009 | No | 0.081 |
| Absences | Subgroup - Grade 2 | 0.054 | 0.018 | 15 | 0.009 | No | 0.095 |
| Absences | Overall | 0.050 | 0.024 | 16 | 0.010 | No | 0.122 |
| PALS K | Subgroup - Black | 0.187 | 0.029 | 17 | 0.011 | No | 0.137 |
| Absences | Subgroup - EL | 0.084 | 0.033 | 18 | 0.011 | No | 0.145 |
| Absences | Subgroup - Female | 0.046 | 0.044 | 19 | 0.012 | No | 0.184 |
| Absences | Subgroup - Grade 1 | 0.047 | 0.044 | 20 | 0.013 | No | 0.175 |
| OSS | Subgroup - FRL | -0.005 | 0.046 | 21 | 0.013 | No | 0.174 |
| Absences | Subgroup - Hispanic | 0.054 | 0.046 | 22 | 0.014 | No | 0.167 |
| Math | Subgroup - Grade 3 | -0.039 | 0.056 | 23 | 0.014 | No | 0.196 |
| OSS | Subgroup - Special Ed. | -0.009 | 0.057 | 24 | 0.015 | No | 0.190 |
| Absences | Subgroup - Special Ed. | 0.051 | 0.059 | 25 | 0.016 | No | 0.189 |
| PALS K | Subgroup - White | 0.053 | 0.071 | 26 | 0.016 | No | 0.219 |
| OSS | Overall | -0.004 | 0.089 | 27 | 0.017 | No | 0.262 |
| OSS | Subgroup - Black | -0.007 | 0.107 | 28 | 0.018 | No | 0.307 |
| Absences | Subgroup - Grade K | 0.036 | 0.123 | 29 | 0.018 | No | 0.339 |
| OSS | Subgroup - Urban | -0.005 | 0.136 | 30 | 0.019 | No | 0.361 |
| Math | Subgroup - Black | -0.058 | 0.150 | 31 | 0.019 | No | 0.386 |
| OSS | Subgroup - Race Other | -0.009 | 0.155 | 32 | 0.020 | No | 0.387 |
| OSS | Subgroup - Female | -0.002 | 0.156 | 33 | 0.021 | No | 0.378 |
| OSS | Subgroup - Grade 2 | -0.004 | 0.156 | 34 | 0.021 | No | 0.368 |
| OSS | Subgroup - Grade 1 | -0.004 | 0.163 | 35 | 0.022 | No | 0.374 |
| PALS K | Subgroup - Special Ed. | 0.083 | 0.172 | 36 | 0.023 | No | 0.383 |
| Absences | Subgroup - FRL | 0.030 | 0.227 | 37 | 0.023 | No | 0.490 |
| Absences | Subgroup - White | 0.035 | 0.246 | 38 | 0.024 | No | 0.518 |

