# *Eureka Math: Research Brief*
Winter 2023

*This report presents quasi-experimental estimates of the impact of Eureka Math on 4th and 5th grade test scores within a school district that implemented Eureka Math over a three-year period. Estimates indicate Year 1 Eureka pilot schools experienced an initial average gain of 3.28 percentiles in 4th and 5th grades, and approximately half a percentile annual gain thereafter. In contrast, schools that implemented Eureka Math in Year 2 experienced an initial drop of approximately half a percentile on average in 4th and 5th grades, but scores then recovered by an average of 0.75 percentile annually thereafter. Estimates for the schools that implemented Eureka Math in Year 3 are incomplete. Specifically, the schools that implemented Eureka Math in the final pilot year experienced an initial significant drop of 2.65 percentiles, but COVID disruptions impeded estimated effects on students test scores over time.*

## *Research Overview*

Research suggests that the choice of instructional materials can both improve teaching practice and improve student learning (Chingos & Whitehurst, 2012). For example, Agodini and Harris (2010) conducted a study in which they randomly assigned one of four math curricula to teachers across 10 states over three school years. The authors found that classrooms using two of the curricula outperformed those using the other two curricula by achievement increases equivalent to 9 to 12 percentiles.

Another study provides evidence that both access to, and training on, high-quality curriculum is especially effective. Specifically, a randomized control trial conducted by Jackson and Makarin (2018) found that simply providing middle school teachers with curricular materials alone produced student math achievement gains (i.e., a o.o6 standard deviation increase). However, when math teachers received both the materials and instructional support, students experienced larger math achievement gains (o.o9 standard deviation increase). In addition, the researchers noted that the largest achievement gains were found among students who were placed with weaker teachers or first-year teachers.

The two studies described above illustrate the potential benefits of curriculum on teaching practice and student achievement. Nevertheless, not all materials are equal in quality, nor do they necessarily have immediate impacts. The implementation process can take more than one school year to produce positive effects. For example, a multi-year, randomized control study of a curriculum in first through third grade found no statistically significant differences in student achievement gains during the first year of implementation (Eddy et al., 2014). However, by the second year, students who received the curriculum for two consecutive years outperformed their control group counterparts on measures of student achievement. Thus, a single school year may simply not be enough time to fully evaluate the effect of curricular changes on student learning.

The study presented in this report analyzes the relationship between learning from Eureka Math ("Eureka") and students' math test scores within schools that implemented Eureka for one, two, and three years. Specifically, this study examines a large urban school district in North Carolina that adopted Eureka over a three-year period—three elementary schools adopted the curriculum in the 2016-17 school year; 16 schools adopted the curriculum the following year; and all elementary schools (over 50) in the district adopted Eureka by the 2018-19 school year. When feasible, we present an estimate of the initial impact of Eureka, as well as the estimated impact over time. The findings presented here employ a quasi-experimental methodology (i.e., ESSA Tier II methodology) called a comparative interrupted time series (CITS), that allows us to examine trends over time (Shadish et al., 2001; Somers et al., 2013).

## *Data & Methodology*

This study utilizes longitudinal data reported to the state from the 2012-2013 through 2018-2019 school years. Eureka was first implemented in the district during the 2016-17 school year. Student test scores in 4th and 5th grade from the district that implemented Eureka are compared to the district's pre-Eureka implementation. In addition, test scores from the district that implemented Eureka are also compared to a neighboring district that has the most similar demographic characteristics and economic conditions.

Student learning is measured by the North Carolina End-of-Grade Math test scores. In the elementary grades, the test is administered in 3rd through 5th grade and measures students' operations and algebraic thinking; number and operations in base ten; number and operations (fractions); measurement and data; and geometry. The test includes 40 items and contains multiple choice, numeric entry, and technology-enhanced type questions. All tests are administered within the last 10 days of the school year, providing a measure of students' year-end math performance.

The estimates presented in this report were produced using quasi-experimental design methodology (i.e., ESSA Tier II evidence). Specifically, we first identified a neighboring district with similar school, demographic, and economic conditions. Next, we estimated propensity scores using test score and demographic data from the years before the Eureka implementation to match each school in the Eureka implementing district with the most similar school in the neighboring district.

Next, we used CITS to estimate the impact of Eureka. We present two main CITS estimates: how student test scores changed with the introduction to Eureka (i.e., the initial impact) and how test scores changed over time when schools continued to use Eureka. The estimates compare if students' state test scores have changed from students' earlier test scores (i.e., from before the Eureka implementation to using Eureka) by more than the comparison group (i.e., students in the neighboring district). In addition, the estimates control for differences in students' prior test scores, demographic characteristics, and school size.

The data include district- and school-level characteristics, as well as students' test scores and demographic characteristics (e.g., race, sex, and special education status). Table 1 displays 4th and 5th grade student characteristics from the Eureka implementing district, both the students that begin learning from Eureka (labeled as "EM") and those using the prior math curricula ("No

EM"). In addition, the characteristics of the students from the neighboring district matched to the students using Eureka are shown in the comparison district. The data is displayed for the years during the Eureka implementation.

