



## RESEARCH MEMO

# Examining *Bridges in Mathematics* and Differential Effects Among English Language Learners

### Authors

Garret J. Hall  
Wisconsin Center for Education Research

Patti Schaefer  
Madison Metropolitan School District

Teri Hedges  
Madison Metropolitan School District

Eric Grodsky  
Wisconsin Center for Education Research

### Content Contact

Garret Hall  
Ghall3@wisc.edu

### Suggested Citation:

Hall, G., Schaefer, P., Hedges, T., & Grodsky, E. (2020). Examining *Bridges in Mathematics* and Differential Effects Among English Language Learners. Madison, WI: Madison Education Partnership. pp. 1-35.\*

\*Reissued with correction, February 2020

## Executive Summary

Wisconsin's Madison Metropolitan School District introduced *Bridges in Mathematics* in phases across three years beginning in 2016 for students in kindergarten through fifth grade. In this report, we discuss findings from a study of whether the use of *Bridges* correlates to gains in student math assessment scores in fifth grade. Additionally, we investigated whether this effect of *Bridges* was stronger for English language learners across levels of English language proficiency (ELP). To understand the effect of *Bridges* on student math achievement, we investigated the following three research questions:

- Did fifth-grade students who received *Bridges* initially in school years 2016-17 and 2017-18 show greater Measures of Academic Progress (MAP) math gains from fall to spring than students receiving the previous curriculum (*Investigations*)?
- For students learning math through *Bridges*, do ELLs who are mid-proficient in the English language or low in ELP improve their math assessment scores more than students who are proficient in the English language?
- To what extent are MAP math performance and ELP scores related in third, fourth, and fifth grade?

We used data from 5,193 fifth-grade students across three years (2015-16, 2016-17, and 2017-18) to investigate whether students who received *Bridges* showed greater fifth-grade MAP math gains compared to students who received the previous curriculum, *Investigations*. Fifth-grade students using *Bridges* in 2016-17 or 2017-18 made up the *Bridges* group ( $n = 1,839$ ). To form our non-*Bridges* comparison group, we used the remaining students from 2015-16 and 2017-18 ( $n = 3,354$ ). In the full sample, 77% of students had ELP of 6 or 7, 14% had ELP levels of 4 or 5, and 8% had ELP levels of 1 to 3. These percentages of ELP levels varied somewhat across treatment groups, but this was due primarily to enrollment year differences.

We find that:

- Students receiving *Bridges* enjoyed greater fall to spring gains in math proficiency than those receiving the previous curriculum (on the order of a quarter of a standard deviation).
- English language learners and English proficient students do not show significantly different math gains under *Bridges*. Math gains in *Bridges* are also similar across ELP levels.
- ELP scores are moderately positively related to math performance in third, fourth, and fifth grade. Obtaining a higher score on the ELP measure relates to having higher scores on MAP math.

## Conclusions

These findings suggest that MMSD implementation of *Bridges* is likely in part responsible for greater student gains on the fifth-grade math assessment. For English language learners, ELP level is an important factor in MAP math performance, though ELP level does not meaningfully alter the extent to which students gained on MAP under *Bridges*. Importantly, our analysis indicates that *Bridges* may have caused these larger gains among students receiving the curriculum; however, we are unable to conclusively state that implementing *Bridges* alone caused these differences in achievement gains. Nevertheless, our study provides promising results for the potential positive impacts of *Bridges* on student math achievement growth. We suspect that the effect of *Bridges* we observe in this study is due to a combination of curriculum content as well as professional development and implementation support for teachers.

## Acknowledgements

The authors thank the MMSD staff that provided additional help and support for this project, including Silvia Romero-Johnson and Ben Kollasch in the Office of Multilingual and Global Education. The authors also thank Karen FASTER at the Wisconsin Center for Education Research for providing detailed feedback and edits on drafts of this work.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

## Introduction

In the 2016-17 school year, the Madison Metropolitan School District (MMSD) in Wisconsin began implementation of a mathematics curriculum called *Bridges in Mathematics*, which is fully aligned to the Common Core State Standards in Mathematics, unlike the prior curriculum, *Investigations*. Designed for students in kindergarten through fifth grade, *Bridges* materials and instructional sequences facilitate efficient implementation of the standards in classroom instruction. MMSD teachers and administrators received intensive professional development and implementation support during the summer prior to implementation and the first year of *Bridges* implementation. District staff and representatives from *Bridges*' publisher, The Math Learning Center, conducted professional development to support adherence to the curriculum sequence and troubleshoot implementation barriers throughout the year.

Ten schools began implementing *Bridges* in 2016-17. Fourteen schools launched the curriculum in the second implementation phase in 2017-18, and the third phase of implementation began in eight schools in 2018-19. As highlighted in the 2016-17 *Annual Report on the MMSD Strategic Framework* (MMSD, 2017), students in the initial 10 schools showed promising gains in mathematics proficiency based on their scores on the Measures of Academic Progress (MAP) assessment. In October 2017, the *Capital Times* reported that “[t]he gradual approach to implementation allows the district to gather feedback from teachers, change course as needed, and establish best practices,” former MMSD math coordinator Ken Davis told the *Capital Times* (Walker, 2017). “‘Constant feedback took place between the (central office) math team and the teachers,’ Davis said. ‘It’s been a really good, collaborative effort around planning and preparing to teach students’” (Walker, 2017).

## The *Bridges* Curriculum

The Math Learning Center says *Bridges* is a widely used, research-based curriculum structured for 80 minutes of math per day. Students use *Bridges* to “solve problems using visual models and manipulatives; make and test conjectures while recording their thinking; [and] talk and move about the classroom as they actively engage in their learning” (The Math Learning Center, n.d.). Teachers implementing the curriculum are to “[e]ncourage students to be responsible for their own learning; [u]se good questioning strategies and draw out student thinking; [and] [p]romote discourse while creating a safe learning environment” (The Math Learning Center, n.d.). Generally, *Bridges* presents “material that is as linguistically, visually, and kinesthetically rich as it is mathematically powerful” (The Math Learning Center, 2019).

The curriculum has received positive reviews and has high national and international take-up. EdReports (2018), an agency that evaluates educational curricula, gave *Bridges* its highest ratings of “Meets Expectations” for all areas of the curriculum across all grade levels, including alignment with standards and usability in practice. *Bridges* is used in numerous U.S. states as well as in Africa, Asia, and Central and South America. The curriculum may produce gains of 7 and 8 percentile points (or 0.18 and 0.19 standard deviations) among fourth- and fifth-graders, respectively, compared to students in non-*Bridges* schools (SEG Measurement, 2018).

## English Language Learners in MMSD

With the explicit focus of *Bridges* on mathematics language and visualization (The Math Learning Center, 2019) coupled with increased teacher professional development to aid implementation, *Bridges* may be particularly beneficial for English language learners (ELLs) across English language proficiency levels. Teaching mathematics language and providing visual, accessible representations of mathematics information may help ELLs acquire and apply mathematics knowledge (Kersaint et al., 2013). The MMSD curriculum change offers a valuable opportunity to investigate whether English language proficiency (ELP) is associated with differential response to the curriculum change.

More than 25% of MMSD's fifth-grade ELLs had limited ELP (2018-19 figures) (Wisconsin Department of Public Instruction, 2019), more than twice the fifth-grade national average of 10.2% in 2016 (McFarland et al., 2019) and more than three times greater than the Wisconsin average of 7.6% in 2018-2019 (Wisconsin Department of Public Instruction, 2019). Most MMSD ELLs in MMSD speak Spanish as their first language, but many others speak other languages as well.

## Relationships of Language to Mathematics

Language is related to mathematics achievement for all students, regardless of whether a student is (or has been) classified as an ELL. Multiple studies have shown that language is related to mathematics proficiency in many different ways. Expressive and receptive vocabulary (LeFevre et al., 2010; Vukovic & Lesaux, 2013) as well as knowledge of the structure of language (e.g., syntax) (Chow & Eckholm, 2019) are positively related to mathematics proficiency in elementary school. More research is needed to understand whether second language proficiency is related to the extent to which students improve their math test scores as a result of instruction. In prior studies, researchers have investigated the relationship of ELL status to student response to math intervention (Doabler et al., 2016; Doabler et al., 2019). In general, however, few studies have taken into account the ELP continuum, Doabler et al. (2019) being one of them. ELP is frequently employed as a binary—yes or no—variable, when in fact, sociodemographic characteristics and linguistic proficiency vary significantly among ELLs.

## The Present Study

We had three goals for the current study. Our first was to investigate the average effect of *Bridges* on improvement in math for fall to spring for fifth-grade students in MMSD. Second, given the limited research on the effects of standards-based math curricula on ELL math achievement, we investigated whether this potential impact of the *Bridges* curriculum varied across ELP levels. Last, we examined the pattern of correlations between ELP and math test scores in third, fourth, and fifth grade (regardless of whether students received *Bridges* or not).

To understand the effect of *Bridges* on student achievement, we investigated the following three research questions:

- Did fifth-grade students who received *Bridges* initially in school years 2016-17 and 2017-18 show greater Measures of Academic Progress (MAP) math gains from fall to spring than students receiving the previous curriculum (*Investigations*)?

- For students learning math through *Bridges*, do ELLs who are mid-proficient in the English language or low in ELP improve their math assessment scores more than students who are proficient in the English language?
- To what extent are MAP math performance and ELP scores related in third, fourth, and fifth grade?

## Method

Thanks to the staggered implementation of the *Bridges* curriculum, we used multiple cohorts of students to create a treatment group (i.e., students in schools that received *Bridges*) comprising two implementation phases of *Bridges*. Students in schools that were not using *Bridges* at the same time, as well as schools from previous years, make up our “control” group.

## Participants

We obtained student-level administrative data from MMSD for 5,193 fifth-grade students across three school years (2015-16, 2016-17, and 2017-18) to analyze whether *Bridges* students ( $n = 1,839$ ) showed greater math gains from fall to spring of fifth grade compared to students who used *Investigations* ( $n = 3,354$ ). To employ the staggered implementation design, we used achievement and demographic data from three cohorts of students, each based on the year the students entered third grade (whether they were in the district at that time or not). Table 1 displays the grades and years in which students were exposed to *Bridges*.

**Table 1.**  
*Grade and Year by Bridges Implementation Phase*

Grade by Academic Year					<i>Bridges</i> Implementation Phase
2013-14	2014-15	2015-16	2016-17	2017-18	
3	4	5			
3	4	5			
3	4	5			
	3	4	5		1 (2016-17)
	3	4	5		2 (2017-18)
	3	4	5		3 (2018-19)
		3	4	5	1 (2016-17)
		3	4	5	2 (2017-18)
		3	4	5	3 (2018-19)

*Note.* Shaded boxes indicate when *Bridges* was implemented.