# Table 57, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Math | Subgroup - FRL | -0.025 | 0.268 | 39 | 0.024 | No | 0.549 |
| Reading | Subgroup - Grade 2 | 0.021 | 0.276 | 40 | 0.025 | No | 0.552 |
| Reading | Subgroup - Black | -0.036 | 0.283 | 41 | 0.026 | No | 0.552 |
| Math | Subgroup - Hispanic | 0.027 | 0.283 | 42 | 0.026 | No | 0.539 |
| Reading | Subgroup - White | 0.016 | 0.294 | 43 | 0.027 | No | 0.547 |
| Math | Subgroup - Grade 2 | 0.025 | 0.352 | 44 | 0.028 | No | 0.639 |
| OSS | Subgroup - Grade 3 | -0.003 | 0.391 | 45 | 0.028 | No | 0.695 |
| Absences | Subgroup - Asian | 0.059 | 0.391 | 46 | 0.029 | No | 0.680 |
| Reading | Subgroup - Hispanic | 0.015 | 0.454 | 47 | 0.029 | No | 0.773 |
| Math | Subgroup - Asian | -0.038 | 0.456 | 48 | 0.030 | No | 0.760 |
| Absences | Subgroup - Race Other | 0.036 | 0.461 | 49 | 0.031 | No | 0.753 |
| Reading | Subgroup - Race Other | 0.016 | 0.478 | 50 | 0.031 | No | 0.765 |
| PALS I | Subgroup - Black | 0.041 | 0.535 | 51 | 0.032 | No | 0.839 |
| PALS I | Subgroup - EL | 0.056 | 0.538 | 52 | 0.033 | No | 0.828 |
| Math | Subgroup - Urban | -0.017 | 0.557 | 53 | 0.033 | No | 0.840 |
| Math | Overall | -0.011 | 0.559 | 54 | 0.034 | No | 0.827 |
| PALS I | Subgroup - Hispanic | 0.033 | 0.622 | 55 | 0.034 | No | 0.905 |
| Reading | Subgroup - EL | 0.012 | 0.633 | 56 | 0.035 | No | 0.904 |
| PALS I | Overall | 0.016 | 0.649 | 57 | 0.036 | No | 0.912 |
| PALS I | Subgroup - FRL | 0.017 | 0.657 | 58 | 0.036 | No | 0.906 |
| Reading | Subgroup - Urban | -0.010 | 0.665 | 59 | 0.037 | No | 0.902 |
| PALS I | Subgroup - Urban | 0.023 | 0.668 | 60 | 0.038 | No | 0.890 |
| OSS | Subgroup - Asian | 0.000 | 0.676 | 61 | 0.038 | No | 0.886 |
| PALS I | Subgroup - Female | 0.014 | 0.678 | 62 | 0.039 | No | 0.875 |
| Reading | Subgroup - Grade 3 | -0.006 | 0.687 | 63 | 0.039 | No | 0.873 |
| PALS I | Subgroup - Asian | -0.025 | 0.689 | 64 | 0.040 | No | 0.861 |
| PALS I | Subgroup - Special Ed. | 0.013 | 0.722 | 65 | 0.041 | No | 0.888 |
| Math | Subgroup - Race Other | -0.009 | 0.729 | 66 | 0.041 | No | 0.883 |

# Table 57, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Reading | Subgroup - Female | 0.005 | 0.765 | 67 | 0.042 | No | 0.913 |
| Reading | Subgroup - Asian | -0.008 | 0.805 | 68 | 0.043 | No | 0.947 |
| Math | Subgroup - EL | 0.009 | 0.810 | 69 | 0.043 | No | 0.939 |
| Reading | Overall | 0.003 | 0.826 | 70 | 0.044 | No | 0.944 |
| Math | Subgroup - Female | -0.004 | 0.829 | 71 | 0.044 | No | 0.934 |
| Reading | Subgroup - Special Ed. | -0.004 | 0.849 | 72 | 0.045 | No | 0.943 |
| Reading | Subgroup - FRL | 0.003 | 0.852 | 73 | 0.046 | No | 0.934 |
| PALS I | Subgroup - White | 0.005 | 0.878 | 74 | 0.046 | No | 0.949 |
| Math | Subgroup - Grade 1 | -0.005 | 0.882 | 75 | 0.047 | No | 0.941 |
| Reading | Subgroup - Grade 1 | -0.004 | 0.900 | 76 | 0.048 | No | 0.948 |
| PALS I | Subgroup - Race Other | -0.005 | 0.907 | 77 | 0.048 | No | 0.942 |

In addition, we apply the Benjamini-Hochberg procedure to estimates of the differences in impacts between AGR and SAGE. Table 58 displays these results. Similar to the AGR impacts in Table 57, few of the AGR-SAGE comparisons remain statistically significant post-procedure.

# Table 58

## Results of the Benjamini-Hochberg Procedure for Multiple Comparisons–AGR-SAGE Comparisons

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| PALS K | Subgroup - EL | 0.401 | 0.000 | 1 | 0.001 | Yes | 0.000 |
| PALS K | Subgroup - Hispanic | 0.294 | 0.000 | 2 | 0.001 | Yes | 0.000 |
| Absences | Subgroup - White | -0.092 | 0.000 | 3 | 0.002 | Yes | 0.010 |
| PALS K | Subgroup - Urban | 0.174 | 0.000 | 4 | 0.003 | Yes | 0.010 |
| PALS K | Subgroup - FRL | 0.114 | 0.003 | 5 | 0.003 | Yes | 0.042 |

