Virtual 1:1 Literacy Tutoring in Oakland Unified School District: Implementation and Effectiveness of a Pilot at Scale

Mary Laski & Calen Clifton

NOVEMBER 2025

OAKLAND UNIFIED SCHOOL DISTRICT
Community Schools, Thriving Students

CRPE
REINVENTING PUBLIC EDUCATION

ASU Mary Lou Fulton College for Teaching and Learning Innovation
Arizona State University

EXECUTIVE SUMMARY

In response to dramatic COVID-19 learning loss and growing concerns about a "literacy crisis," many education agencies have turned to tutoring to provide students with additional support. In-person tutoring has proven to be effective, but can be logistically difficult to scale. Virtual tutoring could mediate these difficulties, and recent experimental work demonstrates that virtual tutoring can improve student learning. However, experimental studies rarely replicate the real-world experiences of school and district leaders, who are often navigating competing interests, choosing between multiple providers, and deciding how best to allocate limited resources.

This study examines a virtual literacy tutoring pilot implemented in the Oakland Unified School District (OUSD), designed to accelerate early literacy gains among emerging readers. OUSD partnered with three providers to offer a limited number of high-dosage virtual, 1:1 tutoring slots to schools throughout the district. OUSD chose these partners in collaboration with district literacy staff and in alignment with existing research on effective tutoring practices. Each school was assigned a provider, and OUSD directed providers to focus on phonics. OUSD asked schools to target below-grade-level students and gave specific guidelines for participation. School staff exercised some discretion over selection. The district then collected data on tutoring implementation and student literacy development. We assess how the district implemented tutoring and estimate the effectiveness of the pilot on student growth across several literacy outcomes. Our analyses focus on tutoring overall, and we note meaningful differences between providers when relevant.

We find that:

- **Implementation fidelity was high.**
  - Schools generally followed district guidelines for student selection.
  - Students received relatively high doses of tutoring. Over 80% of tutored students received at least 20 sessions. 58% received more than 900 minutes (15 hours) of tutoring.
  - Implementing schools were generally satisfied with tutoring provid

- **Tutored students made stronger gains than similar non-tutored students in foundational reading skills like phonics and accuracy.**
  - On the i-Ready Phonics assessment, tutored students grew significantly more than their peers—equivalent to 1.3 additional months of learning.
  - On the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral

* Cover photo courtesy of Oakland Unified School District

Reading Fluency-Accuracy assessment, tutored students grew much more than their peers—about 9 months of additional learning. However, this estimate is based on a smaller subsample of tutored students and should be interpreted with caution.

- **Tutored students grew in fluency and overall literacy development, but these gains were not significantly greater than expected.**
  - On the DIBELS Oral Reading Fluency assessment, tutored students saw slightly less growth than their peers.
  - On the i-Ready Overall assessment, tutored students overall grew just as much as their non-tutored peers.

- **Higher tutoring dosage was associated with greater gains.**
  - Students who received at least 900 minutes (15 hours) of tutoring saw larger gains than students with less tutoring. On i-Ready Phonics, these students saw gains equivalent to 1.6 additional months of learning.

- **Differences across providers were generally minor**, with one important exception: students in Ignite Reading performed significantly better than their peers on i-Ready Overall, with growth equivalent to 1.5 additional months of learning.

These associations between tutoring and foundational literacy skills are promising and meaningful, particularly given Oakland's unique context. Significantly greater literacy gains are notable given that the tutoring in this study was both **virtual** and implemented **at scale** throughout the district—conditions often associated with diminished tutoring effects. Further, the average tutored student received about 18 hours of tutoring over the course of the year. **These 18 hours translated to 1.3 additional months of learning in phonics.**[1] In short, Oakland's pilot may provide a rare real-world, non-experimental example of effective high-dosage tutoring at scale.

1. Months of learning are calculated using the "Rule of 27," assuming roughly three grade levels per standard deviation and nine months in a school year. As such, a change measured in standard deviations can be multiplied by 27 to roughly estimate months of growth: 1 standard deviation is 27 months, or 3 years, of learning. The vertical link that underlies this rule is detailed further here.

We note three key limitations to the current work:

- **The analysis is not causal.** Tutoring assignments were not random. We account for measurable differences across students, but our effect estimates remain descriptive. Unmeasured differences between tutored and non-tutored students may still influence the results. These results tell us whether tutored students experienced greater gains than non-tutored students, not whether tutoring independently led to those gains.

- **Our sample is limited and differs across outcomes.** Each tutoring provider had roughly 500 students, but our analytic samples are smaller, as not every student took both beginning- and end-of-year assessments. This is especially true for DIBELS, where testing was inconsistent across schools. We encourage caution when generalizing our findings, especially on the DIBELS measures. The Technical Appendix includes a table noting differences across samples

- **Our understanding of implementation, including the experiences of non-tutored and tutored students, is limited.** This analysis utilizes one year of administrative data collected by OUSD and a brief school-level survey. We are limited in what we can measure: for example, we do not know which (if any) students, tutored or non-tutored, participated in other interventions, including other in-school tutoring programs. Future work would benefit from additional quantitative and qualitative data to better understand implementation on the ground, as well as continued work to estimate long-term benefits.

CONTEXT: VIRTUAL 1:1 LITERACY TUTORING IN OAKLAND

OUSD piloted three different virtual 1:1 literacy tutoring programs in elementary schools throughout the district: Hoot, Ignite Reading, and OpenLiteracy. After an initial "mini-pilot" of 18 schools in 2023-24, tutoring expanded to 33 schools in 2024-2025 for "Pilot 2.0," examined in this study. The district assigned each provider 11 schools; providers offered roughly 500 tutoring seats across their assigned schools. Schools selected students based in part on district guidance, detailed below. Table 1 provides sample sizes for the participation and assessment score samples. Note that the total number of tutored students in each program exceeds 500 because schools reassigned some seats to different students over the course of the year.

All three tutoring programs advertise explicit instruction in the foundational skills of phonics, fluency, and decoding, aligned with the district's overall literacy vision. All three programs also feature built-in assessments for progress monitoring. However, anticipated session frequency and duration differ notably across models. Each week, Hoot, Ignite Reading, and OpenLiteracy intend to deliver three 25-minute sessions (75 minutes), five 15-minute sessions a week (75 minutes), and three 30-minute sessions (90 minutes), respectively. Hoot and OpenLiteracy employ teachers with classroom experience, while Ignite internally certifies its own tutors. OUSD encouraged providers to focus on phonics, the knowledge of letter-sound relationships that allows readers to decode unfamiliar words.

Table 1: Regression sample sizes

		i-Ready				DIBELS			
		Overall		Phonics		Composite		ORF & ORF-A	
	Total tutored	Non-tutored	Tutored	Non-tutored	Tutored	Non-tutored	Tutored	Non-tutored	Tutored
Hoot	592	10779	494	6296	459	721	195	1876	281
Ignite	705	10780	606	6298	557	716	218	1875	303
Open Literacy	610	10780	559	6299	522	720	212	1878	333
Tutoring	1907	10781	1659	1659	1538	727	625	1883	917

Note: Cells include the number of individual students in each subgroup included in our analysis regressions. In other words, these are students with beginning- to end-of-year growth scores that do not have missing values for any of our included controls. The non-tutored subsamples include all students in the same grade as tutored students.

Oakland's Multi-Tiered Systems of Support (MTSS) incorporates tutoring, with an explicit push to align Tier 3 literacy instruction throughout the district. OUSD encouraged schools to implement tutoring during the school day but outside of core instructional times, though schools had flexibility in their scheduling. Schools were encouraged to select students to receive tutoring based on specific district criteria, but also had discretion in their selection. OUSD directed schools to prioritize first- and second-grade students, as well as students who were two or more grade levels behind on i-Ready Phonics (or one grade behind for first

graders). The district suggested that 1) at least 50% of slots should go to these priority students in first and second grade, and 2) up to 10% of tutoring slots should go to students with individualized education programs (IEPs). Each school designated a program liaison to manage tutoring logistics and communicate with providers.

ANALYSIS

Data

We use anonymized data provided by OUSD. To identify tutored students, we use data on individual tutoring sessions attended by each student, including the date, start time, and end time for each session, which each tutoring provider shared with the district. We link this attendance information to administrative data from the district, including student demographics and classroom assignments, as well as scores on i-Ready and DIBELS assessments.

We look at three key outcomes of interest:

1. **Tutoring assignment and dosage**. To assess how tutoring was implemented, we look both at how tutoring slots were assigned to students and the frequency and duration of tutoring sessions, based on detailed data on individual tutoring sessions.

2. **Satisfaction of school-level providers.** In late fall 2024, OUSD surveyed school-level staff supporting tutoring implementation, assessing their satisfaction with tutoring and ease of implementation. We use this survey data to compare in-school experiences of tutoring between providers about halfway through the school year.

3. **Student assessment data**. To measure the relative effectiveness of tutoring programs, we study students' growth on i-Ready and DIBELS assessments, using regression models to compare tutored students' within-year growth to that of similar non-tutored students. For both of these assessments, we look at both broad measures of growth in overall literacy and specific measures of growth in foundational literacy skills. For DIBELS, we focus on Oral Reading Fluency ("ORF") and Oral Reading Fluency-Accuracy ("ORF-A"). For i-Ready, we focus on Phonics. We selected these subtests in dialogue with OUSD because they most closely track the stated goals of the tutoring initiative.

