Effects of Large-Scale Early Math Interventions on Student Outcomes: Evidence from Kentucky's Math Achievement Fund

Zeyu Xu
Umut Ozek
Jesse Levin
Dong Hoon Lee

March 2023

WORKING PAPER No. 279-0323

CALDER
National Center for Analysis of Longitudinal Data in Education Research

AIR®

Effects of Large-Scale Early Math Interventions on Student Outcomes: Evidence from Kentucky's Math Achievement Fund

Zeyu Xu
American Institutes for Research /CALDER

Umut Ozek
RAND

Jesse Levin
American Institutes for Research

Dong Hoon Lee
American Institutes for Research

Contents

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2000483. The authors would like to thank their research partners Karen Dodd, Erin Chavez, Aaron Butler, and Hannah Poquette from the Kentucky Department of Education; Barrett Ross from the Kentucky Center for Statistics; Kelly DeLong from the Kentucky Center for Mathematics; and Mary Lee Glore from Northern Kentucky University. The authors would also like to thank Kirk Walters, Heather Hill, and participants of the SREE conference for valuable feedback. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Kentucky Department of Education, the Kentucky Center for Statistics, the Kentucky Center for Mathematics, or Northern Kentucky University.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

Effects of Large-Scale Early Math Interventions on Student Outcomes: Evidence from Kentucky's Math Achievement Fund

Zeyu Xu, Umut Özek, Jesse Levin, and Dong Hoon Lee

CALDER Working Paper No. 279-0323
March 2023

Abstract

Reading has been at the forefront of early-grade educational interventions, but addressing the educational needs of students in math early on is also critical given that early gaps in math skills widen further over the course of schooling. In this study, we examine the effects of Kentucky's Math Achievement Fund – a unique state-level program that combines targeted interventions, peer-coaching, and close collaboration among teachers to improve math achievement in grades K-3 – on student outcomes and the costs associated with this policy. We find significant positive effects of the program not only on math achievement, but also on test scores in reading and non-test outcomes including student attendance and disciplinary incidents. The benefits exist across racial/ethnic groups and students from different socioeconomic statuses, and they are slightly higher for racial minorities. These findings, along with the cost estimate of the program, suggest that this program could provide a cost-effective blueprint to address the educational needs of students in math in early grades.

1. Introduction

A long line of research demonstrates the economic value of mathematics skills to both individuals and the economy as a whole (e.g., Behrman, Ross, & Sabot, 2008; Glewwe, 1996; Hershbein & Kearney, 2014). Based on a review of evidence from the United States (e.g., Altonji & Pierret, 2001; Lazear, 2003; Murnane et al., 2000), Hanushek and Woessmann (2008) conclude that one standard deviation increase in mathematics performance at the end of high school translates into 12% higher annual earnings.

Adolescent mathematics performance, in turn, is strongly predicted by students' early-grade mathematics skills (Austin et al., 2021; Goldhaber et al, 2021; Watts, Duncan, Siegler, & Davis-Kean, 2014). Although the association between children's early skills and their later development is present in many domains, such association is twice as strong in mathematics as in reading after controlling for pre-school entry cognitive and socioemotional skills and family background (Duncan et al., 2007). More importantly, children who exhibit early delays in mathematical development also demonstrate lower growth rates in mathematics skills later (Morgan, Farkas & Wu, 2011; Baumert, Nagy, & Lehmann, 2012). Consequently, early gaps in mathematics skills widen further over the course of schooling (Geary et al., 2013; Loeb & Bassok, 2007; Magnuson & Duncan, 2006). How to diagnose and correct mathematics deficiencies early, therefore, is a challenge that our K–12 schools must confront *systematically*. That said, as pointed out in a National Academies report, attention to early mathematics learning does not match its importance, and reading has been the focus of early childhood education (Cross, Woods & Schweingruber, 2009).[1]

[1] Although nearly half of all states have specific requirements for early literacy assessments and interventions, only three states at the time of writing have policies that mandate additional mathematics support for K–3 students based on screening test results (Georgia Kindergarten Inventory of Developing Skills, Maine's pilot Numeracy4ME, and

In this study, we examine the effects of Kentucky's Math Achievement Fund (MAF) — a prime example of a state's effort to intervene early to improve student mathematics achievement— on student outcomes using a difference-in-differences (DiD) design. To the best of our knowledge, MAF was the only fully implemented statewide K-3 mathematics intervention program at the time of this study. Kentucky's early-grade mathematics performance mirrors the national trend, with roughly half of third-grade students proficient in mathematics in 2017. To help students in grades K–3 who struggle to meet grade-level expectations, MAF adopts a Response to Intervention (RtI) approach that consists of a universal screener, tiered interventions with increasing intensity, and regular progress monitoring. It also requires professional development for both the intervention teacher and classroom teachers in each grantee school. The goal of MAF is to identify mathematics deficiencies or the risks of mathematics difficulties early and provide remedial or preventative interventions before it becomes too late.

Our DiD estimates suggest no statistically significant difference in mathematics achievement trends between MAF and non-MAF schools (i.e., schools that have never implemented MAF) in pre-treatment periods, yet the trends started to diverge gradually after MAF designation. For example, we find no effect of MAF designation on mathematics scores at the end of the first year of implementation, but MAF raised student mathematics test scores by 0.05 standard deviations after two years, 0.08 standard deviations after three years, and 0.09 standard deviations after four years. We also find positive spillover effects of the program on reading achievement, with students in MAF schools outperforming non-MAF school students by 0.06 standard deviations after four years. Additionally, MAF significantly reduced disciplinary incidents and absences. Using detailed information obtained from administrative records and

Kentucky's Mathematics Achievement Fund which is the focus of the current study). A handful of other states require numeracy screening before Grade 3, but these states do not stipulate interventions based on test results.

surveys administered to school MAF teams and coordinators, we find that the estimated four-year per-pupil cost of the MAF intervention is $750 in 2018-19 dollars, which corresponds to an increase in math test scores by 0.09σ. As we discuss at the end of this study, closing gaps in mathematics skills early has important economic and equity implications. But even without considering those long-term benefits, these findings suggest that state-led early mathematics intervention programs like MAF could provide a cost-effective blueprint to raise mathematics achievement in early grades.

2. Policy Background

As provisioned in Kentucky statute KRS 158.844, the main objective of MAF is to provide teacher professional development and intervention services that address the needs of students in primary grades (Grades K-3) who are struggling to meet the grade level expectations for mathematics. The intervention is supposed to supplement, not replace, regular classroom instruction. In addition, MAF was also designed to help schools build capacity by overhauling mathematics instructional strategy and practice and strengthening collaboration among teachers. Through renewable, two-year local grants to approximately 90 schools per year, MAF awards an estimated $50,000 per school to support the hiring of one full-time mathematics intervention teacher (MIT) and training for the MIT and two classroom mathematics teachers (known as the Plus 2 teachers). MAF requires a common structure of intervention with a set of prescribed components, but grantee schools also have discretion over how specific components are implemented, as described below.

2.1 Student selection

All K-3 students in MAF schools are required to take a screener test to determine who is struggling to meet grade level or benchmark expectations for mathematics and thus needs supplemental support. In the 2018–19 school year, MAF schools in Kentucky primarily used 3 screeners: Measures of Academic Progress (MAP) in 55 schools, iReady in 18 schools, and STAR in 15 schools. Screeners were typically administered three times a year. Students were added to and dismissed from intervention services throughout the year based on whether they have met grade-level expectations. Along with screener scores, several factors (e.g., teacher and parental input, scores on alternative tests, need for remediation in reading) were used in the selection of students for the intervention. In the first cohort of MAF schools (schools designated as MAF schools in the first grant cycle that started in 2015–16 school year), the share of students in grades K-3 chosen for MAF intervention ranged from 2 percent to 34 percent (with a median of 13 percent) in 2015–16 school year.[2]

In addition to universal screeners, easyCBM was used to monitor progress among intervention students. Kindergarten students took the Numbers and Operations Measurement, while students in Grades 1-3 took the Numbers and Operations Measurement and the Numbers and Operations in Algebra Measurement. easyCBM was administered three times a year (October, January, and May), and results were reported in a centralized database.

2.2 Requirements for teachers

The MIT must be a highly trained, certified teacher with at least three years of teaching experience, with preference given to teachers with at least three years of primary teaching experience or training in mathematics intervention services for primary students. At least one-half of the MIT's time is required to be spent delivering mathematics interventions to lower-performing primary-grade students. Administrative records show that these interventions were typically provided in small groups of 4-6 students with each session lasting about 30-45 minutes. Typically, 5-6 sessions were offered every day for 5 days a week. Importantly, these sessions did not replace regular mathematics instruction, a key requirement of the MAF grant. A review of MIT weekly schedules demonstrates

[2] MAF excludes students with Individualized Educational Plans (IEPs) related to math goals. Students with IEPs receive specialized instruction from the special education teachers.

that MITs scheduled interventions around homeroom teachers' schedules so that students were not pulled during core instruction. In addition to direct instruction for targeted students, the MIT spent about 5.6 hours per week coteaching mathematics with regular classroom teachers and about 3.8 hours on lesson planning.

The MIT is also expected to engage in a significant number of professional development activities. Based on grantee performance reports, MITs typically participated in 8-12 events per year including webinars, multi-day training sessions, conferences, and courses offered by programs such as Add+VantageMR. The MIT is also expected to share instructional activities, books, and strategies obtained from these professional development opportunities with building staff during weekly team planning and professional learning communities (PLCs). MIT daily schedules suggest that the MITs spent about 2.3 hours per week on PLC activities.

In addition to the MIT, MAF requires grantee schools to select two classroom teachers every year as Plus 2 teachers. Plus 2 teachers receive 10 days of professional development and attend, at the expense of the grantee schools, at least one state mathematics conference every year. Plus 2 teachers are required to be available for collaboration and co-teaching with the MIT throughout the school year. Along with the MIT, Plus 2 teachers are expected to lead professional learning with additional teachers to build capacity in the building.

2.3 Mathematics intervention packages

Grantee schools are required to use one of the four MAF-approved primary mathematics intervention packages: Add+VantageMR® (AVMR), Assessing Math Concepts®, Do the Math®, and Math Recovery Intervention Specialist® (MRIS). These packages are used to provide interventions for targeted students, professional development for classroom teachers, and professional learning for intervention teachers. In the 2018–19 school year, out of the 97 MAF grantee schools, the overwhelming majority chose to use AVMR (76 schools) and MRIS (17 schools) to facilitate the training of intervention teachers and the delivery of intervention services.

2.4 Channels of change

While improving the mathematics achievement of low-performing students is a key objective of MAF, program administrators and grantee performance reports indicate that the defining feature of MAF is its whole school approach. MAF's emphasis on coteaching, professional development for classroom teachers, and the expectation that the MIT and Plus 2 teachers share what they have learned from professional development activities with other teachers in the building suggest that the program could lead to positive spillover effects on other students in the school. To capture the full impact of MAF, this study examines the schoolwide effects of MAF instead of focusing narrowly on the outcomes of students who were referred to interventions.[3]

MAF could improve student achievement in mathematics through three channels. First, because MAF requires that interventions not replace regular instruction, it results in a net increase in instruction time for intervention students. Empirical evidence shows that education interventions that add instruction time tend to improve student achievement significantly (e.g., Taylor 2014; Cortes, Goodman, & Nomi 2015; Figlio, Holden, & Özek, 2018).

Second, MAF interventions were delivered in small group settings. Small groups allow the interventionist to better adapt instruction to individual student needs, and students are thought to learn better and faster if the content is within their zone of proximal development (Guthrie, 2015; Stipek, 2002; Vygotsky, 1978). Small groups are also thought to help establish emotional and peer support quickly, which is particularly important in the context of early childhood development and learning (Mashburn et al., 2008). As we review later, early childhood mathematics interventions delivered in small group settings have consistently produced positive

[3] Ideally, we would also want to estimate the extent to which MAF improves the outcomes of targeted students. The use of universal screeners in grantee schools lends support to a potential regression discontinuity design. Unfortunately, an examination of program implementation indicates that screener scores were one of the many inputs used in intervention referrals. An empirical analysis of screener scores confirms that the designated cutoffs on universal screeners are not binding, and there is no significant discontinuity in treatment probability at the cutoffs.

improvements in student mathematics achievement (Pellegrini et al., 2018; Slavin, Lake, Davis, & Madden, 2011; Wanzek et al., 2016).

A third feature of MAF that may help improve student mathematics achievement is setting clear expectations for what the MIT should and should not do while encouraging close collaboration among teachers. This design feature protects the interventionist from distractions such as being asked to substitute for regular mathematics teachers in school. At the same time, the learning community approach to professional development has been demonstrated to improve teacher performance (e.g., Buysse, Castro, & Peisner-Feinberg, 2010). Close collaboration between the MIT and classroom teachers also allows the interventionist to provide "just in time" support to children who need help with mathematics, and research has shown that working in a collaborative environment greatly improves teaching quality (Bransford, Brown, & Cocking, 1999).

3. Literature Review

There are hundreds of elementary-grade mathematics interventions and more than ten thousand studies on their efficacy. Less than 1% of those studies are considered rigorous enough to be included in several best-evidence syntheses (Jacobse & Harskamp, 2011; Pellegrini et al., 2018; Savelsbergh et al., 2016; Slavin & Lake, 2008). However, the rigor of early mathematics intervention studies has improved in the last decade, and some patterns start to emerge. For example, two types of interventions are found to produce positive results consistently.

The first is tutoring programs that provide one-to-one or small-group supplemental instruction, with average effect sizes of about 0.30 standard deviations in student test scores.[4]

[4] Examples of tutoring programs include *Mathematics Recovery* (Smith, Cobb, Farran, Cordray & Munter, 2013), *Galaxy Mathematics* (Fuchs et al., 2013), *Pirate Mathematics* (Fuchs et al., 2010), *Number Rockets* (Gersten et al., 2015), and *FocusMATH* (Styers & Baird-Wilkerson, 2011).

Tutoring programs typically require extensive professional development for the interventionists and the intervention tends to be intensive, meeting at least three times per week for at least 12 weeks, with each session lasting at least 30 minutes. The second type of promising interventions focuses on improving instructional process (such as cooperative learning, classroom management and motivation), with average effect sizes around 0.25 standard deviations in student test scores.[5] All these programs pay close attention to student needs and behavior, and target instruction to student proficiency levels.[6] Similar to tutoring programs, teacher professional development is an integral part of instructional process interventions. Peer coaching and teacher collaboration in instructional planning are often featured in teacher professional development. Taken together, reviews of empirical evidence suggest that early mathematics intervention strategies that substantially affect children's daily experiences by emphasizing personalization, engagement, and motivation are likely to produce beneficial results (Pellegrini et al., 2018).

