

Welcome! Please load.

Slide deck

bit.ly/4wtm49L



Live poll

bit.ly/4whmsrl





Building Trustworthy Evidence

Partnering for Proof and Positivity in Innovation

Calen Clifton

Center on Reinventing Public Education

SDP Annual Convening | Cambridge, MA | May 2026

Agenda

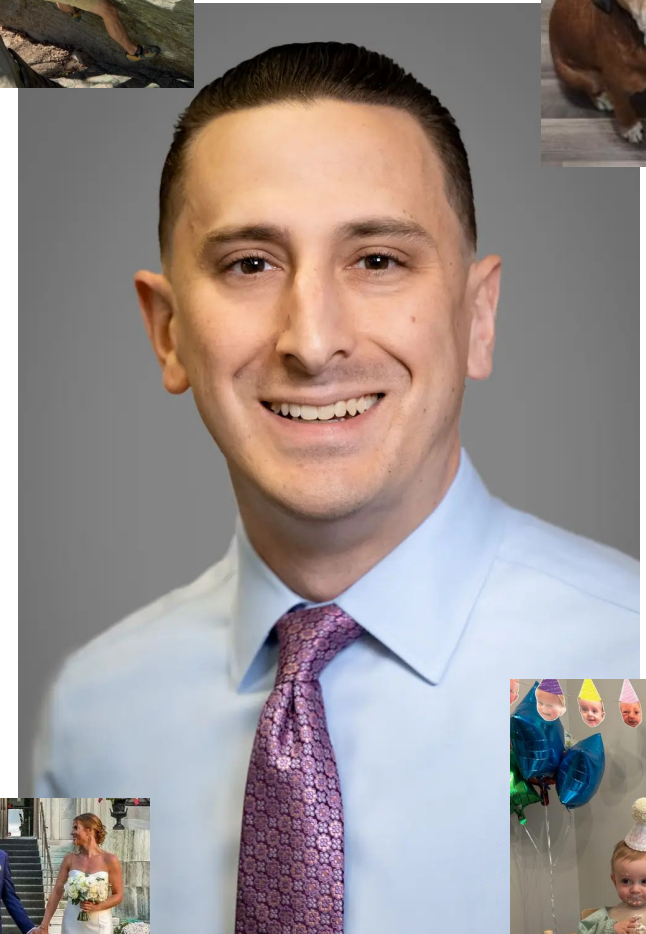
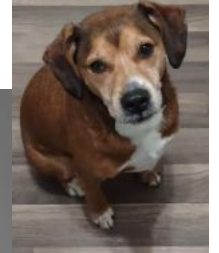
1. Session Introduction (2:45-2:50)
2. The Partnership Context (2:50-2:56)
3. Part 1 (2:56-3:15)
4. Quick Stretch Break (3:15-3:17)
5. Part 2: March 2025-July 2025 (3:17-3:45)
6. Conclusion (3:45-4:00)

Session Introduction

2:45-2:50

About Calen

- Research analyst at the Center on Reinventing Public Education.
- Studied social studies education and history at Appalachian State University (2011) and then taught middle school in Charlotte, NC.
- Transitioned into research after earning an Ed.M. from the Harvard Graduate School of Education (2019).
- Previously an associate in research at Duke University, an SDP Cohort 12 fellow, and a research analyst at the North Carolina Department of Public Instruction.



Content

- This session draws on two evaluations to explore themes of trust across two research partnerships:
 - OUSD virtual literacy tutoring pilot (primary)
 - Next Education Workforce, Mesa Public Schools (quick asides)
- I'm telling this story from my perspective, using paraphrased, summarized, and anonymized emails. Very light creative license, used mostly to protect anonymity.
- I used Claude Pro and NotebookLM to help build this presentation and checked all AI-generated output for accuracy.

Objectives

1. Define “trust.” What is trust?
2. Surface what trust looked like (or didn’t look like) in partnership with OUSD.
3. Reflect on moments in your own work when trust was tested or built, perhaps in unexpected ways.
 - What similarities and differences do you notice?
 - What's different about your context that might influence the development of trust?

RPP elements



1. Building **trust** and cultivating partnership relationships
2. Conducting rigorous research to inform action
3. Supporting the practice organization in achieving its goals
4. Producing knowledge that can inform the field more broadly
5. Building the capacity of researchers and practitioners

Henrick, Ethan C., Paul Cobb, William R. Penuel, Kara Jackson, and Tiffany Clark. *Assessing Research-Practice Partnerships: Five Dimensions of Effectiveness*. New York: William T. Grant Foundation, 2017.

The Partnership Context

2:50-2:56

Partnership Considerations

- Tutoring program and evaluation funded by an anonymous organization
- December 2024-August 2025
 - 31 total meetings
- Four formal deliverables, with several other informal ones
 - Mid-year results (March)
 - End-of-year results (July)
 - Formal grant reporting (August)
 - Final evaluation report (August)
- Initial goal: to help the district select *one* tutoring provider for a much larger future contract.

CRPE at ASU

INTEREST

Produce rigorous, publishable research that protects methodological credibility.

TRUST

Being honest about limitations + partners respecting research independence.

OUSD

INTEREST

Get timely, actionable insights before a major budget decision.

TRUST

Being clear about what you actually need + researchers who respect your constraints.

Intermediary

INTEREST

Keep all parties at the table and surface concerns before crises.

TRUST

Being transparent about tensions + bringing people together when they disagree.

The Funder

INTEREST

Demonstrate that investment produced valuable, defensible evidence.

TRUST

Transparent reporting + knowing research independence was protected.

Providers

INTEREST

Demonstrate effectiveness to win future contracts.

TRUST

Fair, transparent evaluation + understanding how methods affect findings.

Oakland's tutoring initiative

Literacy tutoring in Oakland

- In 2024-25, Oakland Unified School District (OUSD) piloted a virtual 1:1 literacy tutoring program in 33 elementary schools with three different tutoring providers, focused on phonics.
 - Schools were assigned a provider and asked to assign students to tutoring based on specific criteria.
- We analyzed implementation and effectiveness, including differences across tutoring providers.