Methods

For outcomes (1) and (2), we examine simple descriptive differences across tutoring providers. For outcome (3), we use multiple regression models to estimate overall and provider-specific differences between tutored and non-tutored students at the student level. Given OUSD's desire to accelerate learning for tutored students, our outcomes of interest are beginning- to end-of-year growth scores on specific literacy outcomes from i-Ready or DIBELS, defined as the scale point difference between a student's spring and fall semester scores. These assessments allow us to define student-specific learning rates as scale point differences during the school year. This approach measures each student's progress relative to their own prior achievement to determine whether tutored students made more progress than they would have without tutoring, based on comparisons with similar non-tutored students.

Models estimate differences in growth between tutored and non-tutored students or a specific program's students and non-tutored students. They tell us whether students grew *more or less during the year* than similar non-tutored students. Tutored students from other programs are omitted from the comparison group of non-tutored students used to estimate program-specific effects.

All regressions include classroom fixed effects, which inherently account for differences between schools, grades, teachers, and other important classroom resources, including peers and instructional materials. We also include controls for students' school attendance rate and demographics (race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage). Because tests were administered at different times across and within schools, we also include a control for the days between assessments to ensure that any observed tutoring gains are not simply because some students experienced more instruction between assessments. We include all students who ever received at least one tutoring session from any provider, but our results do not meaningfully change if we limit them to students with at least twenty sessions over the course of the year (see additional models included in the Technical Appendix). Students averaged at least 60 sessions across outcome measures. The Technical Appendix includes additional information on our modeling decisions.

Limitations

We emphasize a few key limitations of this work. First and foremost, this analysis is not causal. We cannot confidently say that tutoring caused any of the differences we see in the data. Rather, we measure associations between tutoring and outcomes of interest, after accounting for measurable differences across students. We are limited in what we can measure: for example, we do not know which (if any) students, tutored or non-tutored, participated in other interventions. There may also be inherent differences across tutored and non-tutored students that are unmeasured. As an example, it is possible that students were selected into tutoring based on perceptions of potential growth; OUSD generally emphasized selection among students who "needed the most support" to receive tutoring seats. Any remaining differences between tutored and non-tutored students could bias our results.

Second, our sample is limited and differs across outcomes. Table 1 presents basic sample information. Each program had roughly 500 students, but our analytic samples are often smaller as we cannot calculate growth for students without both beginning- and end-of-year assessments. We hesitate to draw sweeping conclusions across measures because the samples vary meaningfully. This is especially true for DIBELS, where testing was inconsistent across schools, but more likely for tutored students. Smaller samples also limit our ability to measure differences precisely—our estimates often have quite wide confidence intervals—and make it more difficult to measure variation across student groups, providers, or grade levels, as this would require further segmenting the data.

Finally, our understanding of implementation is limited. We use session-level data to describe tutoring frequency and duration, but do not otherwise know "what tutoring looks like" within schools. We capture school-level tutoring perceptions from a survey given to one person per school at a single point in the fall semester. Future work would benefit from additional qualitative data to better understand implementation as it unfolds within schools.

Understanding Differences Across Measures

Literacy development is multifaceted. We look at two measures of general literacy (i-Ready Overall and DIBELS Composite) as well as multiple assessments of specific skills:

- **DIBELS Oral Reading Fluency ("ORF")** measures the total number of words read correctly in a grade-level text in one minute
- **DIBELS Oral Reading Fluency-Accuracy ("ORF-A")** measures the percentage of words that students read correctly in a grade-level text in one minute
- **i-Ready Phonics** measures broader phonics skills in subdomains of alphabetic knowledge, sound-spellings, and decoding

Both i-Ready and DIBELS contain measurements of other literacy skills. These other subtests factor into each assessment's overall measure of general literacy but are unexplored here, given OUSD's goals and tutoring's explicit focus on foundational skills.

In our sample, growth measures within tests are somewhat related. However, correlations between measures from different assessments are generally weak to moderate, suggesting that these assessments define literacy in different ways. We find that growth in DIBELS Composite is strongly related to growth in ORF, but only weakly related to growth in ORF-A. In other words, overall literacy development in DIBELS is driven more strongly by a reader's ability to pronounce words with accuracy *and* speed (fluency), rather than just accuracy. The correlation between ORF and ORF-A is small and negative, suggesting that improvements in accuracy could be slightly detrimental to fluency by negatively affecting reading speed, at least in the short term. Overall i-Ready growth is closely related to growth in i-Ready Phonics. Neither DIBELS Composite nor i-Ready Overall should be interpreted as pure measures of comprehension. These measures define overall reading proficiency as an aggregate of skills including but not limited to comprehension. See Technical Appendix for a table of correlation coefficients.

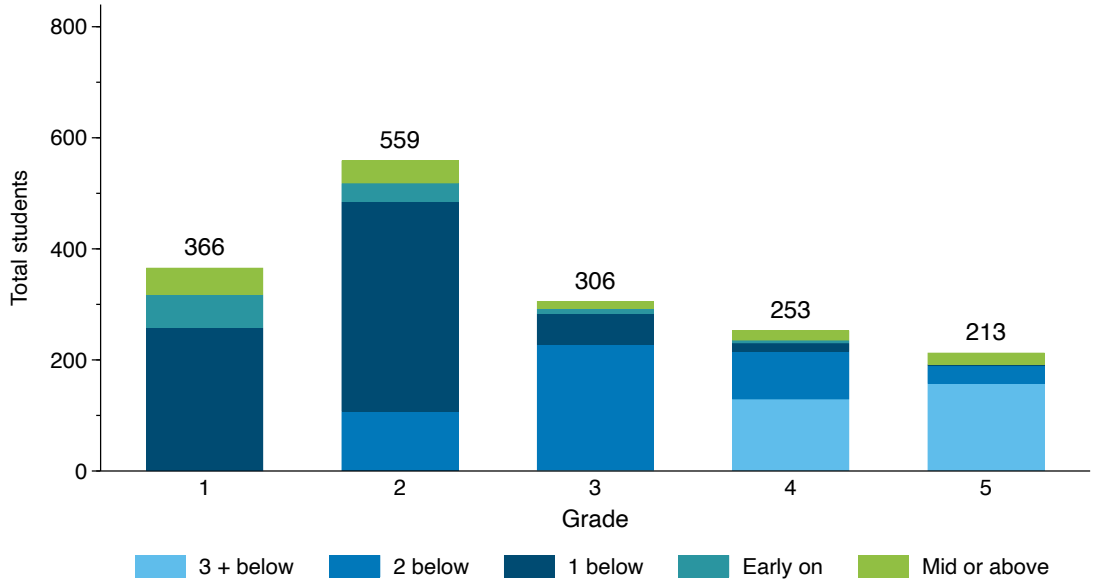
Literacy skills are related but not always developed in simple sequential order. For example, accuracy is a component of fluency, which, along with phonics and other related skills, is a necessary precursor to overall literacy. Phonics also supports fluency by helping students use letters to decode unknown words, increasing accuracy. As such, we might expect students reading below grade level—like those recommended for tutoring—to experience gains in accuracy (ORF-A) and/or phonics *before* gains in fluency (ORF) or overall reading proficiency. Our findings are suggestive of this progression, which OUSD can further study as tutoring continues in future years.

IMPLEMENTATION FINDINGS

Implementation fidelity was high overall, as evidenced by three key findings.

Finding 1: Schools assigned tutoring slots to students as the district intended.

Figure 1: Prior i-Ready Phonics Relative Placement for students receiving tutoring during 2024-2025, by grade



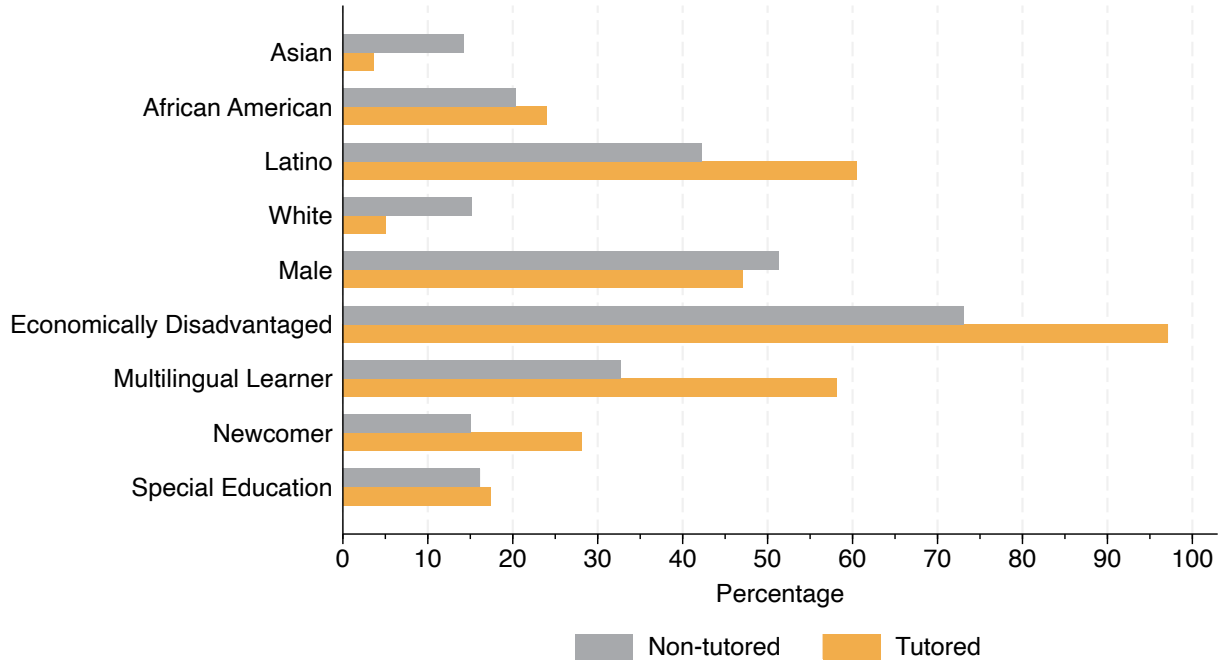
Note: Stacked bars display i-Ready Phonics Relative Placement during spring 2024 for all students, by fall 2024 grade, where the data are available.