As the shaded boxes in Table 1 show, our treatment group comprises *Bridges* phase 1 fifth-graders in 2016-17 and 2017-18 and *Bridges* phase 2 fifth-graders in 2017-18.<sup>1</sup> Our control

<sup>1</sup> A very small percentage of students started *Bridges* in fourth grade in 2016-17 but switched to a non-*Bridges* school for fifth grade. They were labeled as not receiving *Bridges*, as they did not receive *Bridges* instruction in the year of our analysis.

group comprises students in fifth grade in 2015-16, non-*Bridges* fifth graders in 2016-17 and 2017-18, and *Bridges* phase 3 students because they had not yet started the curriculum in 2017-18.

We obtained student-level ELP from MMSD, which relies on results of the *ACCESS for ELLs* assessment. In the full sample of 5,193 fifth-graders, 4,010 (77.2%) had high ELP levels of 6 or 7, another 746 (14.4%) had mid ELP levels of 4 or 5, and 437 (8.4%) had low ELP levels of 1 to 3.

Table 2 displays ELP and other demographic information for students who received *Bridges* or *Investigations*. Our sample is limited to the 5,193 students who had a fall and spring fifth-grade MAP mathematics scores, a fall MAP reading score, a documented ELP level, and complete demographic and special education status data. The far-right column presents statistical test results to determine whether the composition of *Bridges* and *Investigations* groups on these variables are significantly different. *p* values labeled as  $\geq .05$  indicate that there are no significant differences in the composition of each variable across groups. We control for all of the demographic characteristics listed below in our statistical analyses, which helps removes some systematic demographic differences between treatment groups.

**Table 2.***Percentages of Demographic Composition Across Treatment Groups*

Variable	<i>Investigations</i> ( <i>n</i> = 3,354)	<i>Bridges</i> ( <i>n</i> = 1,839)	<i>p</i> -value Significance of Percentage Differences
Student with Individualized Education Plan (IEP)	14%	15%	≥.05
Free or Reduced-Price Lunch Eligible	50%	51%	≥.05
Female	50%	50%	≥.05
<b>Race/Ethnicity</b>			
White	42%	42%	≥.05
Black or African American	17%	18%	≥.05
Hispanic/Latino	22%	23%	≥.05
Asian or Asian American	9%	8%	≥.05
Native Hawaiian/Pacific Island or American Indian/Alaska Native	0%	0%	≥.05
Multiracial	10%	9%	≥.05
<b>Parent Education</b>			
Less than High School Degree	7%	6%	≥.05
High School Degree	18%	19%	≥.05
Some College or Technical Degree	23%	24%	≥.05
Four-Year Degree	17%	19%	<.05
Graduate School/Professional Degree	31%	28%	≥.05
Missing Education Level	5%	4%	≥.05
<b>English Language Proficiency</b>			
Proficient (Levels 6–7)	77%	78%	≥.05
Mid English Proficiency (Levels 4–5)	16%	12%	<.05
Low English Proficiency (Levels 1–3)	7%	11%	<.05

*Note.* Parent education is the highest-reported level for third- fourth-, and fifth graders. Students identified as Native Hawaiian/Pacific Islander or American Indian/Alaskan Native were collapsed into a single category and rounded to 0% due to very small sample sizes. Differences in ELP percentages across groups could be attributed to changes in the ACCESS measure in 2016-2017 that resulted in more students obtaining lower ACCESS scores. Predicting three-category ELP level using *Bridges* membership shows *Bridges* students are more likely to be in the low ELP group (compared to Mid), but this effect is removed when enrollment year indicators are included as control variables. *p*-values of .05 or greater indicate variable does not statistically vary across treatment groups.



Table 3 shows MAP scores in the fall and spring for each group. Importantly, fall MAP math and reading scores did not differ significantly between *Bridges* and *Investigations* groups, indicating groups were similar with respect to their math performance. *p*-values labeled as  $\geq .05$  indicate the differences are not statistically significant.

**Table 3.**  
*Mean Differences of MAP Scores Across Treatment Groups*

Test Score	Mean (Standard Deviation)		<i>p</i> -value Significance of Mean Differences
	<i>Investigations</i>	<i>Bridges</i>	
MAP Math Score Fall Grade 5	209.64 (17.52)	210.72 (17.56)	$\geq .05$
MAP Reading Score Fall Grade 5	206.24 (18.22)	206.72 (17.82)	$\geq .05$
MAP Math Score Spring Grade 5	219.22 (19.01)	221.37 (19.06)	$\geq .05$
MAP Reading RIT Score Spring Grade 5	212.70 (17.33)	213.22 (16.93)	$\geq .05$

*Note.* Standard errors for tests of mean differences were clustered among schools within each enrollment year. 440 students in *Bridges* had ACCESS scores and 927 control students had ACCESS scores. Controlling for enrollment year removes ACCESS score differences. Only 3,345 control and 1,836 *Bridges* students had spring MAP reading data.

Table 4 reports means, standard deviations, as medians of ACCESS proficiency scores. The distribution of ACCESS scores does not exhibit a normal distribution (like MAP data), so means are not necessarily a representative value of the typical ACCESS score. We estimated differences between groups based on the means as well as medians as a result.

**Table 4.**  
*Mean and Median Differences of Grade 4 ACCESS Composite Proficiency Scores Across Treatment Groups*

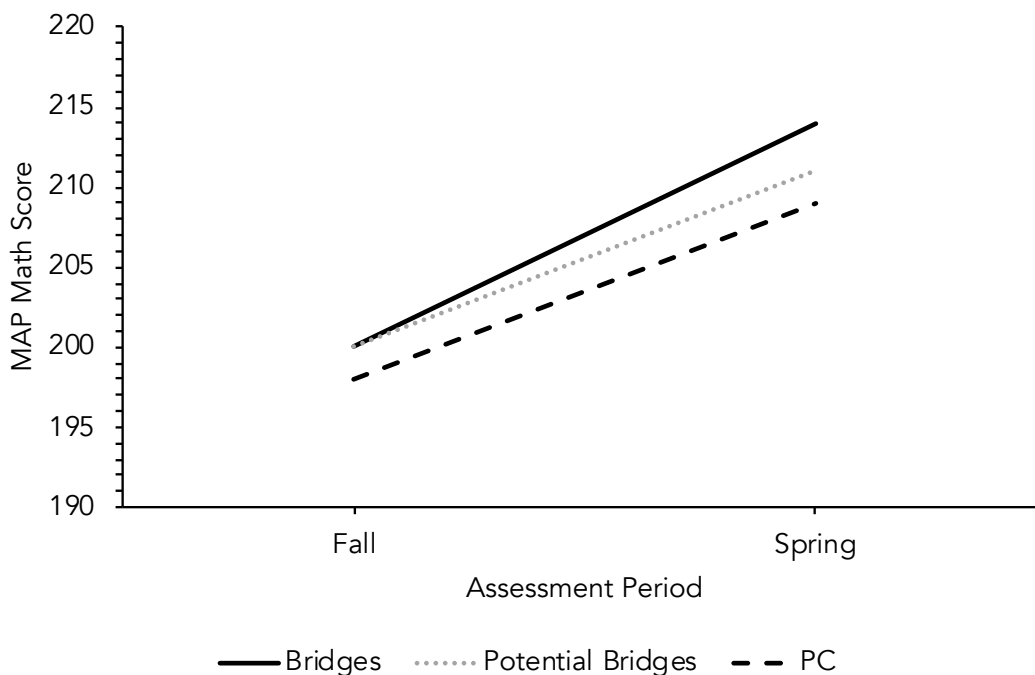
	Mean (Standard Deviation)		Median		<i>p</i> -value Significance of Mean/Median Differences
	<i>Investigations</i>	<i>Bridges</i>	<i>Investigations</i>	<i>Bridges</i>	
ACCESS Composite Proficiency Score	4.7 (0.97)	4.2 (0.95)	4.6	4.1	$< .05$ / $< .05$

*Note.* Median differences produced using median regression with standard errors clustered among schools within enrollment year (Machado, Parente, & Santos Silva, 2011). Controlling for enrollment year removes these differences.

The information in Tables 2, 3, and 4 suggest that few demographic or academic performance differences exist between groups at baseline (fall of fifth grade). Although ELP data appears to vary somewhat between treatment groups, this appears to be largely explained by enrollment year differences rather than meaningful differences between treatment groups.

### Analytic Strategy

With the rich student-level data available to us, we considered multiple approaches to assessing the effect of the *Bridges* curriculum. We looked at the extent to which we could state that the curriculum caused a change in student achievement. To make this claim, many assumptions must be met with the study design and the methods used to the cause-effect relationship of *Bridges* to math achievement. One design frequently used to assess changes in policies, or the implementation of an intervention, is called difference-in-differences (Angrist & Pischke, 2009). In this design, an analyst needs one preintervention measurement and one postintervention measurement to assess whether the pre-post changes within the group exposed to the policy (or treatment, intervention, etc.) are significantly different from the unexposed individuals. A mock visualization of this design is presented in Figure 1.



**Figure 1.** Mock visualization of fall-spring gains for students who received *Bridges* or *Investigations* (PC).

*Note.* PC = Previous curriculum.

In this mock example, *Bridges* students scored 200 in the fall and 214 in the spring (a “change score” of 14). Students using the previous curriculum gained 11 points (from 198 to 209). We assume that, had the *Bridges* group stayed with *Investigations*, they also would have gained 11 points, as shown with potential *Bridges*’ gray dotted line that is parallel to the dashed

*Investigations* line. The difference-in-differences is 3 points (14 minus 11). In Figure 1, *Bridges*' solid line is slightly steeper than the previous curriculum's dashed line, showing the benefit of the new curriculum. The difference in slope between the "potential *Bridges*" line and the actual observed *Bridges* solid line is the 3-point difference-in-difference estimate. Because we assume the potential *Bridges* line is parallel to the previous curriculum line (that the students would score the same on MAP with the same curriculum), the estimate equals the difference in slopes between the *Investigations* and *Bridges* curricula.

Once other factors affecting or related to pre-post changes and the policy implementation are accounted for (if necessary), an analyst may then say that the differences in pre-post changes between "treatment" and "control" groups are due to the causal effects of the new policy, even without the "gold standard" method of randomizing people or units (schools, organizations) to treatment and control groups. That said, without knowing the exact process by which schools opted in to implementing *Bridges* (i.e., "selection bias"), estimates of the impact of *Bridges* are likely imprecise.