Hoot

592 students
11 schools

25 minutes
3x week (75 min)

Emphasis on
decoding, fluency,
and comprehension

Ignite

705 students
11 schools

15 minutes
5x week (75 min)

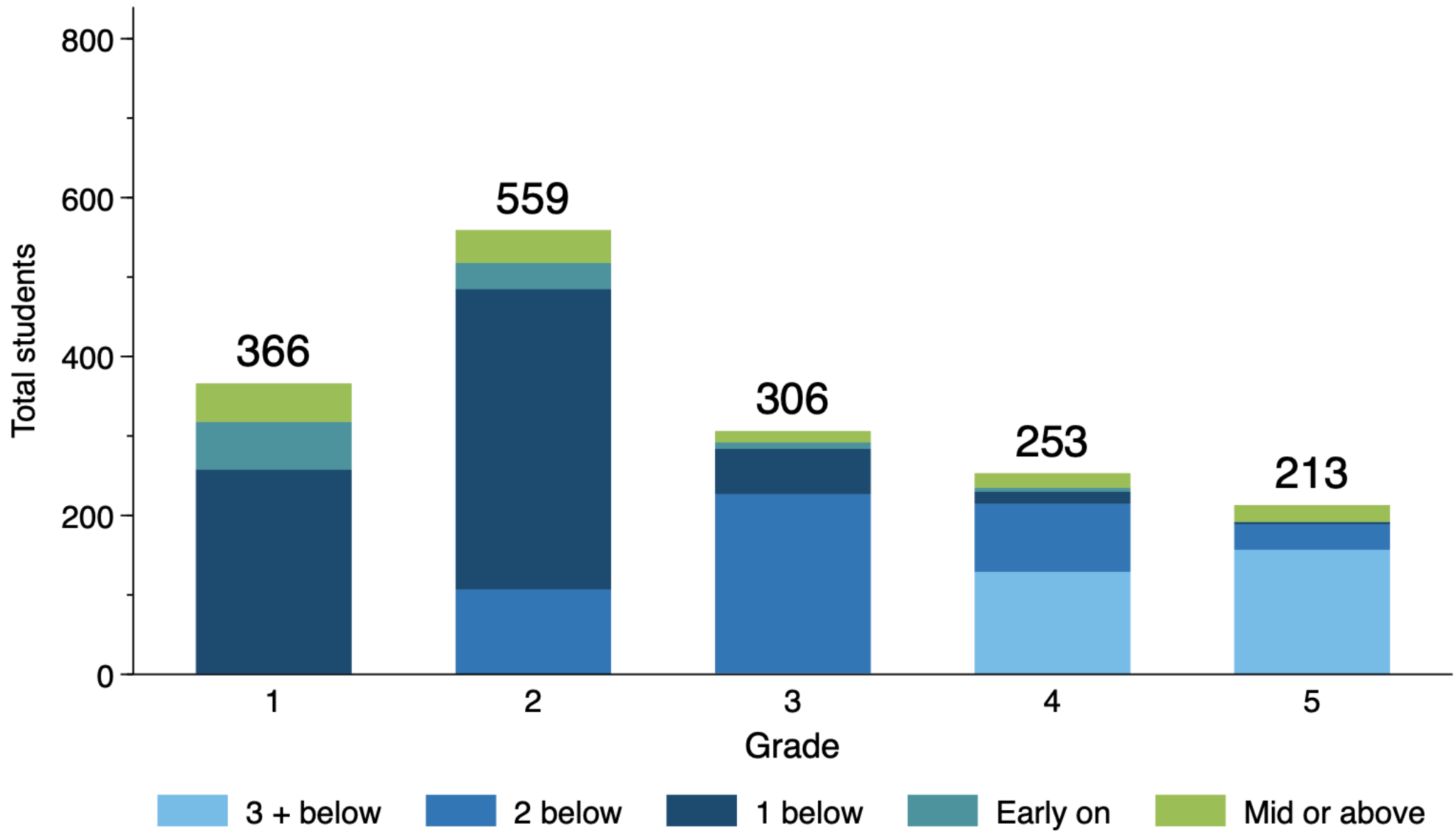