To assess implementation, we first explore which students received tutoring in OUSD. Recall that the district suggested prioritizing first and second graders and students who were two or more grade levels behind on i-Ready Phonics (or

one grade level behind for first graders) when assigning tutoring slots. Figure 1 displays the grade-level and prior year i-Ready relative placements of all students receiving tutoring in 2024-2025. As intended, half (53%) of tutored students were in grades 1 and 2, and over half (58%) of students met OUSD's criteria for 2023-24 i-Ready Phonics performance. Overall, 84% of tutored students demonstrated below-grade-level phonics mastery.

Given the district's intention to focus on early elementary students with below-grade-level skills, we also note that tutored students are demographically different from non-tutored students in the same grades. Figure 2 plots demographic characteristics of tutored students and all non-tutored students in the same grades.

Figure 2: Demographics of tutored and non-tutored students



Note: Non-tutored sample includes all non-tutored students in grades 1 through 5 in OUSD. Bars compare demographics between non-tutored and tutored students receiving at least one session from any provider. We use the demographic terms that OUSD uses in their internal data.

We note meaningful differences across the tutored and non-tutored samples. Relative to non-tutored students, tutored students were:

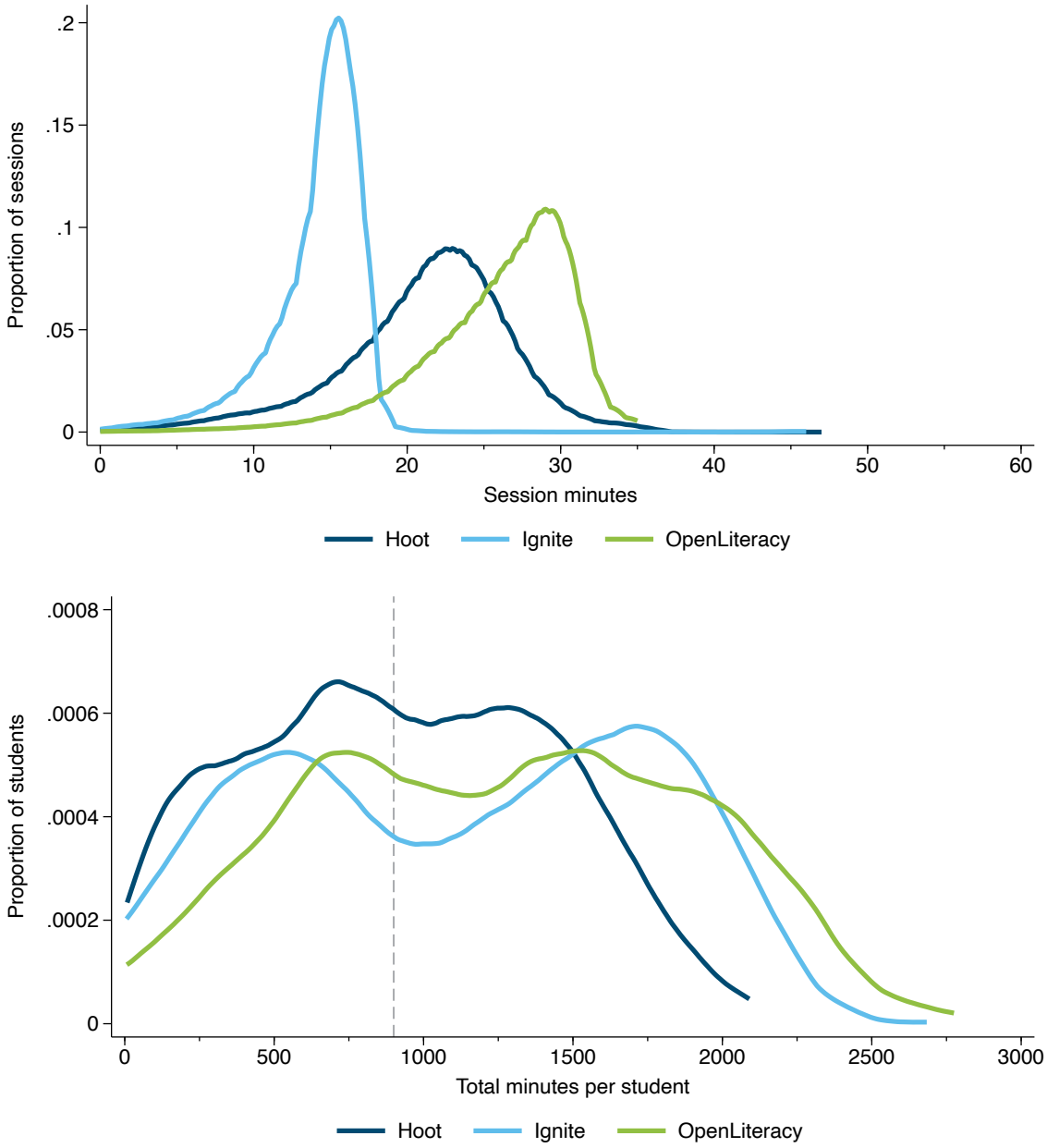
- More likely to be African American or Latino.
- More likely to be economically disadvantaged.
- More likely to be multilingual learners.
- More likely to be identified as a Newcomer to the United States.[2]

2. We use the demographic terms that OUSD uses in their internal data.

Overall, tutoring participants were generally more likely to be from traditionally disadvantaged backgrounds. We also note that tutoring students were about as likely to qualify for special education services as non-tutored students, which again aligns broadly with the district's intended prioritization of student slots.

Finding 2: Students received relatively high doses of tutoring, though there is meaningful variation across students and providers.

Figure 3: Session minutes and total minutes, by provider



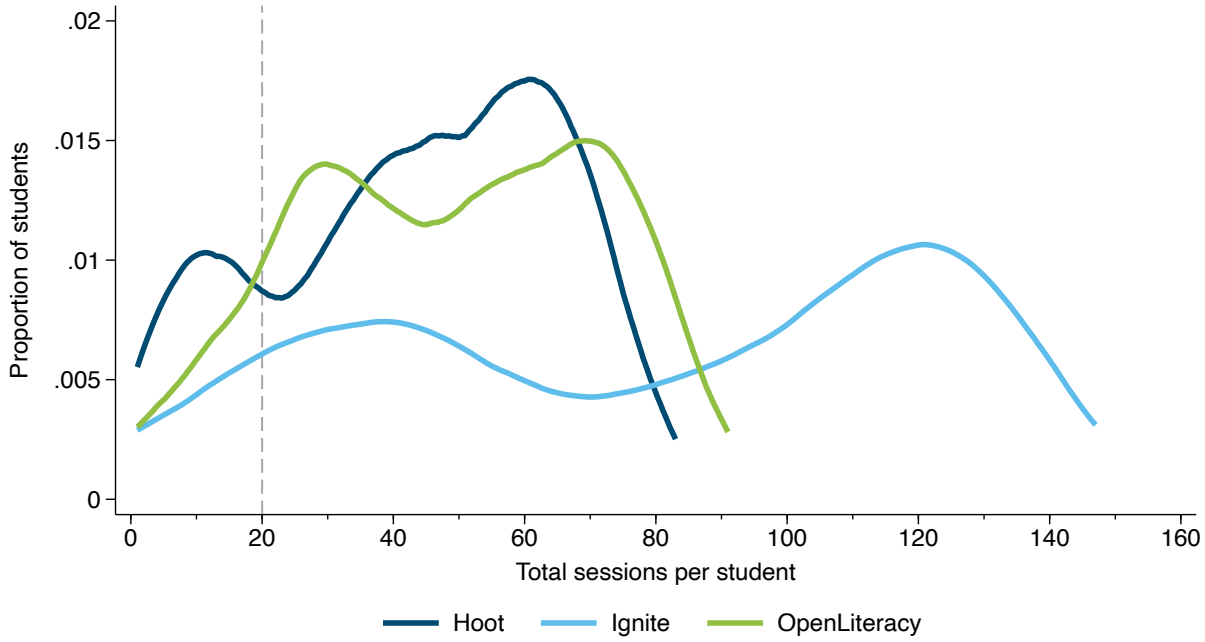
Note: Panels summarize dosage by provider, according to session length (top) and total (bottom) minutes per student. The top panel gives the proportion of sessions at each observed session length, in minutes, while the bottom panel gives the proportion of students who ended the year with a given level of total tutoring minutes. The dashed vertical line at 900 minutes marks the district's minutes goal.

