# The Impacts of Tracking on Accumulated Learning Losses: Experimental Evidence from Brazilian Middle-schoolers Five Years After the Pandemic*

# Pre-analysis Plan

Guilherme Lichand[†], Thiago da Costa[‡] and Susanna Loeb[§]

May 22, 2025

## 1 Study Background

The COVID-19 pandemic has caused significant disruptions in children's learning worldwide, with particularly severe impacts in lower- and middle-income countries, where many children attending public schools lacked access to effective remote learning solutions. In Brazil, the setting of our study, millions of children and adolescents faced prolonged periods without classes – an average of 266 days without in-person classes, the second longest in Latin America. Even before the pandemic, public education in Brazil grappled with numerous challenges; remote learning exacerbated these issues. Middle-school students were more adversely affected (Lichand and Doria, 2024), returning to classrooms far below the expected proficiency levels. As a result, these students now lack the foundational literacy and numeracy skills to learn at grade level.

Such a challenging educational landscape underscores the urgent need for scalable,

*Study conducted in the context of a technical cooperation regulated by a Memorandum of Understanding between the Stanford Lemann Center and the São Paulo State Secretariat of Education.

[†]Stanford Graduate School of Education; glichand@stanford.edu.
[‡]São Paulo School of Economics (FGV/EESP).
[§]Stanford Graduate School of Education.

targeted interventions. Correlational evidence from Brazil suggests that remedial tutoring and individualized learning can accelerate learning recovery in the aftermath of the pandemic (Lichand and Doria, 2024). A growing body of causal research confirms that tutoring significantly improves academic outcomes—particularly for disadvantaged students—and can also enhance attendance and reduce dropout rates (Dietrichson et al., 2017; Nickow et al., 2020; Cabezas et al., 2021). High-impact tutoring models—characterized by frequent sessions, personalized support, and strong curricular alignment—have delivered particularly large gains (Cortes et al., 2023).

However, replicating these models in low- and middle-income countries faces structural constraints. High-impact tutoring often relies on intensive formats (e.g., 1:1 to 1:4 ratios) that are rarely feasible or sustainable in public education systems where class sizes of 30+ are common. Moreover, much of the literature focuses on primary school students with mild learning lags, whereas in many LMICs, middle school students are frequently three or more years behind grade level. Whether similar interventions can succeed under these more demanding conditions remains an open—and crucial—question.

To address these concerns, this study evaluates a large-scale remedial education program implemented across 899 secondary schools in São Paulo State, Brazil. The intervention targets middle school students identified as lagging behind based on standardized assessments and removes them from their regular math and Portuguese classes. These students are reassigned to remedial classrooms with reduced pupil-teacher ratios—larger than high-impact tutoring models but significantly smaller than standard classrooms—and are taught using a simplified, low-tech approach to individualized instruction. Students are initially diagnosed using a short assessment and then progress through curricular content spanning grades 1 to 5 at their own pace, allowing for heterogeneous levels within the same classroom. This design embraces the principles of Teaching at the Right Level (TaRL) without depending on technology—an important consideration in Brazilian public schools, where robust digital infrastructure is often lacking.

A key advantage of our approach is that it avoids the logistical and bureaucratic complexity of after-school tutoring, which typically requires additional staffing, transportation, meals, parental coordination, and often faces resistance from students. By embedding the intervention during regular school hours and removing students from their usual classes, we can directly test whether within-school remediation is an effective, scalable way to reduce learning gaps.

Our evaluation strategy also goes beyond conventional academic metrics. In addition to using standardized government assessments in math and Portuguese, we incorporate three diagnostic tools: ROAR and SMARTE, which assess foundational literacy and numeracy skills, and a computer-based adaptive test (CAT), an adaptive assessment of math and Portuguese that is aligned to the curriculum of each grade, and whose results try to place each student's ability within a specific grade. These tools are particularly valuable for capturing progress among students with severe learning deficits, whose improvements may not be reflected in standard assessments but are evident when measured through foundational skills and adaptive diagnostics.

This study builds on and extends several strands of existing research. Muralidharan et al. (2019) evaluate a blended after-school program in India combining computer-assisted learning (CAL), group tutoring, and additional instruction time. They report large learning gains ($0.37\sigma$ in math, $0.23\sigma$ in Hindi), but their design does not disentangle the individual contributions of each component, and the reliance on after-school access limits its scalability. By contrast, our program is integrated into the regular school schedule and uses a low-tech strategy to personalize instruction, making it more realistic for LMIC systems.

