Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1451

Access is Not Enough: Human Support Improves Engagement with AI Tutoring

Carly D. Robinson
Stanford University

David Gormley
Stanford University

Ana Trindade Ribeiro
Stanford University

Susanna Loeb
Stanford University

AI tutoring platforms offer a promising path to scaling personalized instruction, but only if students use them. We report findings from two randomized controlled trials in which elementary students were assigned to use an AI literacy platform independently or with an in-person tutor focused on engagement, not direct instruction. Despite dedicated session time, nearly half of students in the control group never used the platform, and those who did averaged only 2-5 minutes per week. Working with human tutors increased average weekly platform usage by 1 to 4 minutes and engagement by 71-80%. However, usage remained low, and the intervention did not improve reading achievement. Findings suggest that access alone is insufficient; implementation and human support matter for promoting meaningful engagement with AI-based learning tools.

Acknowledgements: We are grateful to our partners in this research. We received insightful feedback and support from the SCALE Initiative team, in particular Nancy Waymack and Cristina Barnard Gonzalez. We thank Accelerate for the funding for this study. We are also grateful to the Smith Richardson Foundation for their support of our full research program and for the financial support from the National Academy of Education/Spencer Postdoctoral Fellowship program.

VERSION: June 2026

Suggested citation: Robinson, Carly D., David Gormley, Ana Trindade Ribeiro, and Susanna Loeb. (2026). Access is Not Enough: Human Support Improves Engagement with AI Tutoring. (EdWorkingPaper: 26-1451). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/pz7p-p388

Access is Not Enough: Human Support Improves Engagement with AI Tutoring

Carly D. Robinson
David Gormley
Ana Ribeiro
Susanna Loeb
Stanford University, SCALE Initiative

Abstract

AI tutoring platforms offer a promising path to scaling personalized instruction, but only if students use them. We report findings from two randomized controlled trials in which elementary students were assigned to use an AI literacy platform independently or with an in-person tutor focused on engagement, not direct instruction. Despite dedicated session time, nearly half of students in the control group never used the platform, and those who did averaged only 2-5 minutes per week. Working with human tutors increased average weekly platform usage by 1 to 4 minutes and engagement by 71-80%. However, usage remained low, and the intervention did not improve reading achievement. Findings suggest that access alone is insufficient; implementation and human support matter for promoting meaningful engagement with AI-based learning tools.

Introduction

Personalized instruction is a powerful lever for improving student learning, and AI tutoring platforms offer a promising path to scaling it by delivering individualized content and feedback at low cost. Although the evidence base remains limited (Fesler et al., 2026; LearnLM et al., 2025), realizing that promise depends on a precondition: students must consistently engage with the platform. Yet, educational technology has historically struggled with low uptake, particularly among struggling students (Eames et al., 2026; Robinson et al., 2025), limiting the practical impact of even the most promising tools.

Historically, personalized instruction depended almost entirely on human educators, and the educator-student relationship was central to the experience. Advances in educational technology have changed this dynamic: personalized instruction now exists on a spectrum of relational intensity from a consistent 1:1 human tutor to a static computer-assisted platform the student navigates alone (Figure 1). AI tutors may approximate some relational components of personalized instruction, but students may still benefit from motivation, accountability, and care that humans provide.

One response is pairing AI platforms with human support, which can take many forms. At the tech-led, low relational intensity end, humans oversee AI-student interactions from the backend. At the human-led, high relational intensity end, a dedicated tutor uses AI as a supplement to direct instruction. This study examines a model near the midpoint of this spectrum: an engaged, in-person human tutor whose role is to support student engagement with an AI literacy tutor, not to deliver direct instruction.

Prior work on hybrid human-AI tutoring finds that human support generally increases usage and task completion compared to students engaging with the AI alone (Gurung et al., 2025; Thomas et al., 2024). However, these studies used virtual tutors and dedicated all session time to platform use. In contrast, the model studied here devoted time to check-ins and reflection, reducing the minutes available for students to work on the platform. Positive educator-student relationships are associated with greater student motivation and persistence (Roorda et al., 2011). This study examines whether that relational investment increased engagement enough to offset the reduction in platform time.

Across two randomized controlled trials (RCTs), we found that human tutors increased, rather than reduced, students' usage of and engagement with the AI platform. However, baseline usage was so low that the gains amounted to less than two additional hours of platform use across the study, on average. These findings underscore two patterns: access to AI tutoring does not ensure meaningful use, and human support can increase student engagement with AI-based learning tools, though not necessarily enough to generate measurable learning gains.

These findings are directly relevant for schools and policymakers considering AI tutoring investments. AI tutors may personalize instruction, but their potential depends on students using them often enough to benefit. Human support appears to help, but the consistency of that support matters.