Figures 3 and 4 display the distribution of tutoring minutes and sessions by provider. In the top panel of Figure 3, we see that the average session length varies by program, as expected. Ignite sessions are, on average, much shorter than OpenLiteracy sessions, aligning with program expectations. That said, there is also notable variation in session lengths even within providers: for example, many OpenLiteracy sessions are less than 20 minutes. Variation in session duration could be due to many factors; the data collected for this study does not allow us to definitively explain the reasons for this variation.

The bottom panel of Figure 3 presents the total number of tutoring minutes received by each student. On average, OpenLiteracy students spent the most minutes in tutoring (1,252), greater than Ignite (1,121) or Hoot (916). But there is, again, substantial variation across students: the full distribution of tutoring experiences ranges from 1 session and 6 minutes to 147 sessions and 2,775 minutes (roughly 47 hours).

Figure 4 displays the total number of sessions each student received across tutoring programs. Even within the same program, students received very different numbers of sessions. Ignite had the highest number of sessions on average, as we would expect given its stated goal of everyday sessions, but also had the widest range in the number of sessions between students.

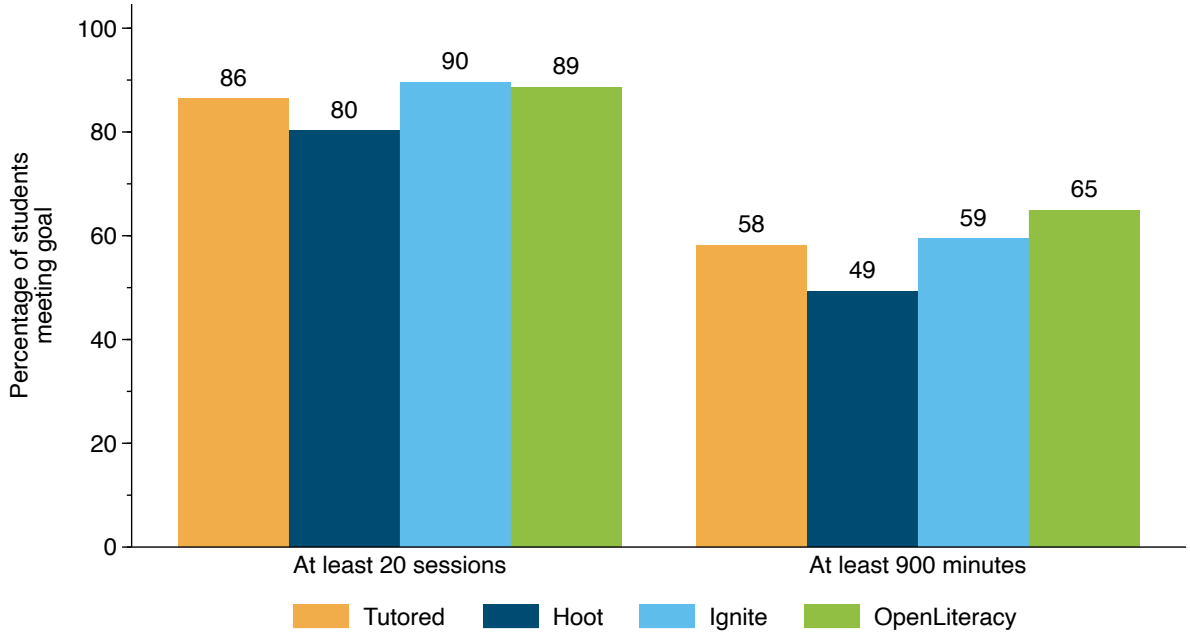
Figure 4: Sessions per student, by provider



Note: Figure presents the proportion of students who ended the year with a given number of total sessions, by provider. The dashed vertical line at 20 sessions marks the district's session goal.

Finally, Figure 5 plots the percentage of tutored students who received at least twenty sessions (left) and those who received at least 900 minutes of tutoring (right). Note that the orange bar includes all tutored students and is essentially a weighted average of the provider-specific bars. Using these thresholds, dosage is generally high: the majority of students (86%) received at least twenty sessions over the course of the year. Additionally, over half of students (58%) met the informal benchmark of at least 900 minutes (15 hours) of tutoring over the year.

Figure 5: Percentage of tutored students meeting district dosage goals, by provider



Note: Figure displays the percentage of students meeting OUSD's dosage goals of 1) at least 20 sessions or 2) at least 900 minutes, by provider. Students may meet one goal without meeting the other; the figure does not display the percentage of students meeting both goals.

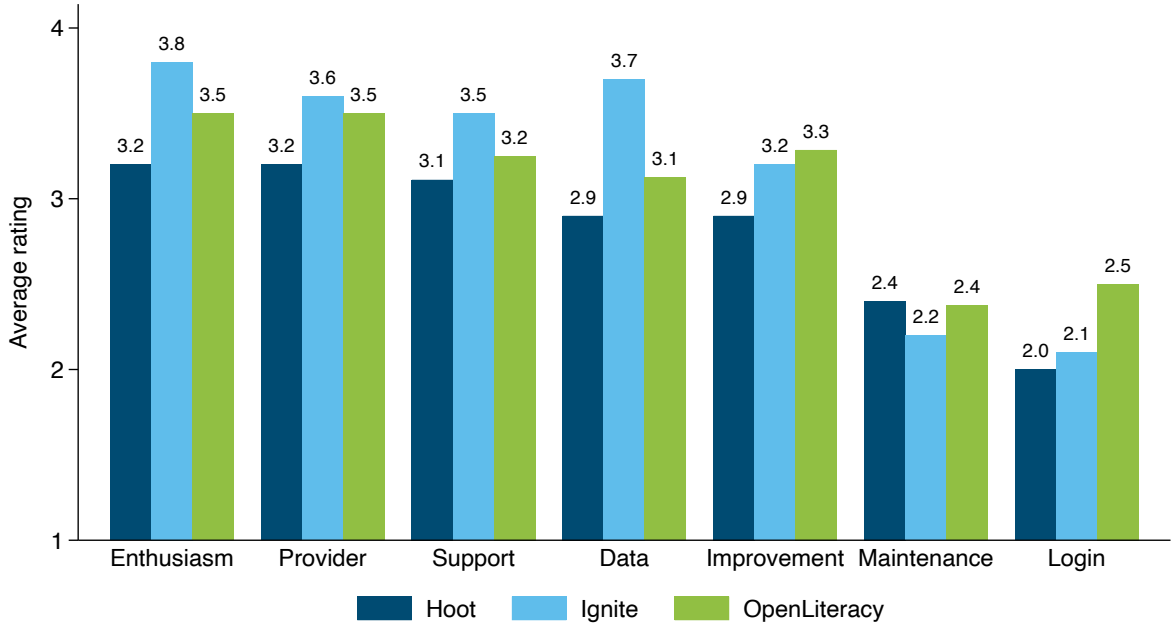
Finding 3: Implementing schools were generally satisfied with tutoring providers.

To further assess implementation, OUSD conducted a brief survey of teachers on special assignment (TSAs) who supported tutoring implementation at schools. Survey responses were submitted in November and December 2024, when tutoring was still somewhat fresh, which likely influenced the results. It is possible that respondents' opinions towards providers changed over the course of the school year, but we only have this single snapshot.

Figure 6 plots average response rates across providers. Survey questions are available in the Technical Appendix. Note that for maintenance and login, lower scores are better. We find that satisfaction is relatively high across

the board. Respondents reported noticeable gains in student motivation ("Improvement") and felt that tutoring was supporting student success in the classroom ("Support"). However, in open-ended responses, nearly a third of respondents described scheduling challenges and indicated that at least some of their students attended tutoring during core instructional times. Some schools also reported difficulties with laptop and platform technology and mentioned persistent log-in challenges, especially at the beginning of the year. Teachers generally managed program delivery within schools, but other staff, such as librarians, teacher assistants, and volunteers, were included as well. Differences across programs are small overall, but respondents were generally most enthusiastic about Ignite ("Enthusiasm") and found the data Ignite provided to school sites the most helpful ("Data").

Figure 6: Mid-year implementation survey results, by provider



Note: Data are from Tier 3 Tutoring "Teacher on Special Assignment" (TSA) Survey. Figure plots average responses on a scale of 1-4. Note that for maintenance and login, lower scores are better. See Technical Appendix for corresponding survey questions. N = 28 respondents. Survey given during November-December of 2024.