Based on this literature, the MAF design—intensive small-group supplemental instruction by full-time MITs, personalized support, teacher peer-coaching, and collaboration between interventionists and regular classroom teachers—is consistent with the characteristics of successful early mathematics interventions. As such, our study complements the existing literature in this context in several important ways.

[5] Examples of instructional process programs include *Team Assisted Individualization (TAI)* (Stevens & Slavin, 1995; Slavin & Karweit, 1985), *PAX Good Behavior Game* (Weis, Osborne, & Dean, 2015), and *Individualizing Student Instruction* (Connor et al., 2018).

[6] By contrast, interventions that focus on mathematics curricula (textbooks), benchmark assessment and teacher professional development for mathematics knowledge or pedagogy have generally produced little effect on early grade mathematics learning (Pellegrini et al., 2018). For example, *Mathematics Solutions* and *Cognitively Guided Instructions*, two widely used mathematics professional development programs that focus on improving teachers' mathematics knowledge and pedagogy, are found to have no discernible effect on student mathematics learning (Jacob, Hill, & Corey, 2017; Schoen, LaVenia, & Tazar, 2018). Mathematics curricula are found to have no detectable relationship with 4th- and 5th-grade students' mathematics achievement growth (Blazar et al., 2019).

First, studies that find positive effects on student mathematics learning tend to be small, typically with no more than a few hundred students carried out in a single district. In small-scale studies, developers and experimenters are closely involved, to an extent that is often not realistic if interventions were scaled up. Our study adds to the existing literature by studying whether promising intervention features will scale successfully in a statewide program. Second, large-scale evaluations of elementary mathematics interventions have produced reasonably solid evidence on what interventions do *not* work, but it remains unclear which intervention strategies may scale successfully. This is because large-scale studies, typically RCTs involving thousands of students across multiple districts or states, have been focusing on curricula, benchmark assessments and PD for mathematics content and pedagogy.[7] To the best of our knowledge, there has been no evaluation of any statewide K-3 mathematics intervention programs. This is primarily because very few states have systematic, coherent early-grade mathematics policies in place. Third, existing mathematics intervention research generally focuses exclusively on the mathematics achievement of students who were targeted for intervention, ignoring potential spillover effects on other students and outcomes. Finally, because there are competing demands for attention to deficiencies in multiple domains like reading and socio-emotional competence (Denton, Germino-Hausken, & West, 2000), our study estimates the cost-effectiveness of MAF to inform the best way to deploy educational resources.

4. Data

In our impact analysis, we use student-level administrative school records obtained from the Kentucky Department of Education (KDE). These data cover all students in grades K-8 between 2011-12 and 2018-19 school years and include reading and mathematics scores of all

[7] These studies generally show null program effects (e.g., Garet et al., 2016; Randel et al., 2016; Newman et al., 2012).

tested students (in grades 3 through 8) on the Kentucky Performance Rating for Educational Progress (K-Prep) test along with demographic information on students, such as race, gender, free- or reduced-price lunch eligibility, English language learner status, exceptional/special education status, student age, and schools attended. We also observe information about student disciplinary problems (number of disciplinary referrals, number of suspensions, and the total length of suspensions), and attendance (including days absent, days present, number of unexcused/excused tardy days, and the number of days possible).

Kentucky's administrative data also include detailed information on MAF interventions. The reporting of this information is mandatory. These data are populated by the mathematics team in each MAF school and checked three times a year by KDE to ensure high quality and timeliness of data entry. The intervention data include *individual student-level* information on start and end date of the intervention, intervention program (e.g., AVMR, MRIS), tier of instruction, duration of intervention (in minutes), frequency of intervention (times per week), intervention staff (the qualification level of the staff that is most directly providing the interventions services to the student), areas of student need (deficiency areas such as number sense, mathematics reasoning, and measurement). We use this information to better understand the type, duration, and intensity of the intervention students receive.

There were two MAF grant cycles between the redesign in 2015-16 and the most recent school year we observe in our data (2018-19 school year). The top panel in Figure 1 presents the number of "ever-MAF" schools (i.e., elementary schools that were ever designated as MAF during this time frame) by MAF status. There were 107 elementary schools in Kentucky that were designated as MAF in the first grant cycle that started in 2015-16 school year. Of these schools, 51 remained as MAF in the second grant cycle that started in 2017-18 whereas 56 left the MAF designation. Thirty-nine schools were newly designated as MAF in the second grant cycle.

Table 1 examines the student characteristics in pre-MAF years (between 2011-12 and 2014-15 school years) for first-cohort MAF schools (schools that were first designated as MAF in 2015-16), second-cohort MAF schools (first time designation in 2017-18), and schools that were never designated as MAF between 2015-16 and 2018-19. The results suggest that the first- and second-cohort MAF schools were comparable along observed student characteristics: students in these schools had similar scores on math tests and were comparable along disciplinary incidents, attendance, race/ethnicity, socioeconomic status (as proxied by subsidized meal eligibility), English learner status, and special education status. That said, students in both first- and second-cohorts of MAF schools had lower math scores, were more likely to be eligible for subsidized meals, more likely to be involved in disciplinary incidents, had more absences, and were more (less) likely to be White (Black) compared to students in schools that were never designated as MAF between the 2015-16 and 2018-19 school years.

In our cost analysis, we use both administrative data and primary data collected through surveys administered to MITs and Plus 2 Teachers who were part of a MAF school intervention team in the 2018-19 school year in order to obtain information on the resource effort devoted to the MAF program. The survey questionnaire was organized around the activities and personnel/nonpersonnel resources associated with the MAF program.

Survey administration took place in May and June of 2021 and was sent out to 97 MITs and 187 Plus2 teachers (P2Ts) from 97 MAF grantee schools that were active in the 2018-19 school year.[8] We received responses from 25 MITs and eight P2Ts representing 30 schools.

[8] Not all the second cohort schools were eligible for the analysis. Out of the 97 schools, 4 schools joined in the second year of the grant cycle, 4 schools did not serve students in grades 3-5 and had no outcome data.

Among the 25 MITs who responded, six are dropped because they are from schools that are not included in the impact analysis.[9] Ten of the remaining 19 MITs are from schools that are also first cohort MAF grant recipients, and nine are from schools that received MAF grants in the second cohort only. Furthermore, because the response rate among P2Ts is low, we decided to base the calculation of costs associated with these staff solely on extant administrative information pertaining to their required participation in professional development as part of the MAF program.

5. Empirical Framework

5.1 Determining Program Effectiveness

To estimate the effects of MAF on student outcomes, we rely on a difference-in-differences (DiD) design. We estimate MAF effects for each cohort separately to circumvent issues in traditional DiD models when the timing of treatment implementation is staggered and the treatment effects are heterogeneous across groups or time (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020; Sun & Abraham, 2021). In our main analysis, we focus on MAF schools in the first cohort, comparing the differences in student outcomes in years after versus before the designation (2015-16 school year) with the same difference in never-MAF schools. Formally, using OLS we estimate the following two-way fixed-effects model:

$$Y_{igst} = \alpha + \beta MAF_s * \delta_t + \delta_t + \mu_s + \gamma_g + \varepsilon_{ist} \quad (1)$$

where Y_{igst} is the outcome of interest (test scores in math and reading standardized to zero mean and unit variance at the grade-by-year level, whether the student received a suspension, and % absent days) for student *i* in grade *g* and school *s* in year *t*, MAF_s is an indicator for MAF schools

[9] Three of the six schools only serve grades K through 2 and therefore have no state standardized test scores, two schools received off-cycle grants (i.e., they joined MAF in the second year of the second MAF grant cycle), and one school does not have complete historical student performance data that are needed to estimate the MAF impact.

in the first cohort, δ_t is year fixed-effects (with 2014-15 school year serving as the omitted category), and μ_s and γ_g are school and grade fixed-effects respectively. Given that MAF is a schoolwide intervention, we cluster our standard errors at the school-level. In this setting, precisely estimated zero coefficients on the interaction between MAF_s and δ_t for years prior to 2014-15 would provide evidence that the parallel trends assumption (that is, treatment and comparison schools would follow similar trajectories in post-treatment years in the absence of the treatment) cannot be rejected, which is a critical identification assumption in the DiD design.

Panel (B) in Figure 1 checks the fidelity of implementation and examines student exposure to MAF in the first cohort MAF schools and never-MAF schools in the years before and after the MAF designation. In particular, this graph compares the share of students who ever received the MAF intervention over time between these two types of schools. The results indicate that by the end of the 4th year after the initial MAF designation, roughly 15 percent of all students in the first-cohort MAF schools ever received the MAF intervention compared to less than 1 percent in never-MAF schools (nearly all of these ever-MAF students in never-MAF schools had transferred from MAF schools).

Another concern in this context is that the first-cohort MAF schools started implementing other interventions at the same time as MAF. For example, during the time frame we examine in this study, Kentucky also started implementing a K-3 reading intervention called Read to Achieve (RTA). If first-cohort MAF schools were more likely to implement RTA simultaneously compared to never-MAF schools, it would become harder to attribute the observed differences in student outcomes between the two types of schools in post-designation years to the causal effect of MAF. Panel (C) in Figure 1 repeats the same exercise in Panel (B), replacing MAF exposure with exposure to other interventions (including RTA). The share of students exposed to other interventions increased in both first-cohort MAF and never-MAF schools after 2014-15 school year, but students in MAF schools were no more likely to receive another intervention compared to students in never-MAF schools.

5.2 Determining Program Costs and Cost-Effectiveness

In order to gain a well-rounded understanding of whether the MAF program is a wise investment we cannot depend on estimates of impact alone. We must also consider the costs of implementing the program and assess these costs in relation to its effects through a cost-effectiveness analysis, which measures program cost per unit of outcome produced.

Our analysis employs the Ingredients Approach (Levin et al., 2018) to identify the costs of implementing the MAF program. The approach isolates the costs associated with the program by identifying the quantities of the personnel and nonpersonnel resources (ingredients) used in the implementation of the program and assigning corresponding prices to calculate their costs. Based on reviews of documentation and extant data, as well as discussions with MAF administrators, key ingredients involved in the implementation of MAF consist of activities related to intervention services, family engagement, assessment and monitoring, professional development, and administration (Table 2). School staff (MITs and P2Ts) surveys were used to quantify each ingredient. For personnel components, we asked staff to provide hours spent for a given program activity along with auxiliary costs related to activities in question, such as transportation and lodging/food costs for attending required formal professional development. We also collected data on staff years of experience and highest level of education attainment in order to apply appropriate compensation rates in the next step. Data for nonpersonnel resources such as software and equipment utilized during MAF-related activities, we asked respondents about the use of commonly used equipment and software.

The next step in the ingredients approach is to assign prices to the resources. The prices for personnel resources are compensation rates (salary plus benefits), which vary by years of experience and the highest level of education attainment. Importantly, the personnel compensation rates used reflect statewide averages, so that any variation in the subsequent costs across sites reflect differences in the qualifications of staff used and not the influence of local labor markets on the price of staff.[10] The unit prices of nonpersonnel resources involved in the MAF program were derived from information posted by major online retailers (e.g., BestBuy, Amazon, and HP).

The final step to calculate costs simply involves multiplying the quantity of each resource used by its corresponding unit price and summing across the resources. The final calculation of the overall program implementation cost for each school sums together the costs of personnel and the annualized costs of nonpersonnel resources. As resource allocation data was only collected for the 2018-19 school year, the costs for this year are translated into present values for each of the previous three years and summed to provide a four-year implementation cost for each school (in 2018-19 dollars) and expressed in per-pupil terms.[11] We calculate an enrollment-weighted average of the four-year program per-pupil costs across the cohort 1 grantee schools (using school K-5 enrollment as the weight), which serves as the key cost metric for the cost-effectiveness analysis.

This four-year weighted average per-student cost is then coupled with the estimated four-year MAF impacts on student outcomes to generate cost-effectiveness ratios. Each ratio shows the cost per unit of outcome produced by the intervention. For example, the cost-effectiveness of

[10] A detailed account of how the staff salaries and benefits that make up the compensation rates used to estimate costs can be found in Appendix A.
[11] Additional details on developing the program cost estimates can be found in Appendix A.

the MAF program with respect to outcome (o) is simply the average per-student cost ($\underline{AverageCostPerStudent_{MAF}}$) divided by the impact (μ_o), as follows:

$$Cost\text{-}Effectiveness_o = \frac{AverageCostPerStudent_{MAF}}{\mu_o}$$

6. Results

6.1 Effectiveness of MAF

Figure 2 presents the event study estimates (along with their 95% confidence intervals) for our four main outcomes of interest: math (Panel A) and reading (Panel B) scores on K-Prep in grades 3 through 5; whether the student was involved in a disciplinary incident in Panel C and percent absent days in Panel D (both estimated separately for grades K-5 and K-3). Table 3 presents these coefficients, estimated without (in columns labeled as (I)) and with (columns labeled as (II)) student covariates (i.e., an indicator for subsidized meal receipt, race/ethnicity, gender, special education status, English learner status, foreign-born indicator, and age).

The main takeaway from this analysis is that MAF had a significant positive effect on both test and non-test outcomes of students, especially beyond the first year after the designation. implementation. In particular, while we do not find any concerning evidence of differential pre-treatment trends between first-cohort MAF and never-MAF schools in the outcomes of interest, we find significant differences in the years after MAF designation. For example, MAF increased student mathematics test scores by 0.05 standard deviation after 2 years, 0.08 standard deviation after 3 years, and 0.09 standard deviation after 4 years. Similar, yet slightly smaller effects emerge for reading scores, with second-year effects of 0.03 standard deviation, third-year effects of 0.04 standard deviation, and fourth-year effects of 0.06 standard deviation. The gradually increasing effect of MAF is in line with (1) more students who received the MAF intervention in grades K-3 being tested in grades 3 through 5 and (2) the gradual implementation of effective practices regarding math instruction throughout the school.

We also find that MAF significantly reduced disciplinary incidents and student absences in participating schools. For example, MAF decreased the likelihood that students were involved in disciplinary incidents by 1 percentage points in the first year (by 25 percent of the dependent variable mean), 2 percentage points in the second year (by roughly 50 percent), and 3 percentage points in the third year (by 75 percent).[12] Similarly, MAF led to a decline in percent absent days of roughly 0.2 to 0.3 percentage points in the first four years (about 5 to 6 percent of the dependent variable mean). Panels (C) and (D) in Figure 2 also present the results for students in earlier grades (K-3) to see whether the effects were larger in the grades directly targeted by MAF interventions. The results suggest that the K-3 effects are mostly in line with the overall effects on non-test outcomes, which provides suggestive evidence that the effects of MAF designation go above and beyond the students chosen for the intervention.

