Part II: Online tutoring - Impact evaluation of a program for 5th grade children in Colombia

1 Introduction

This study investigates the impact of the TuTutor program, an online tutoring initiative offering either Math or English one-to-one tutoring to young students from low socioeconomic backgrounds. Specifically, the aim is to estimate the effect of online tutoring in Math and English on score improvements for these students.

The TuTutor program offers one-to-one online tutoring in Math or English (as a foreign language) to 5th-grade students in public schools. The tutoring sessions last approximately 10 weeks, with each session ranging from 1.5 to 3 hours per week. This initiative uniquely mixes socio-economic strata by pairing tutors, often from high Socio-Economic Status (SES) backgrounds, with tutees from low SES backgrounds. The tutors are high school or university students from top private educational institutions in Bogotá, Cali, Barranquilla, Medellín, and Ibagué. The tutees are fifth graders from underprivileged public schools across 16 regions of Colombia. The online format was first introduced to mitigate the adverse impacts of COVID-19 on children's learning and to foster interactions among socio-economic groups that typically would not interact. The program include a workshop for tutors and provided continuous support throughout their tutoring experience. Each tutoring pair's progress was closely monitored.

The program benefits from Law 115 of 1994 and Decree 1860 of 1994, which require high school students to complete 80 hours of student social service during their final years of high-school (grades 10 or 11) as a condition to university admission. This requirement made it easier to find volunteers for the program. The program therefore offers three options: one of 40 hours, one of 80 hours and one of 100 hours for the most motivated participants. These hours include not only tutoring sessions but also training and workshops designed to enhance their effectiveness

Due to oversubscription, tutees were randomly assigned to either a treatment group or a control group. The subject of tutoring (Math or English) was based on the preference expressed by each tutee. Data for

this study were collected over six distinct cohorts, from the second semester of 2021 (Tututor 3, TT3) to the first semester of 2024 (Tututor 8, TT8).[1]

To evaluate the program's impact, surveys were administered to both tutors and tutees before and after the program. These surveys gathered data on their personal attributes, views on poverty, self-perception, attitudes towards income redistribution, and feedback on the program. Parents were also surveyed to identify preferences for Math or English tutoring and provide insights into household dynamics. In addition, each tutee underwent two performance assessments in Mathematics and English of same difficulty to measure the effectiveness of the program in improving their skills.

This project contributes to the literature on tutoring, a widely used strategy to address learning losses incurred during the pandemic. Research has shown that strong tutor/tutee relationships are among the most effective tools for enhancing both academic and non-academic educational outcomes, particularly for students from lower socioeconomic backgrounds, as demonstrated by DuBois et al., 2011, Bowman-Perrott et al., 2014, Dietrichson et al., 2017, Balfanz and Byrnes, 2018, Christensen et al., 2021, and Nickow et al., 2024. These improvements include greater academic success, school attainment, confidence, pro-social behavior, and career aspirations, as noted in Grossman and Tierney, 1998, Sánchez et al., 2008, Herrera et al., 2011, and M. Kraft et al., 2021.

Our study specifically contributes to the emerging literature on online tutoring. The most closely related studies include M. A. Kraft et al., 2022, which evaluates the impact of one-to-one online tutoring provided by Ivy League volunteers in reading and math to low-income students of color, showing positive results in both subjects. Additionally, Carlana and La Ferrara, 2024 examines a virtual one-on-one tutoring experiment conducted in Italy during the pandemic, finding significant improvements in math and socio-emotional outcomes. Other evaluations of remote interventions implemented during COVID-19 include Gortazar et al., 2024, which assesses a grouped online tutoring program for children aged 12 to 15 in Spain; Hardt et al., 2022, which explores a remote peer mentoring program among university students in Germany; and Hassan et al., 2023, which evaluates a telementoring intervention for primary school-aged children in Bangladesh.