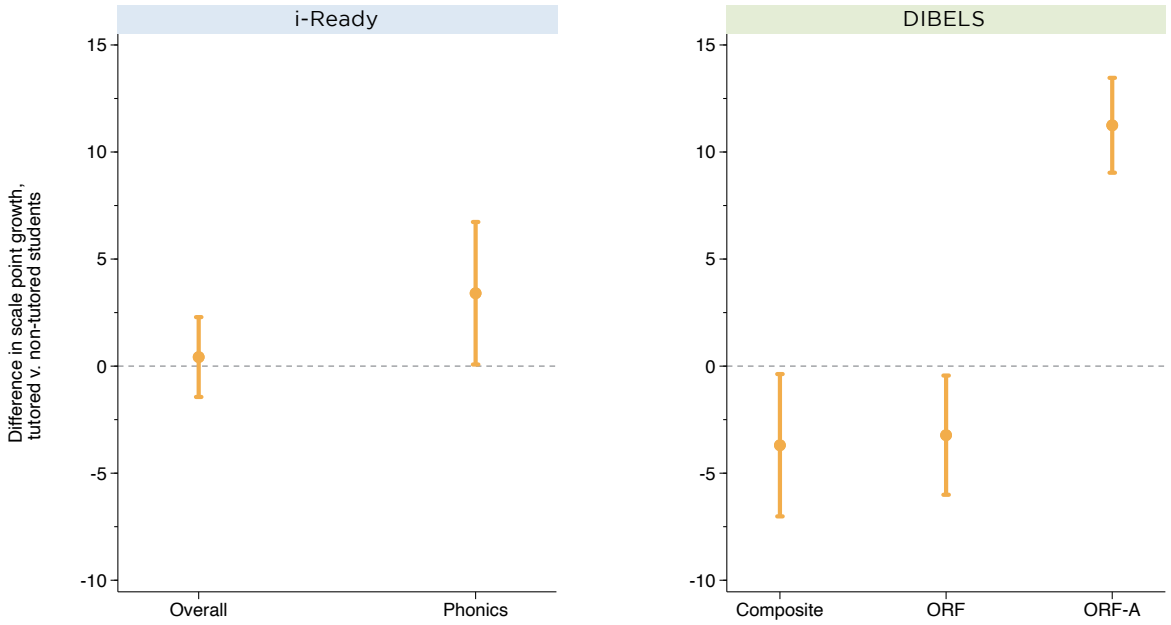
EFFECTIVENESS FINDINGS

Finding 4: Tutored students made stronger gains than their peers in foundational reading skills like phonics and accuracy. Tutoring was not associated with significantly greater gains in fluency or overall literacy development.

We now turn to differences in growth between tutored and non-tutored students. To assess this, we run regression models that account for differences across students, as described in the Methods section. Descriptive differences in overall growth, not accounting for systematic differences between tutored and non-tutored students, are available in the Technical Appendix.

Figure 7 presents estimated differences in growth across all outcomes of interest. Each dot represents the coefficient from a separate regression, estimating the average difference in growth between similar tutored and non-tutored students. A positive value means that, on average, tutored students experienced more growth during the year than similar non-tutored students. A value of 0 does not mean that tutored students did not grow. Rather, it indicates that tutored and similar non-tutored students experienced the same gains. The vertical lines above and below each dot represent each estimate's confidence interval, or the range of plausible values for the estimated average effect. Confidence intervals that do not include 0 are statistically significant.

Figure 7: Differences in growth between tutored and non-tutored students



Note: Each dot represents a coefficient from a separate regression. The numbers are estimated differences in growth between tutored and non-tutored students. The vertical bars are the 95% confidence interval. All regressions include fixed effects for classroom and controls for students' school attendance rate, race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage. Ns vary by outcome. N = 12,440 for i-Ready Overall; 7,841 for i-Ready Phonics; 1,352 for DIBELS Composite; and 2,800 for DIBELS ORF and ORF-A.

First, we find that **tutored students saw significantly more growth in foundational reading skills like phonics** (i-Ready Phonics) **and accuracy** (DIBELS

ORF-A), **but not in fluency** (DIBELS ORF). On i-Ready Phonics, tutored students grew about 3.6 points more than their peers, or about 0.05 standard deviations. **This is a meaningful difference, equivalent to roughly 1.3 additional months of learning on phonics**.

The accuracy difference is even larger: tutored students grew about 11.2 points more than their peers on DIBELS Oral Reading Accuracy, which is equivalent to 0.35 standard deviations. **This is an exceptionally large difference, equivalent to roughly 9 additional months of learning in accuracy**. However, tutored students overall experienced significantly less growth than their non-tutored peers in fluency, though the relative difference is much smaller. We emphasize that the DIBELS samples are much smaller than the i-Ready samples and should be interpreted with caution.

Second, we find that, on average, **tutoring was not associated with significantly greater gains in overall literacy development.** On i-Ready Overall, tutored students grew just as much as comparable non-tutored students: the difference in growth is close to zero and not statistically significant. On the DIBELS Composite, tutored students saw slightly less growth than their peers, and these differences are statistically significant. Given the previously described correlations between assessment measures, this decrease is at least partially driven by tutored students seeing less growth in fluency (ORF), which is strongly correlated with the DIBELS Composite. However, as previously noted, we did not explore all skills measured within i-Ready or DIBELS. Those skills also factor into i-Ready and DIBELS' broader overall literacy measure, and growth between tutored and non-tutored students may vary within those measures as well.

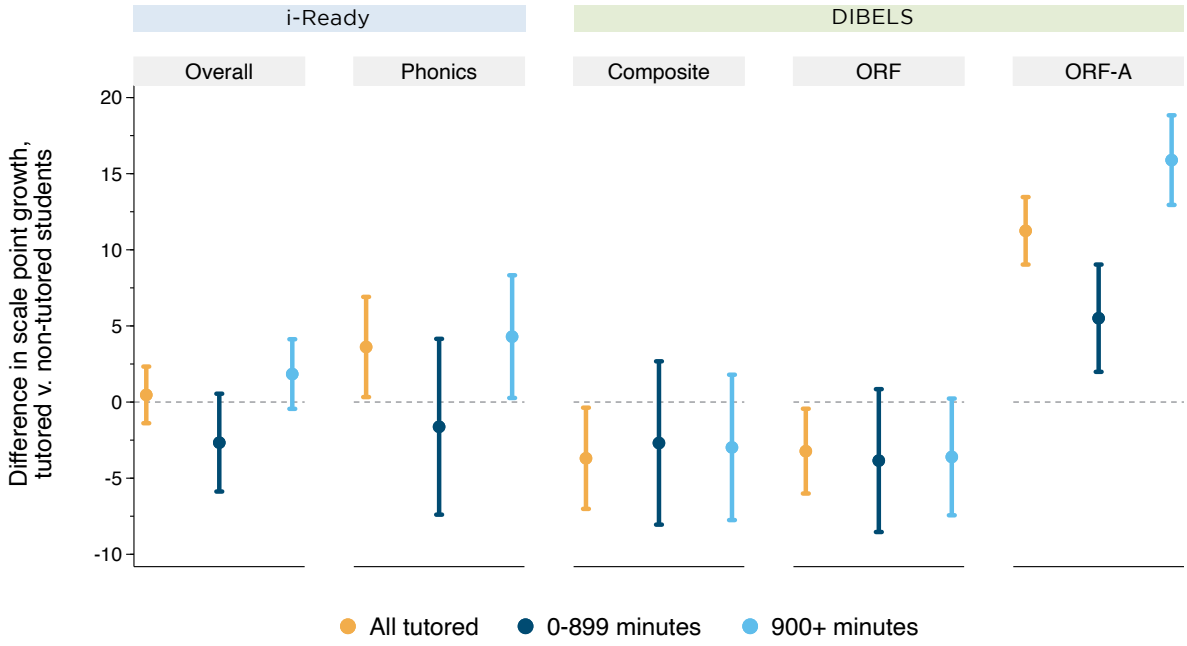
Taken together, the evidence suggests that tutoring led to meaningfully greater gains in students' ability to sound out and accurately decode or recognize words, as assessed on i-Ready Phonics and DIBELS ORF-A. However, even though these skills support higher-level literacy, gains in these skills have not yet translated into gains in fluency, and so tutored students have not yet experienced additional gains in overall reading ability. Tutored students may achieve these gains more quickly now that they have stronger foundational skills, which further study could test in the coming years.

We note again that samples differ across measures. The i-Ready Overall estimates include 12,440 students (1,659 tutored), and the i-Ready Phonics estimates include 7,841 students (1,538 tutored). DIBELS samples are much smaller: 1,352 (625 tutored) for the Composite and 2,800 (917 tutored) for Oral Reading Fluency and Accuracy. We emphasize caution in overinterpreting differences across measures.

Finding 5: Students with higher doses of tutoring saw more growth.

Is greater tutoring dosage associated with greater gains? To assess this, we break up our tutored sample into students receiving a relatively "high dosage" of tutoring (at least 900 minutes, or 15 hours) and students receiving a lower dosage (0-899 minutes). OUSD identified 900 minutes as an informal benchmark based on existing research on effective tutoring practices.

Figure 8: Differences in growth between tutored and non-tutored students, by minutes tutored



Note: Each dot represents a coefficient from a separate regression. The numbers are estimated differences in growth between tutored and non-tutored students. The vertical bars are the 95% confidence interval. All regressions include fixed effects for classroom and controls for students' school attendance rate, race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage. Ns vary by outcome. For all tutored, N = 12,440 for i-Ready Overall; 7,841 for i-Ready Phonics; 1,352 for DIBELS Composite; and 2,800 for DIBELS ORF and ORF-A. Schools that conducted tutoring after school are excluded.