Additionally, a number of assumptions of the difference-in-differences design must be met to detect causal relationships. The primary assumption is that no differences exist in the trend of the outcome between treated and untreated groups prior to the beginning of the treatment (i.e., the parallel trends assumption). In other words, for difference-in-differences to work in this situation, *Bridges* and non-*Bridges* students should have been improving their scores at similar rates on MAP prior to *Bridges* being implemented. Parallel trends strengthen the assumption that *Bridges* students would have made similar gains to students using *Investigations* in the absence of actually receiving *Bridges*.

## Study Variables

**Dependent variable: Change in MAP math score.** MMSD fifth-graders take the MAP math test near the beginning of the fall semester, in September and early October, and again in April. Our outcome measure is the difference in the scores from fall to spring in the 2015-16, 2016-17, and 2017-18 school years. We simply subtracted fall from spring MAP scores to calculate our dependent variable. Recent research suggests these change scores have valuable properties that can improve causal inferences in difference-in-differences designs (Kim & Steiner, 2019).

**Independent variable: *Bridges* exposure.** Fifth-grade students in schools that implemented *Bridges* in phase 1 in 2016-17 or 2017-18 or in phase 2 in 2017-18 were assigned a 1 to denote they received *Bridges* during that year. All other students were assigned a 0 (*Investigations* students). These *Investigations* students included those in *Bridges* phase 3 schools that had not yet implemented *Bridges* as well as students in 2015-16, who never received *Bridges*.

**Independent variable: ELP level.** The Wisconsin Department of Public Instruction classifies student ELP level as ranging from 1 to 7. Students who only spoke English are assigned a 7 to indicate they have never been ELLs; they are not eligible to take the ELP assessment (*ACCESS for ELLs*) or receive services for students with limited ELP. A score of 6 identifies students who obtained a 6 on *ACCESS for ELLs* in the prior assessment window (grade 4) or

were previously classified as having limited ELP.<sup>2</sup> Scores from *ACCESS for ELLs* are reported as proficiency level scores rounded to the nearest tenth. For Department of Public Instruction ELP classification, ACCESS scores of 1 to 5.9 are truncated to the left decimal of the proficiency score (i.e., scores of 1.1 to 1.9 are rounded to 1, 2 to 2.9 rounds to 2, etc.). The department considers students with ELP levels of 5 or below to be of limited ELP, making them eligible to receive language support services as ELLs.<sup>3</sup> However, state criteria for limited ELP slightly vary from criteria in MMSD and other districts, and criteria for ELL classification changed with the realignment of ACCESS proficiency levels in 2016-17 by WIDA, the assessment’s publisher. In our data, students who scored between 5 and 5.9 on ACCESS in 2015, 2016, or 2017 received an ELP level of 5. This does not reflect exactly who was classified as “limited English language proficient” due to changes in classification criteria.

We defined our ELP variable by grouping students into three categories: low (ELP levels 1–3), mid (ELP levels 4–5), and English language proficient (levels 6–7). This categorization does not capture many of the nuances of proficiency; however, it eases our analytic approach. Students who are classified as English language proficient are either former ELLs (ELP level 6) or never ELLs (ELP levels 7), the latter category reflecting English-monolingual students. Examining the effect of *Bridges* across all seven ELP levels would have resulted in problematic small sample sizes, especially at levels 1 and 2. Students with ELP levels of 1–2 do not typically take MAP assessments (Brown, 2017), which is partly why we classify our groups as low, mid, or proficient even though students scoring at 3 show more mid-level proficiency skills based on the *ACCESS for ELLs* measure. We also assessed whether the effect of *Bridges* varied across *ACCESS for ELLs* scores.

**Control variables.** We used a variety of statistical controls to better detect a cause-effect relationship of the curriculum implementation to student math achievement. Adding control variables can help remove additional variability in MAP annual change scores, which then helps us better detect the effect of *Bridges* on MAP scores. We controlled for race/ethnicity, gender, whether the student had an individualized education plan (IEP), free or reduced-priced lunch eligibility, parent education level, fall of fifth grade MAP reading scores, and students’ year cohort (whether the student was in fifth grade in 2015-16, 2016-17, or 2017-18). In addition, we controlled for which implementation phase of the district-wide technology plan each school was assigned because that technology plan might have affected *Bridges* implementation.

The full statistical models we used to analyze the effect of *Bridges* on annual MAP math scores are shown in Appendix A.

---

<sup>2</sup> A small percentage of students were in the district in grades three and five but not four, in which case these students are not included in this analysis as they did not have a fourth-grade ELP level. If third- and fifth-grade data were present in two consecutive years (e.g., due to advancing a grade), ELP from third grade was used.

<sup>3</sup> ACCESS scores may not be the only determinant of ELL status. Other data may inform ELL classification as appropriate. Also, measures such as a screening measure or an alternative assessment for students with severe disabilities may determine ELL status. In our analytic sample, 1.43% of students without ACCESS data had an ELP level of 6 or below.

## Missing Data

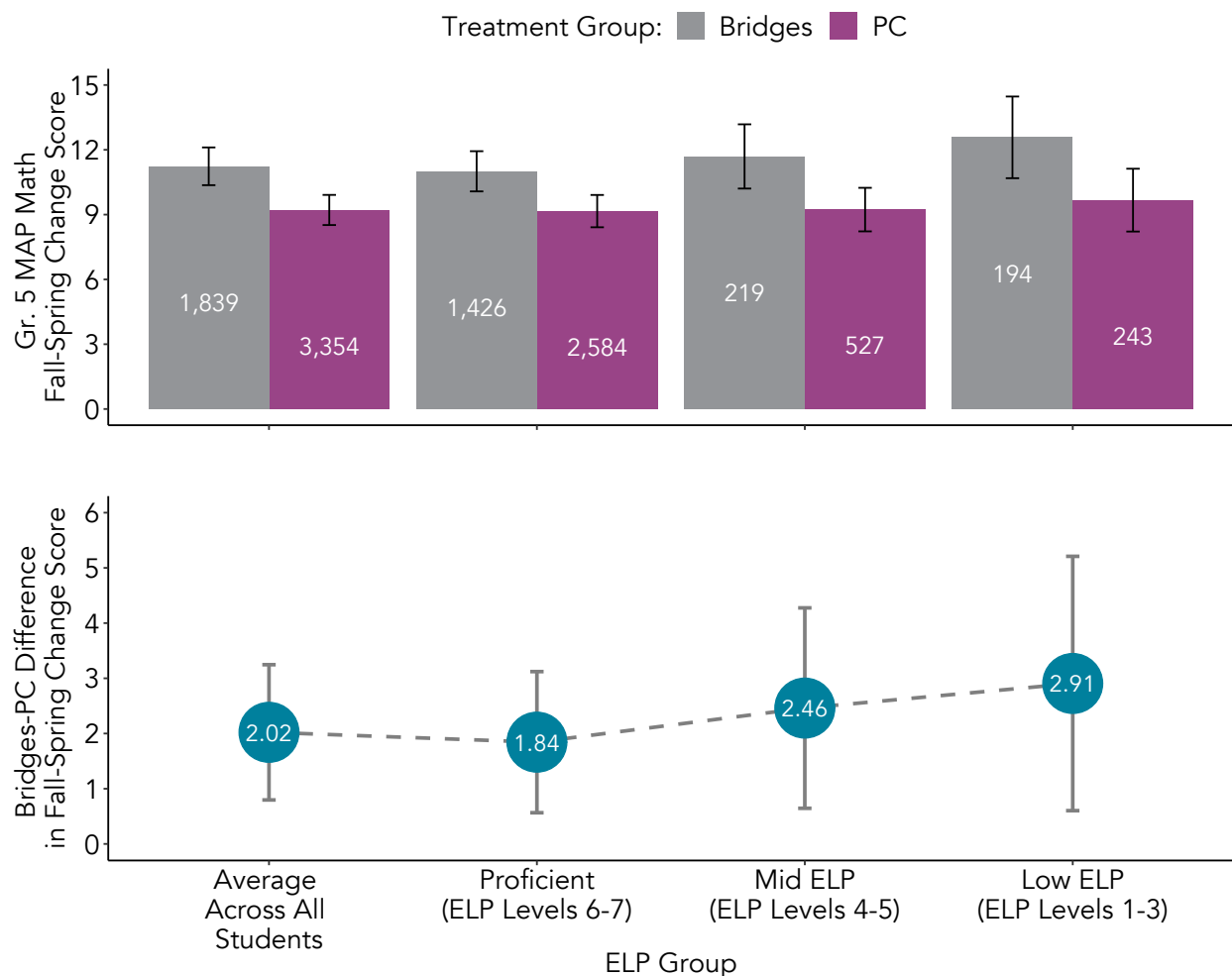
Another important part of our analysis is our treatment of missing data. As mentioned, our analytic sample comprised 5,193 students with complete demographic, ELP, fall MAP reading, and fall and spring MAP math data. However, the sample of fifth graders with complete demographic and fall MAP reading data consisted of 5,359 students, meaning that 3.1% of those students did not have fall and/or spring MAP math data. The reasons students did not have fall and spring MAP math data could be random or it could be related to particular student characteristics. To account for these missing outcome data, we weighted our analyses such that students who were more likely to have both fall and spring MAP mathematics scores were given less weight in the analyses, which reduces potential bias in results due to missing data and sample attrition. More information on this approach, called inverse probability weighting (Seaman & White, 2013), can be found in Appendix B.

## Results

### **RQ1: Did fifth-grade students who received *Bridges* initially in school years 2016-17 and 2017-18 show greater math gains in assessment scores from fall to spring than students learning with *Investigations*?**

Students receiving *Bridges* improved from fall to spring by 2.02 more MAP points compared to students receiving *Investigations*. On average, students in *Bridges* gained 11.24 points whereas students in *Investigations* gained 9.21 points, a statistically significant difference of 0.250 (95% confidence interval [0.193, 0.307]) in standard deviation units (Hedges's *g*). Another way to frame these results is in terms of percentile rank, and we can use the control group to calculate these percentile ranks (Lipsey et al., 2012). In the control group the adjusted mean is 9.21, and 48.2% of students gained 9 points or fewer. The adjusted mean for *Bridges* students is 11.24, and 59.5% of students gained 11 points or fewer in the control group. In percentile terms, the effect size is 11.3 percentile points.

Results from our primary analyses are presented in Figure 2 as average marginal effects, which are the predicted MAP change scores for each group adjusted for each student's observed value on each of the control variables (Williams, 2012).



