Study Information

Hypotheses

Students assigned to participate in the early numeracy tutoring program will have higher achievement on a validated end-of-year assessment (e.g., iReady, aMath, STAR Math, provider assessment) than students assigned to the control condition.

We will supplement analysis of this main hypothesis with exploratory analyses that will address whether the effectiveness of the program varies by district and school features, underlying student characteristics (i.e, student gender, race/ethnicity, grade level, socioeconomic disadvantage, English learner status) or prior achievement and behaviors (e.g., attendance).

Our primary, confirmatory research question is:
To what extent does a math tutoring program for young learners affect students' early math achievement?

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Is there any additional blinding in this study?

Note: While students know if they are participating in tutoring, students will not be aware of the study and randomization procedure.

Study design

The student-level randomization will be stratified by grade and schedule. In each school, the affiliated district coach with support from a program provider coach will identify the available number of intervention slots within schools and share the pool of treatment-eligible students (i.e., below grade level in math and/or in-need by the school) with the provider. Schools will be given flexibility to determine how to spread tutor spots across grade levels. These rosters of students, which will exceed the available space on a tutor's caseload, will then be shared with the Stanford researchers, who will conduct the random assignment. The research team received a waiver of consent from the IRB prior to implementation.

We intend for variation in treatment status to exist within each classroom with students identified for tutoring. The minimum tutoring caseload that the program expects is 24 students among full-time tutors and 14 students among part-time tutors. Some schools may have multiple tutors (i.e., with two full-time tutors, 48 students would be expected to receive treatment). The intended within-school treatment to control ratio is 50-50, although an approximate ratio of 60-40 is permitted if fewer than 24 students can be identified for the comparison group (in which case the minimum treatment assignment would be 16 students). The provider will then update their data management system with tutoring assignments and the district-provider coaching pairs will schedule the tutoring. To establish a clean pre-treatment baseline for both conditions, tutoring will not be implemented until the fall benchmark assessment has been completed.

No files selected

Randomization

See above, in study design.

Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

As of the date of submission of this research plan for pre-registration, the data have not yet been collected, created, or realized.

Data collection procedures

Districts will execute data sharing agreements with Stanford directly. Stanford will collect secondary data including formative assessment (e.g., iReady, aMath, STAR Math) scores and demographic data which the research team will use to study the impact of the program on K-3 students. At the conclusion of the 2025-26 and 2026-27 school years, the research team will request outcome data again to measure the longitudinal effects of participation in tutoring.The district will also provide student demographics, and attendance. Below, we provide more details on the data provided:
Student identifiers: Student ID, school, grade-level
Student math achievement: prior end-of-year, beginning-, middle-, and end-of-year benchmark assessment name, composite score, and subscale score (if applicable).
Student attendance: Student-by-day attendance in school
Student demographics: Race/ethnicity, gender, economically disadvantaged indicator, special education indicator, English Learner designation status, and enroll/exit dates, if applicable
Teacher data: Teacher identifier, and (as available) teacher experience, race/ethnicity, and gender

The provider will also provide baseline and endline data from their own formative assessment, measures of math fact fluency in select grade levels, and information on tutoring fidelity, dosage, and curricular progression.

The experiences of students in the control group will be monitored periodically to document the degree to which those students received supplemental math support outside the scope of Math Corps.

No files selected

Sample size

We estimate a sample size of approximately 1000 students, across approximately 25 schools each of which contain four grade-level randomization blocks.

Sample size rationale

We base the assumptions for our power analysis on current best practices and knowledge from prior and on-going work with math tutoring. We conducted our power analysis using the Power Under Multiplicity Project (Hunter et al., 2022). We anticipated recruiting 15 to 30 schools and are currently projected to recruit 25. Averaging across expected loads for full and part-time tutors, we anticipate an average caseload of 21 students per tutor. We are aiming for a 50/50 split in the share of students in the treatment versus control condition though it is possible treatment students may somewhat outnumber control students if not enough eligible students are identified within school-grade strata.
For our power calculations we assumed 50% of the eligible students will be able to receive the treatment per block (i.e., 12 students). Furthermore, we assume that student-level covariates explain up to 50 percent of the variance in the outcomes (R2) and that intraclass correlations (ICC) within blocks explains up to 15 percent of the variance in the outcomes.
These values are conservative based on the amount of explained variance we observed in our other work that uses a similar experimental design and outcomes (JPAL, 2021). We set the values for alpha and beta at the customary cut-offs of 0.05 and 0.80 respectively.
Given these assumptions, we calculated that our conservative recruitment scenario sample size would have power to estimate a minimum detectable effect size (MDES) of 0.14-standard deviations (SD) on student assessments. This MDES is less than half of the average effect size of math tutoring interventions in the literature delivered by paraprofessionals (ES = 0.28-SD; Nickow et al., 2024) and is smaller than the effects observed in prior Math Corps

evaluations in older grades on STAR Math assessments (ES = 0.18 and 0.17 in Codding et al., 2022 and Parker et al., 2019, respectively). Our MDES will decrease to <0.1 SD if we meet our higher recruitment goal of 30 schools, slightly above our current expected sample of 25 schools.
We anticipate we will have sufficient power to detect an effect after year 1 of the implementation. Our ability to detect longitudinal effects will depend on our sample size after year 2 (i.e., how many schools do not provide the MathCorps program to students eligible during year 1), which we cannot confirm at this time.