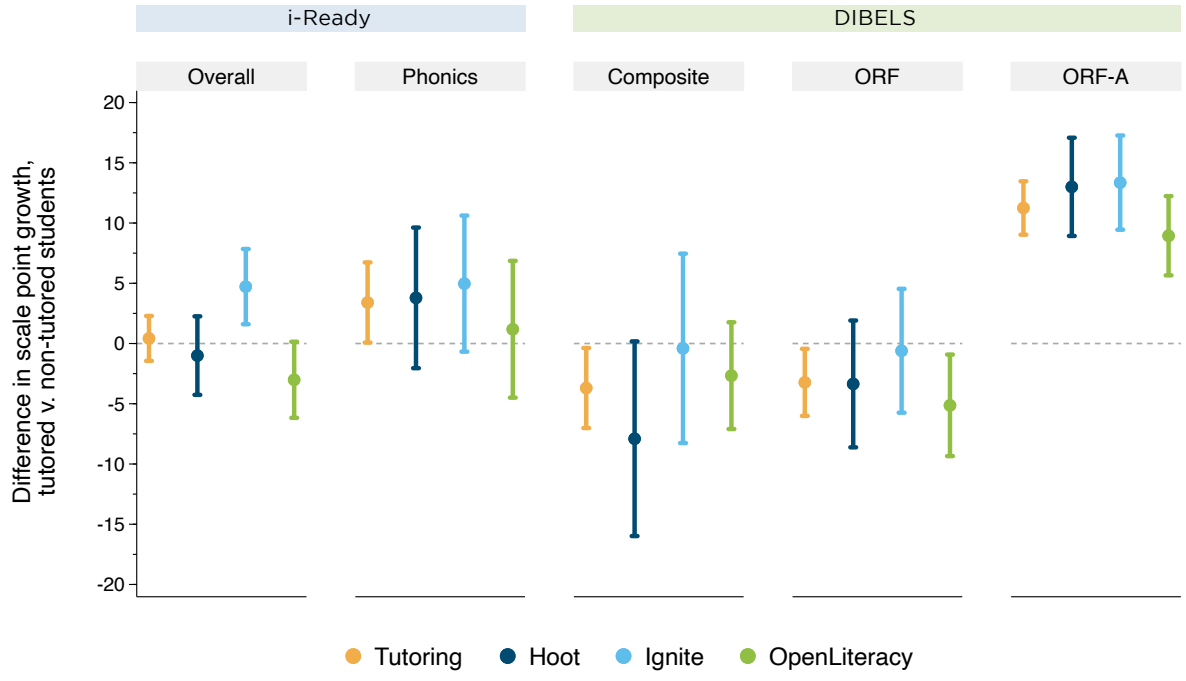
Figure 8 presents estimated differences in growth for tutored students with different dosage levels, relative to no tutoring. The orange dots still represent the overall tutored sample, plotting the same numbers as Figure 7. We then separately display relative growth for tutored students receiving 0-899 minutes (red) and students receiving 900 or more minutes (yellow). For i-Ready Overall, i-Ready Phonics, and DIBELS ORF-A, we see convincing patterns indicating that students who received a higher dosage of tutoring saw notably more growth than students who received less or no tutoring. The numbers are large and meaningful: **on i-Ready Phonics, students who received at least 900 minutes of tutoring grew 4.3 more points than comparable non-tutored students. This difference**

is equivalent to 0.06 standard deviations, or roughly 1.6 additional months of learning. For DIBELS Composite and ORF, the relative growth across doses is basically equal.

While these analyses use a relatively crude cutoff, our general takeaway is that there is broadly a positive association between dosage and growth, and this is most pronounced in the specific subtests where tutored students generally outperformed similar non-tutored students (i-Ready Overall, i-Ready Phonics, and DIBELS ORF-A). However, even tutored students with relatively large tutoring doses did not experience greater fluency or overall literacy gains than their non-tutored peers, which is consistent with OUSD's foundational literacy focus.

Finding 6: Most differences across tutoring providers are minor, with the exception of Ignite Reading and i-Ready Overall.

Figure 9: Differences in growth between tutored and non-tutored students, by provider



Note: Each dot represents a coefficient from a separate regression. The numbers are estimated differences in growth between tutored and non-tutored students. The vertical bars are the 95% confidence interval. All regressions include fixed effects for classroom and controls for students' school attendance rate, race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage. Ns vary by outcome and provider. For the combined tutoring estimates (orange), N = 12,440 for i-Ready Overall; 7,841 for i-Ready Phonics; 1,352 for DIBELS Composite; and 2,800 for DIBELS ORF and ORF-A.

In the majority of our analyses, we find that differences across the three tutoring providers—Hoot, Ignite Reading, and OpenLiteracy—are generally small. Figure

9 presents estimates from our regression models separately for Hoot, Ignite, and OpenLiteracy. In general, differences across providers are small and not statistically significant. The one exception is Ignite: specifically, on i-Ready Overall, students in Ignite Reading saw significantly higher growth than their non-tutored peers. **The difference is large and meaningful: Ignite students saw 4.8 points more growth than similar non-tutored students, which is equivalent to 0.06 standard deviations or about 1.5 additional months of learning**. On DIBELS Composite and ORF, we also find that Ignite students grew just as much as comparable non-tutored students. However, the differences across providers are smaller than the i-Ready Overall difference.

CONCLUSION

This study examines a pilot of three different virtual 1:1 literacy tutoring providers in Oakland schools. We find compelling evidence that tutoring implementation met district goals: tutoring was assigned as intended, students received relatively high doses of tutoring, schools were generally satisfied with the tutoring support they received, and students made significant gains in foundational literacy skills. Tutored students saw significantly more growth than similar non-tutored students in phonics and accuracy, but did not see greater gains in fluency or overall literacy. We find suggestive evidence that increased dosage is associated with greater growth. We also find relatively minor differences across providers, save for the notable exception of Ignite Reading significantly outperforming the other providers on i-Ready Overall. We note again that our analyses are not causal and are limited in various ways.

Looking across assessments, our results suggest that tutoring of this nature can potentially lead to meaningful gains in specific foundational literacy skills—such as phonics and accuracy—without leading to immediate gains in overall literacy development. Given Oakland's emphasis on phonics, this is not surprising. Enhanced foundational literacy skills may help tutored students experience more rapid literacy gains in the coming years. As OUSD continue their tutoring evaluation in the coming year, we look forward to learning more about how these different literacy skills coalesce. As tutoring expands, future work should also delve deeper into differences across subgroups and dosage.

The need for additional support in schools is clear, and virtual tutoring has the potential to drastically change how students learn. We emphasize the need for continued rigorous research to ensure these policies are appropriately supporting students who our current systems have failed.

TECHNICAL APPENDIX

Data & Sample Restrictions

We note a few key definitions and restrictions we use in our analyses. We identify tutored students as any student who received at least one tutoring session between their beginning and end-of-year test administration for a given assessment. Total sessions and minutes are defined between each student's individual assessment window to restrict each student's participation to that which could influence assessment scores. We drop two students who received tutoring sessions from more than one provider.

To calculate growth scores, we subtract beginning-of-year scores from end-of-year scores for each measure, separately. We count the last assessment in April or May 2025 as the end-of-year assessment. We count the first assessment of the year as the beginning-of-year assessment, limited to August or September 2024 for i-Ready and August through October 2024 for DIBELS. This excludes some students with at least two scores during the year. For example, if a student's first assessment was in November 2024, we cannot calculate their growth as we do not have a beginning-of-year score for them. This is a minor problem in i-Ready but more notable in DIBELS, as seen in the sample sizes in Table 1. If students took the same test multiple times in the same day, we take the average score. This is rare, representing less than 1% of our observations.

Eighteen students with i-Ready growth scores were tied to a tutoring provider that was not the provider for that school (for example, an OpenLiteracy school had one student with Ignite sessions). For DIBELS, five students have growth scores, but an inconsistent tutoring provider. These students are not included in our sample; they most likely changed schools during the year. Less than 1% of students in our sample are not in the data we use to connect students to classrooms. Because our regression analyses use classroom fixed effects, these students are not included in our sample.

Modeling Approach

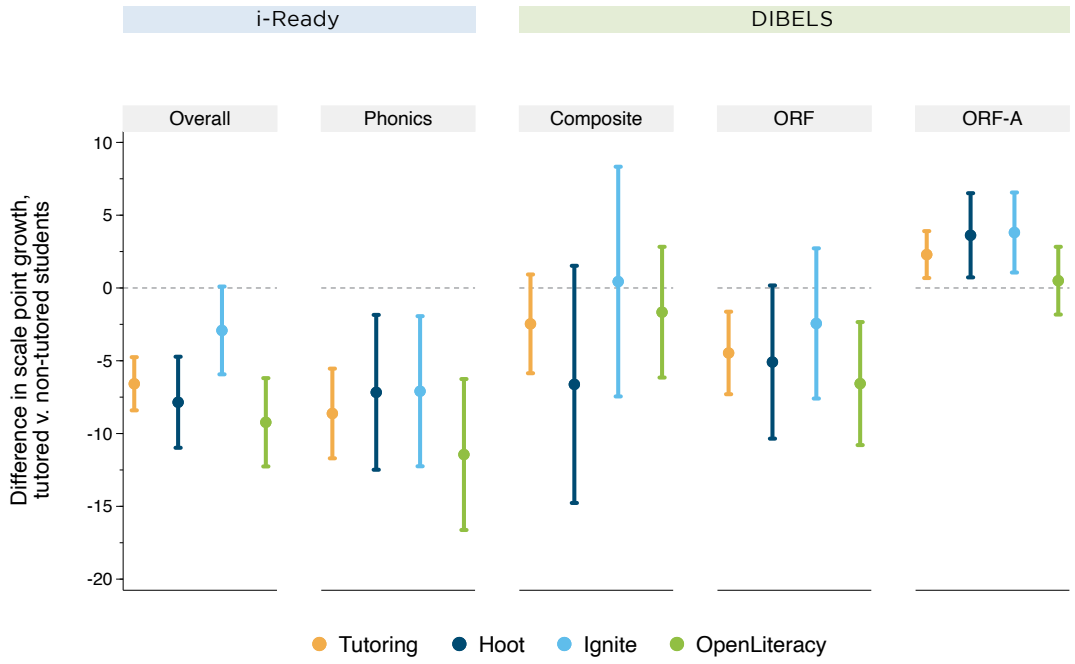
Our models present growth in scale points as an intuitive outcome, in line with OUSD's goals. Alternatively, we could measure differences in growth (or spring scores) between tutored and non-tutored students conditional on fall scores. These models are similar, but they ask different questions, have different expectations, and give different results, a phenomenon well-documented in

the statistical literature. Using this alternative specification, we find that tutored students generally ended the year with lower scores than non-tutored students with the same fall score, except in DIBELS ORF-A. Both i-Ready differences are significantly negative using this approach (see Figure A1).