**Figure 2.** Adjusted effects of *Bridges* by ELP with 95% confidence intervals. Top panel displays adjusted means by treatment group and ELP level. The bottom panel displays differences in *Bridges vs. Investigations* (PC) differences in the top panel adjusted means by ELP level. *Note.* Sample sizes for each group are displayed within the bars in the top panel. PC = Previous curriculum.

**RQ2: For students learning math through *Bridges*, do ELLs who are mid-proficient in the English language or low in ELP improve their math assessment scores more than students who are proficient in the English language?**

The other three columns to the right in Figure 2 display the *Bridges* and *Investigations* means (top panel) and mean differences (bottom panel) by ELP level. The impact effect size among English language-proficient students in standard deviation units ( $g$ ) was 0.226 (95% confidence interval [0.161, 0.291]), among mid-ELP it was 0.336 (95% confidence interval [0.178, 0.494]), and among low-ELP students the effect size was 0.337 (95% confidence interval [0.147, 0.527]).<sup>4</sup> Although the difference in treatment impact between English language proficient, mid ELP, and low ELP levels vary in size, we cannot conclude that these differences

<sup>4</sup> Effect sizes calculated using standard deviations within each ELP group.

are statistically meaningful. There is no statistically significant difference in the effect of *Bridges* between English language-proficient students and combined mid- and low-ELP students. Importantly, The *Bridges* effect at low ELP is representative mostly of students with an ELP level of 3; MMSD policy is that students with ELP levels of 1 or 2 are not required to take MAP tests (Brown, 2017). This allowance may create greater bias and variability in the *Bridges* effect at the low ELP level due to systematic missingness, although some students with ELP levels of 1 or 2 did take MAP.

To further investigate the relationship of ELP level to the effect of *Bridges*, we also tested the effect of *Bridges* among only students who took ACCESS and analyzed whether *Bridges* differed significantly across ACCESS composite proficiency scores. This analysis limited our sample to students who took ACCESS the prior year and stayed in the district through fifth grade ( $n = 1,367$ , *Bridges*  $n = 440$ ). The effect of *Bridges* remained statistically significant in this subsample of students: students who received *Bridges* gained 12.09 points, and comparison peers gained 9.15 – a difference of 2.94. In standard deviation units, this equals a difference of 0.376 (95% confidence interval [0.262, 0.490]). However, the effect of *Bridges* did not vary significantly across ACCESS proficiency scores. See Appendix C for information regarding testing *Bridges* impact variation by ACCESS score.

Appendices D and E show the results of the full statistical models for the primary analysis. Appendix F reports the unadjusted means and standard deviations of change scores for each treatment group as well as the adjusted mean estimates from the analyses reported above. Additionally, because ELP data are constructed from prior year ACCESS scores (and the ACCESS scores in our models above were also from the previous year, fourth grade), we provide results using ELP and ACCESS data from fifth grade in Appendix G. The primary interpretation of the results remains similar: *Bridges* has a statistically significant impact of small magnitude, that does not meaningfully vary across ELP levels. However, the fifth-grade ELP data were measured after *Bridges* began for most students. Predictor variables that were measured after a treatment began may bias the results (Montgomery et al., 2018), so these results should be interpreted cautiously.

### **RQ3: To what extent are MAP math performance and ELP related in third, fourth, and fifth grade?**

To further analyze relationships between MAP math performance and ELP, we correlated the eight available ACCESS ELP scores (four composite and four subtest scores) with MAP math scores in fall and spring. These analyses were for all students in each grade with an ACCESS score, regardless of the math curriculum they received. We present in Table 4 correlations only within-grade, as cross-grade comparisons may be inaccurate due to the recent realignment of the ACCESS ELP scores. The correlations in Table 4 suggest that, on the whole, MAP and ACCESS positively correlate to a small to moderate degree. Obtaining a higher score on ACCESS measures is associated with having higher scores on MAP math. ACCESS overall, literacy, reading, and comprehension scores exhibit the strongest relationships to MAP mathematics performance across grades.

**Table 5.**

Spearman Correlations Between MAP Mathematics and ACCESS Composite and Subtest Scores

MAP Math Assessment Period	ACCESS Composite Scores				ACCESS Subtest Scores			
	Overall (All)	Literacy (Reading + Writing)	Comprehension (Reading + Listening)	Oral Language (Speaking + Listening)	Writing	Reading	Speaking	Listening
Gr. 3 Fall	0.69	0.68	0.60	0.56	0.60	0.61	0.49	0.47
Gr. 3 Spring	0.65	0.65	0.57	0.52	0.57	0.58	0.47	0.42
Gr. 4 Fall	0.69	0.68	0.68	0.58	0.59	0.66	0.45	0.49
Gr. 4 Spring	0.61	0.61	0.61	0.50	0.53	0.59	0.38	0.44
Gr. 5 Fall	0.58	0.57	0.66	0.48	0.40	0.60	0.28	0.43
Gr. 5 Spring	0.55	0.53	0.62	0.46	0.38	0.55	0.26	0.42

*Note.* ACCESS scores are from the same grade year as MAP scores. Missing data are pairwise-deleted (i.e., each correlation is based on available data for the two measures in each correlation). Partial Pearson correlations controlling for enrollment year show similar patterns.



These results suggest that students with higher ELP tend to perform higher on MAP math, though the relationship between ACCESS and math may be somewhat dependent on the specific subtest of ELP. The Oral Language composite score tends to have the weakest relationship to mathematics performance across grades. This finding may indicate that ELP domains pertaining to comprehension and literacy skills tend to be more pertinent to mathematics performance among ELLs than oral language, though the difference is not large. However, these differences should be interpreted cautiously. The overall composite is composed of the other three composites, which are themselves constructed with combinations of subtest scores. As a result, composite scores are not directly comparable to subtest scores, and an overall composite score generally possesses the most desirable measurement properties (Canivez, 2013; Youngstrom, Kogos, & Glutting, 1999).

## Discussion

This study investigated whether the curriculum better aligned to the Common Core State Standards in Mathematics, *Bridges in Mathematics*, produced meaningful mathematics gains compared to the previous curriculum, *Investigations*, for fifth-grade students in the MMSD. We also investigated whether this curriculum differentially benefits ELLs at varying ELP levels. Finally, we explored the relationship between ELP scores on mathematics performance within grades.

Our primary finding is that students receiving *Bridges* show significantly greater annual gains on MAP mathematics compared to students receiving *Investigations*. This effect does not vary significantly across ELP levels. We take this as evidence that *Bridges*, coupled with professional development and implementation support, is an effective curriculum to support student achievement gains and is similarly beneficial across ELP levels.

The effect size of *Bridges* on MAP math gains across all students translates to nearly one-quarter of a standard deviation. An effect size between 0.2 and 0.5 is conventionally considered “small,” though the size of the effect needs to be considered under the circumstances of the actual study (Cohen, 1988). In this case, students receiving the previous curriculum gained 9.21 points from fall to spring, which is near the national average of 9.9 points (Northwest Evaluation Association, 2015). Given that we observe a statistically significant and small effect size of 0.25 even when compared to the *Investigations* (another well-established curriculum), this effect size might still be interpreted as meaningfully impactful and an indication that *Bridges* presents advantages to general student achievement gains over *Investigations*.

Regarding the relationships between ACCESS scores and MAP math performance, the correlations between ACCESS composite and subtest scores and MAP math scores vary by ACCESS test. These correlations indicate that ELP plays an important role in student mathematics proficiency, though it may in part depend on the ELP area.

## Limitations

Our study has several limitations. Because we used administrative, observational data (as opposed to conducting an experiment) from the school district, many potential factors could influence our results. Confounds are variables that we did not measure but may in part

account for the effects we observed. Many possible confounders could affect our results, but without conducting our own experiment and randomizing individuals to treatment and control groups, we are unable to rule out everything necessary to determine an unconfounded estimate of the effect of *Bridges*. In Appendix H, we discuss analyses and adjustments that help strengthen our primary findings. Some of these analyses help rule out potential confounding variables, though this is not an exhaustive list of all possible confounds. We encourage readers to consider alternative explanations to the findings we reported based on their own knowledge.

## Appendix A

To assess the main impact of *Bridges* and variation across ELP levels, we specify our ordinary-least square regression model as the following (Model 1):

$$\Delta MAP Math_{ij} = \beta_{00} + \beta_{01} Bridges_j + \beta_{02} ELP Level_{ij} + \beta_{03} ELP Level_{ij} \times Bridges_j + \beta_{04} Cohort 2_{ij} \times Bridges_j + \beta X_{ij} + \varepsilon_{ij}$$

where  $\Delta MAP Math_{ij}$  is the fall-to-spring change score outcome for student  $i$  in school  $j$ ,  $Bridges_j$  represents the students' exposure to treatment (assigned at the school level),  $ELP Level_{ij}$  represents individual students' three-category ELP level in school  $j$ , and the product term  $ELP Level_{ij} \times Bridges_j$  represents the interaction between treatment exposure and ELP level. Because there are three levels of ELP, this product term breaks into two interaction terms with ELP levels 6-7 (English language proficient) as the reference group.  $\beta_{04} Cohort 2_{ij} \times Bridges_j$  is an interaction between cohort 2 students (fifth-graders in 2016-17) and *Bridges*, as some cohort 2 students received *Bridges* or the previous curriculum (see Table 1).  $X_i$  represents a vector of covariates, including five indicator variables for race/ethnicity with White students as the reference group (Black or African American, Hispanic, Asian, Multiracial, and an indicator that combined American Indian/Alaska Native and Native Hawaiian/Pacific Island students due to very small sample sizes), five indicator variables for parent education level (high school as reference group), whether the student had an individualized education plan, eligibility for free or reduced-price lunch, fall of fifth grade MAP reading performance, enrollment cohort year (2015 or 2016) with 2017 as reference group, and five dummy codes for schools in each of the six implementation cohorts of the district-wide technology plan with cohort 1 as the reference group. Digital literacy cohort was assigned to schools regardless of cohort of this study (i.e., 2015, 2016, or 2017). Finally,  $\varepsilon_{ij}$  is a random error term.  $\varepsilon_{ij}$  is corrected for clustering at the school-by-cohort level, resulting in 87 school-by-cohort clusters following recent recommendations that suggest errors should be corrected at the level of treatment assignment (Abadie et al., 2017).

In Model 2, we specify the effect of *Bridges* among only students who took the ACCESS test (by virtue of having a ELP level of 5 or below) and test the interaction effect between ACCESS proficiency level (1-6 rounded down to nearest tenth) and *Bridges* exposure on yearly MAP gains. This model is specified as:

$$\Delta MAP Math_{ij} = \beta_{00} + \beta_{01} Bridges_j + \beta_{02} ACCESS Composite_{ij} + \beta_{03} ACCESS Composite_{ij} \times Bridges_j + \beta_{04} Cohort 2_{ij} \times Bridges_j + \beta X_{ij} + \varepsilon_{ij}$$

Besides terms  $\beta_{02}$  and  $\beta_{03}$ , this model mirrors Model 1. In Model 2, Hispanic students served as the race/ethnicity reference group and high school served as the parent education reference group.