# Table 58, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Absences | Subgroup - Grade K | -0.077 | 0.003 | 6 | 0.004 | Yes | 0.041 |
| Math | Subgroup - Special Ed. | -0.060 | 0.005 | 7 | 0.004 | No | 0.058 |
| PALS K | Subgroup - Asian | 0.180 | 0.006 | 8 | 0.005 | No | 0.063 |
| Absences | Subgroup - Female | -0.065 | 0.007 | 9 | 0.006 | No | 0.066 |
| Math | Subgroup - Grade 3 | -0.050 | 0.007 | 10 | 0.006 | No | 0.060 |
| Reading | Subgroup - Black | -0.065 | 0.008 | 11 | 0.007 | No | 0.056 |
| Absences | Overall | -0.063 | 0.009 | 12 | 0.008 | No | 0.059 |
| PALS K | Subgroup - Female | 0.087 | 0.011 | 13 | 0.008 | No | 0.065 |
| PALS 1 | Subgroup - EL | 0.164 | 0.011 | 14 | 0.009 | No | 0.062 |
| PALS 1 | Subgroup - Hispanic | 0.127 | 0.011 | 15 | 0.009 | No | 0.059 |
| Math | Subgroup - Black | -0.068 | 0.011 | 16 | 0.010 | No | 0.056 |
| PALS K | Overall | 0.087 | 0.011 | 17 | 0.011 | No | 0.053 |
| Reading | Subgroup - Urban | -0.048 | 0.011 | 18 | 0.011 | No | 0.051 |
| Absences | Subgroup - Grade 1 | -0.062 | 0.013 | 19 | 0.012 | No | 0.053 |
| Absences | Subgroup - Grade 2 | -0.059 | 0.014 | 20 | 0.013 | No | 0.055 |
| Reading | Subgroup - Hispanic | -0.045 | 0.024 | 21 | 0.013 | No | 0.091 |
| Absences | Subgroup - Grade 3 | -0.056 | 0.024 | 22 | 0.014 | No | 0.089 |
| Absences | Subgroup - Race Other | -0.064 | 0.026 | 23 | 0.014 | No | 0.092 |
| Absences | Subgroup - FRL | -0.050 | 0.034 | 24 | 0.015 | No | 0.112 |
| Reading | Subgroup - Grade 3 | -0.033 | 0.036 | 25 | 0.016 | No | 0.114 |
| OSS | Subgroup - Grade 3 | 0.005 | 0.036 | 26 | 0.016 | No | 0.111 |
| Absences | Subgroup - Special Ed. | -0.050 | 0.046 | 27 | 0.017 | No | 0.136 |
| Math | Subgroup - Grade 2 | 0.040 | 0.064 | 28 | 0.018 | No | 0.184 |
| PALS 1 | Subgroup - White | -0.051 | 0.098 | 29 | 0.018 | No | 0.270 |
| Math | Subgroup - EL | 0.035 | 0.160 | 30 | 0.019 | No | 0.427 |
| PALS K | Subgroup - Black | 0.070 | 0.164 | 31 | 0.019 | No | 0.424 |
| PALS K | Subgroup - White | 0.042 | 0.180 | 32 | 0.020 | No | 0.450 |
| Math | Subgroup - Grade 1 | 0.033 | 0.199 | 33 | 0.021 | No | 0.484 |