Our findings also indicate that the MAF effects persist even after schools leave the program, which provides further evidence that systemic, schoolwide shifts in MAF schools are an important driver behind the observed benefits. In particular, Figure 3 breaks down the analysis in Figure 2 by whether the school remained as MAF in the second grant cycle (that started in the 2017-18 school year) or left the MAF program. The estimated MAF effects are comparable for the two types of schools in the first two years of their designation (2015-16 and 2016-17 school years). What is more interesting is that we find significant benefits of the program on student outcomes in the third and fourth years after their designation when they no longer receive MAF support from the state. For example, for these schools, we find MAF effects of 0.1 standard

[12] It is important to note that this decline could be driven by two factors: (1) change in student behavior and/or (2) increased leniency among teachers and school administrators for student misbehavior after MAF designation.

deviations in math, 0.06 standard deviations in reading, and significant reductions in disciplinary incidents and absences in the fourth year.

Do certain student groups benefit more from MAF? Figures 4A-4D examines this question and breaks down the analysis in Figure 2 by student grade (in tested grades) in Figure 4A, by whether the student received subsidized meals in Figure 4B, by student gender in Figure 4C, and by student race/ethnicity in Figure 4D. Table 4 presents the traditional, static DiD estimates (after MAF versus before MAF, in first-cohort MAF schools versus never MAF schools) obtained using the outcomes of interest in Figure 2 (with K-5 incidents and absence rates) overall and by student gender, subsidized meal receipt, and race/ethnicity. This table also presents the DiD results using an "ever-MAF" indicator that equals 1 if the student ever received the MAF intervention as the outcome to assess whether the MAF designation has a differential effect on the likelihood of receiving the intervention for different student groups.

The overarching conclusion from this analysis is that the benefits of MAF are widespread. In particular, we do not find any differences in MAF effects on outcomes of interest across grades, by student gender, or by student subsidized meal receipt although we find that MAF designation has a significantly larger effect on ever receiving the intervention for students who receive subsidized meals (an effect of 15 percentage points versus 8 for students who do not receive subsidized meals).

That said, we find significant differences between White and non-White (mostly Black and Hispanic students). While MAF designation does not have a differential effect on the likelihood of receiving the intervention for these two student groups, the benefits on test scores and non-test outcomes are significantly larger for racial minorities. For example, we find MAF effects of 0.13 standard deviation in math for racial minorities (compared to 0.03 standard deviation for White students), 0.10 standard deviations in reading for racial minorities (compared to 0.02 standard deviation for White students), a decline in disciplinary incidents of six percentage points for racial minorities (equivalent to 67 percent of the dependent variable mean for the student group) compared a decline of 0.8 percentage points for White students (26 percent of the dependent variable mean), and a decline of 0.28 percentage points in absence rates for racial minorities (compared to 0.10 percentage points for White students. The estimated coefficients for White and non-White students are statistically different at the 10 percent level (or lower) for outcomes of interest other than absence rates.

What about differential effects by the share of students receiving the MAF intervention? For example, if the observed MAF effect is primarily driven by the effect of the intervention on treated students, then one would expect the effects to be larger in schools with a larger share of students receiving the intervention. The primary challenge in this analysis is that the number of students who directly receive intervention services under MAF in a school is endogenous and it depends partly on the effectiveness of the intervention in prior years (i.e., an effective intervention in the prior year would improve students' mathematics skills in the current year and hence reducing the share of students receiving intervention services). Therefore, in Figure 5 we break down the analysis in Figure 2 by the share of K-3 students who received direct interventions in first-cohort MAF schools in 2015-16 school year (first year of implementation). We find no significant differences in MAF effects for MAF schools with higher (above median) and lower (below median) share of K-3 students identified for MAF in the first year of implementation. This finding once again provides evidence that the policy impacts student outcomes in ways beyond its effect on students directly receiving intervention services.

An alternative explanation for the observed MAF effect is the possible effect of MAF designation on student composition. For example, if the designation leads to lower-performing students leaving these schools or higher-performing students entering, it could lead to improved outcomes in MAF schools on average compared to other schools. Appendix Figure 1 examines this possibility and checks the effects of MAF designation on observed student characteristics (kindergarten readiness for students in kindergarten, whether the student receives subsidized meals, race/ethnicity, special education status, English learner status, gender, and immigrant status). Overall, we do not find evidence of shifts in student composition large enough to explain the observed effects of MAF on student outcomes. This is consistent with the finding presented in Table 3 which shows that the estimated effects of MAF remain virtually unchanged when we include student covariates in the regressions.

6.2 Effects for the Second MAF Cohort

How do these estimates compare with the effects on the second cohort of MAF schools? Appendix Figure 2 repeats the analysis in Panels (B) and (C) in Figure 1 for the schools that were first designated as MAF in the 2017-18 school year and Appendix Figure 3 presents the event study estimates obtained using these schools as the treatment group and never-MAF schools as the comparison group. Similar to the first cohort, MAF designation increases exposure to the MAF intervention and we find no significant differences in exposure to other interventions between first-time second-cohort MAF schools and never-MAF schools.

In terms of MAF effects on student outcomes, we find results that are somewhat in line with the findings using the first-cohort MAF schools in the first two years after their designation. We find no significant effects on math scores and absence rates after two years although it is important to note that these coefficients are less precisely estimated as we have fewer treatment schools in this exercise and there is a slight upward trend in MAF effects on math scores in years after the designation. In contrast, we find significant benefits of MAF designation on reading scores and disciplinary incidents in the second year. In all cases, we find no concerning differences in pre-treatment trends between treatment and comparison groups.

6.3 Cost-Effectiveness Estimates

The following section reports the results of both the cost and cost-effectiveness analyses. The findings provided for the main cost analysis are based on those first-cohort MAF schools that were two-time grant recipients and therefore implemented the program for four years from 2015-16 to 2018-19. The findings for the cost-effectiveness analysis are based on all schools in the impact analysis sample, which includes both cohort one MAF schools that were only grant recipients for two years and those that persisted in the program for four years. As reported earlier, MAF impact estimates are very similar between grantee schools that received grants for two and four years (see Figure 3).

The overall four-year per-pupil cost of MAF is $750 (Figure 6). As is the case with most education programs, most of the cost ($726 or 96%) is attributable to personnel resources. The breakout of costs by program activities in Figure 7 shows that over half of the overall program cost ($431 or 57%) is dedicated to providing direct intervention services to low-performing students, while just over a quarter ($194 or 26%) is spent on professional development for program teachers. Smaller shares of the program cost are associated with assessing students and monitoring their progress ($72 or 10%), program administration ($25 or 3%), and engaging families ($28 or 4%).

Figure 8 shows considerable variation in program cost across the grantee schools. We divide MAF grantee schools into terciles based on per-pupil costs. It depicts the average per-pupil costs for low-, medium-, and high-cost MAF schools, as well as how these costs break out across the program activities.[13] The average program cost per pupil for MAF schools in the top tercile is $1,062, nearly twice as much as the per pupil cost for MAF schools in the bottom tercile ($604). The cost breakout shows that the additional cost associated with the high-cost group can be attributed to greater spending across all activities, but especially on intervention services and administration.

Figure 9 illustrates how per-pupil program costs in 2018-19 vary with the scale at which MAF schools operate.[14] It shows a clear pattern consistent with economies of scale where the per-student cost of schools operating with a smaller enrollment tends to be higher than those with larger enrollments. For instance, schools with fewer than 300 K-5 students have annual per-pupil costs around $350, while the per-pupil costs of those with more than 600 students are close to $150.

Extrapolating from the cost estimates and the estimated impacts of MAF, Figure 10 provides cost-effectiveness estimates for the key student outcomes of interest: achievement scores on mathematics and reading tests, likelihood of behavioral incidents, and student absences. The program is estimated to cost $798 to produce a 0.1 standard deviation increase in mathematics test scores. The spillover effect of the program on reading suggests that the cost of a similar increase of 0.1 standard deviations in reading test scores is $1,317. The cost it takes to half the likelihood of a disciplinary incident from 4% to 2% is $580. Finally, reducing the average number of student absences over the year by half (from 8.9 to 4.4 days) costs $6,943.

---

[13] Specifically, we calculated the K-5 enrollment-weighted averages of per-pupil program costs for the seven least costly, six most costly schools, and six medium cost MAF schools from which resource allocation data was collected from. Then these averages were proportioned to the average cost of overall analysis sample cost of $750.
[14] The 19 schools are composed of 10 first-cohort MAF schools and 9 second-cohort MAF schools. The 9 second-cohort schools were included in Figure 9 because we included them in the imputation model for the 89 first-cohort schools with missing cost data. However the 9 second-cohort MAF schools were not included in the impact analysis sample which consists only of the first-cohort schools (2015-16 school year).

7. Concluding Remarks

Children starting their K–12 education today are more diverse in their knowledge and skills than before (Diamond et al., 2013), and K–12 schools are called upon to make up for deficiencies in foundational competencies that many entering children display due to their preschool experiences. Gaps in early-grade mathematics proficiency are stubbornly difficult to close, and early delays in mathematical development leads to lower growth rates in mathematics skills later. The MIT from one of the grantee schools wrote:

> "The reason we wrote the grant originally was because of our fall to fall data. It always showed no growth, and actually seemed to increase in need as the students got older. We were doing a great job of working really hard to reduce the number of tier 3 students from fall to spring, but only to have an increase in the number of students needing math interventions steadily climb as students got older."

Other MITs shared similar frustrations. How a state should design a coherent program that remedies and prevents mathematics development delays before they become intractable is an urgent question for state education policymakers to consider.

Findings from this study confirm that key elements of successful early mathematics interventions—small-group instruction, personalized support for students, peer-coaching, and close collaboration among teachers—can be scaled up moderately without losing effectiveness. To put the size of the MAF effect into perspective, the 4th year effect on mathematics achievement is equivalent to a class size reduction of 4-5 students (Angrist & Lavy, 1999) or replacing all novice teachers with 3-5 years of experience (Xu, Özek, & Hansen, 2015). But at a per-pupil cost of $750 for four years of intervention, MAF could be considerably cheaper than these other policy alternatives.

Importantly, the MAF benefits were found across grade levels and sustained even after grantee schools stopped receiving MAF support. One possible explanation for the sustained impact is the combination of targeted support for K-3 students and attention to improving mathematics instruction in regular classrooms. Researchers have pointed out that mathematics interventions—even when successful—must be paired with classroom practices to maintain the gains in mathematics skills achieved during the interventions (Clements et al., 2013; Smith et al., 2013; and Watts, Duncan, Clements & Sarama, 2017). Because of the critical role that post-intervention learning environments play, it is essential to improve regular classroom teaching even if the sole objective of a program is to help students who are struggling with mathematics.

MAF produced similar benefits for all students regardless of sex, subsidized meal status, or race/ethnicity. For some outcomes, the impact is larger for students who are eligible for subsidized meals and for racial minority students. These findings, if replicable in more schools and states, may hold the key to addressing inequalities in later life outcomes. This is because efforts to address achievement gaps in secondary schools tend to be too little, too late.[15] Some of the proposed policy options are likely met with more challenges in practice than early-grade interventions.[16]

The extent to which a program like MAF can be further scaled up remains an open question. The success of MAF, in our view, is partially attributable to the management of the program. With just over 90 grantee schools (fewer than 1/10 of all primary schools in Kentucky), MAF was closely managed by an experienced mathematics teacher at KDE. A centralized database was used to collect and monitor the number of intervention students and their progress

[15] Transition interventions, for example, are designed to help secondary school students get ready for college. Even though 39 states have offered transition intervention programs as of 2017, empirical evidence suggests that they had no effect on helping students succeed in college.

[16] For example, academic tracking (e.g., only advanced mathematics students can take algebra 2 in 8th grade) is shown to be related to within-school segregation of racial/ethnic groups (e.g., Clotfelter et al., 2021). But tracking, especially in mathematics, reflects differential skills accumulation over many years. Policies that call for opening access to advanced secondary mathematics courses to all students raise concerns about instructional challenges, diminished rigor, and extra workload for underprepared students to simultaneously learn advanced topics and close preexisting gaps in mathematics skills. Addressing gaps in mathematics skills early appears to be a more sensible policy choice.

multiple times a year. Individual MITs were contacted whenever discrepancy or data anomaly occurred. Before each school year, proposed daily schedules for all MITs were collected and approved. At the end of the school year, surveys were administered to collect feedback from MITs about their teaching and learning experiences. And at the end of each grant cycle, grantees that wished to renew were evaluated using a 14-point rubric. This level of involvement will be challenging when more schools are involved, and the quality of management will likely vary when multiple program managers are needed.

Even at the current scale, grantee schools reported several implementation challenges. These include difficulty in recruiting qualified MITs, turnovers of MIT and Plus 2 teachers, insufficient number of training slots, and difficulty in creating an intervention schedule that did not conflict with any core classes. These issues are likely to become more acute if more schools were to be included in programs like MAF.

In addition to how well programs like MAF could be further scaled up, future research should also investigate how accurately existing K-3 mathematics screeners reflect and predict mathematics skills development. Some MITs reported concerns that students identified for mathematics intervention might have deficiencies in language instead of mathematic skills. Although we find no research evidence for early-grade mathematics screeners, a study on interim assessments used for reading screening in North Carolina concluded that its K-2 screeners did not adequately identify students at risk of scoring below proficient on the state reading assessment at the end of grade 3 (Koon et al., 2020). There is a need for more research so that districts can make informed choices of early-grade mathematics screeners that are based on their reliability, validity, and classification accuracy (Petscher et al., 2019).

References

Altonji, J. G., & Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics, 116*(1): 313–50.

Austin, W., Figlio, D., Goldhaber, D., Hanushek, E., Kilbride, T., Koedel, C., … & Strunk, K.O. (2021). Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students. CALDER Working Paper No. 227-0821-2.

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools?. *Child Development, 83*(4), 1347-1367.

Behrman, J. R., Ross, D. R., & Sabot, R. (2008). Improving Quality versus Increasing the Quantity of Schooling: Estimates of Rates of Return from Rural Pakistan. *Journal of Development Economics, 85*(1–2): 94–104.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). The design of learning environments. *How people learn: Brain, mind, experience, and school*, 119-142.

Buysse, V., Castro, D.C., and Peisner-Feinberg, E. (2010). Effects of a Professional Development Program on Classroom Practices and Outcomes for Latino Dual Language Learners. *Early Childhood Research Quarterly, 25*(1), 94-206.

Chambers, J. (1999). Measuring Resources in Education: From Accounting to the Resource Cost Model Approach. Working Paper No. 1999-16. U.S. Department of Education. National Center for Education Statistics.