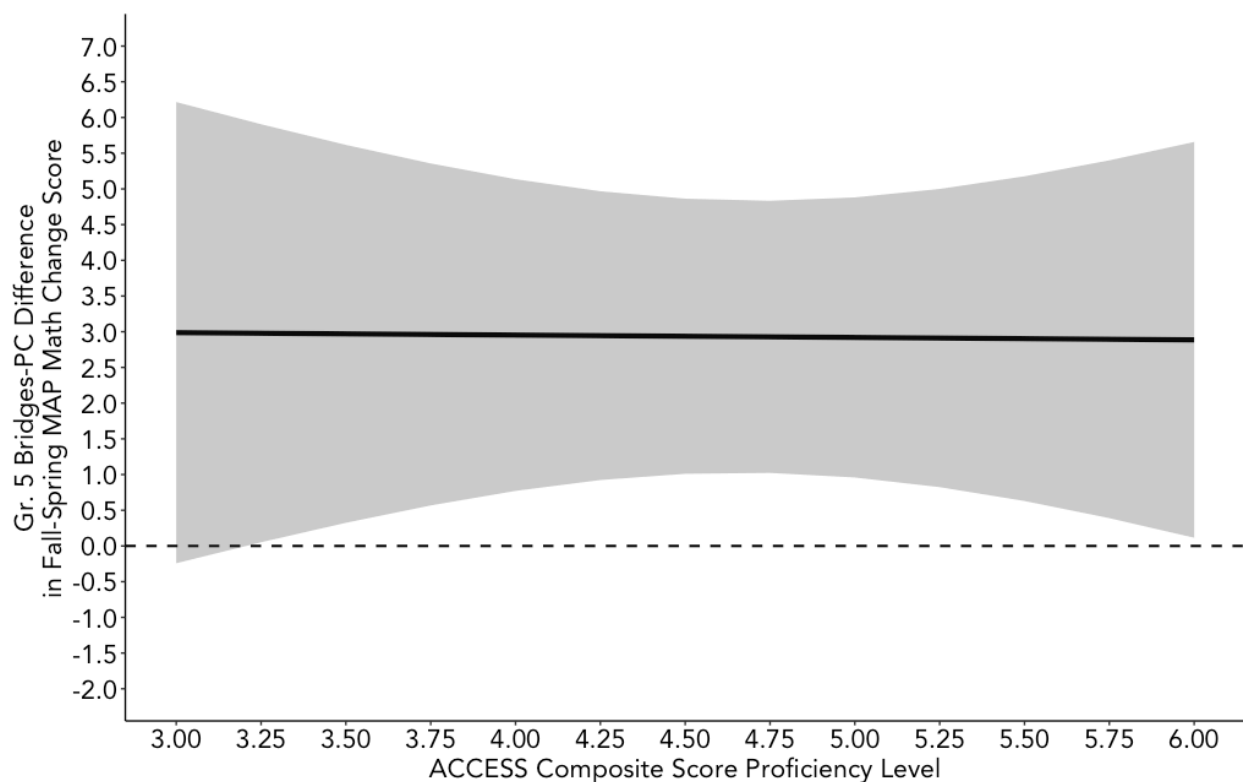
## Appendix B

We used the following model to construct inverse probability weights to account for missing data in our analysis of the impact of *Bridges* on student MAP math scores:

$$\begin{aligned} \text{logit Observed MAP Change Score}_{ij} = & \beta_{00} + \beta_{01} \text{Bridges}_j + \beta_{02} \text{ELP Level}_{ij} + \\ & \beta_{03} \text{Race/Ethnicity}_{ij} + \beta_{04} \text{Fall Gr. 5 MAP Reading}_{ij} + \\ & \beta_{05} \text{Special Ed.} + \beta_{06} \text{FRL}_{ij} + \beta_{07} \text{Female}_{ij} + \\ & \beta_{08} \text{Parent Education Level}_{ij} + \beta_{09} \text{Cohort Year}_{ij} + \\ & \beta_{10} \text{Digital Literacy Implementation Cohort}_j + \beta_{11} \text{Bridges}_j \times \mathbf{X}_{ij} + \varepsilon_{ij} \end{aligned}$$

where race/ethnicity is a six-category indicator variable code with White students used as the reference group. We combined students reporting their race as Native Hawaiian/Pacific Islander or American Indian/Alaskan Native into a single value due to very small sample sizes that perfectly predicted missingness. Parent education level is a six-category variable (no high school, high school, some college/technical degree, college degree, graduate/professional degree, missing) and was the highest-reported level from grades 3 to 5. Cohort year is a three-category indicator variable representing which year students were enrolled in fifth grade (2015, 2016, or 2017) with cohort 3 (2017) as the reference group. Digital literacy cohort is a six-category indicator variable to indicate the district-wide technology plan implementation cohort to which schools were assigned. Cohort 1 was used as the reference group.  $\beta_{11}$  is an interaction term between the treatment and each of the covariates except and digital literacy implementation cohort (since this was assigned to schools regardless of the cohort year). Additionally, *Bridges* is interacted with only the cohort 2 indicator variable, as no cohort 1 students received *Bridges*. We then constructed inverse probability weights by obtaining predicted probabilities from the logit model using the *predict* command in Stata 15.1 and dividing 1 by the predicted probabilities. The weights sum to the sample of individuals with complete data on the covariates ( $n = 5,359$ ). Weights from this model were also used to account for missingness in Model 2.

## Appendix C



The plot above displays the *Bridges-Investigations* difference and 95% confidence band at 0.25 intervals of ACCESS composite score proficiency level. The statistical test of whether the *Bridges* effect significantly varies across ELP levels is not statistically significant ( $b = -0.101$ ,  $SE = 2.325$ ,  $p = .965$ ). However, the impact estimate of *Bridges* remained statistically significant in this subsample ( $b = 2.935$ ,  $SE = 0.964$ ,  $p = .003$ ).

## Appendix D

Unstandardized Regression Coefficients for *Bridges* Main Impact Model (Model 1)

Predictor	Est.	Robust SE	<i>t</i>	<i>p</i>	95% Confidence Interval	
					Low	High
<i>Bridges</i>	2.379	0.772	3.080	.003	0.844	3.914
Mid ELP	0.070	0.508	0.140	.890	-0.939	1.080
Low ELP	0.508	0.725	0.700	.485	-0.934	1.950
<i>Bridges</i> X Mid ELP	0.617	0.839	0.740	.464	-1.051	2.285
<i>Bridges</i> X Low ELP	1.063	1.072	0.990	.324	-1.068	3.194
Some High School	-0.092	0.476	0.190	.847	-1.038	0.853
Some College/Tech Degree	0.284	0.372	0.760	.447	-0.455	1.024
Bachelor's	0.592	0.463	1.280	.204	-0.328	1.513
Graduate/Professional Degree	1.129	0.421	2.680	.009	0.293	1.965
Missing Education Level	-0.428	0.625	0.680	.496	-1.670	0.815
Black or African American	-0.919	0.416	2.210	.030	-1.747	-0.092
Hispanic/Latino	-0.505	0.437	1.150	.252	-1.374	0.365
Asian	0.206	0.437	0.470	.639	-0.663	1.074
Multiracial	-1.339	0.476	2.810	.006	-2.285	-0.392
Native Hawaiian/Pacific Island or American Indian/Alaskan Native	-0.023	2.005	0.010	.991	-4.009	3.962
Fall Gr. 5 MAP Reading	0.015	0.010	1.470	.146	-0.005	0.036
Free/Reduced Price Lunch	-0.179	0.371	0.480	.631	-0.917	0.559
Student with Individualized Education Plan	-0.924	0.584	1.580	.117	-2.086	0.237
Female	-0.210	0.225	0.940	.352	-0.657	0.236
Cohort 1 (15-16)	1.455	0.613	2.370	.020	0.237	2.674
Cohort 2 (16-17)	2.782	0.713	3.900	.000	1.365	4.200
<i>Bridges</i> X Cohort 2	-1.632	0.956	1.710	.092	-3.533	0.269
Digital Literacy Cohort 2	1.497	1.013	1.480	.143	-0.517	3.511

Digital Literacy Cohort 3	-0.640	0.653	0.980	.330	-1.939	0.659
Digital Literacy Cohort 4	-1.287	0.868	1.480	.142	-3.013	0.439
Digital Literacy Cohort 5	-0.068	0.780	0.090	.931	-1.619	1.484
Digital Literacy Cohort 6	-0.482	0.889	0.540	.589	-2.249	1.286
Intercept	5.348	2.206	2.420	.017	0.963	9.732

---

*Note.*  $n = 5,193$ . Standard errors (SEs) corrected for 87 school-by-cohort clusters. ELP reference group is never/former ELLs. Parent education level reference group is High School.

## Appendix E

Unstandardized Regression Coefficients for Secondary *Bridges* Impact Model (Model 2)

Predictor	Est.	Robust SE	<i>t</i>	<i>p</i>	95% Confidence Interval	
					Low	High
<i>Bridges</i>	4.521	3.570	1.270	.209	-2.578	11.619
ELP	-0.114	0.382	0.300	.766	-0.873	0.645
<i>Bridges</i> X ELP	-0.034	0.775	0.040	.965	-1.574	1.507
Some High School	-0.353	0.512	0.690	.492	-1.372	0.665
Some College/Tech Degree	0.268	0.521	0.510	.608	-0.768	1.303
Bachelor's	-0.077	0.937	0.080	.935	-1.940	1.786
Graduate/Professional Degree	1.216	0.802	1.520	.133	-0.378	2.810
Missing Education Level	-1.001	0.873	1.150	.255	-2.738	0.736
White	2.074	1.131	1.830	.070	-0.174	4.323
Black/African American	0.621	0.958	0.650	.519	-1.284	2.526
Asian	0.661	0.678	0.970	.333	-0.687	2.009
Multiracial, Native Hawaiian/Pacific Island, or American Indian/Alaska Native	2.220	1.239	1.790	.077	-0.244	4.683
Free/Reduced Price Lunch	1.281	0.544	2.350	.021	0.199	2.362
Student with Individualized Education Plan	-2.283	0.912	2.500	.014	-4.096	-0.470
Fall Grade 5 MAP Reading	-0.012	0.023	0.540	.588	-0.058	0.033
Female	-0.222	0.404	0.550	.585	-1.025	0.582
Cohort 1 (15-16)	2.640	1.118	2.360	.020	0.417	4.862
Cohort 2 (16-17)	4.103	1.154	3.560	.001	1.809	6.397
<i>Bridges</i> X Cohort 2	-3.814	1.674	2.280	.025	-7.143	-0.485
Digital Literacy Cohort 2	0.465	1.038	0.450	.656	-1.600	2.529
Digital Literacy Cohort 3	-0.491	1.128	0.440	.664	-2.733	1.751
Digital Literacy Cohort 4	-0.415	0.948	0.440	.663	-2.301	1.470
Digital Literacy Cohort 5	0.303	0.899	0.340	.736	-1.483	2.090
Digital Literacy Cohort 6	0.467	1.161	0.400	.689	-1.842	2.776



---

Intercept	8.594	3.942	2.180	.032	0.757	16.431
-----------	-------	-------	-------	------	-------	--------

---

Note.  $n = 1,367$ . Standard errors (SEs) corrected for 86 school-by-cohort clusters. Parent education level reference group is High School. Multiracial students and Native Hawaiian/Pacific Island or American Indian/Alaska Native students collapsed into single indicator variable due to very small sample sizes. ELP = ACCESS composite proficiency score.