2 Intention to Treat

In this initial phase, we assess the impact of being in the treated group on English or Math scores. To ensure comparability, we analyze the effect on the percentage of correct responses, accounting for the different number of questions in each subject. This effect, referred to as Intention to Treat (ITT), evaluates the impact of being assigned to the treatment group rather than the specific effect of receiving tutoring in Math or English. As this approach does not differentiate between Math and English tutoring, it likely underestimates the improvements due to the specific effects of each type of tutoring.

Since tutee assignment was randomized, this experiment qualifies as a Randomized Controlled Trial (RCT).

[1] Cohorts TT1 and TT2 were used to calibrate the program and the evaluation process.

Tutees who passed the pre-exam were randomly assigned to either the control or treatment groups. However, to study the change in scores between the pre- and post-tests, we can only include tutees with complete pre- and post-tutoring scores, which introduces a potential risk of non-comparability if the two groups, initially comparable, differ at the final stage. This concern can be dismissed for several reasons. First, Table 1 shows that the control and treatment groups with complete pre- and post-tutoring scores have similar baseline characteristics. Notably, the only significant difference at the 10% level is the percentage of women, which likely reflects the sample size rather than a substantial issue. Second, our attrition analysis, presented in Table 2, indicates no significant correlation between group assignment (treatment vs. control) and attrition rates. Thus, the control and treatment groups of tutees with complete pre- and post-tutoring scores appear comparable, permitting the use of standard RCT estimators like mean differences. Nevertheless, our study benefits from having two measures of tutee performance before and after the tutoring program. In this context, a Difference-in-Differences (DiD) estimator can be implemented to estimate the effect of tutoring. This method is more robust than using simple mean differences. Unlike mean differences, DiD analysis does not rely on the assumption of perfect randomness but rather on the less stringent common trend assumption, which posits that the evolution of score differences, rather than their absolute values, would have been the same in the absence of tutoring. Furthermore, if randomness is perfect, DiD will estimate the same parameter as mean differences: the Average Treatment Effect (ATE).

Using a Difference-in-Differences approach to compare treated and untreated groups before and after tutoring, Table 3 shows a statistically significant improvement in scores for those in the treated group (ATT) at the 10% significance level in both subjects. This increase is more pronounced in English than in Math, with a higher estimated coefficient.

In this context, the DiD approach relies on the common trend assumption, which posits that both the treated and control groups would have evolved similarly between the pre- and post-tutoring periods, on average. Unfortunately, we cannot conduct a plausibility test for this assumption because we lack prior tests before tutoring to check pre-trends, or other time-varying variables unaffected by tutoring to perform covariate tests. However, this hypothesis, which is less stringent than full randomness, should be considered valid since we have demonstrated that the two groups are comparable and that the randomization was properly executed.

3 Treatment Effect

In this section, we aim to evaluate the impact of participating in either English or Math tutoring on student scores. A challenge arises because the assignment of tutoring subjects was based on students' preferences expressed in a pre-survey, rather than being randomly allocated. For instance, in our randomized controlled trial (RCT), students who preferred Math were randomly assigned either to receive Math tutoring or not. Given that we have preference information for both the control and treated groups, we can still assess the treatment effect within each subgroup. If the randomization was conducted correctly, these subgroups should have similar characteristics on average, allowing us to evaluate the treatment's impact effectively.

In this context, our estimation depends on the common trend assumption, which posits that the evolution of scores without treatment in Math/English would have been similar for individuals who preferred Math/English tutoring and were assigned to treatment, compared to those who were not assigned to treatment. As discussed in Section 2, this hypothesis should be valid since we have demonstrated that the two groups are comparable and that the randomization was properly executed.

In Table 4 and Table 5, we estimate the treatment effects for each subpopulation:

- First, we observe that these treatment effects are larger than the Intent-to-Treat effects for each subject, as expected. For instance, Math tutoring increases the percentage of correct responses in the post-test by 4.4 percentage points among those who preferred Math, compared to a 3.1 percentage point increase observed in the ITT analysis. Similarly, English tutoring results in a 12.2 percentage point increase in the percentage of correct responses in English post-tests, compared to a 8.3 percentage point increase in the ITT results.
- Second, once again, there is a notable difference in effect sizes between English and Math treatments. Students who preferred English and received English tutoring experienced a significantly greater improvement in their English scores compared to the improvements in Math scores for students who preferred Math and received Math tutoring.
- Third, another noteworthy finding is that students who received Math tutoring also showed an increase in English scores of a similar magnitude, though this increase was not statistically significant at the 10% level.[2] This suggests that Math tutoring may have broader benefits, potentially enhancing English performance as well.