Methods

Sample and Design

To test the impact of pairing elementary school students with a human tutor while working on an AI literacy platform, we partnered with two school districts. The implementations differed in setting and tutor type. In District A, students in grades 1-5 received tutoring at five after-school program sites with program staff as tutors. In District B, students in grades 1-3 received tutoring in-school at two K-8 schools with middle school students as tutors.

In both implementations, students were randomly assigned to either work on the AI platform independently (control group) or to work on the AI platform with a human tutor (treatment group). We stratified the randomization by site x grade/classroom. In District A, we assigned 84 students to treatment and 90 students to control. In District B, we assigned 36 students to treatment and 145 students to control. The unequal treatment and control group sizes reflect site-level variation in tutoring capacity, as each site could support only a predetermined number of students with a human tutor. Tables A1 and A2 provide baseline information on the sample, and show baseline equivalence between conditions. In both studies, the majority of students were identified as economically disadvantaged (>80%). We did not receive background information on human tutors. In District B, middle school tutors were selected because they performed well in their own English Language Arts (ELA) class and had a free intervention block in their school day.

Procedure and Intervention Details

In both conditions, students were expected to receive time to use the AI platform over the course of the intervention (range: 14-31 weeks; Table A3). Students were expected to participate in at least two 30-minute sessions per week. According to the platform provider, students typically begin to experience academic benefits after accumulating at least 30 minutes per week of platform use.

Students in the treatment condition were assigned a human tutor to support them as they used the platform. Tutors met with small groups of 2-5 students at a time. Prior to the intervention, tutors completed a standardized training designed to help them support students while using the AI platform. Tutors did not provide direct reading instruction. Instead, the platform delivered individualized literacy instruction and feedback while tutors focused on supporting student engagement, motivation, participation, troubleshooting technology issues, and establishing group norms. Sessions with treatment students followed a consistent structure that included a brief check-in (10 minutes), time for students to work on the platform (15 minutes), and a reflection (2-5 minutes). Control students were expected to work on the platform for the full 30 minutes (Figure A1). Given these structural differences, one might expect students in the treatment condition to spend less time actively using the platform during each session than students in the control condition.

Data

Our preregistered outcomes were platform usage and engagement (Link to preregistration). We defined usage as the number of minutes spent on the AI platform and engagement as the number of stories students read during their sessions. While usage reflects how long the student was logged in, engagement reflects the amount of content students completed. Each site had different study durations, so we calculated average usage and engagement per week. We also measured whether students *ever* used the platform.

Additionally, we collected student-level administrative data, including English learner (EL) status, an indicator of economic disadvantage, special education status, race/ethnicity, gender, and student achievement. Beginning-of-year (BOY) assessments served as controls for prior achievement and end-of-year (EOY) assessments served as the student achievement outcome. District A administered the i-Ready assessment and District B administered the STAR assessment (BOY) and the aimsWeb assessment (EOY). See the Technical Appendix for more details.

Analytic Approach

To evaluate the impact of being assigned to work with a human tutor, we used OLS regressions to generate intent-to-treat estimates. Our estimates are from a model that includes strata (site x grade) fixed effects, student background measures (EL status, economic disadvantaged indicator, special education status, race/ethnicity, and gender), BOY student achievement, and robust standard errors.

Results

Control group usage and student take-up

First, we examined platform usage among students assigned to the control group (Table A4). Despite having access to the AI tutor and scheduled time to use it, only 60.7% and 53.3% of students ever used the AI tutor in District A and District B, respectively. This translated to average weekly usage of 2.18 minutes in District A and 5.23 minutes in District B. When students used the platform, they averaged 13.2 and 25.8 minutes weekly in Districts A and B, respectively; but most weeks students did not use the platform at all. The low average weekly usage reflected limited participation across the intervention period: students used the AI platform during only four to five weeks, on average, despite interventions lasting 14 to 31 weeks.

Platform use also differed by student demographics. Students who used the AI tutor were less likely to receive special education services and more likely to be higher achieving (Table A5).

Usage and Engagement Effects

We present the ITT results in Table 1. We did not find evidence that the treatment increased the likelihood that students ever used the AI tutor (Column 1), suggesting that subsequent treatment effects are not driven by one-time log-ins.

We found positive effects on both preregistered outcomes: usage (minutes per week) and engagement (stories read per week). In District A, working with a human tutor increased usage of the AI platform by an average of one minute per week. In District B, the treatment effect was more pronounced, increasing usage by about 4.4 minutes per week (Table 1, Column 2). Although statistically significant, these increases did little to close the gap between observed usage and the 30 minutes of weekly platform use recommended by the provider for measurable reading gains. The results were robust to alternative specifications (Tables A6 and A7).

Figure 3 illustrates usage patterns over time. Usage remained low throughout the intervention in District A, whereas treatment students in District B consistently used the platform more than control students. Treatment effects also varied across sites (Table A8). Several sites showed little or no effect, while others experienced meaningful increases in platform usage, suggesting that local implementation conditions may influence the extent to which human support increases student usage.