## Appendix F

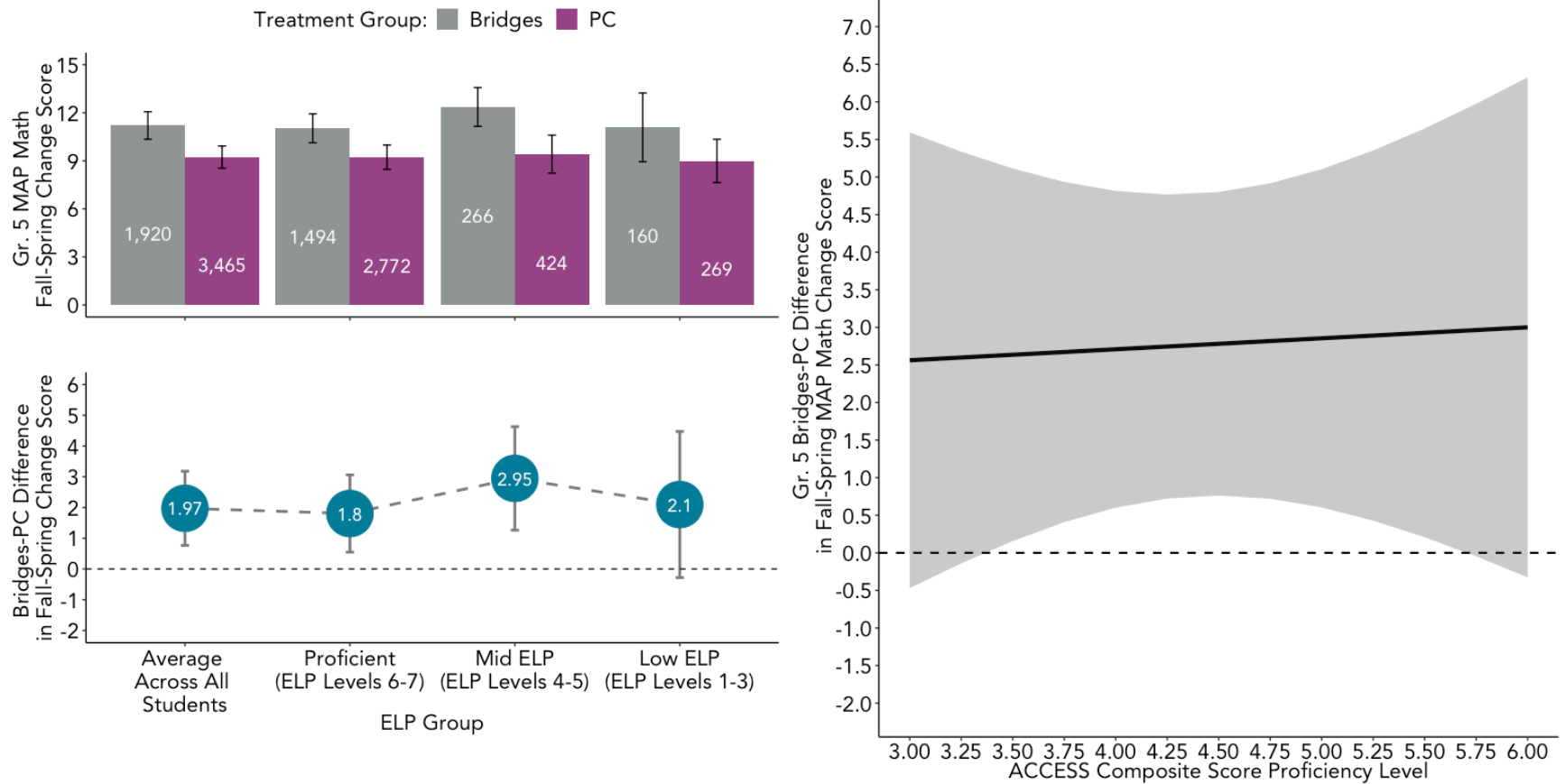
Table F1

*Unadjusted and adjusted change score estimates by treatment and ELP group*

	Unadjusted Unweighted Means ( <i>SD</i> )	Adjusted Weighted Means	<i>n</i>	Effect Size (Hedges's <i>g</i> )
<b>All Students</b>				
Bridges	10.647 (7.882)	11.236	1,893	0.250
PC	9.58 (8.198)	9.215	3,354	
<b>English Language Proficient</b>				
Bridges	10.553 (7.843)	11.005	1,426	0.226
PC	9.66 (8.334)	9.162	2,584	
<b>Mid ELP</b>				
Bridges	10.909 (7.170)	11.693	219	0.336
PC	9.493 (7.375)	9.232	527	
<b>Low ELP</b>				
Bridges	11.036 (8.896)	12.576	194	0.337
PC	8.914 (8.372)	9.670	243	
<b>Subgroup with ACCESS Scores<sup>a</sup></b>				
Bridges	11.172 (7.960)	12.089	440	0.376
PC	9.605 (7.719)	9.154	927	

*Note.* <sup>a</sup>Adjusted estimates from model 2. Each effect size calculated using *SDs* from each treatment group within each ELP subgroup.

## Appendix G



The figure above shows the results from our models using grade 5 ELP levels and ACCESS scores as opposed to grade 4, similar to our primary models presented in the main text. These results show that the primary impact of *Bridges* and the impact for students with ELP levels of 6 or 7 remain similar. The differences in effects between proficient, mid ELP, and low ELP were not statistically significant, and the effect did not significantly vary across ACCESS proficiency levels (graph on the right).

However, the effects vary in their size and confidence bounds compared to the primary results in the text. The variability in the estimates for mid ELP and low ELP students could occur for many reasons. First, this variability could be due to random variation in the effect estimate in part because the sample size being larger is slightly larger ( $n = 5,385$ ) when using the grade five ELP data. The difference in sample size likely contributes to the variation in point estimates. A third reason for the differences in these estimates is that ELP levels entail very different levels of instructional access and academic language skills at different grade levels. Students with lower ELP levels in fifth grade possess different levels of ELP (quantitatively and qualitatively) compared to the same ELP levels in fourth grade. These differences in the knowledge and skills underlying ELP at different grades may entail much different academic and language skills as well as different levels of access to core curricula. A last reason is that using fifth-grade ELP data means that there was one more cohort of students classified under the revised ACCESS ELP levels, which changed the distribution of proficiency level scores and the criteria to meet certain ELP levels. We control for cohort, which in part removes differences in these distributions due to cohort changes, though the 2016-17 standard setting is an important characteristic to keep in mind.

## Appendix H

### Multiple Testing Adjustment

Testing multiple correlated statistical tests increases the likelihood of finding an effect that may not actually be statistically significant under the conventional  $p < .05$  criterion for statistical significance. We had seven primary statistical tests for RQs 1 and 2. There were five tests based off Model 1: the main effect of *Bridges* in the full sample and comparisons of the size of the *Bridges* effect between each ELP level (three tests) as well as between Never/Former ELLs and low/mid ELP combined. Two were tests based on Model 2: the main effect of *Bridges* among students who took ACCESS, and the test of whether the effect of *Bridges* varied across ACCESS proficiency scores. We applied a correction called the Benjamini-Hochberg (BH) correction, and it helps control the false discovery rate (FDR) for statistical significance. What Works Clearinghouse (WWC) recommends this approach in their Group Design Standards, Version 4.1 (2020). Although WWC states that it is unclear whether  $p$ -value corrections are necessary for comparing multiple groups, we present this adjustment as an additional precaution against falsely finding a significant effect. Results of this procedure are presented in Table H1. The “ $p$ -value” column is the observed  $p$ -value from each statistical test for RQs 1 and 2. The column “BH Adjusted Critical Value” is the adjusted value below which the observed  $p$ -value must fall to remain statistically significant. The two significant effects we observed remain significant after adjustment.

Table H1.

