Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1590

Why aren't online learning programs working better in schools: Maybe they need to be more integrated with what is actually taught during whole classroom instruction?

Ethan Scherer
Harvard University

James S. Kim
Harvard University

Computer-assisted learning (CAL)—educational software providing instruction to help students learn concepts and skills—is ubiquitous in US schools. Despite meta-analyses suggesting benefits, national test scores have declined during rapid CAL adoption. This may reflect insufficient research on how and when CAL complements instruction. Using a fixed effect model, we find 60 additional minutes of CAL usage when integrated with classroom instruction improves standardized literacy outcomes by 0.061 standard deviations at a cost of $2.28 per student. Yet a commonly used "naïve" statistical model when not using standardized outcomes biases results upward fourfold, suggesting purported CAL benefits may sometimes reflect biased estimates. We conclude the field should move toward a more nuanced perspective by identifying CAL's comparative advantage rather than whether to use it.

VERSION: September 2026

Suggested citation: Scherer, Ethan, and James S. Kim. (2026). Why aren't online learning programs working better in schools: Maybe they need to be more integrated with what is actually taught during whole classroom instruction?. (EdWorkingPaper: 26-1590). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/nk7c-7b36

Why aren't online learning programs working better in schools: Maybe they need to be more integrated with what is actually taught during whole classroom instruction?

Ethan Scherer

James S. Kim

Harvard University

Computer-assisted learning (CAL)—educational software providing instruction to help students learn concepts and skills—is ubiquitous in US schools. Despite meta-analyses suggesting benefits, national test scores have declined during rapid CAL adoption. This may reflect insufficient research on how and when CAL complements instruction. Using a fixed effect model, we find 60 additional minutes of CAL usage when integrated with classroom instruction improves standardized literacy outcomes by 0.061 standard deviations at a cost of $2.28 per student. Yet a commonly used "naïve" statistical model when not using standardized outcomes biases results upward fourfold, suggesting purported CAL benefits may sometimes reflect biased estimates. We conclude the field should move toward a more nuanced perspective by identifying CAL's comparative advantage rather than whether to use it.

Acknowledgements: The authors thank Catherine Asher, Chris Cleveland, and Doug Mosher for providing valuable feedback on an earlier version of this manuscript. We are also extremely grateful to Ben Lamb for excellent research assistantship. The opinions expressed are those of the authors and do not represent the views of their institution.

U.S. school districts have made substantial investments in educational technology. In the 2024-25 school year, 88% of districts reported providing every student a school-issued device (US DOE, 2025). This infrastructure expansion has accelerated adoption of computer-assisted learning (CAL)—educational software that provides targeted instruction to help students learn and apply concepts and skills (Escueta et al., 2020). Between 2018-19 and 2025-26, districts increased their use of educational technology tools more than fourfold (Instructure, 2024; Instructure, 2026). This growth is driven by CAL's three main advantages: (1) personalized learning at students' own pace, (2) automatic progress tracking that enables targeted grouping by need, and (3) scalable, equitable implementation across districts.

However, evidence on CAL's real-world effectiveness is mixed. While meta-analyses show that CAL improves student outcomes in tightly controlled settings, substantial variation exists across studies (Gilbert & Kim, 2026). More importantly, these gains inconsistently translate when programs are implemented in less controlled, real-world classroom environments (Pane et al., 2010, 2014; Roschelle et al., 2016). Meta-regressions reveal important patterns in this heterogeneity. For example, effects are more than twice as large when measured by researcher-developed or company-developed assessments than when using standardized outcomes (Gilbert & Kim, 2026)—a pattern consistent with findings outside educational technology (Kraft, 2020). An additional key insight is that CAL is more effective for constrained skills (skills with limited content and clear mastery endpoints) than unconstrained skills (skills without ceilings that develop throughout life). It is also possible that CAL effectiveness depends on contextual factors beyond the technology itself.

Implementation challenges in real-world settings likely explain some of this variation as well. Technology malfunctions, competing instructional time demands, and teacher skepticism all reduce effectiveness. A recent international study demonstrated this point by comparing outcomes with and without a full-time, on-site CAL coordinator. The coordinator reduced implementation barriers and improved program dosage, resulting in math gains of approximately half a standard deviation (Oreopoulos et al., 2026). However, this solution increases costs. An alternative approach is to more deeply integrate CAL into classroom instruction rather than treating it as a separate initiative.