# Table 58, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Reading | Subgroup - FRL | -0.020 | 0.225 | 34 | 0.021 | No | 0.529 |
| PALS K | Subgroup - Special Ed. | 0.055 | 0.237 | 35 | 0.022 | No | 0.543 |
| Absences | Subgroup - Asian | -0.037 | 0.241 | 36 | 0.023 | No | 0.535 |
| Math | Subgroup - FRL | -0.020 | 0.278 | 37 | 0.023 | No | 0.602 |
| Reading | Subgroup - Grade 1 | 0.029 | 0.293 | 38 | 0.024 | No | 0.616 |
| PALS 1 | Subgroup - Urban | 0.043 | 0.299 | 39 | 0.024 | No | 0.613 |
| PALS K | Subgroup - Race Other | 0.046 | 0.320 | 40 | 0.025 | No | 0.640 |
| Absences | Subgroup - Urban | -0.023 | 0.344 | 41 | 0.026 | No | 0.671 |
| Reading | Overall | -0.014 | 0.359 | 42 | 0.026 | No | 0.685 |
| PALS 1 | Subgroup - Black | -0.042 | 0.371 | 43 | 0.027 | No | 0.691 |
| Reading | Subgroup - White | 0.014 | 0.372 | 44 | 0.028 | No | 0.676 |
| OSS | Subgroup - Urban | 0.002 | 0.396 | 45 | 0.028 | No | 0.704 |
| OSS | Subgroup - White | 0.001 | 0.409 | 46 | 0.029 | No | 0.710 |
| Math | Subgroup - White | 0.014 | 0.426 | 47 | 0.029 | No | 0.726 |
| Math | Subgroup - Asian | 0.028 | 0.433 | 48 | 0.030 | No | 0.722 |
| Math | Subgroup - Race Other | 0.021 | 0.439 | 49 | 0.031 | No | 0.717 |
| OSS | Subgroup - EL | -0.001 | 0.443 | 50 | 0.031 | No | 0.708 |
| PALS 1 | Subgroup - Race Other | -0.032 | 0.451 | 51 | 0.032 | No | 0.707 |
| Math | Subgroup - Urban | -0.015 | 0.454 | 52 | 0.033 | No | 0.699 |
| Reading | Subgroup - Female | -0.011 | 0.488 | 53 | 0.033 | No | 0.736 |
| OSS | Subgroup - Black | 0.002 | 0.490 | 54 | 0.034 | No | 0.725 |
| OSS | Overall | 0.001 | 0.565 | 55 | 0.034 | No | 0.821 |
| PALS 1 | Overall | -0.016 | 0.607 | 56 | 0.035 | No | 0.867 |
| OSS | Subgroup - FRL | 0.001 | 0.612 | 57 | 0.036 | No | 0.859 |
| Reading | Subgroup - Grade 2 | -0.009 | 0.621 | 58 | 0.036 | No | 0.857 |
| Absences | Subgroup - EL | -0.014 | 0.629 | 59 | 0.037 | No | 0.852 |
| PALS 1 | Subgroup - Asian | 0.024 | 0.639 | 60 | 0.038 | No | 0.852 |
| Reading | Subgroup - EL | -0.010 | 0.664 | 61 | 0.038 | No | 0.871 |

# Table 58, continued

| OUTCOME | MODEL | COEFF. | P-VALUE | RANK | CRITICAL VALUE | STAT. SIG. AFTER MULTIPLE COMPAR. CORRECTION | ADJ. P-VALUE |
|---|---|---|---|---|---|---|---|
| Absences | Subgroup - Black | -0.011 | 0.678 | 62 | 0.039 | No | 0.875 |
| Reading | Subgroup - Race Other | 0.009 | 0.680 | 63 | 0.039 | No | 0.864 |
| OSS | Subgroup - Grade 1 | -0.001 | 0.691 | 64 | 0.040 | No | 0.864 |
| Math | Subgroup - Hispanic | 0.009 | 0.696 | 65 | 0.041 | No | 0.857 |
| OSS | Subgroup - Race Other | -0.001 | 0.709 | 66 | 0.041 | No | 0.860 |
| Reading | Subgroup - Asian | -0.010 | 0.758 | 67 | 0.042 | No | 0.905 |
| OSS | Subgroup - Asian | 0.000 | 0.759 | 68 | 0.043 | No | 0.893 |
| OSS | Subgroup - Hispanic | 0.000 | 0.760 | 69 | 0.043 | No | 0.881 |
| PALS I | Subgroup - Female | -0.008 | 0.793 | 70 | 0.044 | No | 0.906 |
| Reading | Subgroup - Special Ed. | 0.005 | 0.796 | 71 | 0.044 | No | 0.897 |
| Math | Overall | -0.004 | 0.797 | 72 | 0.045 | No | 0.886 |
| OSS | Subgroup - Grade K | 0.000 | 0.820 | 73 | 0.046 | No | 0.898 |
| Absences | Subgroup - Hispanic | -0.006 | 0.822 | 74 | 0.046 | No | 0.889 |
| OSS | Subgroup - Grade 2 | 0.000 | 0.836 | 75 | 0.047 | No | 0.891 |
| PALS I | Subgroup - FRL | -0.006 | 0.864 | 76 | 0.048 | No | 0.910 |
| PALS I | Subgroup - Special Ed. | -0.005 | 0.879 | 77 | 0.048 | No | 0.914 |
| OSS | Subgroup - Female | 0.000 | 0.905 | 78 | 0.049 | No | 0.928 |
| OSS | Subgroup - Special Ed. | 0.000 | 0.953 | 79 | 0.049 | No | 0.965 |
| Math | Subgroup - Female | -0.001 | 0.970 | 80 | 0.050 | No | 0.970 |

Finally, we also conducted the Benjamini-Hochberg procedure to estimates of AGR by strategy. Due to size, a table containing these results is available upon request.