Emphasis on
decoding and
fluency

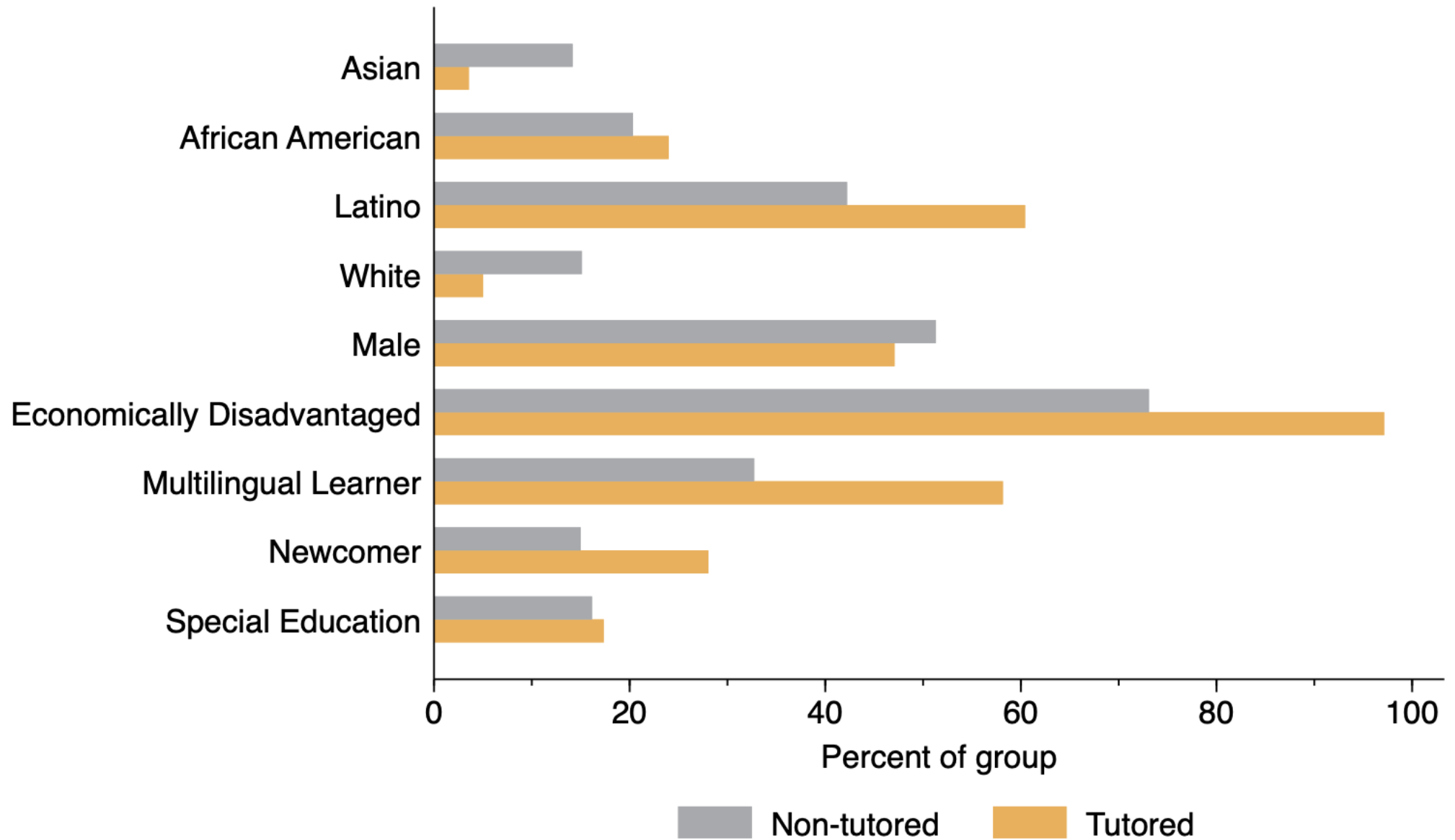
OpenLiteracy

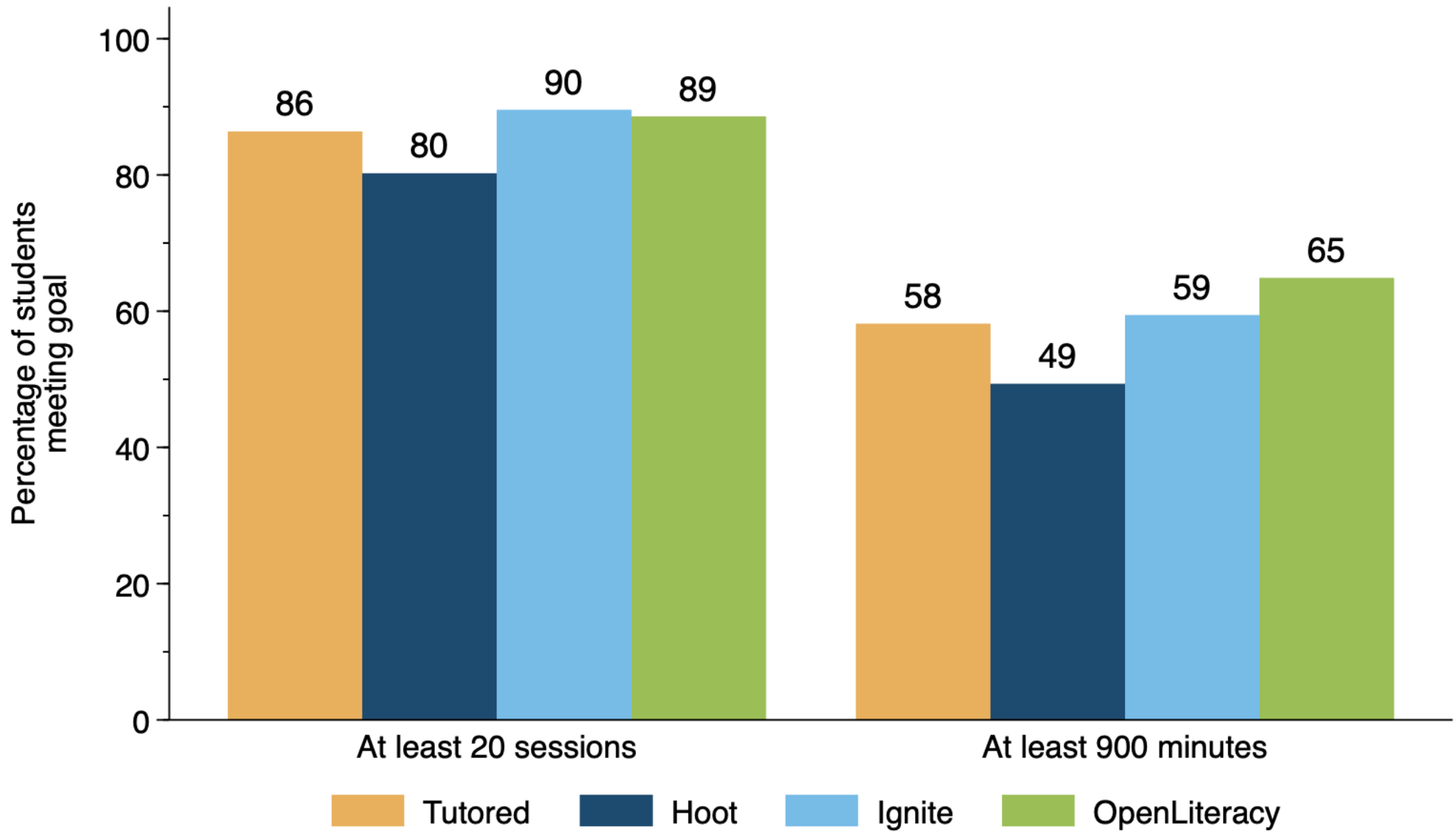
610 students
11 schools

30 minutes
3x week (90 min)