We prefer to explicitly model differences in growth, as our main research question asks whether tutored students experienced greater than expected gains, relative to *their own* prior achievement. It's also likely that tutored students are systematically disadvantaged in ways that influence achievement during the school year but that are not captured by prior achievement, as OUSD emphasized targeting the students who needed the most extra support *during the year*. If this is the case, then spring score comparisons between tutored and non-tutored students with the same fall score would be biased downward, as non-tutored students faced fewer challenges during the year.

Taking these two sets of models together, we find that tutored students made more progress over the year than non-tutored students, on average, but still have lower end-of-year scores than their peers. In short, tutoring likely helped mitigate preexisting achievement disparities without eliminating them entirely.

Figure A1: Differences in growth between tutored and non-tutored students, controlling for beginning-of-year score



Note: Each dot represents a coefficient from a separate regression. The numbers are estimated differences in growth between tutored and non-tutored students. The vertical bars are the 95% confidence interval. All regressions include fixed effects for classroom and controls for beginning-of-year score, students' school attendance rate, race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage. Ns vary by outcome and provider. For the combined tutoring estimates (orange), N = 12,440 for i-Ready Overall; 7,841 for i-Ready Phonics; 1,352 for DIBELS Composite; and 2,800 for DIBELS ORF and ORF-A.

Questions from OUSD's mid-year tutoring implementation survey

- ENTHUSIASM: "How enthusiastic are you about continuing to work with this provider next year?"
  - 1 = not at all
  - 4 = very
- PROVIDER: "Was your provider flexible and accommodating with your scheduling needs?"
  - 1 = not at all accommodating
  - 4 = very accommodating
- SUPPORT: "Do the skills students are learning during their tutoring sessions support their success and engagement in classroom reading instruction (SIPPS, EI Education, Benchmark)?"
  - 1 = not at all
  - 4 = very much
- DATA: "Is the data you are receiving from the provider helpful in monitoring and supporting student growth?"
  - 1 = not at all helpful
  - 4 = very helpful
- IMPROVEMENT: "Do you or your teachers see improvement in student motivation, confidence, and attitudes towards reading?"
  - 1= no improvement
  - 4 = significant improvement
- *MAINTENANCE: "Does the ongoing maintenance of the Tier 3 tutoring program (outside of the original launch) impact your ability to perform other aspects of your role in supporting literacy instruction?"
  - 1 = no impact
  - 4 = significant impact
- *LOGIN: "How much support do students generally need to login to their tutoring sessions? "
  - 1 = little to no support
  - 4 = extensive support

* For MAINTENANCE and LOGIN, lower numbers are better.

Table A1: Differences between tutored students included and excluded from regressions

	i-Ready				DIBELS			
	Overall		Phonics		Composite		ORF & ORF-A	
	Not in analyses	In analyses	Not in analyses	In analyses	Not in analyses	In analyses	Not in analyses	In analyses
Asian	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.04
African American	0.30	0.24	0.26	0.24	0.27	0.18*	0.27	0.21*
Latino	0.54	0.62*	0.59	0.61	0.58	0.67*	0.59	0.63
White	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05
Male	0.48	0.47	0.47	0.47	0.48	0.45	0.48	0.45
Economically Disadvantaged	0.96	0.97	0.97	0.98	0.98	0.96	0.98	0.97
Multilingual Learner	0.52	0.58	0.51	0.60*	0.55	0.65*	0.55	0.61*
Special Education	0.11	0.18*	0.11	0.19*	0.18	0.16	0.17	0.18
Tutoring Sessions	39.73	61.78*	46.03	62.32*	53.51	71.68*	49.87	69.15*
Tutoring Minutes	702.28	1174.44*	838.69	1185.93*	1016.76	1349.67*	912.78	1338.87*
N	248	1659	369	1538	1282	625	990	917

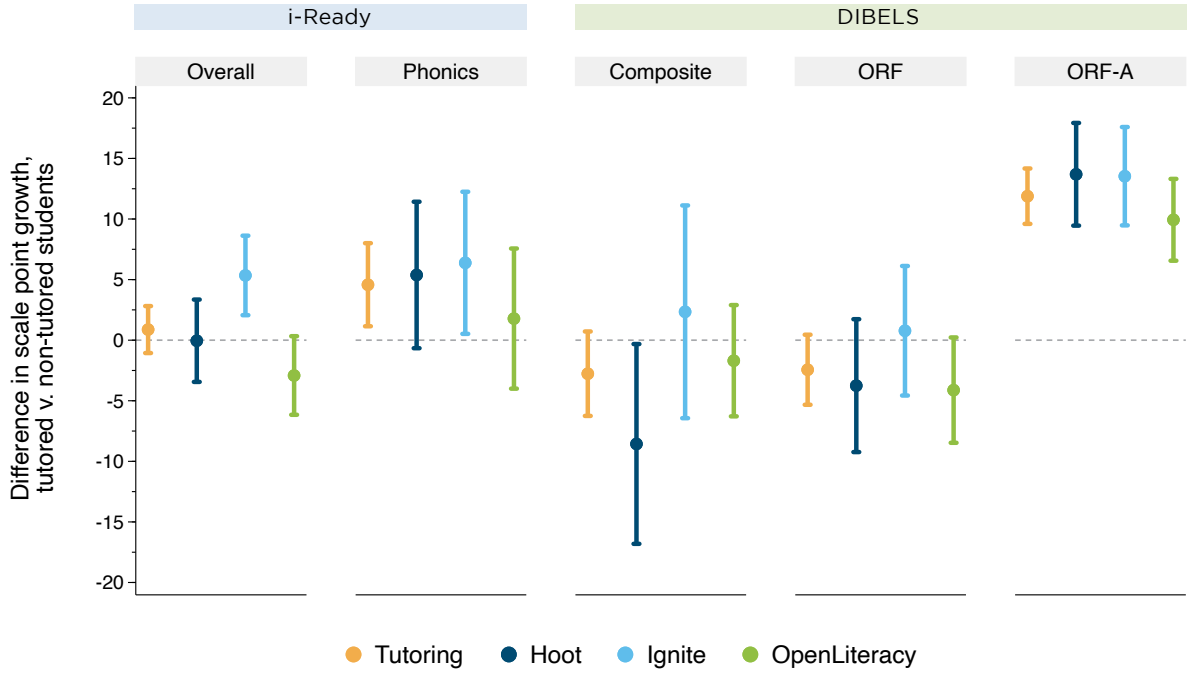
Note: Each cell contains the average value for each characteristic, comparing tutored students who are and are not included in our analyses. Significant differences are indicated with asterisks (*p < 0.05).

Table A2: Correlations between growth measures

	i-Ready Overall	i-Ready Phonics	DIBELS Composite	DIBELS ORF
i-Ready Phonics	0.62			
DIBELS Composite	0.30	0.21		
DIBELS ORF	0.26	0.16	0.81	
DIBELS ORF-A	0.08	0.12	0.12	-0.04

