

Investigating the Use and Effectiveness of Math Tutoring Practices with Human and AI Tutors

Julian Bernado, Ana Ribeiro, Xander Beberman, Carly Robinson, and Susanna Loeb
May 6th, 2026

Stanford | SCALE Initiative
Accelerator for Learning

Investigating Language in Tutoring Sessions:



Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise

Rose E. Wang, Ana Trindade Ribeiro, Carly D. Robinson, Susanna Loeb, Dorottya Demszky

Educator Attention: How computational tools can systematically identify the distribution of a key resource for students

Qingyang Zhang, Rose E. Wang, Ana Trindade Ribeiro, Dorottya Demszky, Susanna Loeb

Understanding Disruptions: Causes of and Variation in Lost Instructional Time

Xander Beberman
Stanford University

Sarah Novicoff
Public Policy Institute of California

Ana Trindade Ribeiro
Stanford University

Carly D. Robinson
Stanford University

Susanna Loeb
Stanford University

The Power of Personalized Attention: Comparing Pedagogical Approaches in Small Group and One-on-One Early Literacy Tutoring

Hsiaolin Hsieh
Stanford University

David Gormley
Stanford University

Carly D. Robinson
Stanford University

Susanna Loeb
Stanford University

What we are working on now

- **What pedagogical moves matter for math tutoring?**
 - Important for training humans and designing AI tutors
 - Lack of evidence about effectiveness for the most part
- **RQ: “When tutors use move X, does student engagement/learning increase?”**

First Step: Measure Creation

- **We're interested in a wide range of behaviors that arise in the tutoring context**
 - Learning engineering
 - Psychology
 - Qualitative education
 - Natural language processing

- **We want a look into both tutors and students**
 - What are tutors doing or not? What can be improved?
 - What do successful sessions look like for students, beyond exit tickets?

Evaluating Tutor Moves

- Multiple sessions per week/month/school year:
 - o Short cycle experiments
 - o Interventions between or within session
 - o Estimate effect at the session level or turn level
- Current approaches:
 - o AI Tutor: Finest grain control over move implementation
 - o Human Tutor:
 - Platform-embedded move recommendations
 - AI coaching platform

Proof of Concept: TutorCoach

AI-Coach platform:

- Targeted move recommendations
- Personalized content based on tutor's past sessions

Jacobs et al., 2022; Yun et al., 2024; Demszky et al., 2025; Thomas et al., 2025

Hi, Julian.

View your feedback from this week below. Scroll down to see older feedback.

January 12, 2026: Say the student's name

Your strategy of the week: Say the student's name

Remember to use the student's name during the session. Using the student's name will help you make the student feel like you care about them and will motivate them to connect with you and stay more engaged throughout the tutoring session.

Click on the boxes below to see examples of this strategy in action.

Your usage is **moderate**.



You're doing OK, but try to use this strategy a little more.

Here are some examples of how to use this strategy effectively:

How to use this move

Tutor: Great job on that problem, Maria!

How to use this move

Tutor: Linda, can you walk me through how you solved that?

Here's an example where this strategy could have been used:

What not to do

Tutor: You're so close, try one more time!

The tutor could have said:

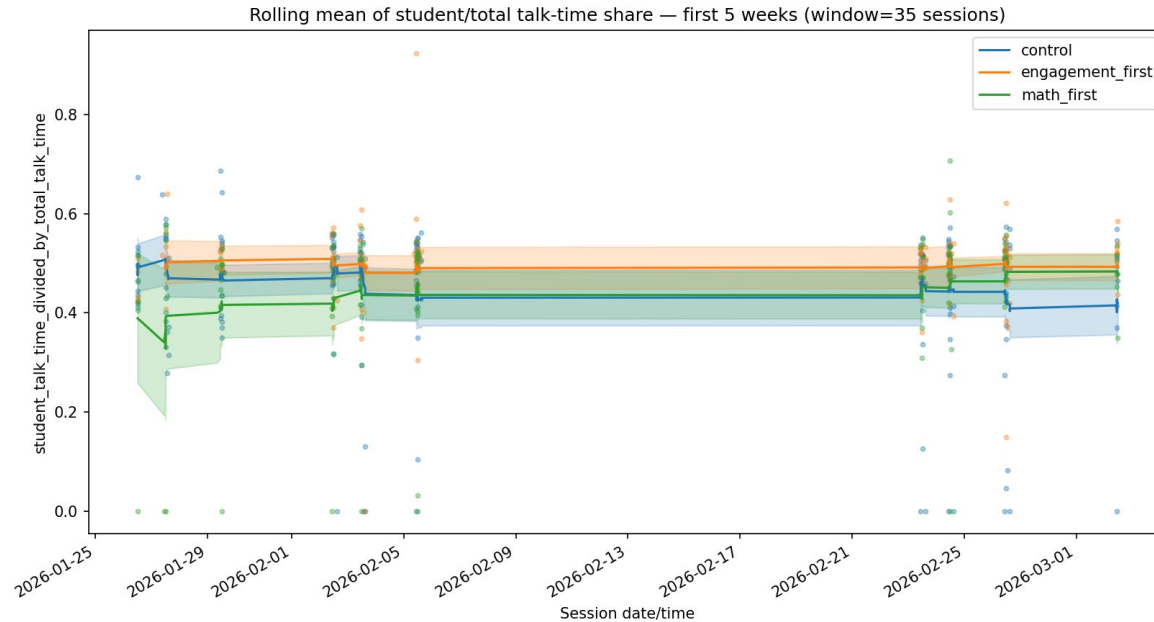
Tutor: You're so close Marcus, try one more time!

TutorCoach Implementation

- Move of the week: email on Monday, reminder on Wednesday
- Treatment: Alternating 3-week sequences of
 - Building Rapport: Using student name, asking about interests, customizing lesson
 - Math Fluency: Writing on the whiteboard, using math nomenclature, explaining reasoning
- Randomization:
 - Placebo
 - Start with Building Rapport
 - Start with Math Fluency

TutorCoach - Early Findings

- First 5 weeks of intervention
- Outcome: talk time as measure of engagement



Next Steps

- Alternative A/B tests with more control (AI, platform-embedded)
- Quasi-experimental approaches
- Assess predictive value of proximal measures of engagement and learning
- What moves to prioritize?