Emphasis on
decoding, fluency,
and comprehension



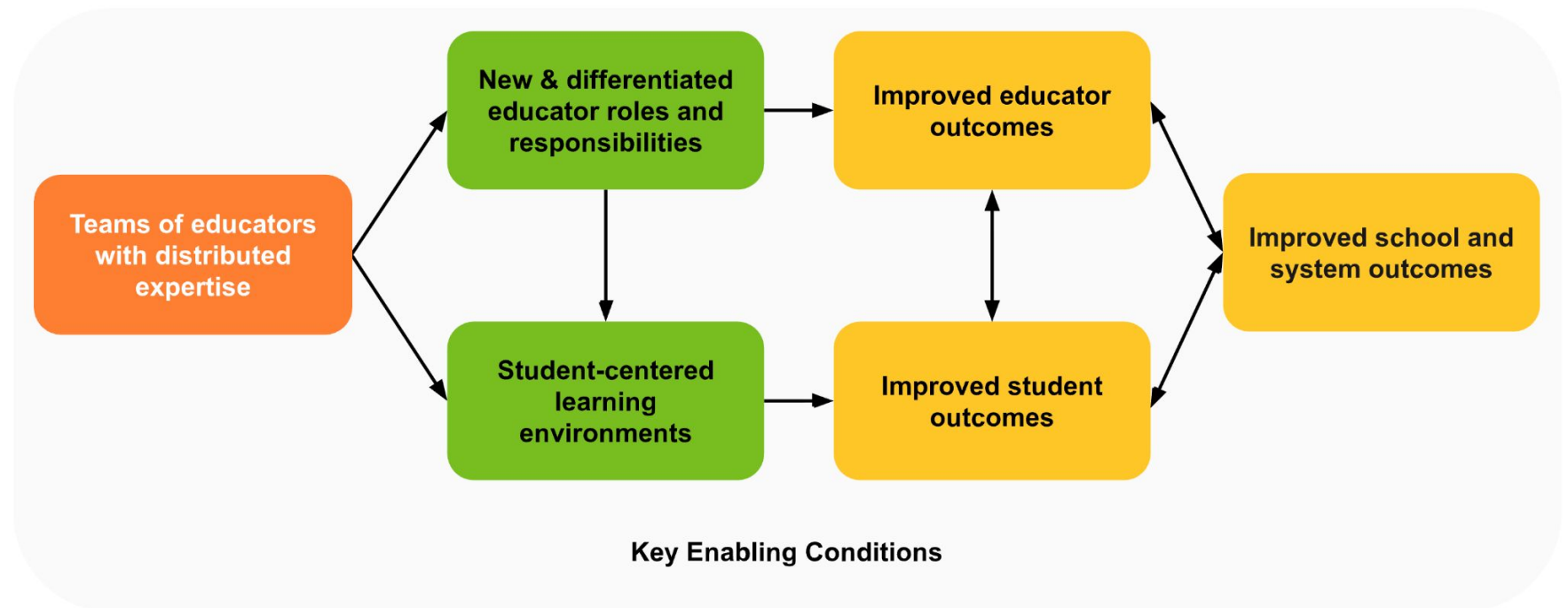




Next Education Workforce

- What is the treatment?
- Teaming.
- What is “teaming?”

Next Education Workforce™ Theory of Action



Part I: December 2024-March 2025

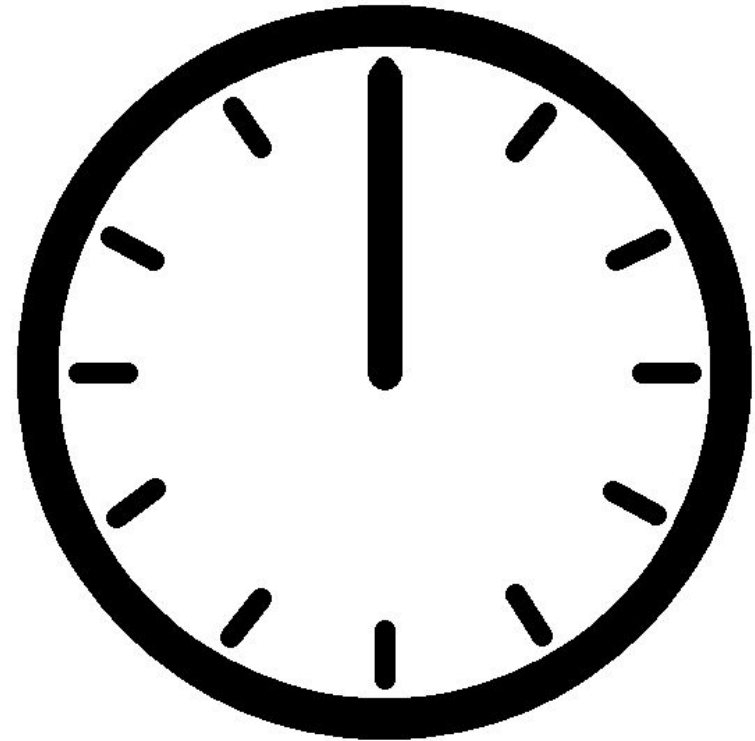
2:56-3:15

Getting started

We began work in January.

Our agreement included mid-year results by the first week of March.

That's not a lot of time.



Later, we will read some simulated emails, looking for any evidence of trust. First, I will model this process using two examples from the beginning of the project.

January 10, 2025

From: District Partner 1

To: Research Team

I'm sorry for the delay — I didn't realize the data sharing agreements aren't yet complete. I told District Partner 2 so that they know.

There's a hiccup with the research application that we are working through. I'll continue pulling the data together so that as soon as the application is final I can send it over.