The magnitude of the results in Math is consistent with findings from Carlana and La Ferrara, 2024, which reported that three hours of individual tutoring per week improved Math performance by 0.23 standard deviations (SD) in 2020 and 0.20 SD in 2022 for underprivileged middle school students. The results for English are particularly notable compared to existing literature: meta-analyses of tutoring programs (Ritter et al., 2009, Dietrichson et al., 2017, Robinson et al., 2021, Nickow et al., 2024) generally find tutoring effects averaging about one-third of a SD. [3] To my knowledge, these are the first results on the effect of online tutoring in foreign languages. The substantial findings can be attributed to the relatively low English proficiency of many teachers, as reflected in the lower baseline percentage of correct responses in English compared to Math among tutees, as shown in Table 1. With the initial proficiency level being relatively low, tutors — who typically have a high level of English proficiency — are able to add more value in English tutoring compared to Math.

Math tutoring might offer broader benefits beyond its primary focus, potentially improving performance in English as well. This observation is consistent with existing literature on spillover effects from educational interventions. For example, Gilraine and Penney, 2023 found that intensive test preparation in either mathematics or English led to a significant increase of 0.013 SD in test scores in the other subject a year

[2]Despite the lack of statistical significance, the observed 4.0 percentage point increase is notable and may be influenced by the relatively small sample size.

[3]These are substantial effects, as the average test score difference between students in the top and bottom 15% of the PISA index of economic, social, and cultural status is approximately 0.7 to 0.8 SD in OECD member countries.

later. Possible explanations for these spillover effects include:

1. Increased self-confidence, motivation, concentration, and determination resulting from tutoring may contribute to improved performance in additional subjects. Our analysis reveals that tutees in the treated group experience a significant increase in happiness and a greater desire to study following tutoring. This finding aligns with Carlana and La Ferrara, 2024, which illustrates that the tutor/tutee relationship often extends beyond academic content, providing mentorship that significantly impacts students' educational aspirations (+0.19 SD), socio-emotional skills such as perseverance, grit, and locus of control (+0.16 SD), and psychological well-being, including happiness and depression (+0.16 SD). Similarly, Gortazar et al., 2024 found that the intervention not only improved test scores but also raised tutees' aspirations.
2. Tutoring may not only impart subject-specific knowledge but also teach effective study methods that enhance performance in other subjects.
3. It is also possible that some tutors, possibly due to a misunderstanding of the program's focus, provided support in both Math and English subjects.

4 Heterogeneity of the Intention-To-Treat

In this section, we try to identify the characteristics of the tutoring that determine its effectiveness. All the results are presented in Table 6

A first hypothesis is that having tutors of the same sex as the tutee could enhance dynamics, improve communication, and ultimately lead to more effective tutoring. Our empirical results do not support this hypothesis. An analysis of the Intent-to-Treat effects in Math and English for same-sex versus different-sex tutor-tutee pairs reveals no significant differences. Specifically, while score improvements in Math are higher for same-sex pairs, they are lower for English compared to different-sex pairs.

A second hypothesis is that the tutoring effect may vary depending on whether the tutor is a university student or a high school student. This hypothesis appears to be supported by the data: Table 6 reveals a stronger tutoring effect from university student tutors. This effect can be attributed to several factors:

1. University students often exhibit higher motivation for tutoring compared to high school students. Unlike high school students who participate in tutoring as part of a mandatory 80-hour social service program, university students engage voluntarily. Table 7 illustrates that while 37% of high school students are involved in the 80-hour program (a number that likely underestimates the true extent due to 35% of high school students having missing information), only 13% of university students participate in this program.
2. University students are generally better equipped to provide effective tutoring due to their maturity, greater experience, and higher academic proficiency. Supporting this, 65% of university students

had previous tutoring experience before joining the Tututor program, compared to only 32% of high school students.