*Table 1*

Student Characteristics of Sample

| | 2016-17 | | | 2017-18 | | | 2018-19 | |
|---|---|---|---|---|---|---|---|---|
| | Comparison | Eureka District | | Comparison | Eureka District | | Comparison | Eureka District |
| | | EM | No EM | | EM | No EM | | EM |
| Female (%) | 0.49 | 0.49 | 0.49 | 0.48 | 0.50 | 0.48 | 0.52 | 0.49 |
| African American (%) | 0.37 | 0.25 | 0.31 | 0.41 | 0.23 | 0.32 | 0.37 | 0.30 |
| Hispanic (%) | 0.12 | 0.53 | 0.41 | 0.14 | 0.48 | 0.38 | 0.14 | 0.41 |
| Other (%) | 0.06 | 0.13 | 0.18 | 0.08 | 0.18 | 0.18 | 0.10 | 0.18 |
| White (%) | 0.44 | 0.10 | 0.11 | 0.38 | 0.11 | 0.11 | 0.38 | 0.11 |
| Disabilities (%) | 0.15 | 0.11 | 0.15 | 0.15 | 0.15 | 0.14 | 0.13 | 0.15 |
| Number of Students | **601** | **409** | **9,438** | **2,895** | **2,605** | **7,524** | **12,965** | **10,158** |

Table 1 provides a description of the students who learned from Eureka and students in the comparison district. Note that there are some demographic differences between the two districts (e.g., the percentage of white and Hispanic students). However, this district represents the most demographically and economically similar neighboring district, an important consideration for selecting the comparison schools (Hallberg et al., 2018; Somers et al., 2013). In addition, the table shows that the early adopters of Eureka served students with different demographic characteristics than the average elementary schools in the district. For example, the first schools that adopted Eureka had fewer students with disabilities and African American students, and more Hispanic students. In addition, the table shows the increasing number of 4$^{th}$ and 5$^{th}$ graders learning from Eureka across the implementation years.

## *Results*

The figures, below, provide estimates of the initial impact of Eureka implementation on the left side, and then the annual percentile change in test scores as schools continue to use Eureka on the right.

A cohort of three schools were the first Eureka adopters in the district during the 2016-17 school year. Figure 1 shows the 4$^{th}$ and 5$^{th}$ grade estimates of the treatment effect, or impact of Eureka adoption during the first year in the left column. Specifically, average grade-level scores within these early adoption schools increased by 3.28 percentiles. For example, if average test scores at the 4$^{th}$ grade were in the 50$^{th}$ percentile, average grade-level scores moved to above the 53$^{rd}$ percentile. This difference is marginally statistically significant ($p<0.14$).

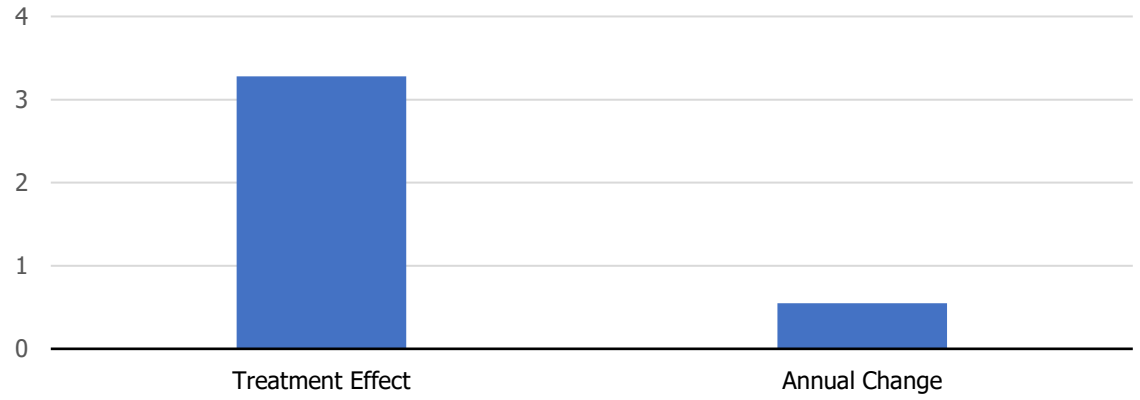
The right column shows the average annual change over time at the grade level. Figure 1 shows that average grade-level test scores increased by approximately half a percentile (0.55) annually.

For example, a movement from the 53rd percentile in 2016-17 school years to the 54th percentile in the 2018-19 school years. This estimate is not statistically significant.

These results suggest that, among the first adoption schools, Eureka caused immediate average gain in 4th and 5th grade students' math scores, and that average math scores continued to increase as these schools continued to use Eureka.