Working with a human tutor also increased student engagement with the platform. In District A, treatment students read 0.20 more stories per week than control students, a 71% increase relative to the control mean of 0.28 stories per week and equivalent to roughly 4.4 additional stories over the intervention. In District B, treatment students read 0.92 more stories per week, an 80% increase relative to the control mean of 1.14 stories per week. These results were robust to alternative specifications (Tables A9 and A10).

Although the magnitude of these effects differed across districts, the overall pattern was consistent. Students assigned a human tutor spent more time using the AI platform and completed more content than students assigned to work independently. At the same time, overall usage remained low, even among treatment students. Human support increased engagement but usage remained well below the levels of platform use recommended for measurable reading gains.

Exploratory Results: Achievement

We did not find evidence that working with a human tutor improved reading achievement in either district (Table 1, Column 4). In both districts, the estimated effects on students' spring ELA scores were not statistically significant and negative in magnitude.

These null results are consistent with the dosage levels actually achieved. The additional reading practice generated by the intervention was far below the threshold typically associated with measurable literacy gains. In District A, the treatment added approximately one additional minute per week, or roughly 22 extra minutes of platform use over the full intervention. Dosage was higher in District B, where the treatment added approximately 98 additional minutes across the intervention. Even there, however, usage remained low. The school with the highest usage averaged 18.3 minutes per week in the control group, but subgroup estimates should be interpreted cautiously given the small sample size.

Discussion

Across two RCTs, students assigned a human tutor used and engaged with an AI literacy platform more than students assigned to work independently. These gains occurred even though tutors devoted part of each session to relationship-building activities such as check-ins and reflection, reducing the time available for platform use. The findings suggest that the combination of relational and organizational support provided by tutors can increase engagement enough to offset this loss of instructional time. However, baseline usage was so low that the resulting gains amounted to less than two additional hours of platform use over the course of the intervention and did not translate into measurable improvements in reading achievement.

The low usage is itself a policy-relevant finding for schools considering AI tutoring investments. Providing access to an AI tutor, whether after school or during school hours, did not ensure meaningful engagement. Most students used the platform for only a small fraction of the intervention period, and many never used it at all. This pattern is consistent with broader evidence that access is a necessary but insufficient condition for technology-enabled learning (Robinson et al., 2025). Discussions about AI tutoring often focus on the quality of the technology itself, but these findings suggest that implementation and student engagement may be equally important determinants of impact.

Take-up also varied by student characteristics. Among control students, those who engaged with the AI platform were more likely to be higher achieving and less likely to receive special education services. Struggling students, who may have benefited the most from additional reading practice, were the least likely to use the tool. This may reflect purposeful accommodation decisions made on behalf of individual students; however, at the time of study launch, all students in the sample were expected to use the AI tutor. The pattern raises equity considerations that matter for deployment decisions. If AI tutoring tools are disproportionally used by students who are already better positioned to succeed, then technology alone may not reduce educational inequality and could potentially reinforce existing participation gaps. Human support may be especially important for engaging students who are least likely to access or persist with AI-based learning opportunities on their own.

The findings also highlight the tradeoffs involved in different points along the spectrum of human-AI personalized instruction. Fully AI-led models maximize scalability and minimize cost, but may struggle to generate consistent student engagement. More human intensive models may increase participation and accountability but require additional personnel, training, and coordination. The model studied here occupied a middle position on this spectrum: tutors did not provide direct instruction but instead focused on relationship-building, motivation, and troubleshooting. The fact that this support increased usage suggests that human support remains an important component of personalized learning even when instruction is delivered primarily by AI. Future research should compare models across the spectrum to identify when additional relational support generates benefits large enough to justify the added costs.

Rather than replacing human relationships, AI-based personalized instruction may still depend on them. Across both studies, human tutors increased engagement with the AI platform despite reducing the time available for direct platform use, suggesting that motivation, accountability, and interpersonal connection remain important drivers of student participation. At the same time, the small gains in usage underscore the limits of relational support when overall implementation remains weak. Consistent with broader evidence that dosage is a central determinant of tutoring effectiveness (Bhatt et al., 2025; Huffaker et al., 2025), students must engage with AI tutors at meaningful levels before academic benefits can emerge. AI tutoring platforms remove some common barriers to dosage. For example, sessions are not canceled because a tutor is absent, and students can potentially access the platform whenever a device is available. Yet our findings suggest that removing these logistical barriers is not sufficient to generate consistent engagement. Even when students were provided access and dedicated time to use the platform, usage remained low. As AI tutoring becomes more common in schools, understanding how different forms of human support shape student engagement will be critical for determining when and how these tools can contribute to student learning.