3. University students often share closer social ties with tutees, particularly as many tutors are beneficiaries of SPP/GEE programs with similar backgrounds. Table 8 highlights that 37% of high school students belong to the highest socioeconomic strata (5/6), whereas this percentage is 13% among university students.

It's worth noting that this result could theoretically be explained by a potential dosage effect, where university students might provide more tutoring hours on average compared to high school students. However, our findings show the opposite trend: while most high school students fulfill their social service requirements with the 80-hour formula, university students typically opt for the 40-hour formula (Table 7).

A third hypothesis is that students who received more hours of tutoring should perform better (dosage effect). Empirical results, however, reveal an initial lack of dosage effect, with no significant improvement in Maths between the 40-hour and 80-hour tutoring programs. This unexpected result is due to the 80-hour social service requirement, which leads tutors in the 80-hour program to be predominantly high school students, who are on average less committed and less experienced. Further analysis of the dosage effect among high school students - the largest group of tutors in our study - reveals a significant dosage effect, consistent with the existing literature (Carlana and La Ferrara, 2024, M. A. Kraft et al., 2022).[4]

Finally, another hypothesis is that university students participating in tutoring to fulfill mandatory social service requirements for their scholarships may differ in motivation compared to other students, potentially leading to varying tutoring effects. Our observations support this hypothesis, showing significant tutoring effects among university students, but with greater effects among those not fulfilling mandatory obligations. This finding aligns with previous results indicating that voluntary participants generally exhibit higher motivation, leading to more substantial tutoring effects.

5 Conclusion

To summarize, this study finds that online tutoring in Math and English for 5th-grade students from low socioeconomic backgrounds consistently improves student outcomes. The magnitude of these improvements varies by subject, with English showing particularly notable gains, likely due to initially lower proficiency among English teachers and a larger potential for improvement. The study also identifies broader benefits beyond subject-specific gains, with Math tutoring showing positive results on English performance. Importantly, the effectiveness of tutoring varies a lot based on tutor characteristics. Specifically, the experience and motivation of tutors are crucial, as university tutors facilitate greater improvements in both subjects

[4]To assess the dosage effect, we use the duration preference announced by tutors before the program, as actual hours completed are not available for all. For cohorts TT7 and TT8, where this information is available, it aligns with the announced duration in over three-quarters of cases. When discrepancies occur, it is more common for completed hours to exceed those announced, suggesting that our estimation of the dosage effect may be an underestimation of the true effect.

for their tutees. These results align with existing research on the importance of teacher experience, described in Kini and Podolsky, 2016. This underscores the critical role of tutor quality in the success of the tutoring program, which appears to be even more important than the duration of the program.

These findings have important policy implications. First, they suggest that the tutoring program could be scaled up, given its effectiveness and relatively low cost. Second, while the program generally produces positive outcomes, adopting more tailored approaches could enhance its impact. The evidence underscores the crucial role of selecting qualified tutors, with more experienced and motivated individuals yielding significantly better tutees performance. To optimize the program's effectiveness, policymakers should prioritize strategies that emphasize the recruitment, training, and ongoing support of high-quality tutors. Additionally, refining the matching process between tutors and tutees could further ensure that the program's benefits are maximized and more equitably distributed among students.

Variable	All	Control	Treatment	P-value	Std diff.
Percentage women	52.6%	55.6%	50.1%	0.08	-0.11
Have a Sisbén score	75.3%	73.9%	76.4%	0.38	0.06
Household size	4.37	4.30	4.43	0.12	0.10
Percentage repeated	12.4%	11.7%	13.0%	0.53	0.04
Percentage educated father	22.1%	22.0%	22.3%	0.94	0.00
Percentage educated mother	37.8%	38.8%	37.0%	0.28	0.07
Baseline Percentage of good responses in Math	54.3%	54.2%	54.3%	0.94	0.00
Baseline Percentage of good responses in English	44.5%	44.1%	44.9%	0.55	0.04