February 10, 2025

From: Research Team

To: District Partners

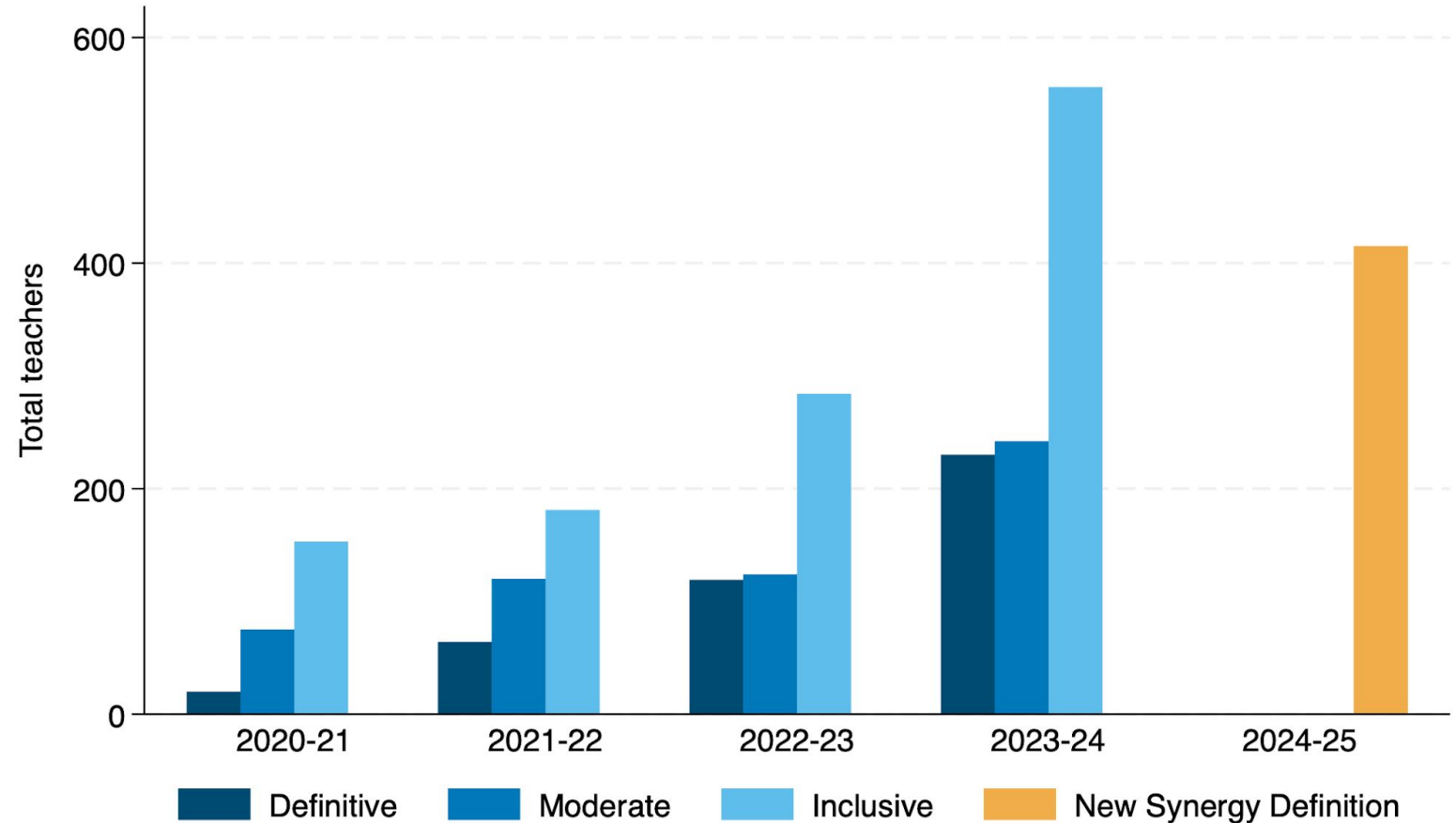
Cc: Intermediary

Thanks for this! When we download the new intervention attendance file you sent, it only has 58 students, all in Program A. Maybe an error in the pull or transfer?

The detail on the roster swap is really helpful context — thank you for flagging it. We will need to think carefully about how to handle those students analytically.

Next Education Workforce

- Who participates?
- Teachers.
- *Which* teachers?



Mid-year presentation

Limitations of this analysis

01

This analysis is not causal.

We cannot confidently say tutoring **caused** any of the differences we see in the data. Rather, we do our best to account for measurable differences across students and **compare the growth of similar students that did and did not receive tutoring.**

02

Our sample is limited.

Each program had roughly 500 students, but our samples were smaller as not every student took both beginning and mid-year assessments. The smaller sample **limits our ability to measure differences precisely**—thus, few differences are statistically significant. It also makes it **more difficult to measure variation** across students (such as differences across schools or grades), as that requires further segmenting the data.

03

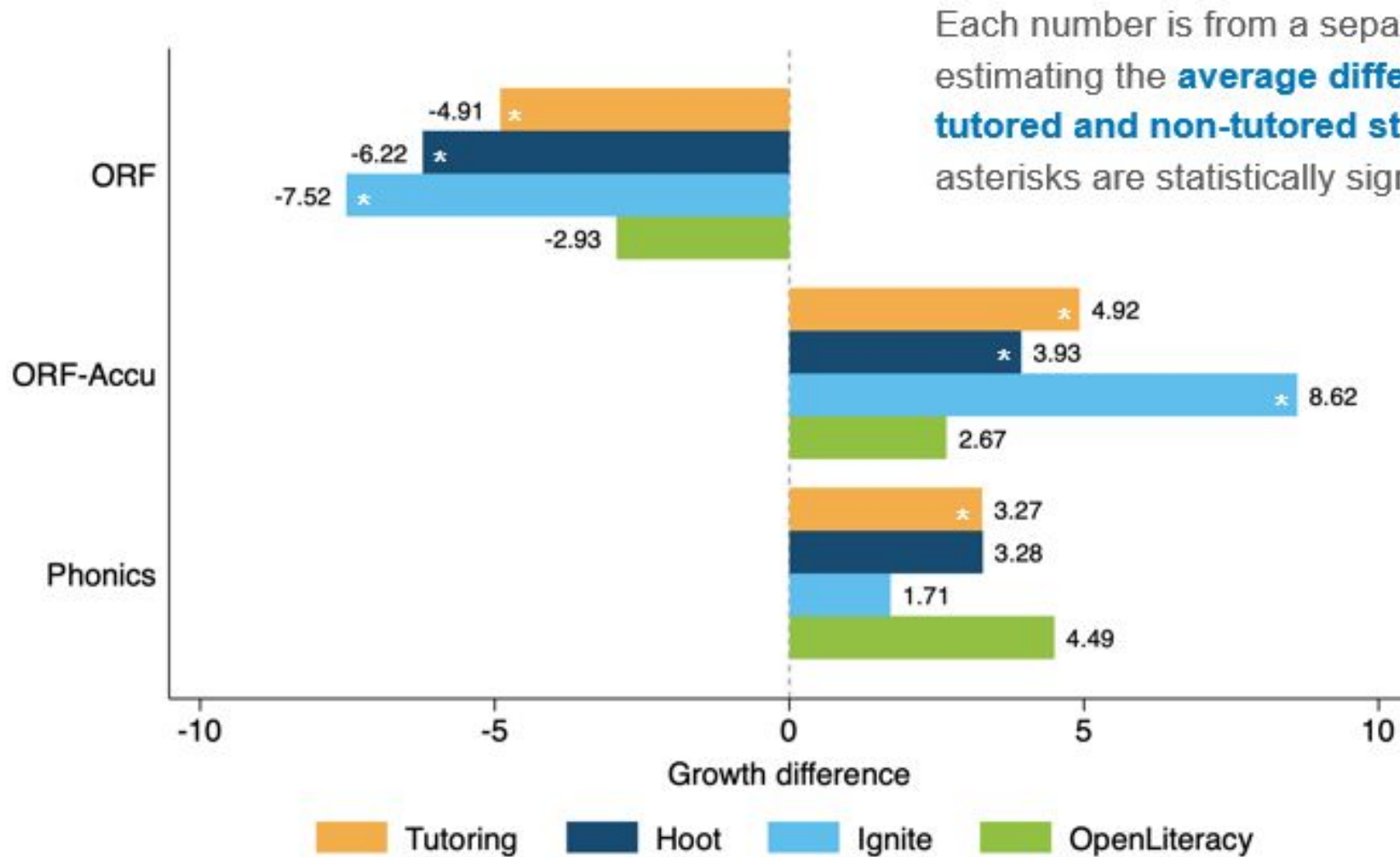
Our time frame is limited.

We examined differences in assessments after roughly four months of tutoring. The average tutored student in our sample received **less than 9 hours of tutoring between assessments.** We expect we will see more growth and have a much more nuanced analysis this summer, when we will have data on a full year of tutoring.

Analytic approach

- We focus on three outcomes: **i-Ready phonics**, **DIBELS ORF**, and **DIBELS ORF-Accu**.
- For each outcome, we measure the **growth** for each student between their mid-year and beginning-of-year assessment.
- We **compare growth of tutoring students and non-tutored students**.
 - We use **controls** to adjust these comparisons, to ensure we are **comparing similar students**.
- We present **differences in growth** from regression models.
 - We present alternate models, including cohort-matched models, in the appendix.
 - Numbers change slightly, but **our overall findings are consistent across model specifications**.