References

Bhatt, M. P., Chau, T., Condliffe, B., Davis, R., Grossman, J., Guryan, J., Ludwig, J., Magnaricotte, M., Mattera, S., Momeni, F., Oreopoulos, P., & Stoddard, G. (2025). *Personalized Learning Initiative Interim Report: Findings from 2023-24*. https://educationlab.uchicago.edu/resources/personalized-learning-initiative-interim-report-findings-from-2023-24/

Eames, T., Brunskill, E., Yamkovenko, B., Weatherholtz, K., & Oreopoulos, P. (2026). Computer-assisted learning in the real world: How Khan Academy influences student math learning. *Proceedings of the National Academy of Sciences*, *123*(1), e2507708123. https://doi.org/10.1073/pnas.2507708123

Fesler, L., Martinez, J. P., Agnew, C., & Loeb, S. (2026). *The Evidence Base on AI in K-12: A 2026 Review* [AI Hub for Education of the SCALE Initiative]. Stanford University. https://scale.stanford.edu/sites/default/files/The%20Evidence%20Base%20on%20AI%20in%20K-12%20Report.pdf

Gurung, A., Lin, J., Gutterman, J., Thomas, D. R., Houk, A., Gupta, S., Brunskill, E., Branstetter, L., Aleven, V., & Koedinger, K. (2025). Human Tutoring Improves the Impact of AI Tutor Use on Learning Outcomes. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), *Artificial Intelligence in Education* (pp. 393–407). Springer Nature Switzerland.

Huffaker, E., Lee, M. G., Zhou, H., Robinson, C. D., & Loeb, S. (2025). *Beyond the One-Teacher Model: Experimental Evidence on Using Embedded Paraprofessionals as Personalized Instructors*. edworkingpapers.com. https://doi.org/10.26300/PZY2-WR51

LearnLM, Eedi, Rysbek, A., Huber, A., Nambiar, A., Kenolty, A., Caulfield, B., Lilley-Draper, B., Groot, B., Veprek, B., Burdett, C., Willis, C., Barton, C., Smith, D., Mu, G., Walters, H., Jurenka, I., Hulls, I., Stalley-Moores, J., … Brazão, V. (2025). *AI tutoring can safely and effectively support students: An exploratory RCT in UK classrooms* (arXiv:2512.23633). arXiv. https://doi.org/10.48550/arXiv.2512.23633

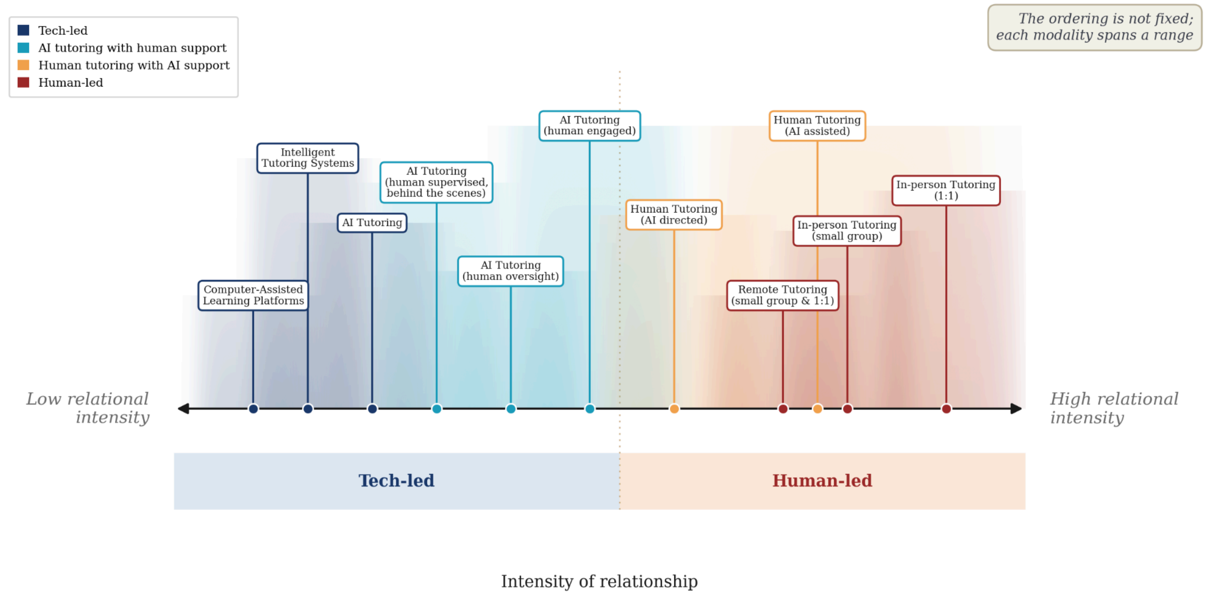
Robinson, C. D., Bisht, B., & Loeb, S. (2025). The Inequity of Opt-in Educational Resources and an Intervention to increase Equitable Access. *Educational Researcher*, 0013189X251331518. https://doi.org/10.3102/0013189X251331518

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*(4), 493–529.