The Balsakhi program evaluated by Banerjee et al. (2007) provides perhaps the closest structural parallel: struggling primary students were removed from regular classrooms and placed in small remedial groups with community tutors. The intervention delivered sizable gains (up to $0.40\sigma$ for the lowest-performing students). Our program adopts a similar approach but targets middle schoolers, who are typically further behind and potentially face greater stigma when remediated. Whether such an intervention remains effective for older students is an open and policy-relevant question that our study directly addresses.

Barros and Ganimian (2023) isolate the impact of the individualized component of the Mindspark software. Although they find no average effects, students in the bottom quartile experienced meaningful improvements ($0.22\sigma$), emphasizing the importance of precise targeting. Our intervention applies this insight by focusing exclusively on the lowest-performing students, while extending coverage to both math and Portuguese and leveraging diagnostic tools sensitive to foundational skill development.

Finally, research in high-income settings affirms the potential of intensive tutoring, but often with high costs and limited scalability. The Chapter One program in the U.S., for instance, achieved substantial reading gains but required dedicated tutors and consistent scheduling (Cortes et al., 2024). Our framework seeks to retain the pedagogical benefits of individualized instruction while adapting it to the fiscal and administrative realities of large-scale public education systems in middle-income countries. In Latin America, the study by Cabezas et al. (2021) in Chile shows that even short-term tutoring can yield lasting benefits on educational engagement, especially for at-risk students—a broader impact we explore through attendance and engagement measures.

In sum, this study addresses a critical policy question: can individualized learning models be effectively implemented for middle school students who are years behind grade level? These students often fall through the cracks of traditional instruction, and it remains unclear whether scalable, low-cost, in-school interventions can meaningfully support their progress without relying on intensive tutoring or advanced technology. Our findings will offer important evidence on the feasibility and effectiveness of adapting individualized learning approaches to serve students with long-standing learning gaps within resource-constrained public education systems.

# 2 Intervention

In partnership with the São Paulo State Secretariat of Education (SEDUC-SP), this study evaluates a remedial individualized learning program ("Professor Tutor Anos Finais") aimed at addressing Portuguese and math learning gaps among middle school students enrolled in State public schools. The program provides targeted support to students who have not met expected competency levels in these subjects. Selected schools are assigned Portuguese and math tutors who conduct focused sessions with small groups of up to 15 students. The intervention will take place during the 2025 school year, with middle school students withdrawn from regular Portuguese and math classes 3 to 4 times a week for 45-minute tutoring sessions. Each school will have one tutor per subject, overseeing 5-6 groups of up to 15 students each.

Tutors use structured pedagogical materials amenable to differentiated instruction. Concretely, printed materials ("Material Coruja") are divided into five booklets (one for each grade, 1st to 5th), with four progress levels each (red, yellow, green and blue). A standardized diagnostic assessment conducted by schools will both identify students eligible for tutoring and define which booklet they should start with, based on their baseline proficiency level. All students start at the basic (red) level and progress at their own pace, guided by regular assessments. Supplementary digital exercises (on the "Tarefa SP" platform) provide weekly interactive assignments to help tutors and teachers monitor student-level progress. Although designed for younger students (grades 1–5), the material will be used for middle school students to address learning gaps.

The decision to conduct tutoring during regular classes, rather than after school, reflects evidence that in-class tutoring yields larger impacts. Moreover, these students often struggle to follow grade-level content in regular classes.

# 3 Experiment Design

The pilot program, launched in Q3/2024, targeted the 20 lowest-performing school districts ("Diretorias de Ensino" or DEs) in São Paulo State, based on standardized test scores. Eligibility was limited to state schools with more than 150 students. A total of 200 schools were randomly assigned to the treatment group and 100 to the control group.

In 2025, the program will expand to 31 additional DEs, using the same performance criteria. Schools from the pilot program will remain, joined by 300 new treatment schools and 299 control schools, resulting in 500 treated schools and 399 controls. Randomization was stratified by school district (DE). The imbalance reflects the inclusion of pilot schools, district-level caps on total schools, and SEDUC's requirement of 500 treated schools. Where feasible, newly selected DEs include 10 treated and 10 control schools. In treated schools, tutors are typically certified in pedagogy.

Baseline evaluations were conducted at the beginning of the school year; however, due to data access protocols, we do not currently have access to the results. SEDUC is expected to release the first round of data—including both baseline and follow-up assessments—after the end of the first semester, which concludes on June 30, 2025.