Key finding: Relative growth, accounting for controls



Tutored students saw more growth than non-tutored students in ORF-Accu and i-Ready Phonics, on average.

Ignite students saw the highest relative growth on ORF-Accu, while **OpenLiteracy** students saw the highest relative growth on i-Ready phonics and the closest growth to non-tutored students in ORF.

- **Move forward with Hoot and/or OpenLiteracy**, as their students have the most consistently positive growth on assessments.
 - In phonics, OpenLiteracy excels in earlier grades, while **Hoot** excels in later grades.
- **Learn from Ignite's positive survey results** when scaling.
 - For example, respondents consistently found the data they received from Ignite to be helpful. Other providers could adopt some of Ignite's data sharing practices to garner more in-school support
- **Consider scaling in a way that can support further meaningful analysis.**
 - Continue surveying implementers to understand what is happening on the ground.
 - Offer tutoring slots to students below a certain cutoff or with some specific rules that could support rigorous analysis in the future.
 - Consider keeping current assignments where possible (ie, Hoot students stay in Hoot next year) to learn more about multi-year exposure.

We look forward to furthering this work in the coming months.

Spring 2025

- Share findings with tutoring providers
- Continued support of OUSD and ELP in decision-making and communicating results
- Further analysis on variation by dosage and across schools and students

Summer 2025

- Incorporate end-of-year assessments into analysis
- Final published report

OUSD's mid-year response

Click [here](#) for a simulated email exchange between CRPE researchers and OUSD staff.

Reflection questions

1. Where do you see trust being built and/or tested in this exchange?
2. How do the parties show they trust each other's capacity and/or acknowledge its limits?
3. When “the work” doesn’t match what was expected, how do they handle the misalignment? What does that tell you about trust?
4. How openly do they acknowledge uncertainty or limitations, and why might that matter for trust?
5. If trust broke down here, what would these emails look like instead?

Quick Stretch Break

3:15-3:17

Part II: March 2025-July 2025

3:17-3:45

Problem transition

- The previous exchange shows a partnership negotiating what to study.
- The next exchange shows what happens when the actual findings arrive.
- The question shifts from "what should we study?" to "what story do these numbers tell?"

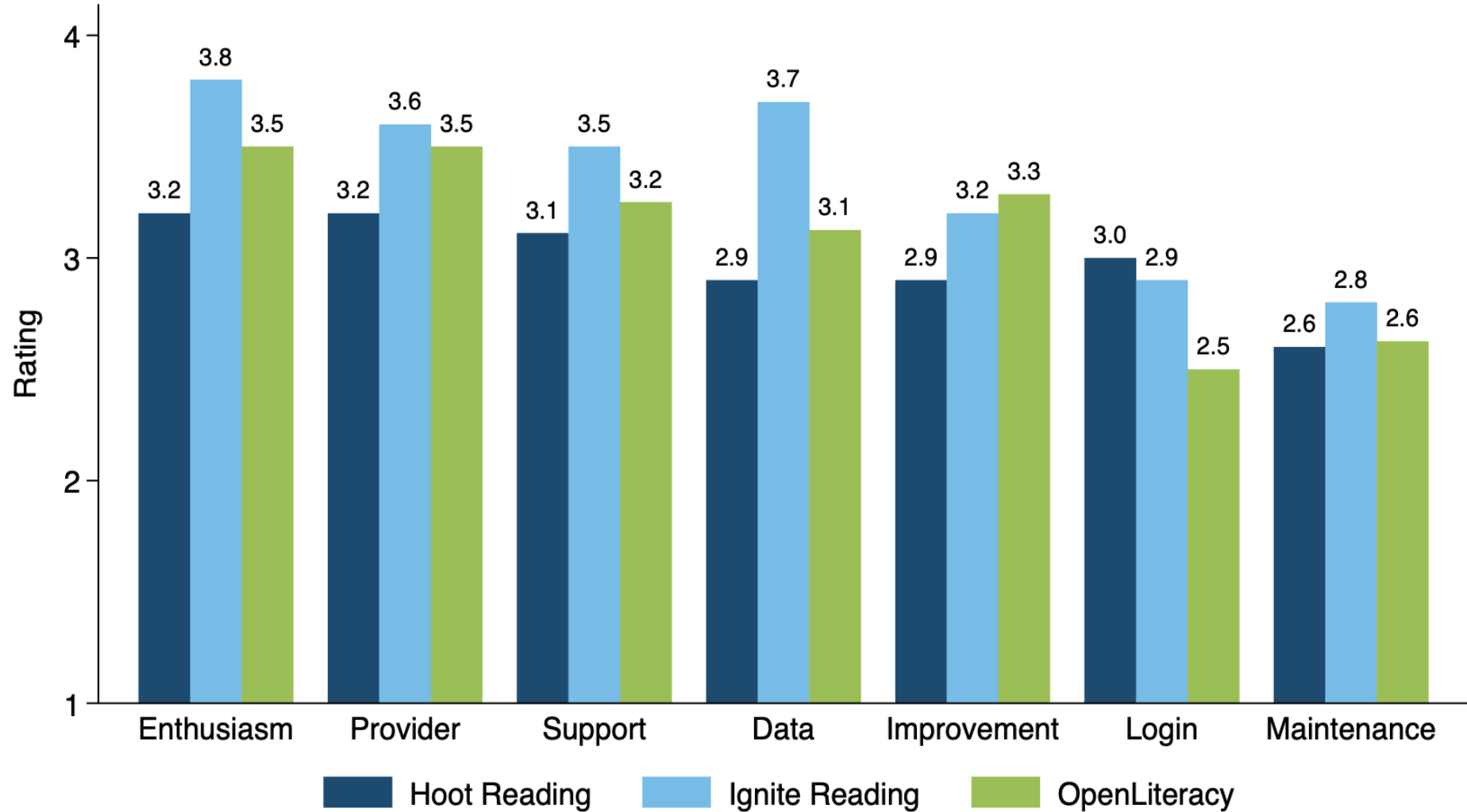
Reflection questions

- What evidence reflects how decisions were being made?
- How are competing views about what the data actually show visible in this exchange, and what does each party seem to prioritize?
- Where do you see alignment between what the research team is comfortable claiming and what the district leader wants to claim?
- Does the district's early decision to exclude certain sessions reflect shared problem-definition, or unilateral choice?
- How openly do the parties acknowledge limits or uncertainties in the data?
- If trust broke down here, what would these emails look like instead?