Thomas, D. R., Lin, J., Gatz, E., Gurung, A., Gupta, S., Norberg, K., Fancsali, S. E., Aleven, V., Branstetter, L., Brunskill, E., & Koedinger, K. R. (2024). Improving Student Learning with Hybrid Human-AI Tutoring: A Three-Study Quasi-Experimental Investigation. *Proceedings of the 14th Learning Analytics and Knowledge Conference, LAK '24*, 404–415. https://doi.org/10.1145/3636555.3636896

Primary Exhibits

Figure 1. Personalized instruction on a spectrum of relational intensity



Caption: A theoretical depiction of models of personalized instruction by relational intensity, from tech-led (low relational intensity) to human-led (high relational intensity). Each modality spans a range, not a point, as indicated by the shaded band, so the ordering is not fixed. For instance, an AI-assisted human tutor–with AI handling logistics so the human invests more in the relationship–can reach or even exceed the relational intensity of in-person tutoring. The vertical position of labels is arbitrary and used only to keep labels legible.

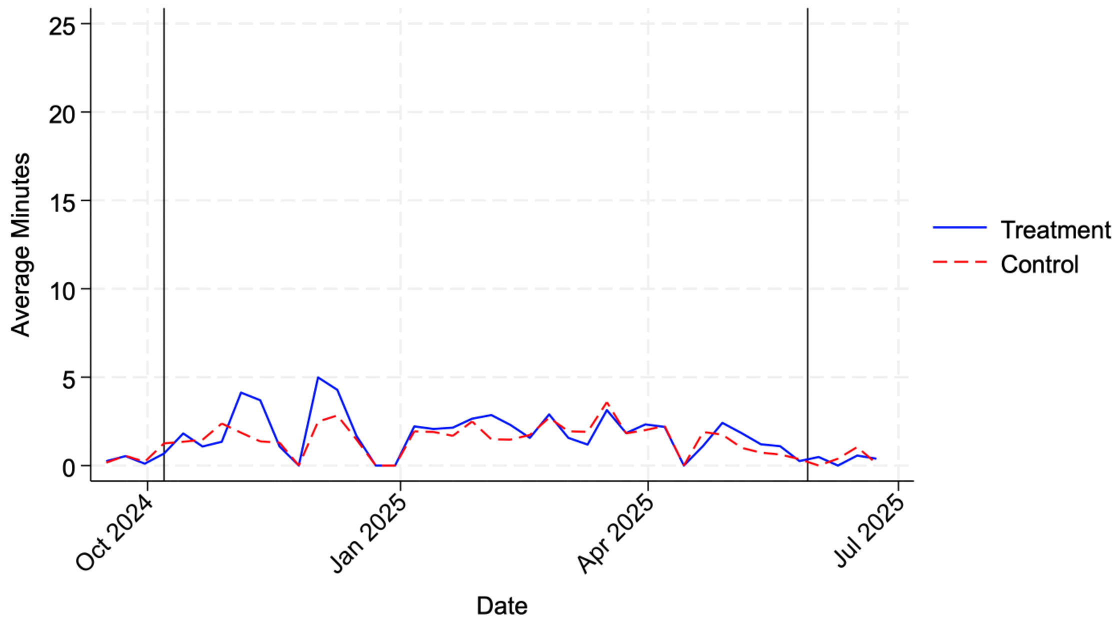
Table 1. Impact of pairing an AI platform with a human tutor on student outcomes

	Usage: Ever used (pp)	Usage: Avg min/week	Engagement: Avg stories read/week	Achievement: Avg EOY ELA score (std)
	(1)	(2)	(3)	(4)
District A				
Treatment	0.075	0.995+	0.201*	-0.052
	(0.065)	(0.578)	(0.093)	(0.086)
Observations	174	174	174	166
Control mean	0.607	2.179	0.280	0.150
District B				
Treatment	-0.013	4.424*	0.916*	-0.129
	(0.084)	(1.879)	(0.373)	(0.143)
Observations	181	181	181	161
Control mean	0.533	5.233	1.138	0.100