Chambers, J., & Parrish, T. (1994). Developing a resource cost database. In W.S. In Barnett (Ed.), Cost analysis for education decisions: Methods and examples (Vol. 4, pp. 7-21). Greenwich, CT: JAI.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal*, *50*(4), 812–850.

Clotfelter, C. T., Ladd, H. F., Clifton, C. R., & Turaeva, M. R. (2021). School segregation at the classroom level in a southern 'New destination' state. *Race and Social Problems, 13*(2), 131–160.

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment long-run impacts of double-dose algebra. *Journal of Human Resources, 50*(1), 108-158.

Cross, C. T., Woods, T. A., & Schweingruber, H. E. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. National Academies Press.

de Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*(9), 2964–2696.

Diamond, K.E., Justice, L.M., Siegler, R.S., & Snyder, P.A. (2013). *Synthesis of IES Research on Early Intervention and Early Childhood Education*. (NCSER 2013-3001). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology*, 43(6), 1428.

Figlio, D., Holden, K. L., & Ozek, U. (2018). Do students benefit from longer school days? Regression discontinuity evidence from Florida's additional hour of literacy instruction. *Economics of Education Review, 67*, 171-183.

Geary, D.C., Hoard, M. K., Nugent, L., & Bailey, H. D. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. PLoS ONE, 8(1): e54651.

Glewwe, P. (1996). The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development economics, 51*(2), 267-290.

Goldhaber, D., Wolff, M., & Daly, T. (2021). Assessing the Accuracy of Elementary School Test Scores as Predictors of Students' High School Outcomes. CALDER Working Paper No. 235-0821-2

Guthrie, G. (2015). The formalistic education paradigm in Papua New Guinea. *Contemporary PNG Studies, 22*, 33.

Hanushek, E., & Woessmann, L. (2008). The role of cognitive skills in economic development. Journal of Economic Literature, 46(3), 607-668.

Hershbein, B., Harris, B., & Kearney, M. (2014). *Major decisions: Graduates' earnings growth and debt repayment*. Washington, DC: Hamilton Project, Brookings Institution. http://www.hamiltonproject.org/assets/files/major_decisions_graduates_earnings_growth_debt_repayment.pdf.

Jacobse, A. E., & Harskamp, E. G. (2011). *A meta-analysis of the effects of instructional interventions on students' mathematics achievement*. Groningen: GION, Gronings

Instituut voor Onderzoek van Onderwijs, Opvoeding en Ontwikkeling, Rijksuniversiteit Groningen.

Koon, S., Foorman, B., & Galloway, T. (2020). *Identifying North Carolina students at risk of scoring below proficient in reading at the end of grade 3* (REL 2020–030). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://files.eric.ed.gov/fulltext/ED607256.pdf.

Lazear, Edward P. 2003. "Teacher Incentives." Swedish Economic Policy Review, 10(3): 179–214.

Levin, H., McEwan, P., Belfield, C., Bowden, B., Shand, R., (2018). Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis. Thousand Oaks, CA: Sage Publications.

Loeb, S., & Bassok, D. (2007). Early childhood and the achievement gap. *Handbook of research in education finance and policy*, 517-534.

Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D., Burchinal, M. Early, D.M., and Howes, C. (2008). *Measures of Classroom Quality in Pre-Kindergarten and Children's Development of Academic, Language, and Social Skills. Child Development, 79*(3), 732-749.

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind?. *Journal of learning disabilities, 44*(5), 472-488.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y. & Tyler, J. (2000). How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings. *Journal of Policy Analysis and Management*, 19(4): 547–68.

Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018, October). Effective programs in elementary mathematics: A best-evidence synthesis. In *annual meeting of the Society for Research on Educational Effectiveness*, Washington, DC.

Petscher, Y., Fien, H., Stanley, C., Gearin, B., Gaab, N., Fletcher, J.M., & Johnson, E. (2019). *Screening for Dyslexia*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy.

Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review, 19*, 158–172.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1–26.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating math recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal*, *50*(2), 397–428.

Stipek, D. (2002). At what age should children enter kindergarten? A question for policy makers and parents. *Society for Research in Child Development Policy Report, 16*, 13–16.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. Journal of Public Economics, 117, 162–181.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, 28(3), 551–576.

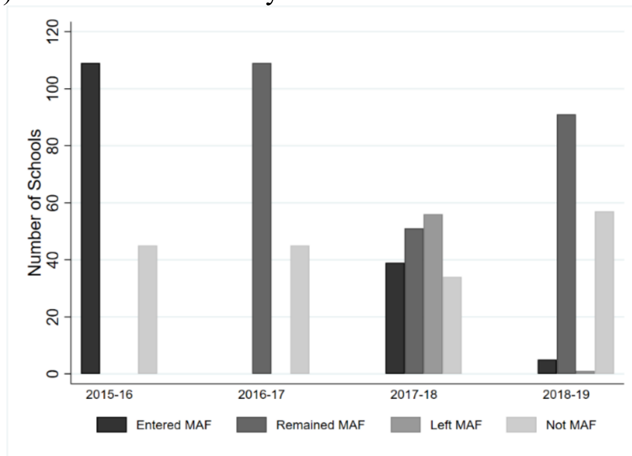
Watts, T. W., Clements, D. H., Sarama, J., Wolfe, C. B., Spitler, M. E., & Bailey, D. H. (2017). Does Early Mathematics Intervention Change the Processes Underlying Children's Learning?. *Journal of Research on Educational Effectiveness*, *10*(1), 96–115.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352-360.

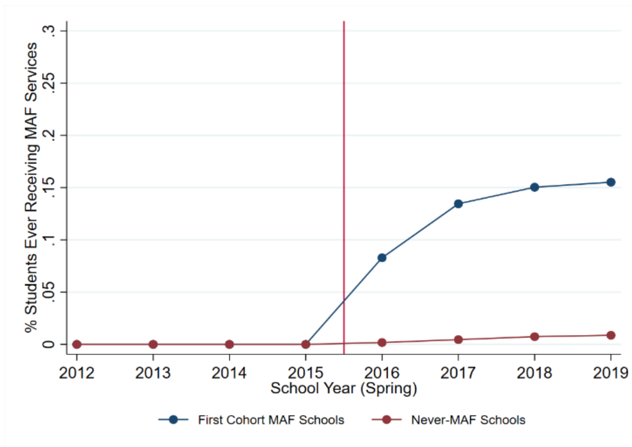
Figures and Tables

Figure 1. Exposure to MAF at the School and Student-Level

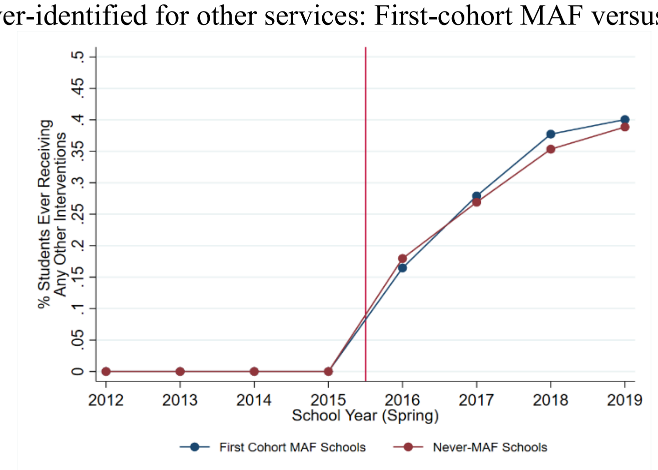
(A) Number of schools by MAF Status: Ever-MAF Schools



(B) % of ever-MAF students: First-cohort MAF versus never-MAF schools

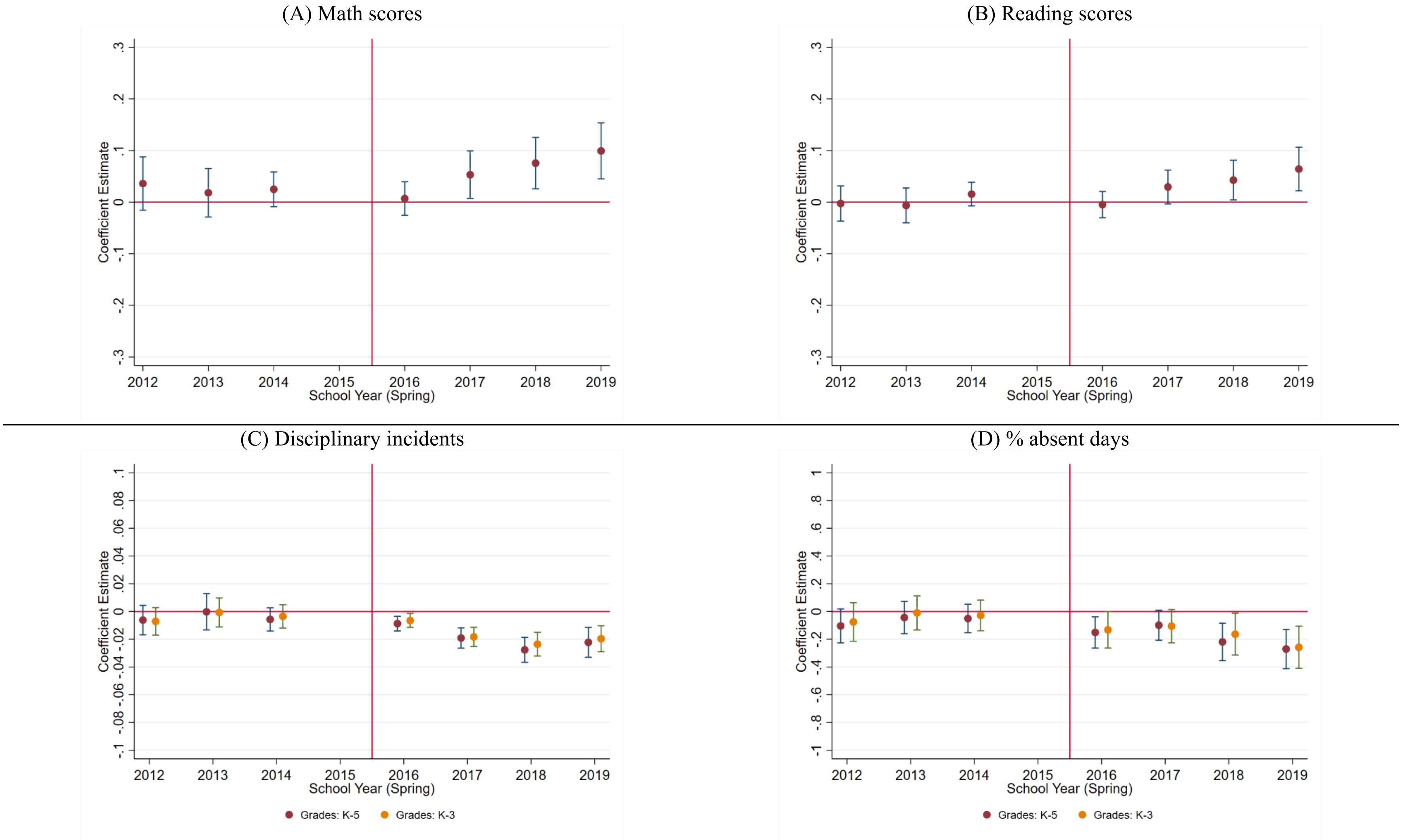


(C) % of students ever-identified for other services: First-cohort MAF versus never-MAF schools



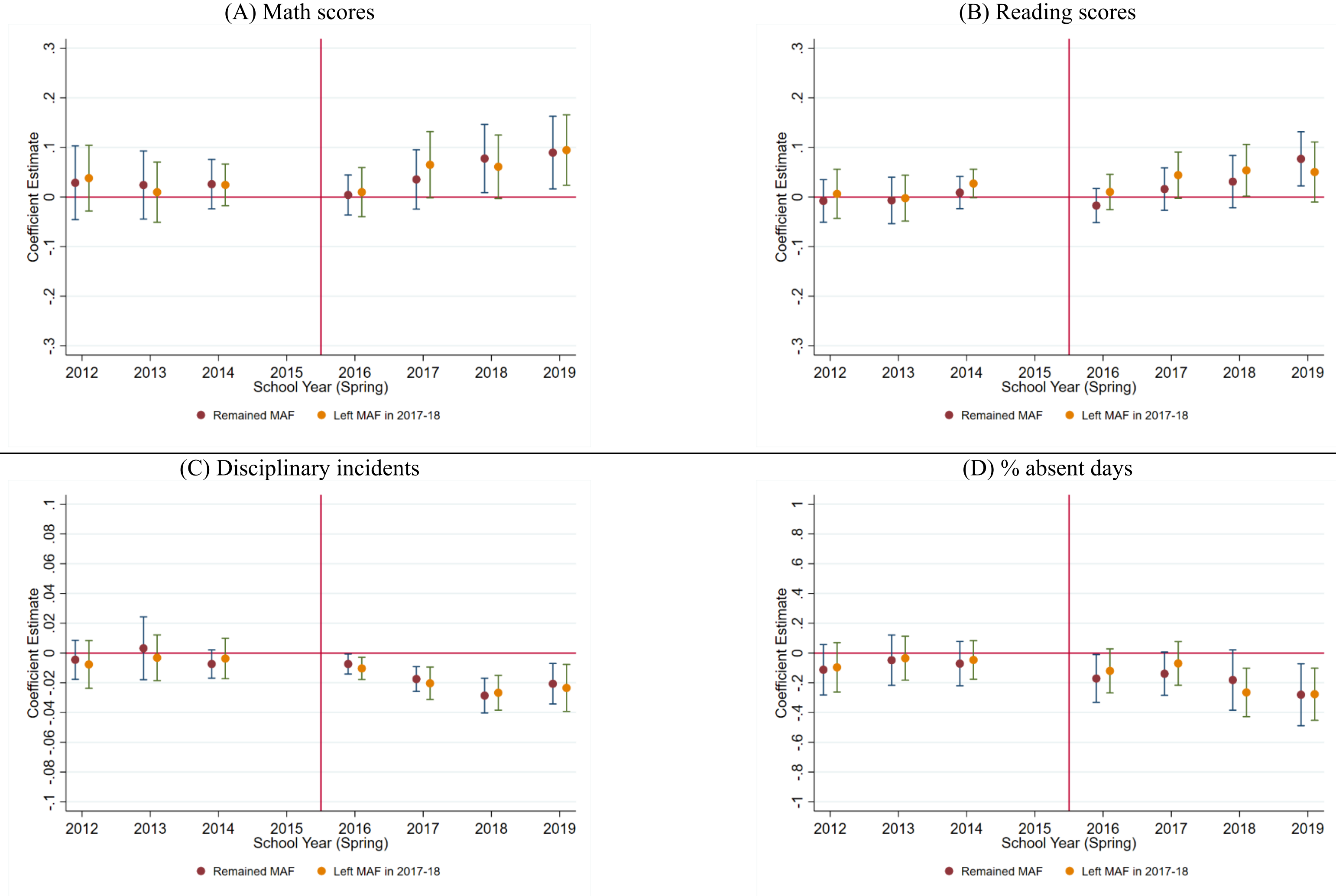
Notes: The top panel presents the number of elementary schools that were ever identified as MAF after 2015-16 by MAF status (entered, remained, left, and not MAF) by school year between 2015-16 and 2018-19 school years whereas the bottom two panels presents the percentage of students in first-cohort MAF schools (elementary schools that were first designated as MAF in 2015-16 school year) and never-MAF schools who have ever received MAF services (panel B) or other services (panel C) up to (and including) that school year.