### 3.1 Student-level eligibility

In the pilot, student selection for the RCT was based on an average score from two standardized assessments: SARESP 2023 and Prova Paulista 2024 (using results from Q1 and Q2). SARESP is São Paulo's annual end-of-year assessment in Portuguese and math, while Prova Paulista evaluates Portuguese, math, sciences, and English for grades 5–9. The 150 lowest-scoring students from each school were selected, creating different cutoffs across schools and logistical challenges in allocating students to tutoring.

In 2025, SEDUC will introduce a clearer and standardized rule for student selection. First, all 6th-grade students and all 7th-grade students with math and Portuguese scores below 2.5 and 3, respectively, will take the baseline standardized assessment to determine eligility for individualized learning. If there are still slots available (i.e., if less than 90 eligible students have been identified in the baseline assessments), 8th and 9th-grade students with scores below 2 in math and 2.5 in Portuguese will also be screened. The cutoff criteria outlined above will be based solely on SARESP 2024 scores and applied consistently across all treated schools. We are able to track comparable students in control schools by following the same criteria.

Below, balance checks for the experiment (Table 1) demonstrate that treatment and control groups are well-balanced across key criteria. "Meta Ouro" refers to an school-specific goal set by SEDUC-SP with the goal of incentivizing improvements in test scores. The target is calculated based on the school's current IDEB (Basic Education Development Index) score from 2021, its structural conditions and the racial profile of its students.

# 4 Outcomes

We aim to evaluate the effects of the individualized learning intervention on student effort—measured by attendance in Math and Portuguese classes—on learning outcomes, including report card grades, standardized state assessments, third-party computer-based adaptive diagnostics, and foundational literacy and numeracy assessments. We also examine impacts on grade progression and dropout rates.

In addition to the standard assessments SARESP and Prova Paulista, we will evaluate improvements in foundational skills, such as basic reading and math skills. Considering

Table (1) Balance Check and F-Test for Overall Balance - Main Experiment

| Variable | Control | Treatment | P-Value |
|---|---|---|---|
| Socioeconomic Level (INSE) | 5.212 | 5.213 | 0.309 |
| Number of Students | 656,183 | 713,290 | 0.996 |
| SARESP Average (2023) | 4.072 | 4.026 | 0.923 |
| SARESP Average (2024) | 4.238 | 4.178 | 0.735 |
| Prop. Meta Ouro Achieved | 0.931 | 0.929 | 0.702 |
| Number of Schools | 399 | 500 | |
| F-Test - F-stat = 0.401 (P-value = 0.848) | | | |

P-values are obtained from regressions of each variable on the treatment indicator, controlling for DE fixed effects and clustering standard errors at the DE level. The F-Test checks the joint significance of all covariates in explaining the treatment assignment.

that some students may be significantly behind grade level, traditional assessments might fail to capture progress, making targeted evaluations necessary. SEDUC is developing its evaluation framework, utilizing the Rapid Online Assessment of Reading (ROAR)[1] and the Stanford Mental Arithmetic Response Time Evaluation (SMARTE)[2], both adapted for Portuguese. ROAR is an automated online tool designed to assess foundational reading skills efficiently and accurately, providing immediate, instructionally relevant feedback to educators. SMARTE is a digital assessment that measures math fluency and arithmetic proficiency, focusing on speed and accuracy in basic operations. Both tools offer precise and scalable solutions to track student progress in essential skills.

In partnership with Parceiros da Educação and Catvante, a computer adaptive testing (CAT) was administered to all students in grades 6 through 9 across the state at the beginning of the school year, with additional rounds planned throughout the year. The assessment follows a multi-stage adaptive design, adjusting the difficulty of questions based on each student's performance. It begins at one grade level below the student's current enrollment (the "reference grade") and moves up or down depending on responses. Students are ultimately placed at the highest grade level for which they consistently demonstrate mastery. Performance is classified into four interpretative bands: Adequate, Intermediate, Critical, and Very Critical. This design allows for a more precise assessment of each student's learning level.

[1] https://roar.stanford.edu

[2] https://edneuroinitiative.stanford.edu/smarte

# 5 Hypotheses

We aim to test the following hypotheses:

1. **Does individualized learning provided during school hours and outside regular classes increase student effort?**

   **Hypothesis**: The intervention increases math and Portuguese attendance of students eligible to individualized learning in treated schools relative to those in control schools.