Observations	All	Control	Treatment
Number of tutees	1007	464	543
Preference in Maths (Treated in Maths)	410	189	221
Preference in English (Treated in English)	527	236	291
Unknown preferences (Unknown treatment)	70	39	31

Table 1: Students' Characteristics at baseline (full sample)

Notes: This table presents the characteristics of all students (column 1), control students (column 2), and treated students (column 3) for all pooled cohorts. Column 4 shows p-values for differences in means, while Column 5 reports the standardized differences between group averages. All variables are measured at baseline. The Sisbén score, a socioeconomic index used by the government to target subsidies and social programs, is not automatically provided and must be requested.

	Estimate	**Std. Error**	**Pr(>|t|)**
Intercept	0.33	0.02	0.00
Treatment	**0.01**	**0.03**	**0.59**
Observations	1334		

Table 2: Attrition difference between treated and control group

Note: This table shows the average correlation between being assigned to the treatment group and the attrition rate for TT3-TT8. Standard errors and p-values presented are robust to heteroscedasticity. Treatment status was assigned to tutees who completed the baseline test. Attrition is defined as the inability to collect results from the post-test.

	Basic (1)		With controls (2)	
Outcome	Estimate	P-Value	Estimate	P-Value
Math	0.017 (0.072)	0.286	0.017 (0.072)	0.293
English	0.082 (0.362)	0.000	0.082 (0.362)	0.000
Observations	1007		982	

Table 3: Impact of Treatment Assignment on English and Math Scores

Note: This table presents the estimates and p-values for the Difference-in-Differences estimation of the effect of belonging to the treated group on the percentage of good responses, with increasing levels of controls. Model (1) presents the basic Difference-in-Differences (DiD) estimation, while Model (2) incorporates demographic controls, cohort fixed effects, and test duration. The effect on the z-score, representing the change in standard deviations, is also shown in parentheses. Note that the p-values remain unchanged in this case as the z-scores are a monotonic transformation of the estimates. Standard errors are clustered at the tutee level. The demographic controls include the following dummies: whether the father or mother has obtained a higher education diploma (university or technical), whether the family has a Sisbén number (the socioeconomic index used by the government to target subsidies and social programs), whether the tutee is female, and whether the tutee has repeated a class.

	Basic (1)		With controls (2)	
Outcome	Estimate	P-Value	Estimate	P-Value
Math	0.034 (0.140)	0.167	0.032 (0.135)	0.191
English	0.033 (0.146)	0.141	0.031 (0.138)	0.166
Observations	410		403	

Table 4: Effect of Math Tutoring on Tutees with a Preference for Math

Note: This table presents the estimates and p-values for the Difference-in-Differences estimation of the effect of Math tutoring on the percentage of good responses among those who preferred Math tutoring, with increasing levels of controls. Model (1) presents the basic Difference-in-Differences (DiD) estimation, while Model (2) incorporates demographic controls, cohort fixed effects, and test duration. The effect on the z-score, representing the change in standard deviations, is also shown in parentheses. Note that the p-values remain unchanged as the z-scores are a monotonic transformation of the estimates. Standard errors are clustered at the tutee level. The demographic controls include the following dummies: whether the father or mother has obtained a higher education diploma (university or technical), whether the family has a Sisbén number (the socioeconomic index used by the government to target subsidies and social programs), whether the tutee is female, and whether the tutee has repeated a class.

	Basic (1)		With controls (2)	
Outcome	Estimate	P-Value	Estimate	P-Value
Math	0.015 (0.064)	0.494	0.016 (0.066)	0.492
English	0.127 (0.564)	0.000	0.130 (0.576)	0.000
Observations	527		511	

Table 5: Effect of English Tutoring on Tutees with a Preference for English

Note: This table presents the estimates and p-values for the Difference-in-Differences estimation of the effect of English tutoring on the percentage of good responses among those who preferred English tutoring, with increasing levels of controls. Model (1) presents the basic Difference-in-Differences (DiD) estimation, while Model (2) incorporates demographic controls, cohort fixed effects, and test duration. The effect on the z-score, representing the change in standard deviations, is also shown in parentheses. Note that the p-values remain unchanged as the z-scores are a monotonic transformation of the estimates. Standard errors are clustered at the tutee level. The demographic controls include the following dummies: whether the father or mother has obtained a higher education diploma (university or technical), whether the family has a Sisbén number (the socioeconomic index used by the government to target subsidies and social programs), whether the tutee is female, and whether the tutee has repeated a class.