Figure 2. Effects of MAF Designation on Student Outcomes: Event Study Estimates



Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

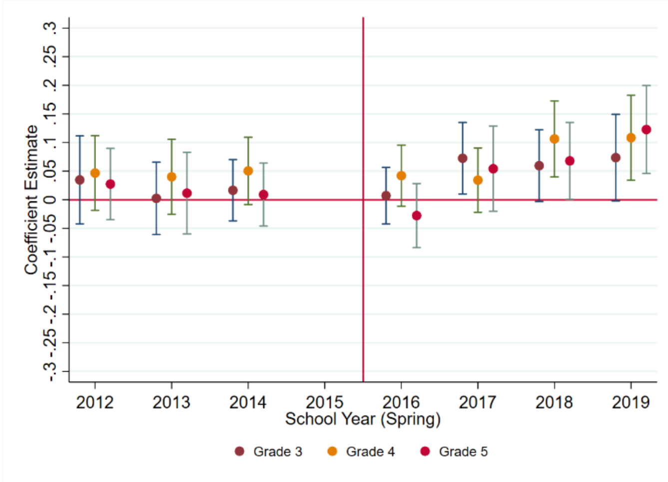
Figure 3. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by School Duration in MAF



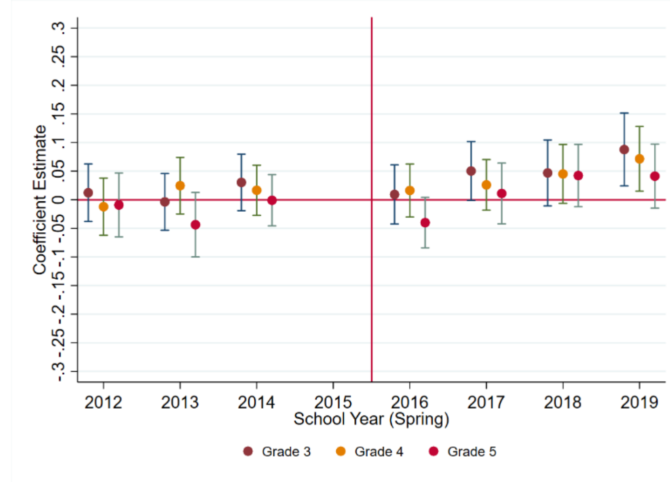
Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools that left MAF at the end of 2017-18 and those that remained as MAF until 2018-19 as the treatment groups and never-MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

__Figure 4A. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by Grade__

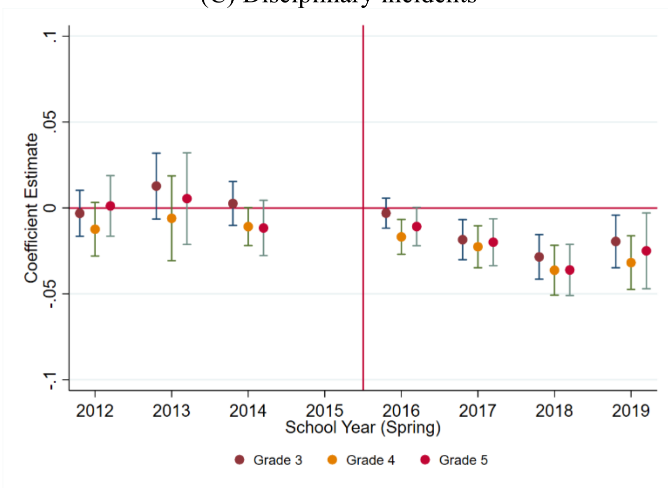
(A) Math scores



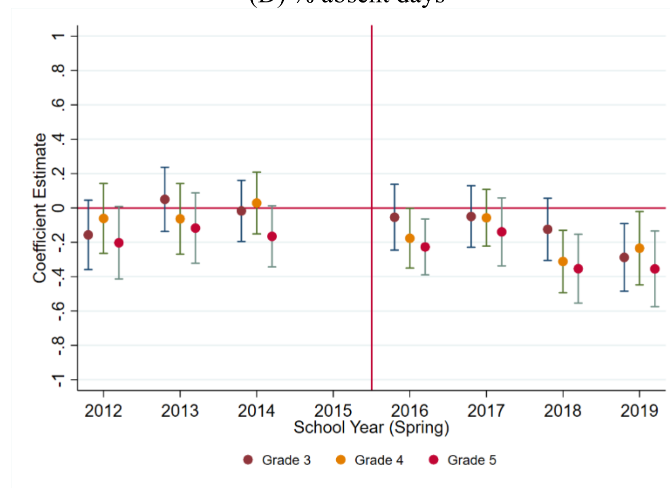
(B) Reading scores



(C) Disciplinary incidents



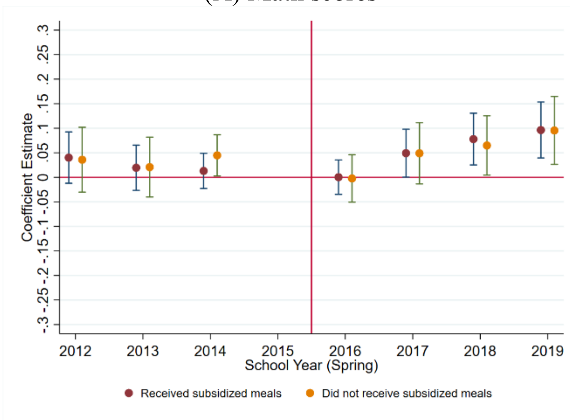
(D) % absent days



Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

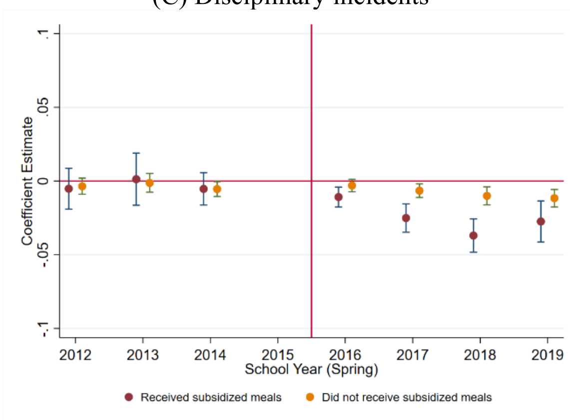
Figure 4B. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by Student Subsidized Meal Receipt

(A) Math scores

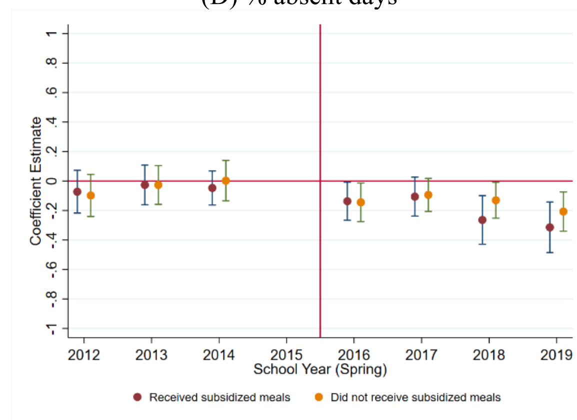


(B) Reading scores

(C) Disciplinary incidents



(D) % absent days

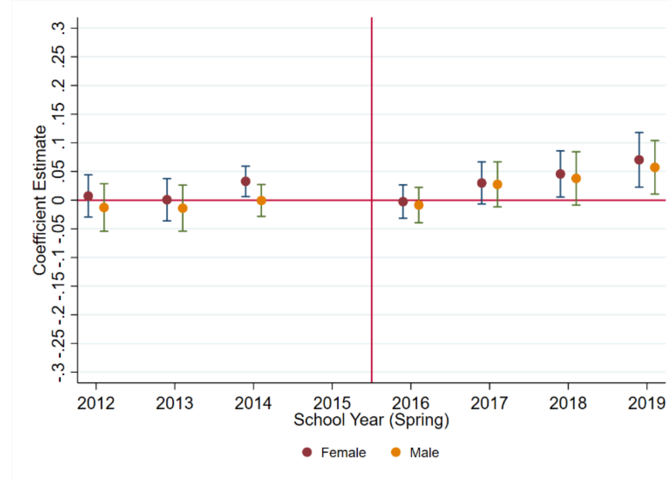


Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

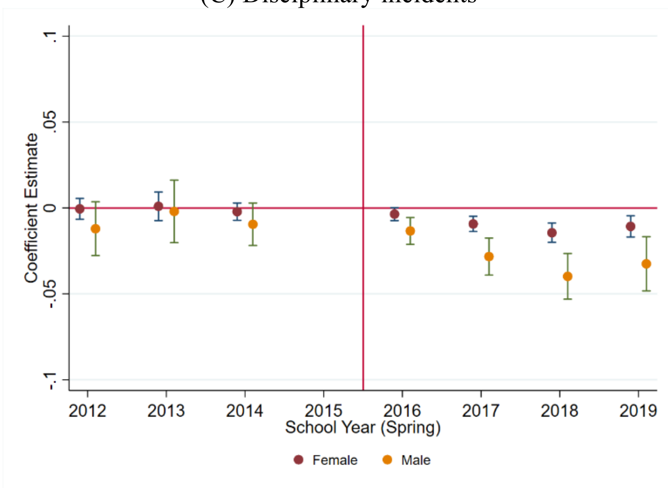
Figure 4C. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by Student Gender