Benjamini-Hochberg (BH) Corrections for Primary Statistical Tests

Term	Observed $p$ -value of Test	Rank	BH Adjusted Critical Value	Significant After Adjustment?
Bridges Main Effect Model 1	.001*	1	0.007	Yes
Bridges Main Effect Model 2	.003*	2	0.014	Yes
Bridges Differential Effects				
Low/Mid Combined vs. Proficient	.273	3	0.021	No
Low ELP vs. Proficient	.324	4	0.029	No
Mid ELP vs. Proficient	.464	5	0.036	No
Low ELP vs. Mid ELP	.724	6	0.043	No
Across ACCESS Composite Proficiency Scores	.964	7	0.050	No

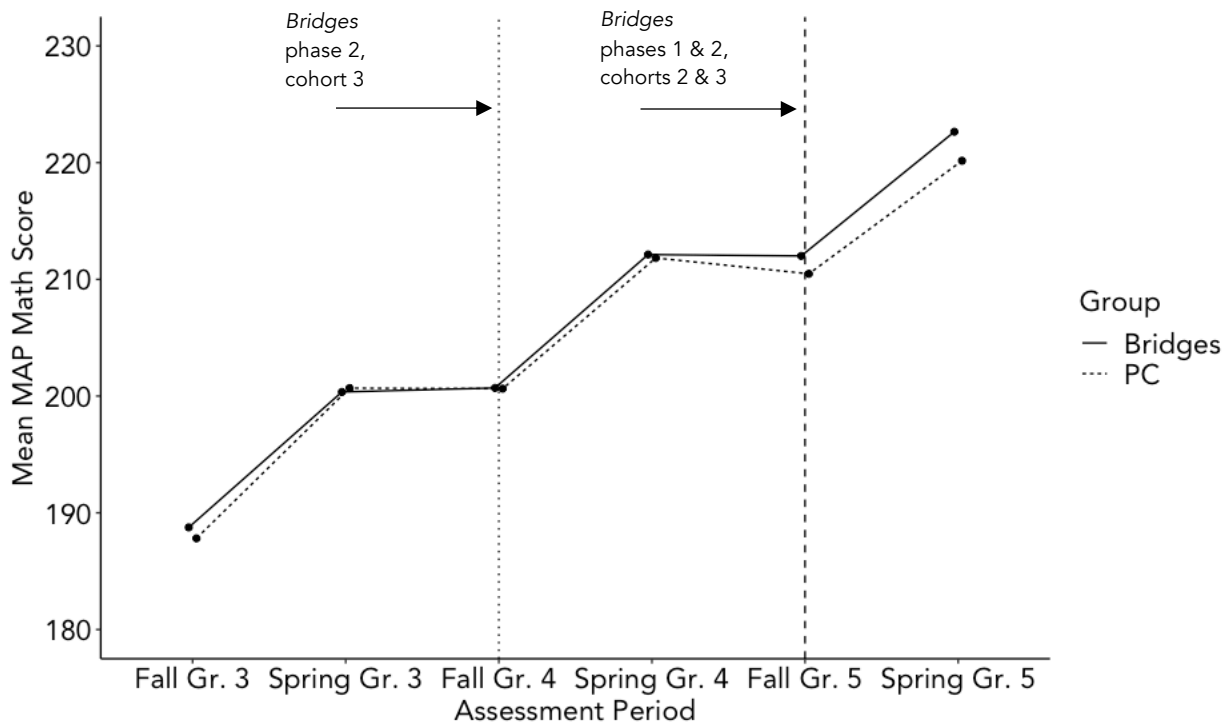
*Note.* The formula for the adjusted critical value is  $.05 * [\text{Rank}/7]$ . \*Significant at .05 level prior to adjustment.

### Assumption Checks for Statistical Tests and Differences-in-Differences Design

We checked the primary assumptions of our ordinary least-squares regression models. The principal assumptions for ordinary least-squares regression are homogeneity of variance (error terms are not correlated with predicted values), normality of residual errors, and no

problematic outliers that may bias or unexpectedly alter the estimates. We dealt with the homogeneity of variance and residual normality assumptions using cluster-robust variance estimation at the school-by-cohort level. We analyzed outliers by assessing the data descriptively prior to modeling as well as measuring residuals, Cook’s distance, and leverage from Models 1 and 2. A very small number of observations were identified as potentially problematic for estimation due to very large positive or negative change scores. Eliminating these observations slightly reduced impact estimates in model 1 and 2 (particularly in the low ELP group), but effects remained significant after trimming them from the dataset. Considering the minimal impact on the results as well as to accurately represent the distribution of student data, we retained these observations.

The primary assumption of the difference-in-differences design is the parallel trends assumption—that the two groups under investigation change in similar patterns prior to the policy or intervention. The two groups look similar in their pretreatment trends across the years, particularly during the academic year.



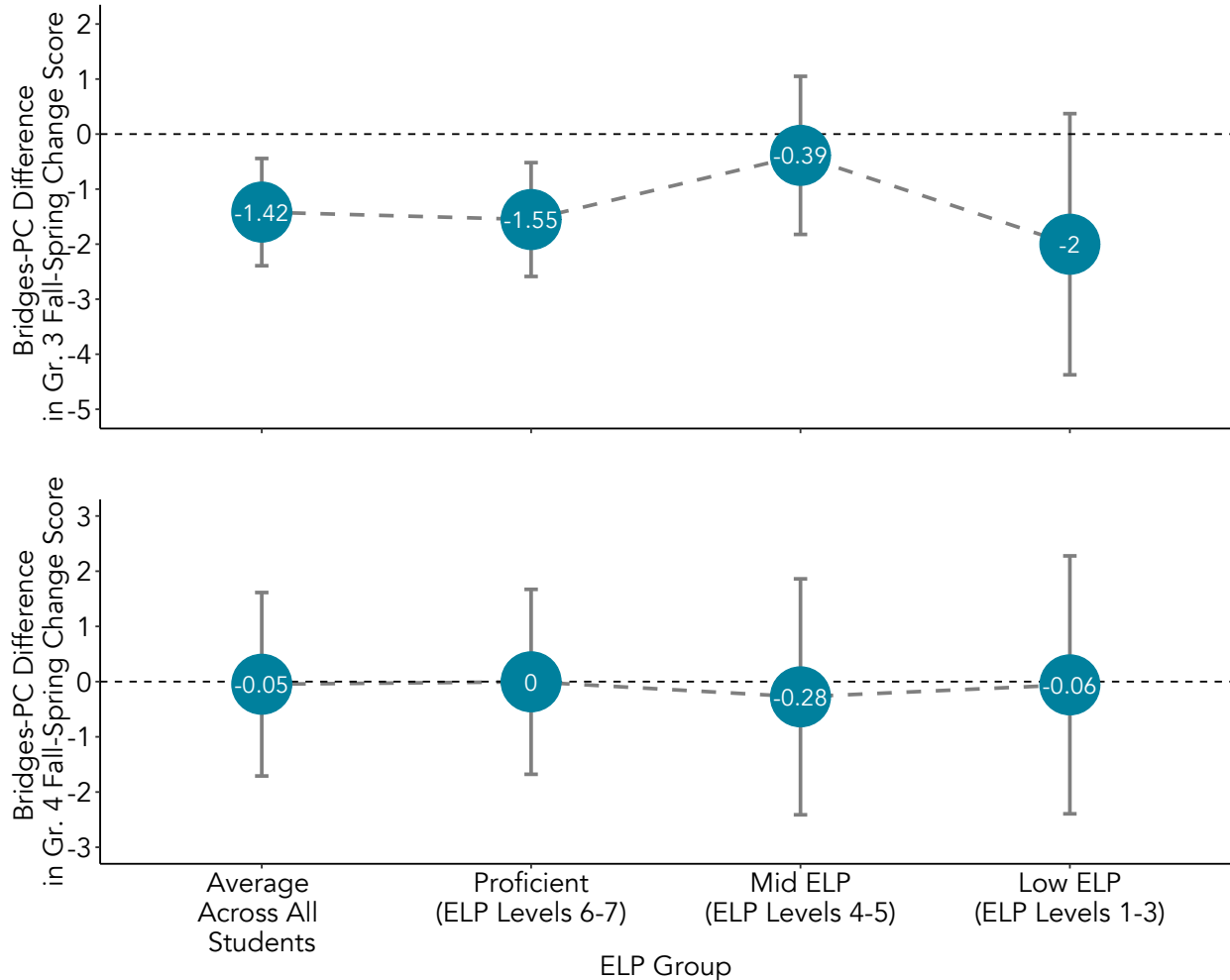
**Figure H1. Visualization of MAP Math Means of Students that Received Bridges of Previous Curriculum in Fifth Grade.**

Note. PC = Previous curriculum. Means based on students with complete data across all waves of assessment (PC  $n = 2,951$ , Bridges  $n = 1,637$ , total  $n = 4,588$ ). Means and trends similar when estimated using multilevel models that control for enrollment year.

### Robustness Checks

Robustness checks can provide greater basis and justification for the inferences and interpretation of the primary results (Furquim, Corral, & Hillman, 2020; St. Clair & Cook, 2015). One robustness check is a falsification test, which entails testing the treatment effect on the outcome in a period in which the intervention was not implemented. This check can also serve

as another test of parallel trends (St. Clair & Cook, 2015). We tested the treatment effect among third and fourth graders controlling for fifth-grade covariates, weighted by the inverse probability of having MAP math data in fall and spring of fifth grade in order to restrict the sample to the students who were observed and weighted in our primary analysis. The results are presented in Figure H2.

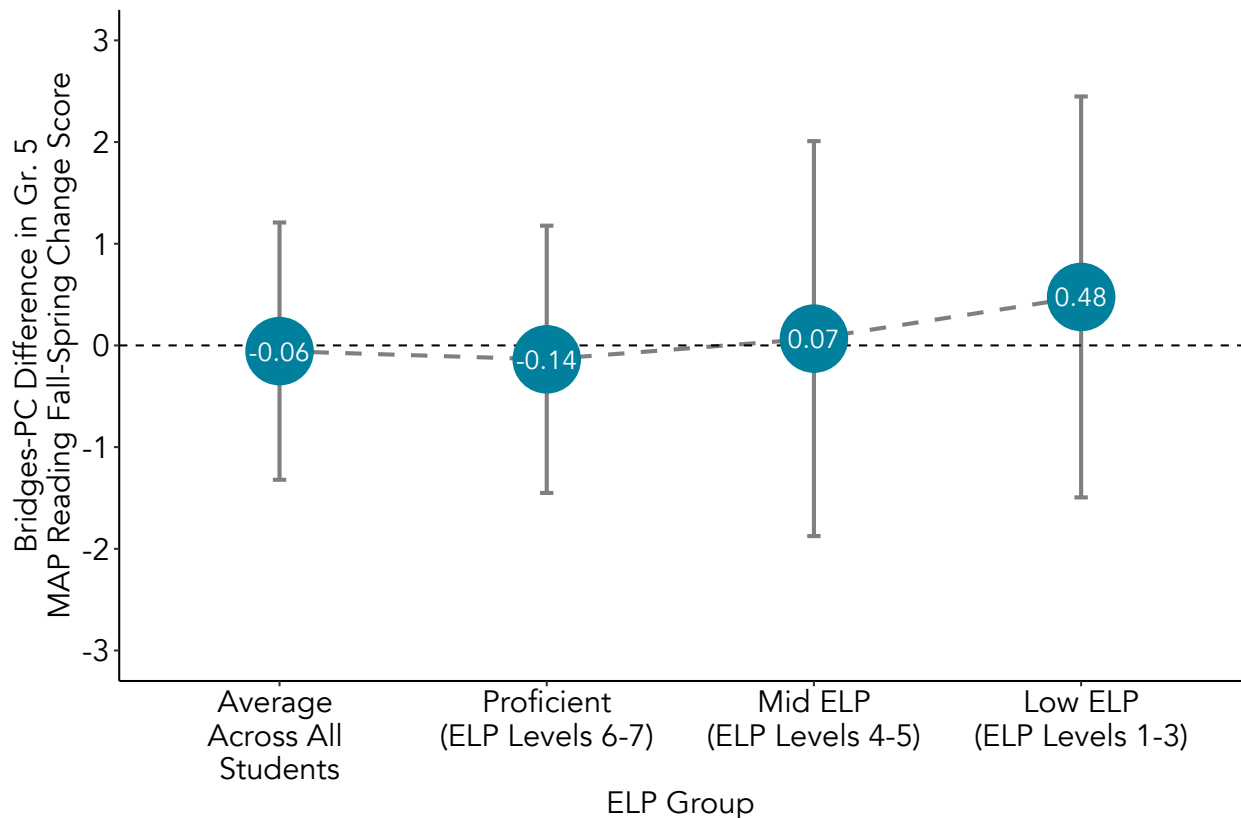


**Figure H2. Falsification Tests of Bridges Effects on MAP Math gains in Third (top) and Fourth (bottom) Grades.**

These results help us conclude that the *Bridges* impacts we observe are not due solely to arbitrary trends that existed prior to the curriculum implementation. Interestingly, the grade 3 model predicted a significantly negative impact of being in a *Bridges* school (similar to what is observed above in the parallel trends analysis). Of course, this effect was not due to the curriculum because it was not implemented yet. However, the trend is nonetheless important to consider in the broader context of the district and this study. Treatment effects are indistinguishable from 0 in grade 4; however, phase 2 fourth graders in 2016-2017 received *Bridges*. Any treatment effects among these students are not detectable in this falsification analysis.