2. **Does individualized learning provided during school hours and outside regular classes mitigate learning gaps?**

   **Hypothesis**: The intervention improves average standardized test scores of students eligible to individualized learning in treated schools relative to those in control schools.

3. **Does individualized learning during school hours, outside regular classes, improve foundational skills?**

   **Hypothesis**: The intervention improves ROAR and SMARTE scores of students eligible to individualized learning in treated schools relative to those in control schools.

# 6 Analyses

## 6.1 Estimation

The cutoff rule designates all students below the threshold as eligible for individualized learning. However, student allocation to individualized learning sessions is managed at the school level, and capacity constraints or logistical challenges may prevent all eligible students from participating. Each school can have a maximum of six groups with 15 students each, limiting the number of treated students to 90 per school. Consequently, if a school has more than 90 students below the cutoff, not all eligible students will receive individualized learning.

Given these compliance limitations, we will conduct an intention-to-treat (ITT) analysis, complemented with an LATE estimation that uses random assignment as an instrumental variable (IV).

1. **Intention-to-Treat (ITT) Analysis**: We will begin by estimating the ITT effects of the randomized intervention. This analysis compares the outcomes between treatment and control groups based solely on the randomized assignment, regardless of whether students actually received individualized learning. The ITT analysis provides an unbiased estimate of the policy's overall effect but may underestimate the true treatment effect due to noncompliance.

2. **Instrumental Variables (IV) Analysis**: To address compliance issues and estimate the Local Average Treatment Effect (LATE) for students who comply with treatment assignment, we will then use an IV approach. Here, treatment assignment at the school level will serve as the instrument for actual treatment receipt.

To prevent inflated test sizes from multiple hypotheses testing due to family-wise error rates across Mathematics and Portuguese treatment effects, we will estimate a summary measure as our main outcome of analysis. If there are many missing values (since standardized tests are not mandatory and only medium stakes for high-school students), we might estimate seemingly unrelated regressions (SUR) instead to account for the correlated error structure across subjects and improves estimation efficiency.

While we focus on aggregate treatment effects, we are also specifically interested in estimating effects by grade and by subject.

## 6.2 Power calculations

While the list of eligible students is not available yet, SEDUC-SP estimates that, on average, 30% of middle school students will qualify for the baseline assessments. Using school data on the number of middle school students, we can estimate the average compliance among schools. However, due to the lack of student data, we are unable to estimate an intra-cluster correlation coefficient (ICC) for both subjects. However, using data on the pilot program, we will present MDE estimations using estimated ICC from the pilot program and from standard values, such as 0.10.

We perform power calculations to estimate the minimum detectable effect (MDE) based on the experiment design. The study includes 500 treatment schools and 399 control schools, resulting in a total of 899 clusters. Standard errors will be clustered at the school level to account for this design.

For all MDE calculations, we will use a 5% significance level ($\alpha$) and a power level of 80% ($1-\kappa$). Additionally, as noted earlier, we account for partial compliance in our calculations. The formula for the Minimum Detectable Effect, adjusted for partial compliance, is given by:

$$\text{MDE} = \left(t_{(1-\kappa)} + t_{\frac{\alpha}{2}}\right)\sqrt{\frac{1}{P(1-P)} \cdot \frac{1}{N} \cdot (1+(m-1)\cdot \text{ICC})}\frac{1}{c-s} \cdot \sigma \tag{1}$$

where $\kappa$ is the number of clusters, $m$ is the average number of units per cluster, $t_{(1-\kappa)}$ and $t_{\frac{\alpha}{2}}$ are critical values for the power and significance levels, $P$ is the proportion of the sample assigned to the treatment group, $\sigma$ is the outcome's standard deviation, $N$ is the sample size, and $c-s$ adjusts for the degree of compliance, and the $ICC$ is given by:

$$\text{ICC} = \frac{\sigma^2_{\text{between-cluster}}}{\sigma^2_{\text{between-cluster}} + \sigma^2_{\text{within-cluster}}} \tag{2}$$

Table 2 compiles the MDEs for different ICC values and compliance scenarios. Partial compliance, estimated at 55%, is based on the assumption that 30% of students per school are eligible and adjusted by each school's middle school enrollment. MDEs are given in standard deviations.

The ICCs include a benchmark value from the literature (0.10) and estimates from a 2024 pilot study using SARESP grades in math (0.18) and Portuguese (0.27). Higher ICCs reduce statistical power by increasing the MDE. Under full compliance, the MDE rises from around 0.06 for the literature ICC to 0.09 for the Portuguese pilot ICC. The effect is more pronounced with partial compliance, where the MDE increases to around 0.17 under the highest ICC.