(A) Math scores

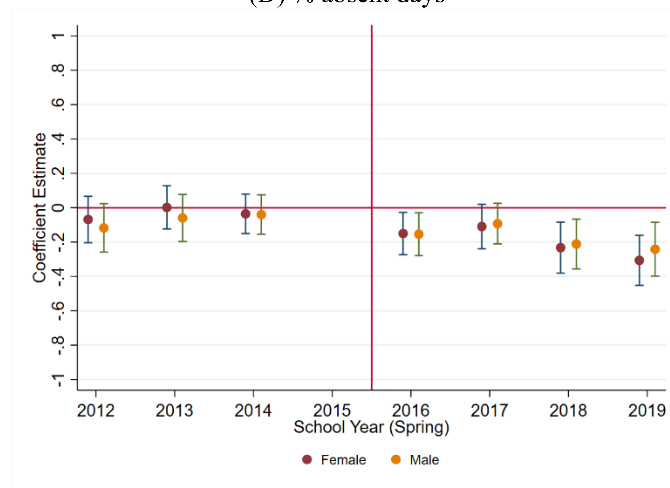
(B) Reading scores



(C) Disciplinary incidents

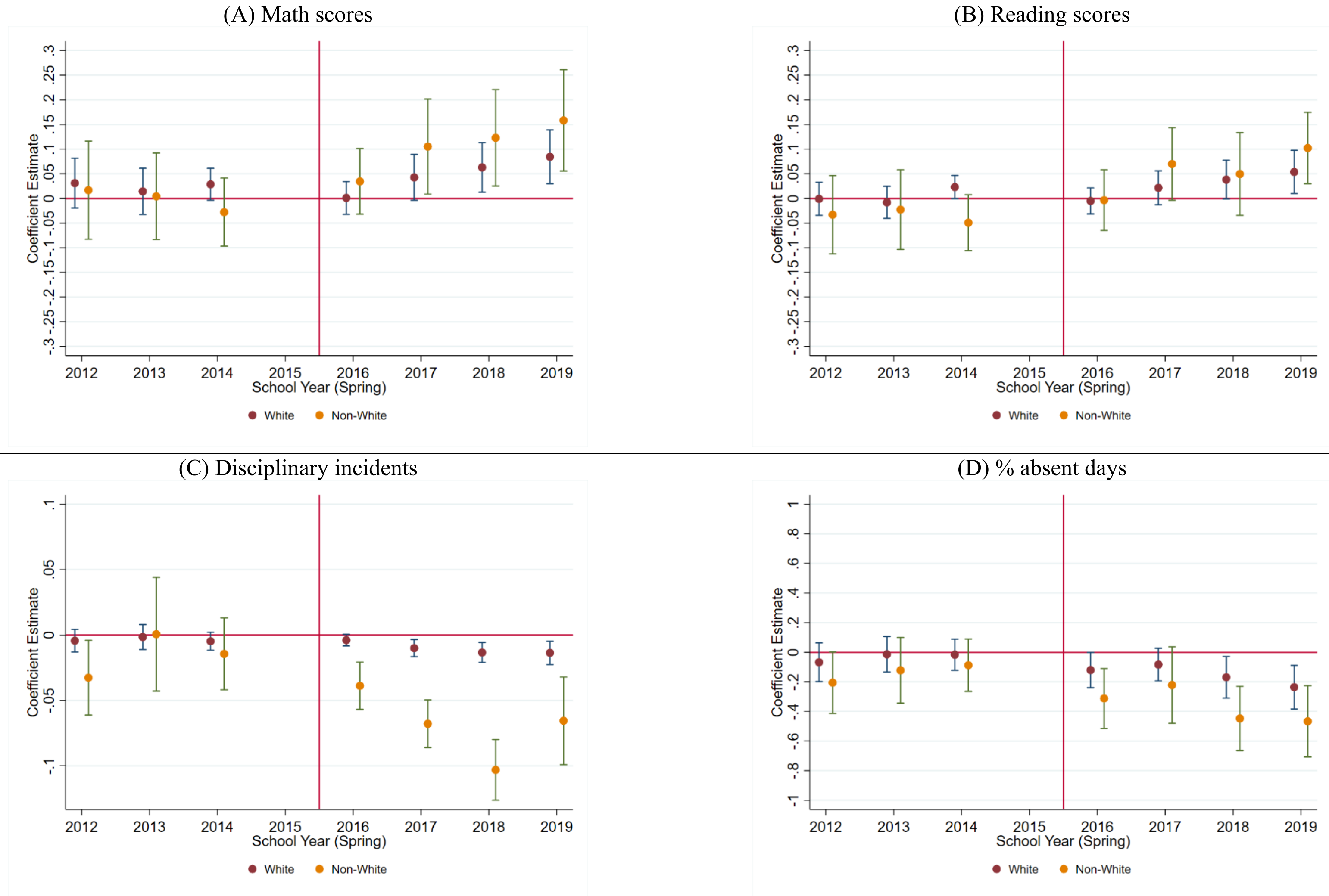


(D) % absent days



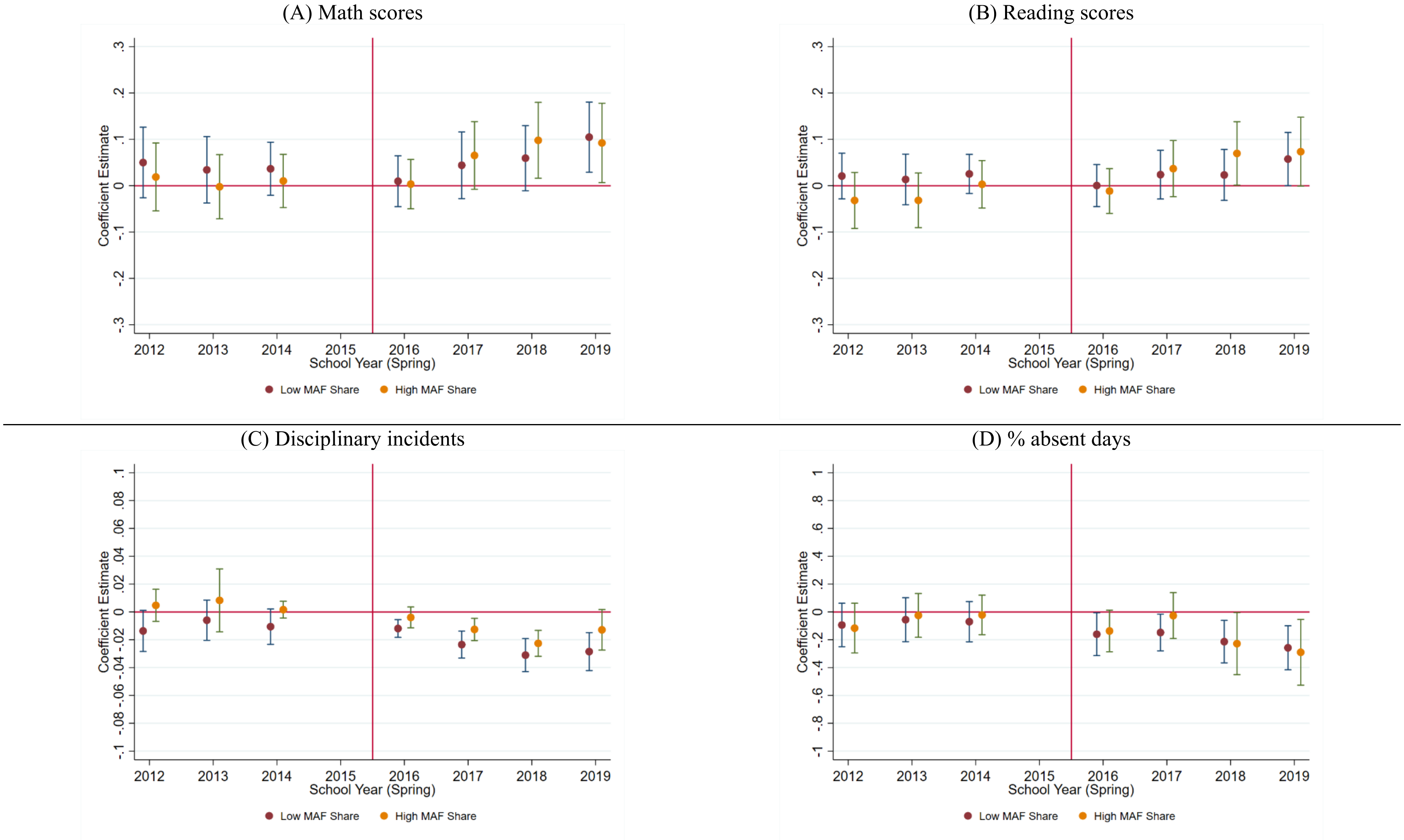
Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

Figure 4D. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by Student Race/Ethnicity



Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

Figure 5. Effects of MAF Designation on Student Outcomes: Event Study Estimates, by the Share of K-3 Students Receiving MAF Intervention in 2015-16 in First Cohort MAF Schools



Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

Figure 6. Average MAF Cost Per K-5 Pupil by Personnel and Nonpersonnel for MAF Implementation Schools (in 2018-19 Dollars)

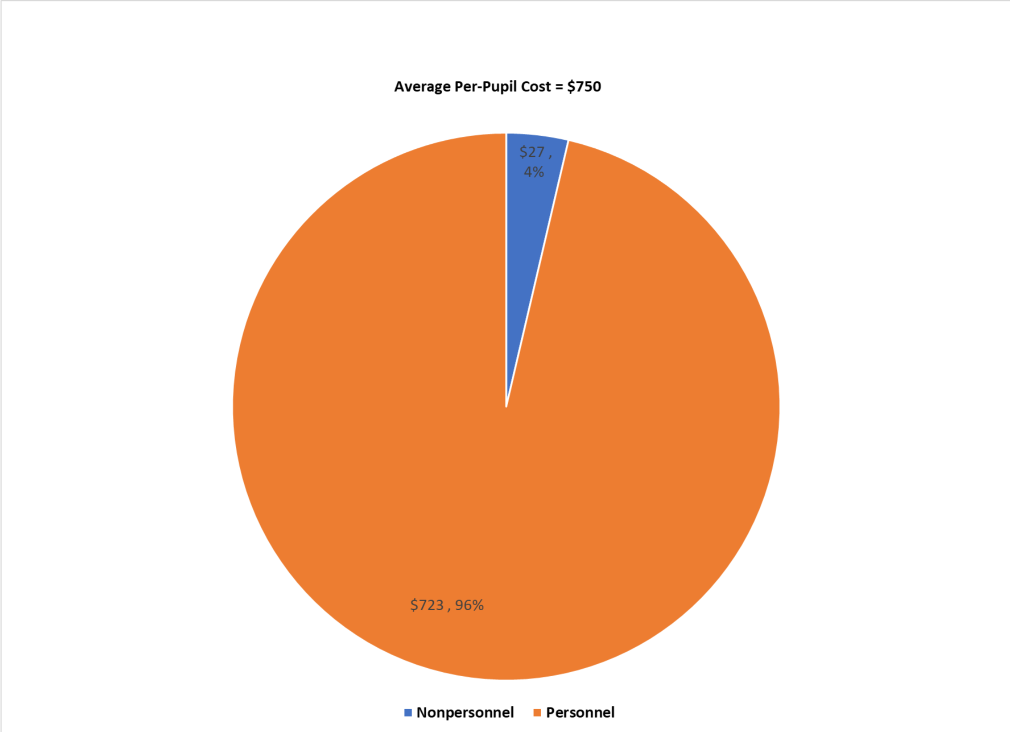


Figure 7. Average MAF Cost Per K-5 Pupil by Program Activity for MAF Implementation Schools (in 2018-19 Dollars)

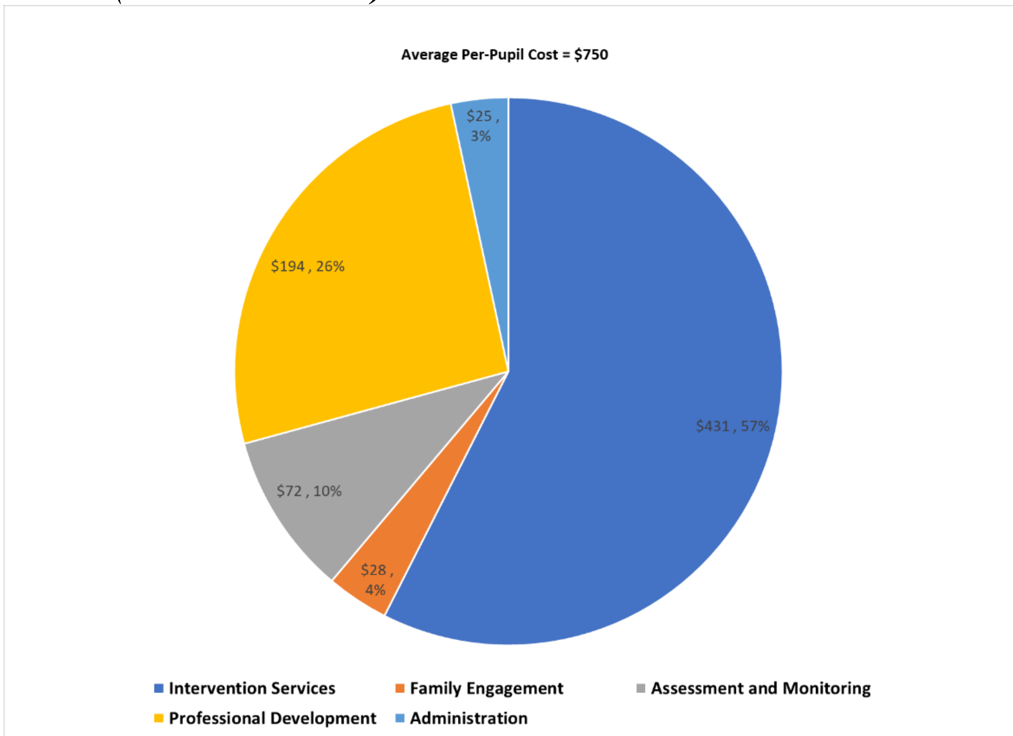


Figure 8. Average MAF Program Cost Per Pupil by Activity in High, Medium and Low Cost MAF Implementation Schools (in 2018-19 Dollars)

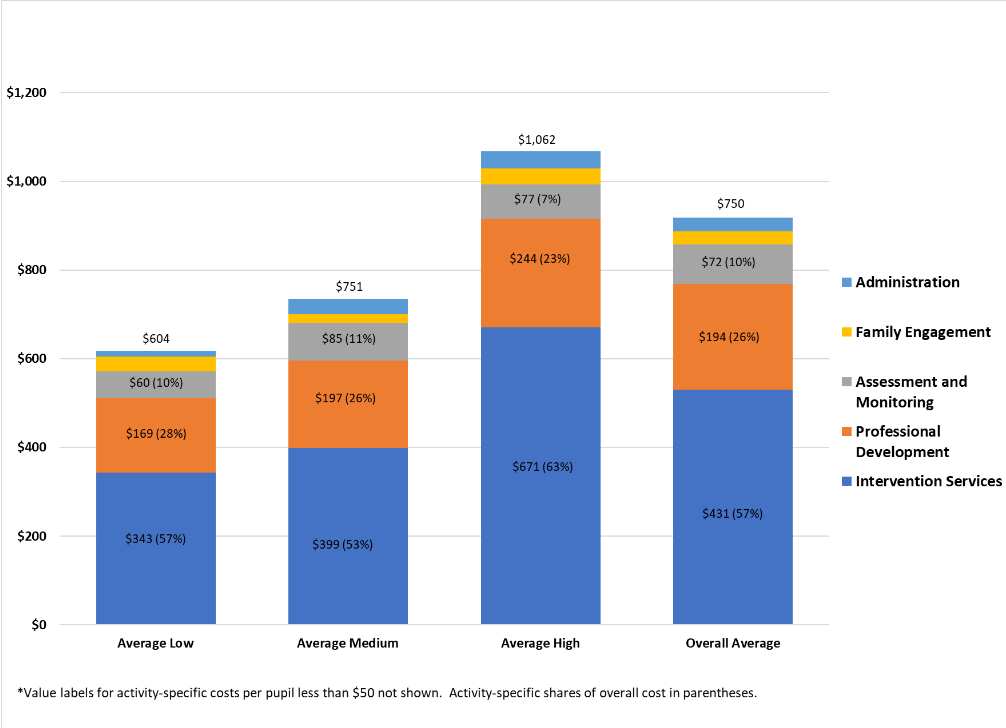


Figure 9. Relationship Between Annual Program Cost Per K-5 Pupil and K-5 Enrollment for MAF Implementation Schools (in 2018-19 Dollars)

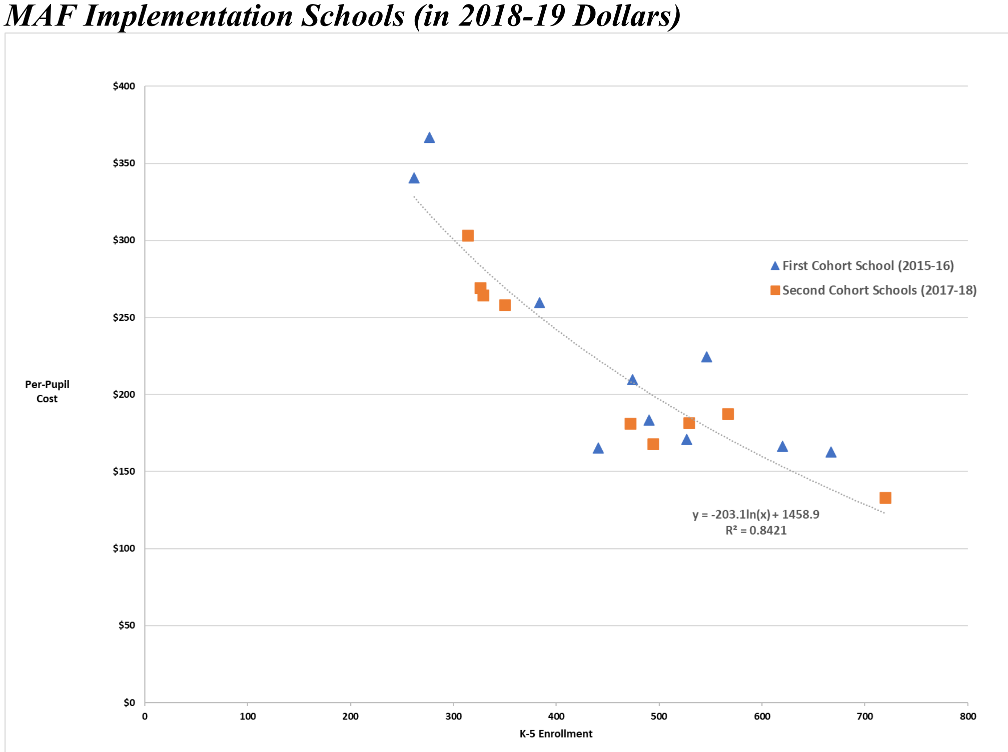


Figure 10. Cost-Effectiveness of MAF Program in Terms of Achievement Test Scores, Likelihood of Behavioral Incidents and Absence Days (in 2018-19 Dollars)