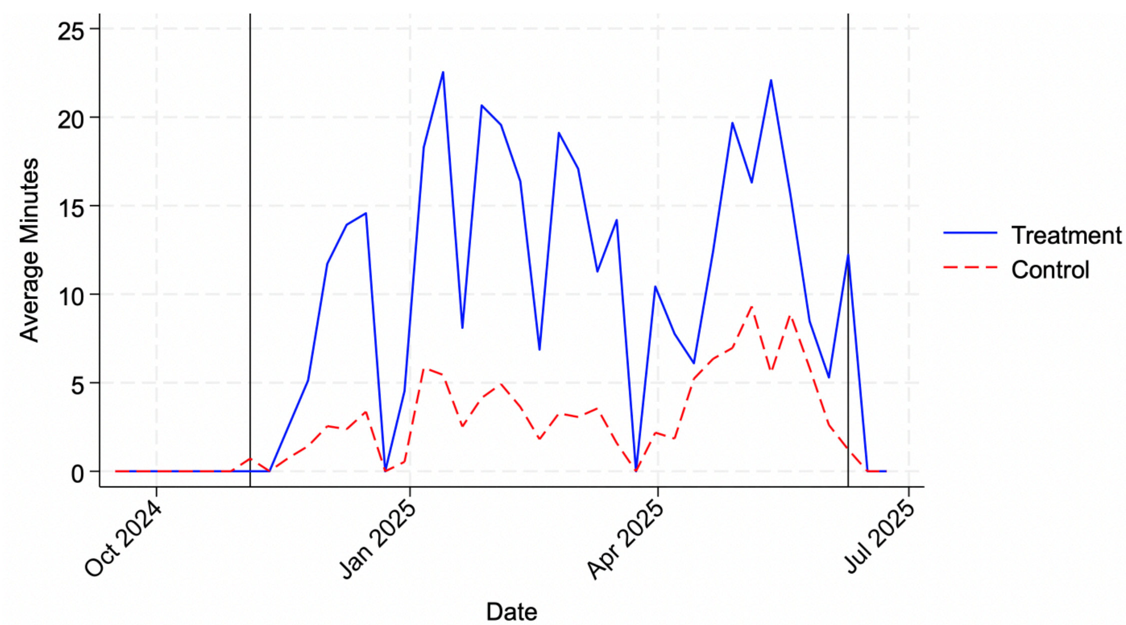
All models control for student covariates, baseline achievement, and include a fixed effect for randomization strata.
Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table 3. Average weekly minutes spent on the AI tutoring platform by study and condition

District A



District B



Note: Vertical lines denote study start and end dates for each district.

Appendix

Technical Appendix

Details on the Sample.

Our initial preregistration was based on a pilot study (N = 236) that was ultimately not implemented as planned. The following year we conducted the study in two school districts (explaining the change in sample and procedures) but we retained all other components, including the study design, conditions, hypotheses, and analytic plans.

Details on Data.

Student demographic data.
When students were missing baseline categorical demographic data, we imputed a constant value and added a “missing” indicator for that variable.

Student achievement data.
Because of the small sample sizes in each grade, we standardized BOY and EOY scores with grades based on all available student data in participating sites/schools. When BOY scores were missing, we imputed their baseline score based on the mean score for their grade and school (9% and 6% in District A and B, respectively). In our preregistration we said we would use an imputed score predicted from non-missing demographic and outcome data, conditional on treatment assignment; however the strata samples were quite small so we chose a more conservative route. The findings are robust to whether we include the sample that have imputed baseline scores.

Details on the Analytic Approach

Our estimates are from a model that includes strata (site x grade/clasroom) fixed effects, student background measures (EL status, economic disadvantaged indicator, special education status, race/ethnicity, and gender), BOY student achievement, and robust standard errors.

In our preregistration, we noted we would include the baseline covariate relative to the outcome of interest, if available (i.e., prior number of minutes spent on the platform). However, we did not receive students’ baseline usage or engagement data, so we were not able to control for students’ baseline in those scenarios.

Appendix Tables

Table A1. District A: Sample characteristics and randomization balance check

Variable Name	Overall Mean	Overall N	Treatment Mean	Treatment N	Control Mean	Control N	*p*	
Female	0.51	174	0.52	84	0.50	90	0.81	
Economic Disadvantage	0.89	174	0.87	84	0.90	90	0.61	
Special Education	0.12	174	0.11	84	0.13	90	0.50	
English Language Learner	0.17	174	0.21	84	0.13	90	0.10	+
Race								
Black	0.22	174	0.26	84	0.19	90	0.28	
White	0.40	174	0.35	84	0.44	90	0.24	
Hispanic	0.57	174	0.58	84	0.57	90	0.69	
Multi/Other	0.38	174	0.39	84	0.37	90	0.78	
Prior Achievement								
BOY Score	0.10 (1.01)	170	0.08 (0.91)	84	0.12 (1.11)	86	0.88	
Missing BOY Score	0.09 (1.00)	174	0.08 (0.91)	84	0.11 (1.09)	90	0.88	
Imputed BOY Score	0.02	174	0.00	84	0.04	90	0.04	*
Grade Level								
Grade 1	0.05	174	0.04	84	0.06	90	0.47	
Grade 2	0.11	174	0.12	84	0.10	90	0.47	
Grade 3	0.24	174	0.24	84	0.24	90	0.30	
Grade 4	0.24	174	0.24	84	0.24	90	0.30	
Grade 5	0.21	174	0.19	84	0.23	90	0.49	
Grade 6	0.15	174	0.18	84	0.12	90	0.30	