## 6.3 Sample restrictions

To maintain comparability between treatment and control groups, we will exclude the following students from the analysis:

Table (2) Minimum Detectable Effect (MDE) by Compliance Scenario and ICC

| | Full Compliance | | Partial Compliance (55%) | |
|---|---|---|---|---|
| | MDE | ICC | MDE | ICC |
| Literature ICC | 0.0624 | 0.1000 | 0.1109 | 0.1000 |
| Pilot ICC (Math) | 0.0809 | 0.1756 | 0.1437 | 0.1756 |
| Pilot ICC (Portuguese) | 0.0984 | 0.2655 | 0.1749 | 0.2655 |

- Students who changed schools during the study period.
- Students in non-sampled schools.
- Students in treated schools who were not assigned treatment but still received it.

These exclusions are necessary because treatment assignment was based on the performance cutoff. Including students outside these parameters would undermine the comparability across groups and potentially introduce bias.

Additionally, we might run additional analyses dropping schools that violate treatment assignment during the school year (i.e., initially assigned to the control group but which ultimately end up offering individualized learning, or the other way around) in case doing so does not violate balance or selective attrition tests.

## 6.4 Intention-to-Treat (ITT) specification

The ITT model is specified as:

$$Y_{is}^{port} = \alpha_0^{port} + \alpha_1^{port} Z_s + \epsilon_{is}^{port}$$

$$Y_{is}^{math} = \alpha_0^{math} + \alpha_1^{math} Z_s + \epsilon_{is}^{math}$$

Where:

- $Y_{is}^{port}$ and $Y_{is}^{math}$: SARESP/Prova Paulista/CAT/Foundational Skills Assessment grades in Portuguese and Mathematics for student $i$ in school $s$.
- $Z_s$: Randomized treatment assignment at the school level.

- $\epsilon_{is}^{port}$ and $\epsilon_{is}^{math}$: Error terms clustered at the school level.

The coefficients $\alpha_1^{port}$ and $\alpha_1^{math}$ represent the ITT effects of the intervention on Portuguese and Mathematics grades, respectively.

The decision to use Seemingly Unrelated Regressions (SUR) is motivated by the interconnected nature of the outcome variables in this study. Estimating outcomes separately when they are driven by shared factors may lead to inefficient estimates. In our case, Portuguese and Mathematics grades are likely influenced by overlapping factors, resulting in correlated error terms. SUR explicitly models this correlation, improving efficiency. Additionally, it enables us to perform cross-equation hypothesis testing.

## 6.5 Dependent variables and instruments

The SUR framework uses two dependent variables:

1. SARESP/Prova Paulista/CAT/Foundational Skills assessment grades in Portuguese.
2. SARESP/Prova Paulista/CAT/Foundational Skills assessment grades in Mathematics.

The instrumental variable (IV) is treatment assignment at the school level, while the key covariate of interest is whether the student received individualized learning.

## 6.6 Model Specification

The SUR model with IV is specified as follows:

$$Y_{is}^{port} = \beta_0^{port} + \beta_1^{port} T_{is} + \epsilon_{is}^{port}$$

$$Y_{is}^{math} = \beta_0^{math} + \beta_1^{math} T_{is} + \epsilon_{is}^{math}$$

Where:

- $Y_{is}^{port}$ and $Y_{is}^{math}$: SARESP/Prova Paulista/CAT/Foundational Skills Assessment grades in Portuguese and Mathematics for student $i$ in school $s$.

- $T_{is}$: Indicator of whether the student received individualized learning.
- $\epsilon_{is}^{port}$ and $\epsilon_{is}^{math}$: Error terms clustered at the school level.

The treatment variable $T_{is}$ is instrumented using treatment assignment at the school level ($Z_s$):

$$T_{is} = \gamma_0 + \gamma_1 Z_s + \eta_{is}$$

## 6.7 Assumptions for IV validity

For the IV analysis to yield valid estimates, the following conditions must hold:

1. **Relevance**: Treatment assignment ($Z_s$) must be strongly correlated with actual treatment receipt ($T_{is}$). This condition is satisfied because the lowest-performing students in schools assigned to treatment are significantly more likely to receive individualized learning.
2. **Exogeneity**: Treatment assignment ($Z_s$) must be independent of unobserved factors affecting outcomes ($Y_{is}^{port}$ and $Y_{is}^{math}$). Randomization ensures this condition is met.
3. **Exclusion Restriction**: Treatment assignment ($Z_s$) can only affect outcomes ($Y_{is}^{port}$ and $Y_{is}^{math}$) through its impact on treatment receipt ($T_{is}$). This assumption holds as long as there are no unmeasured pathways linking assignment to outcomes.