Another robustness check is examining the treatment effect among a nonequivalent outcome, in this case MAP reading scores. Although *Bridges* could potentially improve reading scores, we would expect any effect of *Bridges* on MAP reading to be smaller relative to MAP

math since MAP reading would be more distally related to skills and content learned in *Bridges*. We found no evidence of significant effects of *Bridges* on MAP reading scores using the same set of covariates as in all other analyses (except fall MAP reading). Results were also weighted using the weights described in Appendix B. These effects were not of similar magnitude to MAP math in terms of effect size across groups.



**Figure H3. Bridges Effect Estimates on MAP Reading.**

We tested the robustness of our findings to different model specifications. Standard error estimates were similar clustered at the school-by-cohort level (87 clusters; 29 within each year assumed independent across years) or at the school level (29 clusters). Impact estimates were similar using a multilevel model with random intercepts at the school-by-cohort level ( $b = 2.04$ ,  $SE = 0.70$ ,  $p = .004$ ), though with random intercepts at the school level the primary impact estimate was smaller in magnitude ( $b = 1.68$ ,  $SE = 0.57$ ,  $p = .003$ ). Additionally, a single-level analysis of covariance (ANCOVA) with spring MAP math as the outcome (controlling for fall MAP math and covariates identical to model 1) showed a significant effect of *Bridges* ( $b = 2.12$ ,  $SE = 0.59$ ,  $p = .001$ ). Another recommended robustness check is running the difference-in-differences design as an event study, such that there are multiple pre- and post-intervention data points (St. Clair & Cook, 2015; Furquim, Corral, & Hillman, 2020). However, with only one post-*Bridges* data point, we could not effectively assess our results using an event study design.

Our last analysis to establish the rigor of our finding was a sensitivity analysis, which quantifies the amount unobserved confounding that would be necessary to change our finding to a null (or 0) effect of *Bridges*. For this procedure, we used the *Konfound-It!* R Shiny App (Rosenberg et al. 2018) to generate an estimate of the how correlated an omitted variable



would need to be with *Bridges* exposure and MAP math gains to change our conclusions (Frank, 2000). Based on a sample of 5,193, an average marginal effect estimate of 2.022, a cluster-robust standard error of 0.616, and an alpha level of .05, an omitted variable would have to have a correlation of 0.137 with *Bridges* exposure and the MAP math change score. Jointly, the impact of the omitted variable would need to be .019 to change our finding to a null estimate (Frank, 2000).

Finally, because some students received *Bridges* in fourth grade (but we measured the impact in fifth grade), we tested whether the amount of time exposed to *Bridges* altered the impact estimate. We ran an ordinary least-squares regression model among only the students in *Bridges* (along with all the covariates described in Appendix A [except the year indicator 2015 and 2016 year indicator interaction with treatment] as well as the weights described in Appendix B) and compared whether students who were receiving *Bridges* for the second year improved their score at a higher rate in fifth grade compared to first-year *Bridges* students. We found no evidence that second-year *Bridges* students grew significantly differently than first-year students ( $b = -0.28$ ,  $SE = 0.87$ ,  $p = .750$ ). This effect did not vary significantly by ELP level.

## References

- Abadie, A., Athey, S., Imbens, G.W., & Woolridge, J. (2017). *When should you adjust standard errors for clustering?* NBER Working Paper Series No. 24003. National Bureau of Economic Research. Retrieved from: <https://www.nber.org/papers/w24003>
- Angrist, J. D., & Pischke, J-S. (2009). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.
- Brown, G. (2017). *MAP Scores Report 16-17*. Madison, WI: Madison Metropolitan School District. Retrieved from: [https://accountability.madison.k12.wi.us/files/accountability/uploads/2017-10-3\\_map\\_scores\\_report\\_2016-17.pdf](https://accountability.madison.k12.wi.us/files/accountability/uploads/2017-10-3_map_scores_report_2016-17.pdf)
- Canivez, G.L. (2013). Incremental criterion validity of WAIS–IV factor index scores: Relationships with WIAT–II and WIAT–III subtest and composite scores. *Psychological Assessment*, 25, 484-495. doi: 10.1037/h0088996
- Chow, J. C., & Eckholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, 46, 179–186. doi: 10.1016/j.ecresq.2018.02.011
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Doabler, C. T., Clarke, B., Kosty, D. B., Baker, S. K., Smolkowski, K., & Fien, H. (2016). Effects of a core kindergarten mathematics curriculum on the mathematics achievement of Spanish-speaking English language learners. *School Psychology Review*, 45, 343–361. doi: 10.17105/SPR45-3.343-361
- Doabler, C. T., Clarke, B., Kosty, D., Smolkowski, K., Nelson, E., Fien, H., & Baker, S. K. (2019). Building number sense among English learners: A multisite randomized controlled trial of a tier 2 kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 47, 432–444. doi: 10.1016/j.ecresq.2018.08.004

- EdReports. (2018). *Bridges in Mathematics (2015)*. Retrieved from: <https://www.edreports.org/reports/overview/bridges-in-mathematics-2015>
- Frank, K. (2000). Impact of a confounding variable on the inference of a regression coefficient. *Sociological Methods and Research*, 29, 147–194. doi: 10.1177/0049124100029002001
- Furquim, F. F., Corral, D., & Hillman, N. W. (2020). *A primer for interpreting and designing difference-in-differences studies in higher education research*. In L. W. Perna (Ed.), *Higher Education: Handbook of Theory and Research Vol.:XXXV*. Springer.
- Kersaint, G., Thompson, D., & Petkova, M. (2013). *Teaching mathematics to English language learners*. Routledge.
- Kim, Y., & Steiner, P. M. (2019). Change scores revisited: A graphical models perspective. *Sociological Methods and Research*. Advanced online publication. doi: 10.1177/0049124119826155
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., Busick, M. D. (2012). *Translating the statistical Representation of the effects of education interventions into more readily interpretable forms*. (NCSE 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.
- LeFevre, J., Fast, L., Schwarchuk, Smith-Chant, B.L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Development*, 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Machado, A.F., Parente, P.M.D.C., & Santos Silva, J.M.C. (2011). *QREG2: Stata module to perform quantile regression with robust and clustered standard errors* [Computer software]. Retrieved from: <https://ideas.repec.org/c/boc/bocode/s457369.html>
- Madison Metropolitan School District. (2017). *2016-2017 Annual Report on the MMSD Strategic Framework*. Author. Retrieved from: <https://www.madison.k12.wi.us/files/www/uploads/2016-17-mmsd-annual-report.pdf>
- The Math Learning Center. (n.d.). *What you need to know about Bridges* [Handout]. Retrieved from: <https://www.mathlearningcenter.org/sites/default/files/documents/BridgesAssumptions.pdf>
- The Math Learning Center. (2019). *Bridges in Mathematics*. Retrieved from: <https://www.mathlearningcenter.org/bridges>
- McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., ... Purcell, S. (2019). *The condition of education 2019* (NCES 2019-144). U.S. Department of Education. National Center for Education Statistics. Retrieved from <https://nces.ed.gov/pubs2019/2019144.pdf>
- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on post-treatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62, 760–775. doi: 10.1111/ajps.12357
- Northwest Evaluation Association (2015). *Measures of Academic Progress*. Author.
- Rosenberg, J. M., Xu, R., & Frank, K. A. (2018). *Konfound-It!: Quantify the robustness of causal Inferences*. Retrieved from: <http://konfound-it.com>.

- Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22, 278–295. doi: 10.1177/0962280210395740
- SEG Measurement. (2018). *An Evaluation of the Effectiveness of Bridges in Mathematics for Developing Student Math Skills*. Author. Retrieved from: <https://www.mathlearningcenter.org/sites/default/files/documents/Bridges%20in%20Mathematics%20Effectiveness%20Study.pdf>
- St. Clair, T., & Cook, T. D. (2015). Difference-in-difference methods in public finance. *National Tax Journal*, 68, 319–338. doi: 10.17310/ntj.2015.2.04
- Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children’s mathematical development. *Journal of Experimental Child Psychology*, 115, 227–244. doi: 10.1016/j.jecp.2013.02.002
- Walker, A. C. (2017, October 23). Madison elementary school students explore the district's new math curriculum. *The Cap Times*. Retrieved from: [https://madison.com/ct/news/local/education/madison-elementary-school-students-explore-the-district-s-new-math/article\\_507a1743-0bd7-5f4d-b655-ce75c4cc8a2a.html](https://madison.com/ct/news/local/education/madison-elementary-school-students-explore-the-district-s-new-math/article_507a1743-0bd7-5f4d-b655-ce75c4cc8a2a.html)
- What Works Clearinghouse (2020). *Procedures handbook version 4.1*. Author. Retrieved from: <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf>
- Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, 12, 308-331. Retrieved from: <https://journals.sagepub.com/doi/pdf/10.1177/1536867X1201200209>
- Wisconsin Department of Public Instruction. (2019). *ELL status enrollment, single year, grade 5 Madison Metropolitan School District* [Data set]. WISEdash. Retrieved from: <https://wisedash.dpi.wi.gov/Dashboard/portalHome.jsp>
- Youngstrom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly*, 14(1), 26–39. doi: 10.1037/h0088996