Currently, CAL programs are often purchased independently from core curriculum with minimal alignment. Even when vendors provide both curriculum and supplemental CAL software, little effort ensures CAL content connects to that day's classroom lessons or unit. This fragmentation creates disconnected instructional streams. For example, students decode words through a CAL program without connecting those skills to the academic vocabulary taught during literacy instruction. Tight integration of CAL with classroom instruction could reduce this disconnect by increasing perceived relevance, lowering teacher resistance, and maximizing aligned instructional tools within limited school time. Yet few rigorous studies have tested whether tightly coupling CAL with classroom instruction reduces implementation barriers while improving student outcomes.

Recent evidence suggests that building schema awareness is particularly well-suited to efficiently connect classroom and digital instruction. Schemas are mental frameworks that help students acquire, organize, connect, and transfer knowledge within a domain (Anderson & Pearson, 1984;

Alexander, 2003; Graesser & Nakamura, 1982; Rumelhart & Norman, 1981). When CAL and classroom instruction both build schema awareness around the same topics, they reinforce rather than compete for time. Research demonstrates that curricula emphasizing schema awareness help students transfer learning to new contexts (Kim et al., 2024).

This study addresses this gap by using quasi-experimental methods to evaluate whether school-day-integrated CAL improves student outcomes. We investigate three research questions:

- **RQ1:** What is the incremental impact of CAL seamlessly integrated with whole class Tier I curriculum on student literacy outcomes?
- **RQ2:** Which student characteristics moderate these effects?
- **RQ3:** What is the incremental per-student cost of an integrated CAL program?

By testing whether integration reduces implementation barriers and improves outcomes, this work could offer schools and districts a scalable, cost-effective option.

Active learning time is a key constraint in schooling

Active learning time is among schools' most precious and scarce commodities. Kraft and Novicoff (2024) found that elementary students in Providence, RI lost 16% of the school day to absences, suspensions, tardies, and interruptions. This estimate likely understates actual losses, as it excludes nonacademic routines like transitions between activities.

Given this scarcity, schools must make difficult trade-offs in time allocation. Tyner and Kabourek (2020) document that elementary students spend, on average, nearly 40% of the school day on ELA instruction, compared to less than 20% on science and social studies combined. Moreover, school-level decisions often intensify these trade-offs—administrators could replace science and social studies entirely with expanded math and ELA blocks based on perceived needs. These choices reveal how schools already optimize within fixed time constraints, shifting resources based on institutional priorities.

CAL programs purport to solve this problem but could worsen it. In theory, CAL enables structured independent practice on struggling students' areas of need. If students could complete these tasks as homework with automatic grading, valuable classroom time would be freed, and teachers would gain diagnostic feedback to target small-group instruction. However, low-income students often lack reliable home access to technology, forcing schools to allocate dedicated instructional time to ensure adequate program dosage (Robinson et al., 2025)—typically 60-90 minutes per week. This time must come from existing allocations. Without additional classroom hours, schools face a zero-sum time problem. To implement CAL effectively requires identifying genuine synergies across subjects and instructional modalities to avoid simply displacing other content.

Building Schema Awareness – a potential strategy for school-day-time constraints

One effective teaching approach to optimize instructional time is to help students build schema awareness (Anderson & Pearson, 1984; Alexander, 2003; Graesser & Nakamura, 1982; Rumelhart & Norman, 1981). A schema is a mental framework that helps students organize knowledge so that they can more efficiently retrieve and apply it. One way to help students instantiate a schema is to help them understand the patterns and connections across topics. For example, when students learn about how concept words such as *structure, function, skeletal, muscular, nervous, and system* are interrelated, they form a mental map of this knowledge and are more likely to be able to retrieve and apply it to new situations and topics (Gick & Holyoak, 1983). Although building schema awareness takes time, the payoff is substantial—students develop lasting, interconnected knowledge that they can use and build on throughout their academic careers.

A practical way to build schema awareness is through academic vocabulary networks—teaching students how specialized words relate to each other. Academic vocabulary refers to the formal, subject-specific language used in textbooks, lessons, and assessments. (For example, in physics, "force" has a precise meaning: mass × acceleration, rather than the everyday meaning of "causing physical change".) When students learn how related words connect in meaning and context, they understand and remember them better (Perfetti, 2007). As students encounter these related words repeatedly throughout the year and across grades in different readings, discussions, and activities, their understanding deepens and becomes more automatic. Each exposure strengthens both their memory and their grasp of the broader concept (Alexander, 1997). Research confirms that building students' networks of related academic vocabulary supports both content learning and schema development (Mosher & Kim, 2025; McKeown & Beck, 2011).