Heterogeneity modality	Math		English		
	Estimate	P-Value	Estimate	P-Value	Observations
Same sex pair	0.02 (0.10)	0.23	0.08 (0.36)	0.00	690
Different sex pair	0.02 (0.09)	0.28	0.09 (0.39)	0.00	694
University Tutor	0.09 (0.38)	0.00	0.10 (0.44)	0.00	544
High School Tutor	0.00 (0.02)	0.79	0.08 (0.35)	0.00	927
40h program	0.03 (0.12)	0.19	0.07 (0.33)	0.00	643
80h program	0.02 (0.08)	0.38	0.11 (0.47)	0.00	641
40h program (Only High-School)	0.01 (0.03)	0.74	0.05 (0.23)	0.03	589
80h program (Only High-School)	0.01 (0.02)	0.81	0.10 (0.46)	0.00	629
Uni tutors with scholarship	0.06 (0.23)	0.22	0.06 (0.28)	0.24	495
Uni tutors without scholarship	0.11 (0.46)	0.01	0.13 (0.56)	0.00	512
Generous tutor	0.03 (0.14)	0.07	0.09 (0.38)	0.00	786
Non-generous tutor	-0.00 (-0.01)	0.94	0.08 (0.36)	0.00	598

Table 6: Heterogeneity in ITT effects

Note: This table presents the estimates and p-values for the Difference-in-Differences estimation of the effect of belonging to the treated group on the percentage of good responses. Each column shows the effect of being in a different treated group compared to the control group. Since tutors were randomly assigned to tutees, each sub-treated group is comparable to the control group. The effect on the z-score, representing the change in standard deviations, is also shown in parentheses. Note that the p-values remain unchanged as the z-scores are a monotonic transformation of the estimates. The generosity of the tutor is evaluated based on a question in the pre-survey asking whether he would fairly distribute a cake even if others didn't contribute.

Tutoring chosen	High School	University
40h program	125 (27.0%)	54 (67.5%)
80h program	165 (35.6%)	12 (15.0%)
100h program	9 (1.94%)	2 (2.5%)
Other	5 (1.08%)	11 (13.8%)
Missing information	159 (34.3%)	1 (1.2%)

Table 7: Comparison of Tutoring Modalities Selected by High School and University Tutors

Social strata	High School	University
1 or 2	13 (2.4%)	7 (7.5%)
5 or 6	179 (38.7%)	11 (13.8%)
Other	180 (38.7%)	62 (77.5%)
Missing info	93 (20.1%)	1 (1.3%)

Table 8: Comparison of Social Strata Distribution Between High School and University Tutors

References

Balfanz, R., & Byrnes, V. (2018). Using data and the human touch: Evaluating the nyc inter-agency campaign to reduce chronic absenteeism. *Journal of Education for Students Placed at Risk (JESPAR)*, *23*(1-2), 107–121.

Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review*, *43*(3), 260–285.

Carlana, M., & La Ferrara, E. (2024). *Apart but connected: Online tutoring, cognitive outcomes, and soft skills* (tech. rep.). National Bureau of Economic Research.

Christensen, S., Grønbek, T., & Bækdahl, F. (2021). The private tutoring industry in denmark: Market making and modes of moral justification. *ECNU Review of Education*, *4*(3), 520–545.

Dietrichson, J., Bøg, M., Filges, T., & Klint 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*(2), 243–282.

DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J. C. (2011). How effective are mentoring programs for youth? a systematic assessment of the evidence. *Psychological science in the public interest*, *12*(2), 57–91.

Gilraine, M., & Penney, J. (2023). Focused interventions and test score fade-out. *Review of Economics and Statistics*, 1–27.