## 6.8 Interpretation of results

The coefficients $\beta_1^{port}$ and $\beta_1^{math}$ provide estimates of the causal impact of individualized learning on Portuguese and Mathematics grades, respectively. These estimates represent the Local Average Treatment Effect (LATE) for students who comply with the treatment assignment. By leveraging the randomized treatment assignment and addressing compliance issues through the IV framework, this approach ensures robust estimates of the individualized learning program's impact.

## 6.9 Incorporating dose-response effects

If not all schools begin individualized learning in Q1, we might further examine whether the effect of individualized learning depends on the length of exposure to treatment. Treatment dose is computed as:

$$\text{Exposure}_i = \max\{0, \text{Date of Assessment} - \text{Date Treatment Started}_i\}$$

To evaluate the role of exposure length, we include a variable measuring the time each eligible student could have been exposed to individualized learning. This allows us to test whether the treatment effect increases with extended exposure. The updated model is specified as:

$$Y_{is}^{port} = \beta_0^{port} + \beta_1^{port} T_{is} + \beta_2^{port} \text{Exposure}_i + \epsilon_{is}^{port}$$

$$Y_{is}^{math} = \beta_0^{math} + \beta_1^{math} T_{is} + \beta_2^{math} \text{Exposure}_i + \epsilon_{is}^{math}.$$

Here:

- $T_{is}$: Indicator of whether the student received individualized learning.
- $\epsilon_{is}^{port}$ and $\epsilon_{is}^{math}$: Error terms clustered at the school level.

To address potential endogeneity of length of exposure, we instrument the treatment variables as follows:

1. $T_{is}$ (treated status) is instrumented using randomized treatment assignment ($Z_s$).
2. $\text{Exposure}_i$ (exposure duration) is instrumented using the interaction between randomized treatment assignment and the exposure measure ($Z_s \times \text{Exposure}_i$).

The first-stage equations are specified as:

$$T_{is} = \gamma_0 + \gamma_1 Z_s + \eta_{is},$$

$$\text{Exposure}_i = \delta_0 + \delta_1 (Z_s \times \text{Exposure}_i) + \nu_{is}.$$

Where:

- $Z_s$: Randomized treatment assignment at the school level.
- $Z_s \times \text{Exposure}_i$: Interaction between treatment assignment and exposure measure, providing exogenous variation for exposure duration.
- $\eta_{is}$ and $\nu_{is}$: Error terms.

By separately instrumenting treated status and exposure duration, we ensure that the causal effects of both variables are estimated robustly.

The coefficients $\beta_1^{port}$ and $\beta_1^{math}$ provide estimates of the causal impact of being treated (individualized learning) on Portuguese and Mathematics grades, respectively. The coefficients $\beta_2^{port}$ and $\beta_2^{math}$ capture the dose-response effect, representing the additional improvement in grades for each additional week of exposure to individualized learning.

By leveraging separate instruments for treated status and exposure duration, this approach ensures robust and credible estimates of both the binary treatment effect and the dose-response relationship.

# References


Banerjee, A., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Barros, A. and Ganimian, A. (2023). Which students benefit from computer-based individualized instruction? experimental evidence from public schools in india. *Journal of Research on Educational Effectiveness*, 17:1–26.

Cabezas, V., Cuesta, J. I., and Gallego, F. (2021). Does Short-Term School Tutoring have Medium-Term Effects? Experimental Evidence from Chile. Technical report.

Cortes, K., Kortecamp, K., Loeb, S., and Robinson, C. (2023). A scalable approach to high-impact tutoring for young readers: Results of a randomized controlled trial. *sasa*.

Cortes, K., Kortecamp, K., Loeb, S., and Robinson, C. (2024). A scalable approach to high-impact tutoring for young readers: Results of a randomized controlled trial. Working Paper 32039, National Bureau of Economic Research.

Dietrichson, J., Bøg, M., Filges, T., and Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, 87:243–282.

Lichand, G. and Doria, C. (2024). The lasting impacts of remote learning in the absence of remedial policies: Evidence from brazil. *PNAS*, 121(22).

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4):1426–60.

Nickow, A., Oreopoulos, P., and Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. Working Paper 27476, National Bureau of Economic Research.