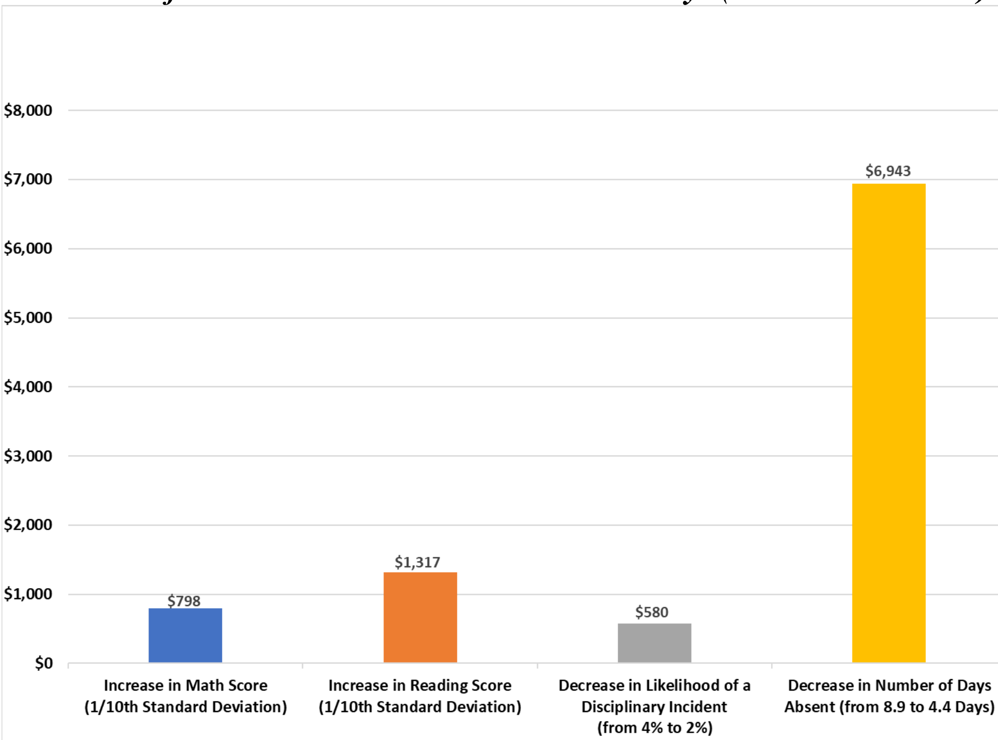


Table 1. Student Characteristics by School MAF Status, Elementary Schools between 2011-12 and 2014-15 School Years

	School in First MAF Cohort	School in Second MAF Cohort	Never MAF School
Math scores (Grades 3-5)	-0.019	-0.019	0.008
the	(0.974)	(0.960)	(1.005)
KG ready	0.351	0.339	0.356
	(0.477)	(0.473)	(0.479)
% absent days	5.052	4.682**	4.641**
	(4.741)	(4.383)	(4.566)
Disciplinary incident	0.034	0.037	0.029
	(0.181)	(0.189)	(0.167)
Received subsidized meals	0.661	0.655	0.602**
	(0.473)	(0.475)	(0.490)
Special education	0.113	0.109	0.103**
	(0.316)	(0.311)	(0.304)
English learner	0.028	0.026	0.035
	(0.165)	(0.159)	(0.183)
White	0.882	0.887	0.828**
	(0.323)	(0.316)	(0.378)
Black	0.070	0.067	0.116***
	(0.255)	(0.250)	(0.320)
Hispanic	0.058	0.059	0.064
	(0.234)	(0.235)	(0.245)
Asian	0.008	0.008	0.015***
	(0.090)	(0.091)	(0.121)
Foreign born	0.001	0.001	0.003***
	(0.034)	(0.038)	(0.055)
Age	8.771	8.808	8.882**
	(1.890)	(2.026)	(1.892)
Number of unique students	80,947	32,591	436,972
Number of unique schools	107	39	720

Notes: Standard deviations are given in parentheses. Math scores are standardized at the grade-year level to zero mean and unit variance. *, **, and *** represent that the means for the corresponding group are statistically different from the first column (students in the first MAF cohort of schools) at 10, 5, and 1 percent, respectively.

Table 2. Key MAF Program Activities and their Definitions Accounted for in the Cost Analysis.

Program Activity	Definition
Intervention Services	Delivery of MAF program instruction to students including preparing lessons, providing direct instruction, and collaborating on instruction with other teachers.
Family Engagement	Engagement with parents through conferences and events.
Assessment and Monitoring	Administration of assessments, identification of MAF students, and monitoring of student progress.
Professional Development	Participation in professional development opportunities related to the MAF program, including formal training provided by Kentucky Center for Mathematics (KCM) and provision of professional development and coaching by Mathematics Intervention Teacher to other staff within schools.[17]
Administration	Planning and budgeting, scheduling service delivery, and entering data on service provision and student progress.

[17] See description of training on KCM website: https://www.kentuckymathematics.org/maf.php.

Table 3. Effects of MAF Designation on Student Outcomes: Event Study Estimates

Treatment: first cohort MAF schools, Comparison: never-MAF								
	Math scores		Reading scores		Disciplinary incidents		% absent days	
	(I)	(II)	(I)	(II)	(I)	(II)	(I)	(II)
MAF*								
2011-12 SY	0.036	0.031	-0.003	-0.008	-0.006	-0.006	-0.104*	-0.100
	(0.026)	(0.026)	(0.017)	(0.017)	(0.005)	(0.005)	(0.062)	(0.062)
2012-13 SY	0.018	0.014	-0.006	-0.010	-0.000	-0.000	-0.044	-0.036
	(0.024)	(0.023)	(0.017)	(0.016)	(0.007)	(0.007)	(0.059)	(0.058)
2013-14 SY	0.025	0.024	0.016	0.015	-0.006	-0.005	-0.051	-0.043
	(0.017)	(0.017)	(0.012)	(0.012)	(0.004)	(0.004)	(0.052)	(0.051)
2015-16 SY	0.007	0.004	-0.005	-0.008	-0.009***	-0.009***	-0.151***	-0.139**
	(0.017)	(0.017)	(0.013)	(0.013)	(0.003)	(0.003)	(0.058)	(0.058)
2016-17 SY	0.053**	0.053**	0.029*	0.030*	-0.019***	-0.019***	-0.099*	-0.094*
	(0.024)	(0.024)	(0.017)	(0.017)	(0.004)	(0.004)	(0.055)	(0.055)
2017-18 SY	0.076***	0.076***	0.043**	0.042**	-0.028***	-0.028***	-0.219***	-0.218***
	(0.025)	(0.025)	(0.020)	(0.020)	(0.005)	(0.005)	(0.069)	(0.068)
2018-19 SY	0.099***	0.094***	0.064***	0.057***	-0.022***	-0.022***	-0.271***	-0.273***
	(0.028)	(0.028)	(0.022)	(0.021)	(0.005)	(0.005)	(0.072)	(0.072)
Mean of Y					0.040	0.040	4.755	4.752
N	1,133,884	1,130,904	1,133,884	1,130,903	2,403,278	2,388,617	2,371,098	2,357,223
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Robust standard errors, clustered at the school level, are given in parentheses. The estimates represent the estimated coefficient on the interaction term given with the 2014-15 school year (the year before the policy took effect) serving as the baseline category. Student covariates include an indicator for subsidized meal receipt, race/ethnicity, gender, special education status, English learner status, foreign-born indicator, and age. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 4. Effects of MAF Designation on Receiving MAF Services and Student Outcomes, DiD Estimates, Overall and by Subgroup

Treatment: first cohort MAF schools, Comparison: never-MAF

	All	Female	Male	Subsidized meals	No subsidized meals	White	Non-White
Ever received MAF services	0.125***	0.133***	0.117***	0.146***	0.078***	0.125***	0.128***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)
90% CI	[0.117, 0.133]	[0.125, 0.142]	[0.109, 0.124]	[0.137, 0.155]	[0.070, 0.086]	[0.117, 0.132]	[0.114, 0.141]
Mean of Y	0.012	0.013	0.011	0.016	0.006	0.013	0.009
Math scores	0.039**	0.035*	0.042**	0.043**	0.025	0.025	0.126***
	(0.020)	(0.020)	(0.021)	(0.020)	(0.026)	(0.020)	(0.034)
90% CI	[0.007, 0.071]	[0.002, 0.068]	[0.008, 0.077]	[0.009, 0.076]	[-0.018, 0.068]	[-0.008, 0.058]	[0.069, 0.182]
Reading scores	0.030*	0.022	0.036**	0.023	0.033*	0.019	0.103***
	(0.016)	(0.017)	(0.018)	(0.018)	(0.019)	(0.016)	(0.035)
90% CI	[0.004, 0.056]	[-0.005, 0.050]	[0.007, 0.065]	[-0.006, 0.052]	[0.001, 0.064]	[-0.007, 0.046]	[0.045, 0.160]
Disciplinary incident	-0.016***	-0.009***	-0.023***	-0.023***	-0.005***	-0.008***	-0.058***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.002)	(0.002)	(0.009)
90% CI	[-0.021, -0.012]	[-0.012 -0.006]	[-0.030, -0.016]	[-0.029, -0.017]	[-0.008, -0.002]	[-0.011, -0.004]	[-0.074, -0.043]
Mean of Y	0.039	0.017	0.061	0.054	0.015	0.031	0.086
% absent days	-0.135***	-0.163***	-0.108**	-0.160***	-0.106**	-0.100**	-0.277***
	(0.049)	(0.050)	(0.054)	(0.060)	(0.045)	(0.050)	(0.085)
90% CI	[-0.215, -0.054]	[-0.245, -0.080]	[-0.196, -0.019]	[-0.258, -0.062]	[-0.180, -0.031]	[-0.182, -0.017]	[-0.417, -0.137]
Mean of Y	4.752	4.759	4.746	5.493	3.472	4.824	4.389
N (tested grades)	1,130,904	554,037	576,863	704,848	426,056	942,064	188,840
N (all grades)	2,388,617	1,159,087	1,229,507	1,500,561	888,056	1,999,506	389,111

Notes: Robust standard errors, clustered at the school level, are given in parentheses. The estimates represent the estimated coefficient on the interaction term given with the 2014-15 school year (the year before the policy took effect) serving as the baseline category. All regressions control for school, year, and grade fixed- effects, and the student covariates including an indicator for subsidized meal receipt, race/ethnicity, gender, special education status, English learner status, foreign-born indicator, and age. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Appendix A. Calculating Costs

The following section describes the methods used to calculate the costs of the personnel and nonpersonnel resources used to implement the MAF program.

Calculating Personnel Costs

The costs of MIT and P2T personnel were calculated by applying the full-time equivalent teacher salaries for teachers predicted from a model using data from the Kentucky Department of Education (KDE) and a benefits rate based on federal data to time allocations obtained from administrative information on formal required professional development and from teacher surveys. Specifically, a statewide dataset of annual teacher salaries and background was used to estimate the following equation:

$$Salary_{id} = \alpha + \sum_{j=2}^{4} \beta_j Rank_{idj} + \gamma Experience_{id} + \delta Experience_{id}^2 + \sum_{l=2}^{5} \epsilon_l SchoolLevel_{idl} + \theta CWIFT_d + \varepsilon_i$$

where *Salary* is the annual fulltime equivalent salary of an elementary school teacher in the 2017-18 school year, *Rank* denotes a measure of career advancement based on educational attainment and certification status,[18] *Experience* is years of experience, and *SchoolLevel* is the schooling level taught,[19] *CWIFT* is the Comparable Wage Index for Teachers,[20] ε is a random error term, and the subscripts *i* and *d* represent individuals and districts. The full compensation rate for each teacher was calculated by multiplying their predicted salary by a constant statewide

[18] For more information on the Kentucky Department of Education rank system, see https://education.ky.gov/edprep/cert/Pages/Rank-System.aspx. Rank I (Master's degree plus 30 credit hours and a regular certificate) serves as the reference group. Indicators for ranks II, III and IV represent teachers with a Master's degree and no additional credit hours and a regular certificate, those with a Bachelor's degree and a regular certificate, and those with less than a Bachelor's degree and an emergency certificate, respectively.

[19] Elementary school serves as the reference group, with additional indicators for prekindergarten, middle, high and other.

[20] The CWIFT is a measure of the differential cost of hiring and retaining teaching staff in different labor markets. For more information on the CWIFT, see https://nces.ed.gov/programs/edge/Economic/TeacherWage.

benefit ratio.[21] These rates were used to calculate the dollar value of teacher time devoted to each MAF program activity.

Calculating Nonpersonnel Costs

Data on nonpersonnel resources such as software and equipment utilized during MAF related activities, was collected through survey items that asked about whether equipment and software items common in elementary classroom settings were used in delivering MAF program services. We calculate the final cost of equipment by measuring the use of each for MAF activities only and converting these to capacity utilization rates that could be projected against the annualized cost of each item.[22] Calculated costs for software subscriptions take into account variation in school enrollment size (i.e., costs are adjusted accordingly to each school's K-5 enrollment numbers). We also calculate the average cost of office space for each MIT where we assume administration and direct instructional services take place by using the national average price of elementary school office space obtained from the Center for Benefit-Cost Studies of Education *CostOut* toolkit.[23]

Calculating Total Costs

We calculate program costs on a per-pupil basis by summing up the personnel and nonpersonnel resource costs for each school and dividing by their K-5 enrollment. Because we only collected data for the 2018-19 school year and the program for the first-cohort schools

---

[21] The ratio of education staff benefits to salaries was obtained from the National Center for Education Statistics "Digest of Education Statistics, 2019," *Table 235.20. Revenues for public elementary and secondary schools, by source of funds and state or jurisdiction: 2016–17*. The compensation rate for each individual teacher *i* is calculated as follows: $Compensation_i = Salary_i \times \left(1 + \frac{Kentucky\ Statewide\ Total\ Benefit\ Expenditure}{Kentucky\ Statewide\ Total\ Salary\ Expenditure}\right)$.

[22] Annualization is used to spread out the cost of resources that remain useful over multiple years and therefore are "consumed" over this period rather than immediately (Levin et al., 2018). The calculation of annualized costs for this study assumed a 3 percent discount rate and different life spans over which different nonpersonnel items remain useful (e.g., a 5-year life span for laptop computers).