*Figure 1*

Eureka Math Treatment Effect and Annual Test Score Change in First Cohort of 4th and 5th Grade Students (Percentiles)



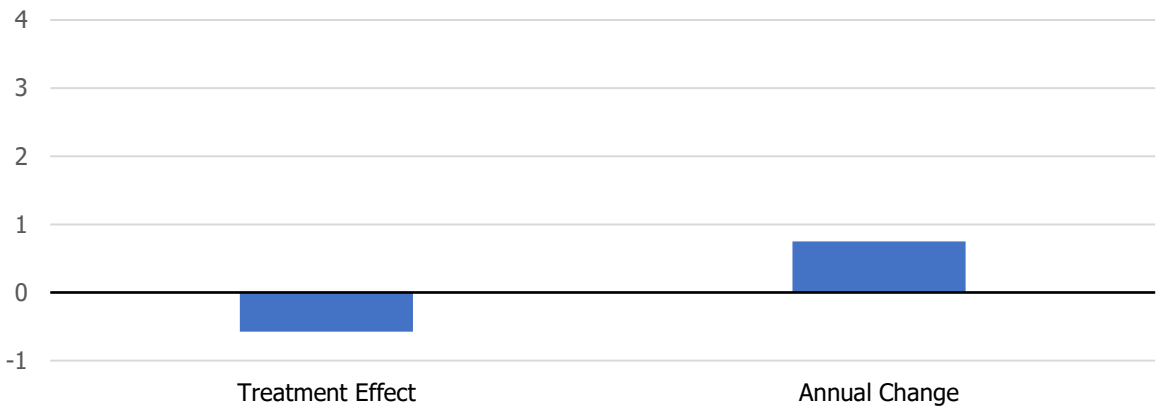

*Notes*—1. The bars on the graphs represent the grade-level average difference in student achievement in the first year of Eureka implementation ("initial change") and the change over time with continued exposure to Eureka, compared to schools' prior no Eureka exposure. 2. Statistical significance is indicated in the following way: ~p<.10, *p<.05, **p<.01, ***p<.001.

During the 2017-18 school years 16 additional schools adopted Eureka. Figure 2 shows the estimated initial impact and change over time of Eureka on the 4th and 5th average grade level scores. Note that math scores dropped slightly in the first year of Eureka implementation for this cohort, but then rose over the next school year. Specifically, the right panel shows that average tests scores decreased by 0.58 percentile after the first year of Eureka implementation. However, over the next year of using Eureka, average grade-level test scores increased by 0.75 percentile. Neither change is statistically significant.

These results suggest that Eureka caused a net gain in math scores among this second cohort of schools, with an initial decrease followed by a larger annual increase in average math scores as these schools continued to use Eureka.

*Figure 2*

Eureka Math Treatment Effect and Annual Test Score Change in Second Cohort of 4th and 5th Grade Students (Percentiles)




1. The bars on the graphs represent the difference in student achievement after 1, 2, and 3 years of Eureka exposure, compared to no Eureka exposure. 2. Statistical significance is indicated in the following way: ~p<.10, *p<.05, **p<.01, ***p<.001.

During the 2018-19 school year, the rest of the elementary schools in the district implemented Eureka Math (i.e., over 50 schools). We are able to estimate the initial impact of Eureka Math but cannot show how students test scores changed over time, as COVID-19 disrupted state testing in the subsequent year. As such, our estimate provides an incomplete picture of Eureka's impact of students' math learning within these schools. Nonetheless, estimate indicate that students experienced a statistically significant decrease in average test scores of 2.65 percentiles with the introduction of Eureka math.

## *Conclusions*

The results of this study suggest that the initial impact of Eureka Math on student learning depends on school conditions. Early adopters of the program experienced relatively large gains. This suggests that the first year of implementation can successfully raise students' test scores, when schools are eager and/or equipped to support the new curriculum. Estimates from the second group of adoption schools show a small average dip in test scores, followed by larger gains. Thus, these schools experienced a net increase in test scores, but required more time to adjust to the new curriculum. Estimates from these schools also show that learning can increase from the use of Eureka. Unfortunately, estimates from the third group of Eureka adopters are truncated. While we can estimate the initial impact of the curriculum on student learning, we do not know how students faired once the schools adjusted to the new curriculum.

## *References*


Agodini, R., & Harris, B. (2010). An Experimental Evaluation of Four Elementary School Math Curricula. *Journal of Research on Educational Effectiveness*, *3*(3), 199–253. https://doi.org/10.1080/19345741003770693

Chingos, M., & Whitehurst, G. (2012). *Choosing Blindly: Instructional Materials, Teacher Effectiveness, and the Common Core*. Brookings Institution. https://www.brookings.edu/research/choosing-blindly-instructional-materials-teacher-effectiveness-and-the-common-core/

Eddy, R. M., Hankel, N., & Murphy, K. (2014). *Houghton Mifflin Harcourt GO Math! Efficacy study year one final report*. Cobbleston Applied Research and Evaluation, Inc.

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short Comparative Interrupted Time Series Using Aggregate School-Level Data in Education Research. *Educational Researcher*, *47*(5), 295–306. https://doi.org/10.3102/0013189X18769302

Jackson, K., & Makarin, A. (2018). Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment. *American Economic Journal: Economic Policy*, *10*(3), 226–254.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Somers, M.-A., Zhu, P., Jacob, R., & Bloom, H. (2013). *The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation.* (MDRC Working Paper on Research Methodology). New York, NY. https://www.mdrc.org/sites/default/files/validity_precision_comparative_interrupted_time_series_design.pdf