The Model of Reading Engagement (MORE) program demonstrates one way schools can build schema awareness efficiently. MORE is a content-focused literacy program that teaches students how to read, write, and think critically within specific subjects like science and history. Rather than memorizing isolated facts or vocabulary lists, MORE helps students see how ideas and words are connected within and across subjects.

Teachers implement MORE through three main instructional strategies. In the MORE curriculum, first, teachers provide direct instruction of academic vocabulary by examining how words relate through shared meanings, word origins, and sounds. This includes using concept maps as visual diagrams highlighting how words and concepts are interrelated. Second, teachers provide read-aloud experiences, modeling fluent reading and comprehension strategies to expose students repeatedly to schema-related vocabulary. Last, the MORE curriculum includes thematically and conceptually coherent texts (Wright et al., 2022) where students read multiple texts on the same topic to deepen their understanding of the content. Following these lessons, students apply their learning through collaborative research projects where they engage in structured discussions, independent reading, and writing to develop deeper expertise on the topic.

Results from two randomized controlled trials have shown that MORE delivers measurable results. Students who received MORE showed improved reading comprehension in science as well as gains on standardized reading and math assessments (Kim et al., 2021; Kim et al., 2024). Notably, MORE required only 30 minutes per day for six weeks. After multiple years of exposure, students demonstrated effect sizes of 0.11 and 0.12 on reading and math standardized measures, respectively—gains that persisted one year later (Kim et al., 2024). By focusing on how knowledge connects within specific content areas, MORE helps students transfer their learning to broader measures of reading and math achievement in a time-efficient manner. Thus, MORE addresses the time-scarcity challenge: building schema awareness doesn't require *adding* instructional time; instead, it optimizes the time already spent.

Integrated CAL

While the MORE program teaches academic vocabulary within science and social studies lessons, students typically learn foundational reading skills like decoding and word patterns separately during literacy instruction. Instruction focusing on the development of language and foundational skills, like decoding, is often unrelated and taught in isolation. Students may learn to decode words in English Language Arts (ELA) without understanding their specific meanings in science. To strengthen this connection, MORE complements its classroom lessons with digital activities designed to develop both code-based and language skills through metalinguistic awareness. Specifically, the activities develop the ability to consciously think about and analyze how language works. This includes recognizing how language's smallest units of sound (phonemes) and meaning (morphemes) function, and how phonemes change predictably when morphemes are combined and modified. By building students' metalinguistic awareness, students simultaneously strengthen their code-related skills (e.g., phonemic patterns, decoding, and spelling) and language-meaning skills (e.g., morphological understanding, semantic knowledge, and vocabulary retention).

In practice, this means that students engage with carefully selected focus words through 36 interactive digital activities. Each activity focuses on a taught word from the MORE lessons or a conceptually related but untaught word that students encountered through reading or listening. Students engage with these words through multiple exercises: analyzing prefixes and suffixes, practicing spelling and pronunciation, and resolving ambiguous meanings using jokes, riddles, and puns (Zipke, 2008). MORE's words are anchored in science and social studies standards and are assessed for their centrality in a semantic network of words on a particular topic. Because these focus words are drawn directly from the MORE unit, students encounter them repeatedly across multiple modalities—classroom reading, teacher read-alouds, and digital activities practice. This repetition strengthens vocabulary retention and deepens understanding.

The selection of focus words is rigorous and strategically purposeful. All words are academic vocabulary—specialized language used across multiple subjects (science, social studies, math, literature) that appears on major vocabulary lists like Marzano's curated research-based academic vocabulary database for K-5 (Marzano, 2004). However, MORE goes further: the words are anchored in state science and social studies standards and reinforced across multiple grade levels to ensure consistency over time. Critically, each word is selected for its centrality in a semantic network where it sits at the hub of a cluster of related words within a topic (see Kim et al., 2023 for details on the selection process). This is essential: when students master a central word, they can transfer that knowledge to related words they haven't explicitly learned (Gilbert et al., 2025). This multiplies the impact of instruction where fewer words are taught yet broader understanding is gained.