Stopping rule

There was no stopping rule.

Variables

Manipulated variables

We randomly assigned students to one of two conditions for within-school math instructional support.
Treated students were assigned to receive 20-minutes of tutoring in pairs, five days a week, in addition to core classroom instruction. All tutor interventions are standardized and characterized by modeling, a high volume of practice opportunities for students, and immediate performance feedback. Within a given grade-level, students progress through a series of units composed of related interventions that build conceptual, procedural, and applied math skills. Each grade level has a unique scope and sequence. Within grades, students progress from one unit to the next using brief mastery assessments to guide service delivery and feedback.
Control: Students not assigned to participate in the tutoring program received “business-as-usual” (BaU) instruction and support during the school day.

No files selected

Measured variables

Covariates:
Beginning-of-year district and provider formative assessment composite and subtest (i.e., numeracy) scores
School, grade, and classroom membership
Student demographics: gender, race/ethnicity, economic disadvantage (e.g., FPRL), and indicators for student classification as an English learner or as having a disability

Primary student outcome:
End-of-year (AY 25-26) district formative assessment composite score

The district composite scores are a composite of the individual tested sub-skill scores. They are calculated by the testing organizations and the formulas are available in their respective technical reports.

Other outcomes:
End-of-year (AY 25-26) provider assessment composite score
Achievement on subscales of EOY 25-26 composite assessments
Fact fluency proficiency from provider assessments
Middle-of-year district assessment composite and subtest (i.e., numeracy) scores
Student school attendance
District formative assessment composite scores year 2 (AY 26-27)
Third grade EOY state math assessment scores

Additional data collected:
Classroom teacher characteristics
Student tutoring session attendance and dosage
Session-level data: duration, content. performance
Tutor characteristics: gender and race/ethnicity, educational level, age prior tutoring experience

No files selected

Indices

N/A

No files selected

Analysis Plan

Statistical models

Our primary model for student-level outcomes will control for strata (classroom) and student covariates (e.g., gender, race, English learner status, grade level, beginning of year scores). To evaluate the impact of being assigned to the treatment condition (versus the control), we will conduct OLS regressions to generate intent-to-treat estimates. We will also run a model that only controls for strata, to assure that the additional controls affect the power but do not substantially change the point estimates. We will use robust standard errors (see Abadie et al., 2023).

See pdf attachment for our full statistical model.

Math Corps Estimating Equation.pdf

Transformations

To compare student achievement across multiple assessment types we will standardize scores within grade, administration period, and assessment type using control groups means and standard deviations.

Inference criteria

To evaluate the impact of being assigned to the treatment condition (versus the control), we will use OLS regressions to generate intent-to-treat (ITT) estimates. We only have two primary outcomes that we anticipate will be correlated with one another, and we will use 95% confidence intervals to evaluate our analysis. This corresponds to a critical p-value of .05.

Data exclusion

We will not trim data for outliers. For our primary ITT analysis we will retain all students identified by the district as eligible to be randomized into either tutoring or control condition.

In additional sensitivity analyses we will exclude students who become ineligible for tutoring services (e.g., due to schedule changes, leaving the school, reclassification) after tutoring has started, if eligibility withdrawal appears to be equal across conditions.

In exploratory analysis of year II outcomes, we will exclude observations if all students (including control students) in a grade-school stratum become eligible for and receive Math Corps.

Missing data

We expect student-level data to be accurate and complete to the extent that the data exists. We will assess and report the extent to which tutors and students attrited from the sample at different rates based on condition assignment. We will remove observations from the analytic sample if they are missing the outcome measures we are evaluating in our analysis.

We will address missing baseline information for participants using What Works Clearinghouse aligned procedures. To retain observations missing information for categorical variables used as covariates (e.g., race/ethnicity, gender), we will impute a constant value and add a “missing” indicator. If students are missing data for a continuous baseline covariate (e.g., beginning-of-year math score), we will conduct analyses using an imputed score predicted from non-missing demographic and outcome data, conditional on treatment assignment. We will explore the sensitivity of key findings to these missing data procedures by also presenting results estimated using only observations that do not require imputation (i.e., complete case analysis), as well as estimates from specifications that exclude baseline controls, across relevant samples. (e.g., with and without “incomplete observations”).

Exploratory analysis

We will explore several exploratory research questions, including:
Exploratory 1: To what extent does a math tutoring program for young learners affect students' early math achievement the year after initial tutoring was offered?
Exploratory 2: How does the intervention effect vary by students' underlying characteristics? (including prior performance (i.e., baseline), race/ethnicity; free and reduced priced lunch (FRPL) status; English Language Learner (ELL) status, special education (SPED) status)
Exploratory 3: How does the intervention effect vary by tutoring session attendance? What are the effects among students who receive the intended dosage?
Exploratory 4: How does assignment to condition impact other student outcomes, including school attendance?

Other

Other

No data