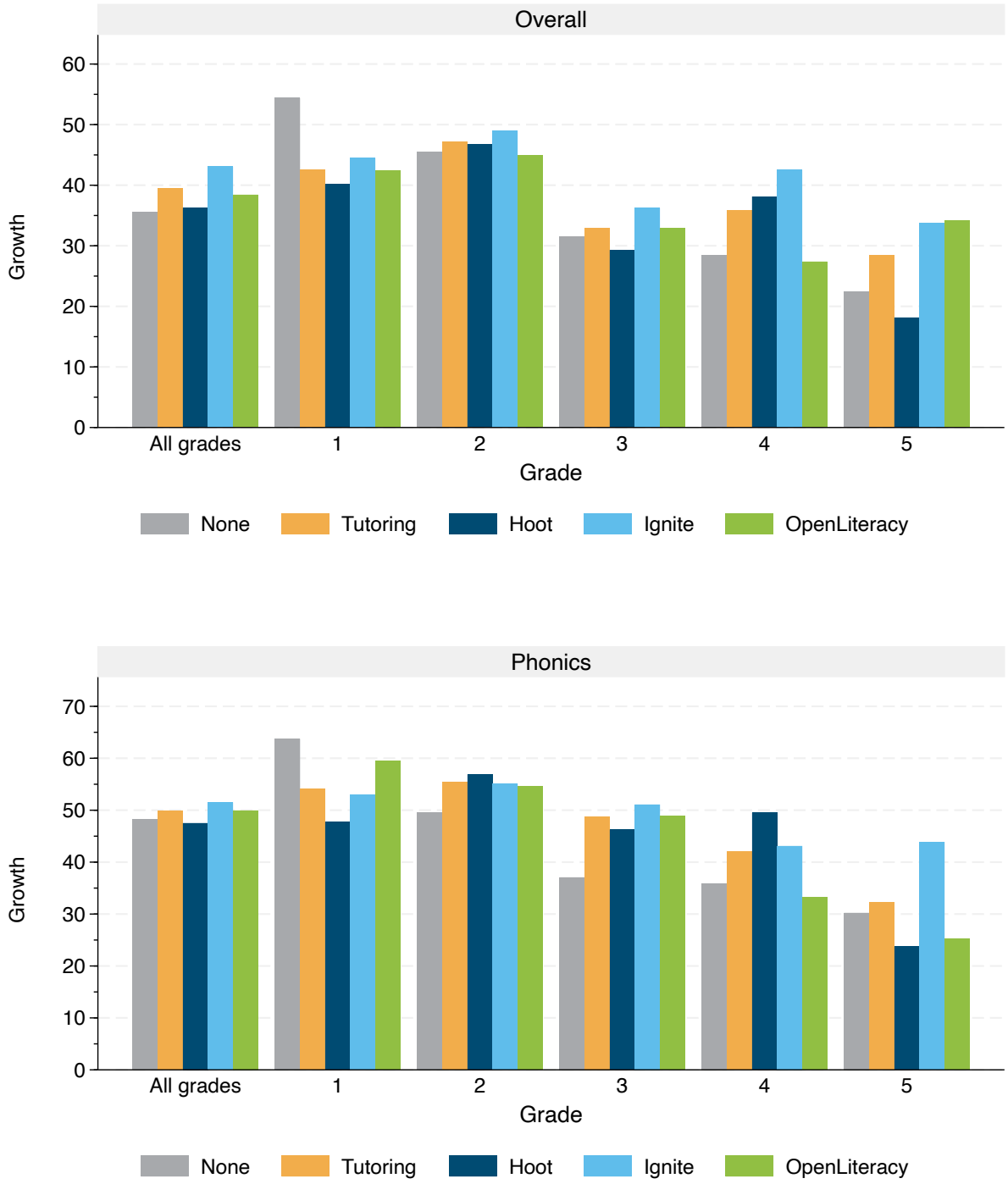
Note: Cells present correlation coefficients between each pair of beginning- to end-of-year growth measures. For example, growth in DIBELS Composite is strongly correlated with growth in DIBELS Oral Reading Fluency (0.81).

Figure A2: Differences in growth between tutored and non-tutored students, limited to 20 or more sessions



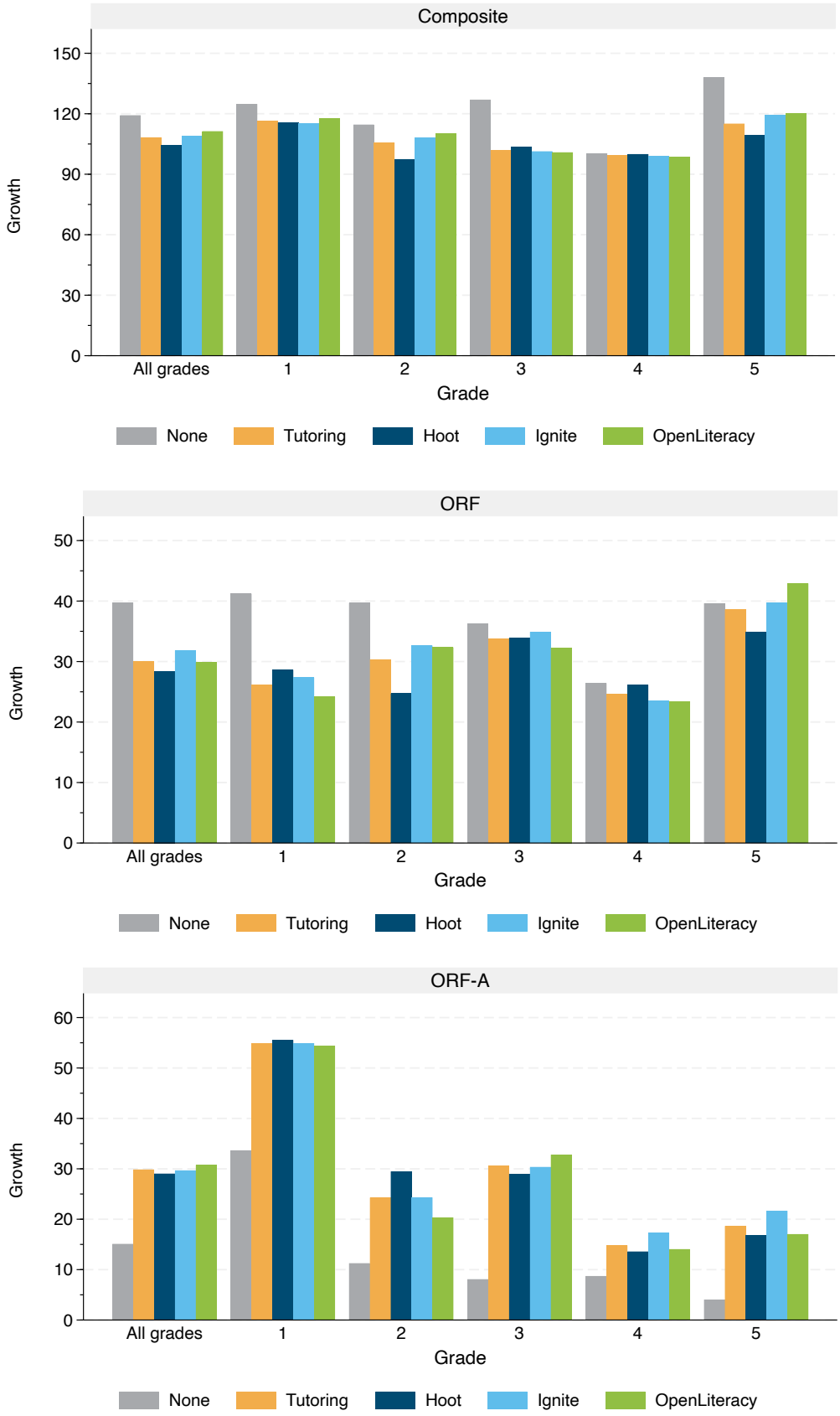
Note: Each dot represents a coefficient from a separate regression. The numbers are estimated differences in gr between tutored and non-tutored students. The vertical bars are the 95% confidence interval. All regressions include fixed effects for classroom and controls for students' school attendance rate, race/ethnicity, gender, multilingual learner status, special education designation, and economic disadvantage. We exclude tutored students who received 19 or fewer tutoring sessions. Ns vary by outcome and provider. For the combined tutoring estimates (orange), N = 12,280 for i-Ready Overall; 7,697 for i-Ready Phonics; 1,317 for DIBELS Composite; and 2,748 for DIBELS ORF and ORF-A.

Figure A3: i-Ready growth by grade and outcome



Note: Bars plot average growth from beginning- to end-of-year scores on i-Ready among first through fifth graders in the district.

Figure A4: DIBELS growth by grade and outcome



Note: Bars plot average growth from beginning- to end-of-year on DIBELS among 1st through 5th graders in the district.

ABOUT THE CENTER ON REINVENTING PUBLIC EDUCATION

The Center on Reinventing Public Education (CRPE) is a nonpartisan K-12 education research center at Arizona State University's Mary Lou Fulton College for Teaching and Learning Innovation. We rigorously examine and test transformative ideas, using our research to inform action. We are truth tellers who combine forward-thinking ideas with empirical rigor. Since 1993, we have been untethered to any one ideology but unwavering in a core belief: public education is a goal—to prepare every child for citizenship, economic independence, and personal fulfillment—and not a particular set of institutions. From that foundation, we work to inform meaningful changes in policy and practice that will drive the public education system to meet the needs of every student.

ACKNOWLEDGEMENTS

We are grateful to our partners at Oakland Unified and Rodriguez Community Group for their support of this work, particularly their thought partnership and generous feedback on analyses in progress. We thank Robin Lake and two anonymous reviewers for providing thoughtful questions and comments that improved our work. Thank you to Melissa Fall, Emily Prymula, Abby Jacobs, and Nadja Michel-Herf for helping us throughout the publication process.

CRPE QUALITY ASSURANCE PROCESS

Independent peer review is an integral part of all CRPE research projects. Before publication, this document was subjected to a quality assurance process to ensure that:

- The problem is well formulated.
- The research approach is well-designed and well-executed.
- The data and assumptions are sound.
- The findings are useful and advance knowledge.
- The implications and recommendations follow logically from the findings and are explained thoroughly.
- The documentation is accurate, understandable, cogent, and balanced in tone.
- The research demonstrates an understanding of related previous studies.
- The research is relevant, objective, and independent.

Research or policy professionals who were not members of the project team conducted peer review for this project.