Note: The p-value is calculated while controlling for strata. Standard deviations are reported parenthetically for achievement scores

Table A2. District B: Sample characteristics and randomization balance check

Variable Name	Overall Mean	Overall N	Treatment Mean	Treatment N	Control Mean	Control N	*p*	
Female	0.47	176	0.54	35	0.45	141	0.29	
Missing Gender	0.03	181	0.03	36	0.03	145	0.32	
Economic Disadvantage	0.87	176	0.91	35	0.86	141	0.60	
Missing Economic Disadvantage	0.03	181	0.03	36	0.03	145	0.32	
Special Education	0.23	176	0.26	35	0.22	141	0.53	
Missing Special Education	0.03	181	0.03	36	0.03	145	0.32	
English Language Learner	0.33	176	0.31	35	0.33	141	0.94	
Missing ELL	0.03	181	0.03	36	0.03	145	0.32	
Race								
African-American	0.45	181	0.67	36	0.40	145	0.02	*
Asian	0.13	181	0.11	36	0.14	145	1.00	
Hispanic	0.22	181	0.14	36	0.24	145	0.51	
Multiple	0.10	181	0.03	36	0.12	145	0.00	**
Native American	0.02	181	0.00	36	0.02	145	0.10	
White	0.05	181	0.03	36	0.06	145	0.65	
Missing	0.03	181	0.03	36	0.03	145	0.32	
Prior Achievement								
BOY Score	0.18 (1.00)	170	0.06 (0.89)	34	0.21 (1.02)	136	0.44	
Missing BOY Score	0.06	181	0.06	36	0.06	145	0.94	
Imputed BOY Score	0.18 (0.97)	181	0.08 (0.87)	36	0.21 (0.99)	145	0.44	
Grade Level								
Grade 1	0.38	181	0.47	36	0.36	145	0.22	
Grade 2	0.27	181	0.22	36	0.28	145	0.45	
Grade 3	0.35	181	0.31	36	0.36	145	0.54	

Note: The p-value is calculated while controlling for strata. Standard deviations are reported parenthetically for achievement scores

Table A3. Intervention length details by site

District	Site	Intervention Start Date	Intervention End Date	Intervention Length (wks)	Avg # wks students used platform
District A	Site A	10/7/2024	5/7/2025	27	3.0
	Site B	10/18/2024	2/5/2025	14	2.7
	Site C	12/6/2024	5/29/2025	23	4.1
	Site D	11/1/2024	5/1/2025	23	5.0
	Site E	10/7/2024	4/9/2025	24	5.1
District B	School A	1/6/25	6/6/25	21	3.6
	School B	11/4/24	6/9/25	30	14.2

Figure A1. Session structure

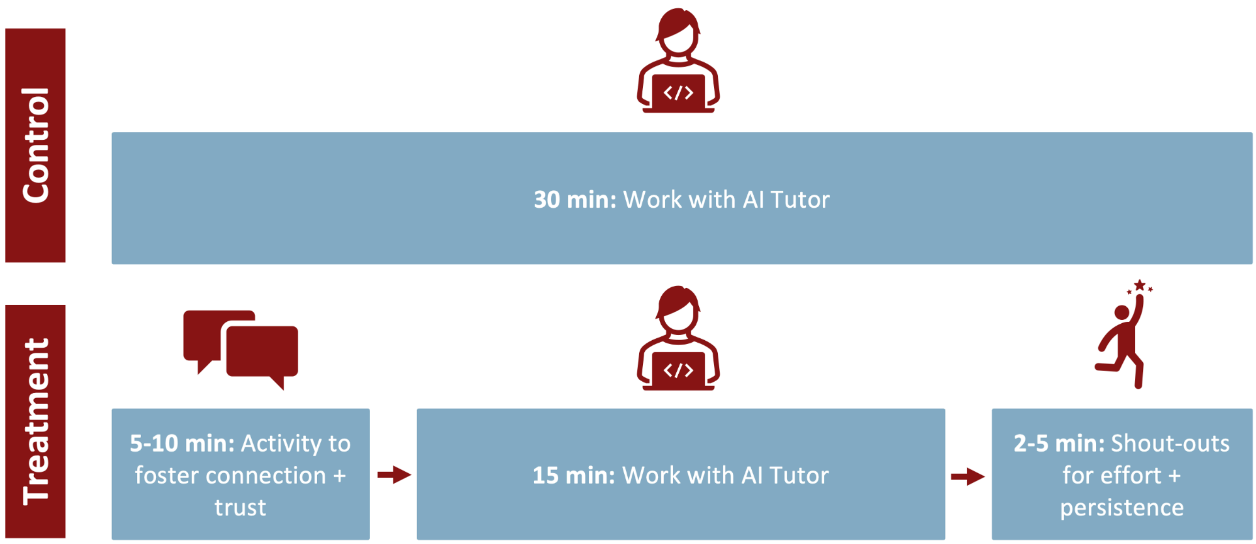


Table A4. Baseline usage of the AI tutoring platform

	% Ever Used	Avg min / week	Avg min / week for wks with > 0 min	Avg # wks used
District A	60.7	2.18	13.2	3.90
District B	53.3	5.23	25.8	4.09

Note: This table only uses data for students in the Control group.