[23] The final calculated cost assumes 150sq ft in size, full capacity utilization equal to 11 hours a day, 180 days of instructional school days, and usage capacity estimated by applying only hours for those program activities requiring the office space.

started three years earlier (in 2015-16), we proxy the costs for the prior three years by calculating the present values of the calculated 2018-19 program cost for each of the previous years using a 3% interest rate. The sum of these year-specific costs yields estimates of the overall per-pupil costs of MAF program implementation over four years of program participation for the schools that responded to the MIT survey.

To facilitate the cost-effectiveness analysis, we impute per-pupil costs for the first-cohort schools that lacked MIT survey responses (and therefore had no data with which to calculate costs) in order to obtain an average cost across all of the MAF schools included in the main 4-year impact analysis.[24] We impute missing costs with predictions from a simple linear regression model using data from the 19 first- and second-cohort schools for which we were able to calculate program costs:

$$Cost_j = \alpha + \beta_j Enrollment + \varepsilon_j$$

where *Cost* is the per-pupil cost in the 2018-19 school year, *Enrollment* is K-5 enrollment, ε is a random error term, and the subscript *j* represents the MIT schools. We also tested a second equation with an indicator for MAF cohort 2 added as a covariate, but because the coefficient for this indicator was not found to be statistically significant, we used the more parsimonious model to perform our imputation.

After imputing for the 2018-19 school year costs and discounting them for the previous years for the schools which we do not have survey data, we only accounted the costs for the first two years from the 51 schools that did not reapply for the second MAF grant cycle when

[24] Surveys were only administered to staff at second-cohort MAF schools that were operational in the 2018-19 school year. To this end, out of 107 schools that participated in the first MAF grant cycle, 51 schools were available for the survey by the 2018-19 school year, of which the survey data for 37 schools were not available due to nonresponse and 4 schools' survey data were not eligible for the impact analysis due to missing outcome data. From the remaining 56 first-cohort schools that did not continue in the second cohort, 3 schools were ineligible for the impact analysis and their cost data did not have to be imputed.

calculating the total costs because they were not operational as MAF schools in the second-cohort (between 2017-18 and 2018-19). The average four-year per-pupil cost for the first cohort MAF schools is calculated using two-year costs from the 51 schools and four-year costs from the other 48 schools that were part of both grant cohorts.

In addition to utilizing the cost data of the 19 first- and second-cohort schools for the imputation of the missing data, we also reflect the cost allocation patterns of those 19 schools on the cost allocations analysis results for all of the MAF schools included in the impact analysis. Using the average per-pupil cost of the impact analysis sample ($750) as the base cost, we calculate the cost shares by function and type to be equivalent to the cost shares of each function and type from the average costs of the 19 schools.

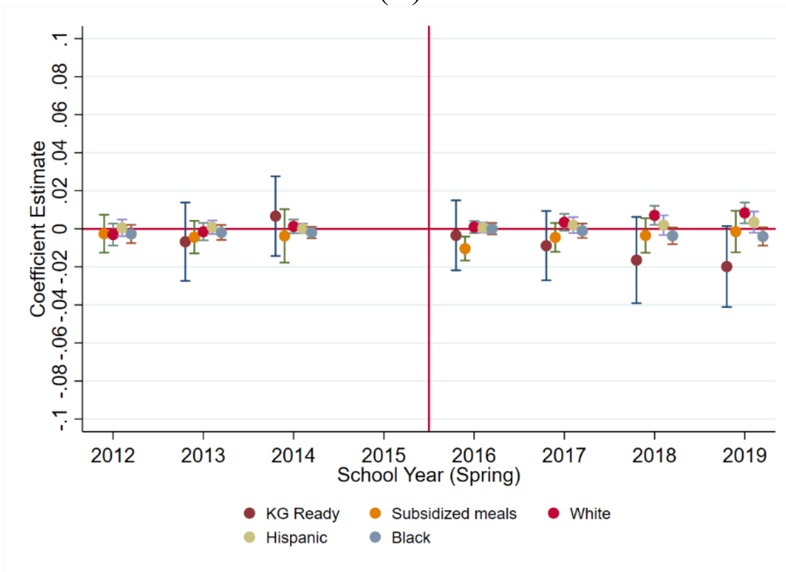
Our use of data from both the first- and second-cohort MAF schools, instead of only using the first-cohort MAF schools that have the necessary data for the main impact analysis, was done for two reasons. First, the combined sample of first and second cohort schools for which we have survey data and therefore could calculate program costs more closely resembles those schools lacking this information with respect to a variety of student characteristics than the first cohort school sample with survey data.[25] Second, we use both cohorts of data to enlarge the sample size for the imputation regression and cost allocation analysis, and increase the precision of the estimated model.

[25] Appendix Table 1 presents aggregated school-level student characteristics across the schools with and without survey data showing almost no statistical difference between the group of first- and second-cohort survey respondents and the first-cohort survey nonrespondents.

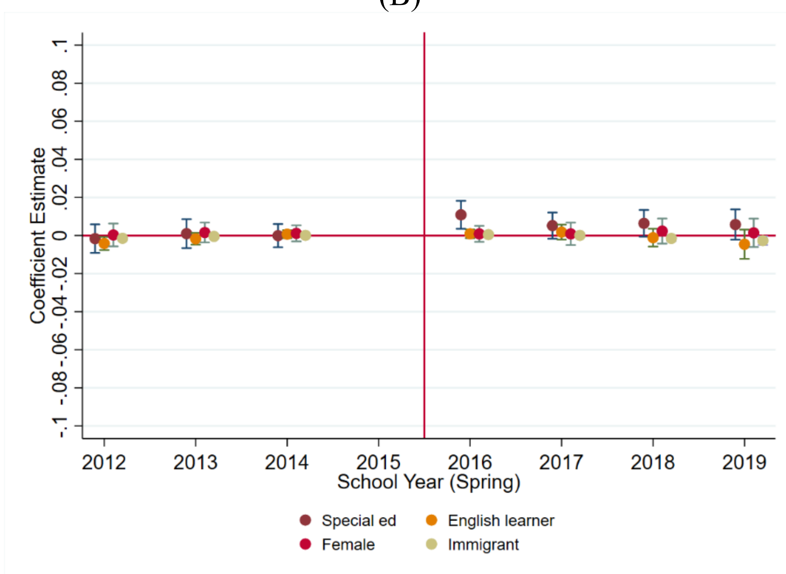
Appendix B. Figures and Tables

Appendix Figure 1. Estimated Effects of MAF Designation on Student Composition: Event Study Estimates

(A)



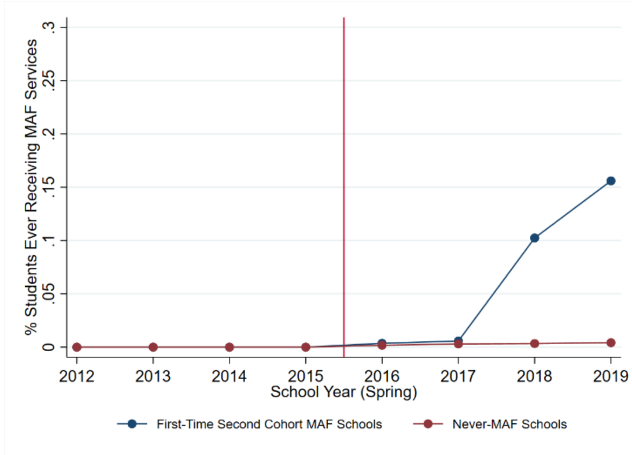
(B)



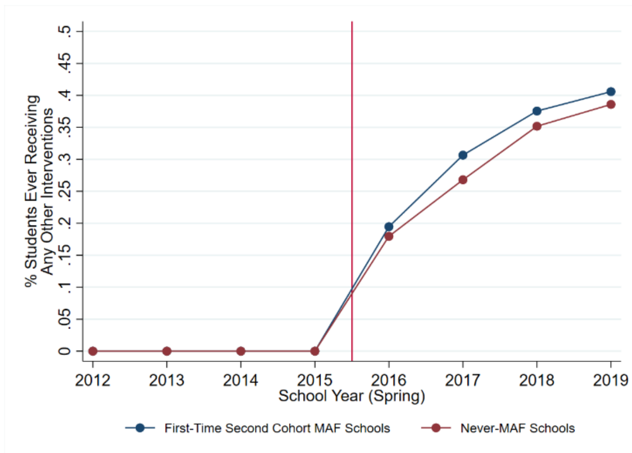
Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

Appendix Figure 2. Exposure to MAF at the School and Student-Level, First-Time Second Cohort MAF Schools

(A) % of ever-MAF students: First-time second cohort MAF versus never-MAF schools



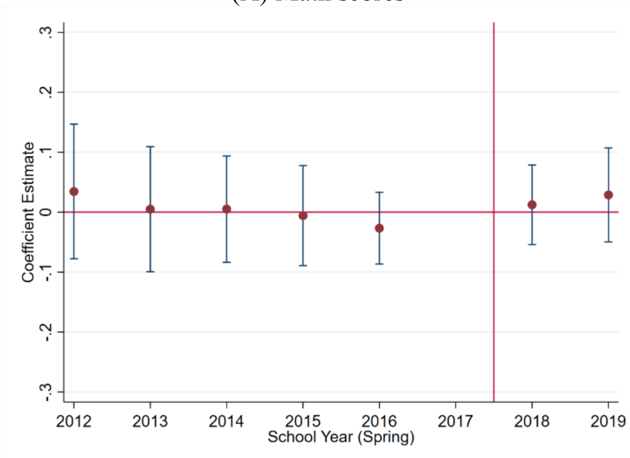
(B) % of students ever-identified for other services: First-time second cohort MAF versus never-MAF schools



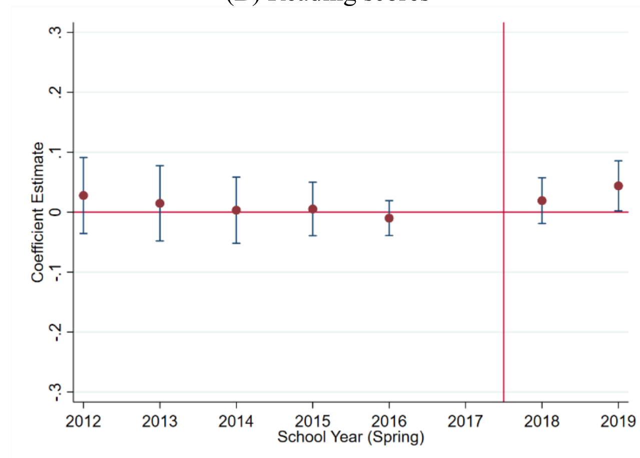
Notes: The two panels present the percentage of students in first-time second cohort MAF schools (elementary schools that were first designated as MAF in 2017-18 school year) and never-MAF schools who have ever received MAF services (panel A) or other services (panel B) up to (and including) that school year.

Appendix Figure 3. Effects of MAF Designation on Student Outcomes: Event Study Estimates, First-Time Second Cohort MAF Schools

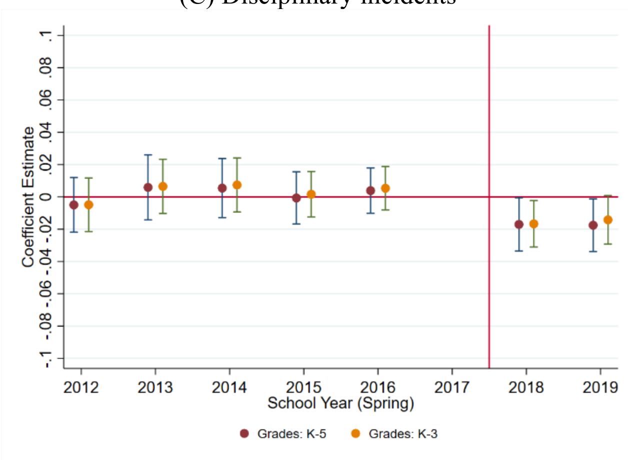
(A) Math scores



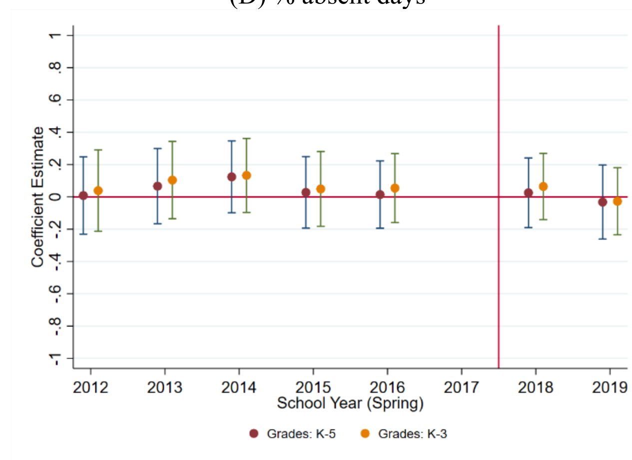
(B) Reading scores



(C) Disciplinary incidents



(D) % absent days



Notes: Each dot presents the estimated coefficient on the interaction term between the MAF indicator and the indicator for the corresponding school year, with the spikes providing the 95% confidence intervals. The estimates are obtained using schools in the first-time second cohort of MAF schools as the treatment and never MAF schools as the comparison group with school, grade, and year fixed-effects. Standard errors are clustered at the school level.

Appendix Table 1. Average School Characteristics by Survey/Cost Data Availability, Elementary Schools in 2018-19 School Year

	First Cohort Survey Respondents	Second Cohort Survey Respondents	First & Second Cohort Survey Respondents	First Cohort Survey Nonrespondents
Eligible for free- or reduced-price meals	0.712	0.679	0.696	0.692
	(0.103)	(0.134)	(0.120)	(0.142)
Special education	0.235	0.185	0.212	0.207
	(0.053)	(0.055)	(0.059)	(0.050)
English learner	0.015**	0.045	0.029	0.058
	(0.023)	(0.046)	(0.038)	(0.086)
White	0.962**	0.867	0.918	0.913
	(0.038)	(0.147)	(0.114)	(0.114)
Black	0.050**	0.163	0.103	0.112
	(0.047)	(0.179)	(0.139)	(0.119)
Hispanic	0.038**	0.058	0.047**	0.079
	(0.028)	(0.042)	(0.037)	(0.086)
Asian	0.007	0.020	0.013	0.013
	(0.008)	(0.020)	(0.016)	(0.019)
Number of enrolled K-5 students	469	456	463	38,148
Number of unique schools	10	9	19	90

Notes: Averages are pupil-weighted using K-5 enrollment from each school. Standard deviations are given in parentheses. ** denote that the means for the corresponding group are statistically different from the sample of first cohort schools from which costs were not collected (final column) at the 5-percent error level.