Section II

# Survey Appendix

# Survey Appendix

As you answer the following questions, please answer individually for each school participating in AGR.

1. Which of the following strategies were used in K-3 classrooms during the school year? (Please select all that apply.) [forced answer]

   - Reduced class size (either 18:1 or 30:2) Yes No
   - One-to-one tutoring Yes No
   - Instructional coaching Yes No

2. [If indicated reduced class size] What percentage of classrooms in each grade have a reduced class size?

| | PERCENT OF CLASSROOMS | | | | |
|---|---|---|---|---|---|
| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
| Kindergarten | | | | | |
| First | | | | | |
| Second | | | | | |
| Third | | | | | |

3. [If indicated reduced class size] Because of the AGR program in your school, what instructional strategies are reduced class size teachers using with students? (Please select all that apply.)

   - We don't use any specific instructional strategies because of smaller class sizes
   - Small-group instruction
   - One-on-one time with the teacher
   - Differentiation of instruction
   - Strategic placement of students in groups
   - Strategic placement of students in classrooms
   - Other ___________________________________
   - Not sure/don't know

4. [If indicated reduced class size] What benefits does the reduced class size provide for your school? ______________________________________________

5. [If indicated one-to-one tutoring] What percentage of classrooms in each grade have one-to-one tutoring?

| | PERCENT OF CLASSROOMS | | | | |
|---|---|---|---|---|---|
| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
| Kindergarten | | | | | |
| First | | | | | |
| Second | | | | | |
| Third | | | | | |

6. [If indicated one-to-one tutoring] On average, how often are AGR one-to-one tutors meeting with students in each of the following areas?

| | MATHEMATICS | READING | OTHER |
|---|---|---|---|
| 3/week or more | | | |
| 2/week | | | |
| Weekly | | | |
| Biweekly | | | |
| Monthly | | | |
| As needed | | | |
| Other | | | |

7. [If indicated one-to-one tutoring] What is the average duration of tutoring sessions with students in each of the following areas?

- Mathematics – [Slider bar from 0 to 120 minutes]
- Reading – [Slider bar from 0 to 120 minutes]
- Other – [Slider bar from 0 120 minutes]

8. [If indicated one-to-one tutoring] Which of the following characteristics do your AGR tutors have? (Please select all that apply.)

- Tutoring training
- Previous tutoring experience
- Not sure/don't know

9. [If indicated one-to-one tutoring] What benefits does one-to-one tutoring provide for your school? ______________________________

10. [If indicated instructional coaching] What percentage of teachers in each grade have received instructional coaching?

| | PERCENT OF CLASSROOMS | | | | |
|---|---|---|---|---|---|
| GRADE | NONE | LESS THAN 25% | 25-50% | 51-75% | MORE THAN 75% |
| Kindergarten | | | | | |
| First | | | | | |
| Second | | | | | |
| Third | | | | | |

11. [If indicated instructional coaching] Which of the following characteristics do your AGR instructional coaches have? (Please select all that apply.)

- Coach training
- Previous instructional coaching experience
- Content specialist in their subject of coaching
- Not sure/don't know

12. [If indicated instructional coaching] On average, how often are AGR instructional coaches meeting with the teachers they coach?

- Weekly
- Monthly
- Quarterly
- Each semester
- As needed
- Other
- Not sure/don't know

13. [If indicated instructional coaching] What is the average duration of instructional coach meetings with the teachers they coach?
    - [Slider bar from 0 to 120 minutes]

14. [If indicated instructional coaching] What topics are instructional coaches covering with teachers? ____________________________________

15. [If indicated instructional coaching] What benefits does the instructional coaching program provide for your school? __________________________

16. Which assessments are you using to measure student progress in each grade during the school year (ex. MAP, PALS, STAR)? (Please list all that apply.)

- Kindergarten ____________
- First ____________
- Second ____________
- Third ____________

# WEC

Supporting PreK-12

Wisconsin Evaluation Collaborative