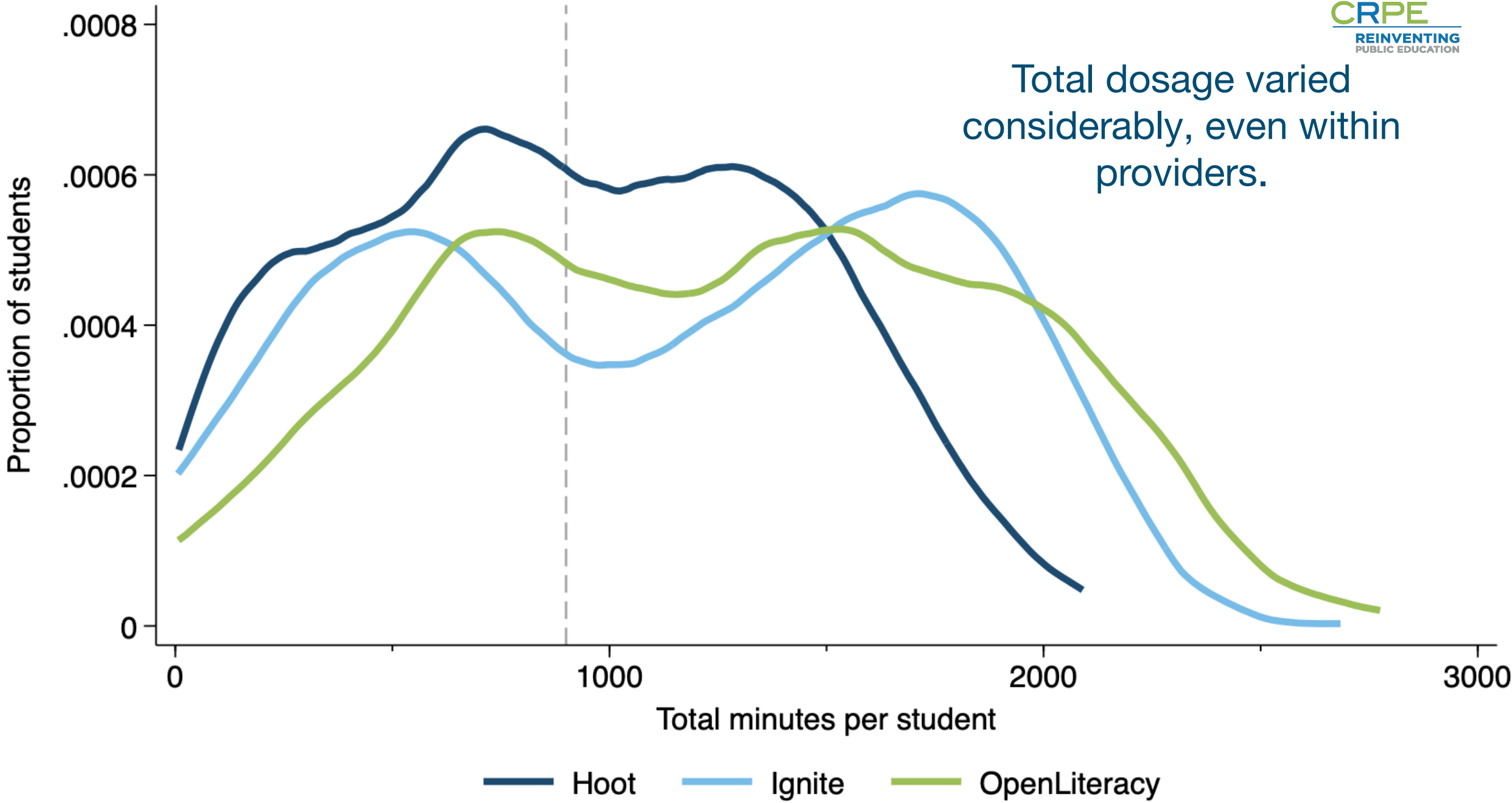
End-of-year results

School feedback

- Schools were generally enthusiastic.
- They reported scheduling and technology issues.