Gortazar, L., Hupkau, C., & Roldán-Monés, A. (2024). Online tutoring works: Experimental evidence from a program with vulnerable children. *Journal of Public Economics*, *232*, 105082.

Grossman, J. B., & Tierney, J. P. (1998). Does mentoring work? an impact study of the big brothers big sisters program. *Evaluation review*, *22*(3), 403–426.

Hardt, D., Nagler, M., & Rincke, J. (2022). Can peer mentoring improve online teaching effectiveness? an rct during the covid-19 pandemic. *Labour Economics*, *78*, 102220.

Hassan, M. H., Islam, A., Siddique, A., Wang, L. C., et al. (2023). *Telementoring and homeschooling during school closures: A randomized experiment in rural bangladesh*. IZA-Institute of Labor Economics.

Herrera, C., Grossman, J. B., Kauh, T. J., & McMaken, J. (2011). Mentoring in schools: An impact study of big brothers big sisters school-based mentoring. *Child development*, *82*(1), 346–361.

Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness? a review of the research. *Learning Policy Institute*.

Kraft, M., Bolves, A., & Hurd, N. (2021). School-based mentoring relationships and human capital formation. *EdWorkingPaper: 21*, *441*.

Kraft, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. *AEA Papers and Proceedings*, *112*, 614–618.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for prek–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, *61*(1), 74–107.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, *79*(1), 3–38.

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). Accelerating student learning with high-dosage tutoring. edresearch for recovery design principles series. *EdResearch for recovery project*.

Sánchez, B., Esparza, P., & Colón, Y. (2008). Natural mentoring under the microscope: An investigation of mentoring relationships and latino adolescents' academic performance. *Journal of Community Psychology*, *36*(4), 468–482.

Appendix

	Basic (1)		With controls (2)		Pre-score control (3)	
Outcome	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Math	0.0183	0.2492	0.0222	0.1788	0.0215	0.1626
English	0.0895	0.0000	0.0914	0.0000	0.0893	0.0000
Observations	1007		982		982	

Table A.1: Impact of Treatment Assignment on English and Math Scores

Note: This table presents the estimates and p-values for the mean-difference estimation of the effect of belonging to the treated group on the percentage of good responses, with increasing levels of controls. Model (1) presents the basic estimation, Model (2) incorporates demographic controls, cohort fixed effects, and test duration, and Model (3) adds a control for the percentage of good response in baseline test. The demographic controls include dummies for parental education, Sisbén status, tutee gender, and class repetition.

	Basic (1)		With controls (2)		Pre-score control (3)	
Outcome	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Math	0.0293	0.2423	0.0281	0.2806	0.0344	0.1466
English	0.0284	0.1996	0.0268	0.2372	0.0312	0.1177
Observations	410		403		403	

Table A.2: Effect of Math Tutoring on Tutees with a Preference for Math

Note: This table presents the estimates and p-values for the mean-difference estimation of the effect of Math tutoring on the percentage of good responses among those who preferred Math tutoring, with increasing levels of controls. Model (1) presents the basic estimation, Model (2) incorporates demographic controls, cohort fixed effects, and test duration, and Model (3) adds a control for the percentage of good response in baseline test. The demographic controls include dummies for parental education, Sisbén status, tutee gender, and class repetition.

	Basic (1)		With controls (2)		Pre-score control (3)	
Outcome	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Math	0.0206	0.3481	0.0306	0.1880	0.0269	0.2122
English	0.1402	0.0000	0.1451	0.0000	0.1425	0.0000
Observations	527		511		511	

Table A.3: Effect of English Tutoring on Tutees with a Preference for English

Note: This table presents the estimates and p-values for the mean-difference estimation of the effect of English tutoring on the percentage of correct responses among those who preferred English tutoring, with increasing levels of controls. Model (1) presents the basic estimation, Model (2) incorporates demographic controls, cohort fixed effects, and test duration, and Model (3) adds a control for the percentage of good response in baseline test. The demographic controls include dummies for parental education, Sisbén status, tutee gender, and class repetition.