Table A5. Baseline Take-up by Student Demographics

Variable Name	Takers Mean	Takers N	Non-Takers Mean	Non-Takers N	*p*	
District A						
Free And Reduced-Price Lunch	0.86	56	0.97	34	0.06	+
Special Education	0.09	56	0.21	34	0.17	
English-Language Learner	0.13	56	0.15	34	0.50	
Beginning of Year Test Score	0.38	55	-0.35	31	0.02	*
District B						
Free And Reduced-Price Lunch	0.89	75	0.82	66	0.86	
Special Education	0.11	75	0.35	66	0.00	***
English-Language Learner	0.35	75	0.32	66	0.49	
Beginning of Year Test Score	0.30	74	0.09	62	0.27	

Note: This table only uses data for students in the Control group

Table A6A. District A - Impact of students working with human tutor on ever using the AI platform

	Ever used AI platform		
	(1)	(2)	(3)
Treatment	0.062	0.074	0.075
	(0.066)	(0.065)	(0.065)
Site-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	174	174	174
Control mean	0.614	0.608	0.607

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A6B. District B - Impact of students working with human tutor on ever using the AI platform

	Ever used AI platform		
	(1)	(2)	(3)
Treatment	-0.005	-0.014	-0.013
	(0.091)	(0.085)	(0.084)
School-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	181	181	181
Control mean	0.531	0.533	0.533

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A7A. District A - Impact of students working with human tutor on average weekly minutes using the AI platform

	Usage: Average minutes per week on AI platform		
	(1)	(2)	(3)
Treatment	0.886	0.983+	0.995+
	(0.560)	(0.577)	(0.578)
Site-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	174	174	174
Control mean	2.231	2.185	2.179

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A7B. District B - Impact of students working with human tutor on average weekly minutes using the AI platform

	Usage: Average minutes per week on AI platform		
	(1)	(2)	(3)
Treatment	4.262*	4.398*	4.424*
	(1.835)	(1.867)	(1.879)
School-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	181	181	181
Control mean	5.265	5.238	5.233

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A8. Effect of Treatment by Study and School Site

District	Site	Control Students	Treat Students	Control Avg Mins	Treat Avg Mins	Treatment Effect: Minutes
A	A	15	15	2.548	1.189	-1.359
A	B	17	17	1.066	6.611	5.545**
A	C	16	16	1.124	2.830	1.706+
A	D	28	21	2.759	2.607	-0.152
A	E	14	15	2.572	2.994	0.422
B	A	132	24	3.221	4.743	1.521
B	B	13	12	18.322	27.436	9.114

Signif.: + p<0.10; * p<0.05; ** p<0.01; *** p<0.001

Table A9A. District A - Impact of students working with human tutor on average weekly stories read on the AI platform

	Engagement: Average stories completed per week on the AI platform		
	(1)	(2)	(3)
Treatment	0.183*	0.199*	0.201*
	(0.092)	(0.093)	(0.093)
Site-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	174	174	174
Control mean	0.289	0.280	0.280

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A9B. District B - Impact of students working with human tutor on average weekly stories read on the AI platform

	Engagement: Average stories completed per week on the AI platform		
	(1)	(2)	(3)
Treatment	0.843*	0.911*	0.916*
	(0.369)	(0.371)	(0.373)
School-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	181	181	181
Control mean	1.152	1.139	1.138

Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A10A. District A - Impact of students working with human tutor on average EOY ELA assessment

	Achievement: Average end-of-year ELA assessment score (standardized)		
	(1)	(2)	(3)
Treatment	-0.125	-0.094	-0.052
	(0.144)	(0.132)	(0.086)
Site-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	166	166	166
Control mean	0.184	0.169	0.150

Note: The ELA outcome was i-Ready. Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A10B. District B - Impact of students working with human tutor on average EOY ELA assessment

	Achievement: Average end-of-year ELA assessment score (standardized)		
	(1)	(2)	(3)
Treatment	-0.134	-0.068	-0.129
	(0.193)	(0.197)	(0.143)
School-Grade FE	Yes	Yes	Yes
Student Covariates	No	Yes	Yes
Prior Achievement	No	No	Yes
Observations	161	161	161
Control mean	0.101	0.088	0.100

Note: The ELA outcome was the Oral Reading Fluency (ORF) from aimsWeb. Standard errors in parentheses
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001