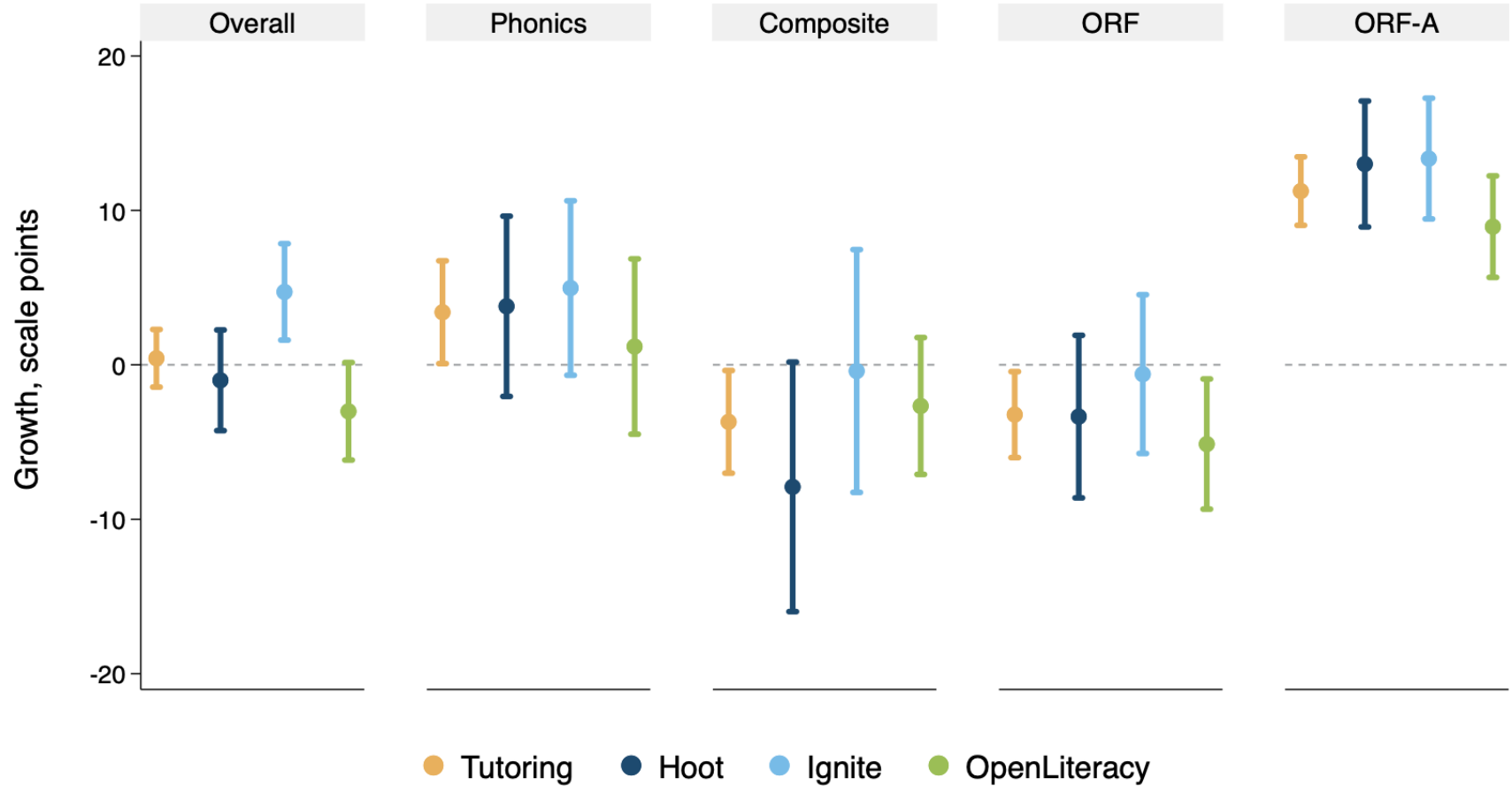


Total dosage varied considerably, even within providers.



Student growth

- Tutored students grew more in phonics and accuracy but not fluency or broader measures.
- Differences between programs were generally small, with one notable exception.



Conclusion

3:45-4:00

OUSD's response to final results

- Their focus was on how to accurately communicate findings and make them and useful for high-stakes decision-making.
 - For example, they requested:
 - More explanation of regression modeling
 - Clearer explanation of what each outcome actually measures
 - Conversions of effects to “months of learning”
- However, they also wanted to influence positioning and framing.
 - They wanted certain results in the executive summary and others in the appendix. They also again asked if it was fair to identify a certain program as the strongest provider.

Initial timeline for final report

- Grant ending July 31
- Meeting with funder to review findings & comms plan July 30
- Current plan: report out by early August
 - Feedback on deck by July 25
 - PI and RA respond to feedback July 28-August 1
 - PI walk through deck/findings with funder July 30
 - PI aim to have near final report by July 31/Aug 1
 - OUSD & funder Final review by Aug 4-6
 - Copy edit, layout, etc August 6-8 Publication targeted for August 11-15

August 20, 2025

From: Intermediary

To: Research Team

Cc: District Partners, Funder

Thank you for working on a cleaned up report to share with the providers! The findings are helping us think through the framing we want to use to communicate headlines from the evaluation, and they're helping us consider how the study sits in the broader body of research on high dosage tutoring.

Again, thank you so much for your thoughtfulness and extra time on this project!

November 16, 2025

From: Funder

To: Research Team

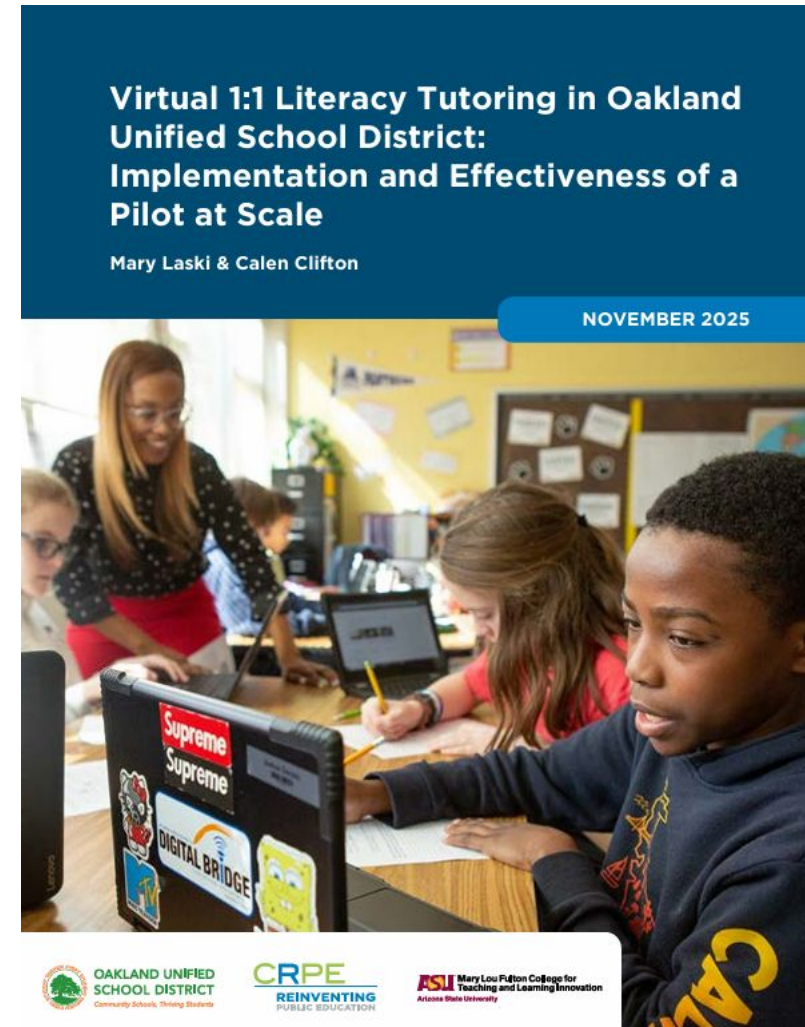
Cc: District Partners, Intermediary

I appreciate all your work to address questions from our team and concerns from the providers. This process has been an education for me in how sensitive this can be for providers. I'm deeply appreciative of the more complex understanding we now have of high dosage tutoring through this evaluation.

Thank you for the work you've done and for your thoughtfulness in navigating these conversations. We appreciate your efforts on this evaluation report and want to affirm that it has been useful for our key internal audiences.

Final report

- Our evaluation report was published in December!
- The core findings were identical to those presented in July-August.
- Further feedback from OUSD and the providers motivated additional analyses and other informative appendix material.



Next Education Workforce

- What are the results?
- Depends.
- Which results are “clear?”



Calen's final thoughts on trust

Trust is expressed in small pieces.

It is tested often.

And it solidifies when we acknowledge uncertainty together.

1. Name your own mistakes.
2. Ask clarifying questions.
3. Stay engaged during disagreement.
4. Fulfill responsibilities promptly.
5. Express genuine gratitude.

Open Q&A and Reflection



Thank You