Implementation of Integrated CAL

This study took place in a large southeastern school district during the 2024-25 school year. The district implemented MORE in elementary schools from February through April, consisting of a 3-week life science unit followed by a 3-week economics unit, both taught during the regular science and social studies instructional block. The accompanying digital activities were suggested to be implemented at the same time during the school year as MORE instruction, and the suggested dosage was approximately 10-15 minutes per week. Teachers had flexibility in scheduling these activities outside the science and social studies block, which was typically during designated skill-practice periods, not during core instructional time.

While the activities focus on code and language related skills, all students receive the teacher-led lesson that focuses on language comprehension. Time-wise, the lessons and researcher-developed formative assessment, the two other classroom tools that are part of the MORE program, account for approximately 94% of the time students spend on MORE. Thus, the digital activities component is only 6% of the school-day time related to MORE.

The implementation window was strategically timed. Teachers implemented MORE immediately after the winter formative assessment, and implementation concluded just before spring end-of-year formative assessment. This timing allowed us to measure learning gains during the MORE implementation period while minimizing confounding factors.

Data and Methods

Sample & Data

All 3rd grade students in the district had access to the MORE program and its digital activities during the 2024-25 school year (N = 10,075). In our sample all students completed DIBELS and iReady Inform at three time points: fall, winter, and spring, thus the students in our analysis sample have a balanced panel of these scores. DIBELS (8th Edition) measures foundational reading skills—specifically phonemic awareness, phonics, and fluency—with less emphasis on comprehension. iReady Inform is an adaptive assessment that has a stronger emphasis on reading comprehension and correlates over 0.80 with state standardized tests. Together, these measures capture both code-based skills (phonics/decoding) and language-based skills (comprehension). To account for differences in assessment scales, we standardized all scores using the fall (beginning-of-year) mean and standard deviation. Table 1 presents the sample characteristics and fall to winter gains for both DIBELS (8^{th} Edition) and iReady assessments.

Approximately 31% of students never engaged with the digital activities. Seventy-two percent of the minute variation is between classrooms, suggesting that the usage of the digital activities was primarily a teacher decision[1]. As shown in Table 1, students with zero exposure differed from their peers: they were more likely to be Hispanic or Black, an English Learner, a student with disability, and come from a high poverty neighborhood. Table 1 also reports the nationally normed percentile change from Fall to Winter on DIBELS and iReady, a key identification assumption in the models described below. While the groups differ in magnitude of the change, the changes are not statistically different than zero when we include teacher fixed effects when we model percentile change on usage and include teacher fixed effects. Because of the large number of zeros and the skewed nature of minutes, we bin time on the digital activities into four categories: 0 minutes, up to 10 minutes, 10 to 60 minutes and more than 60 minutes. These cut-offs correspond to light, moderate, and full recommended usage (10-15 min/week = ~1 hour/year). See Appendix Table A1 for a more detailed distribution of minutes.

---

[1] Note an additional 16% of students used the app's other features exclusively, like reading ebooks, rather than completing digital activities. These students are also conservatively labeled as conducting 0 minutes of digital activities in our analysis because this time did not include deliberate word practice.

Methods
Table 1 shows that students who engaged more with digital activities differed systematically on both observable characteristics (e.g., DIBELS gains, IEP status) and thus likely unobservable characteristics (e.g., motivation, ability). This selection bias creates a problem: standard regression and propensity score matching methods often used by many CAL companies may overestimate treatment effects because they cannot fully account for these differences. To quantify this bias and move toward causal inference, we estimate three progressively more rigorous models. Model 1 replicates the student-level (e.g., one observation per student) **Naïve Model** typical of industry reports.

$$Score_i = \beta_0 + \sum_{j=1}^{3} \beta_j Digital\ Act\ Minutes_i + \beta_4 Score\ (t-1)_i + \varepsilon_i \quad (1)$$

where $Score_i$ is the test score for student i, $Digital\ Act\ Minutes_i$ is a flexible non-parametric measure of time spent on the digital activities for student i, $Score(t-1)_i$ is the linear baseline score for student i at time $t-1$ (i.e., the winter assessment, administered just before MORE began), and ε_i is the residual. The minutes bins correspondent to the light, moderate, and full recommended usage noted in the data section, with the omitted category being 0 minutes. The coefficient β_j estimates the effect of the number of minutes on student test scores. It's important to keep in mind the minimum "recommended" dosage is an hour of class time *for the entire year*, which is relatively small compared to most CAL programs. Many CAL companies report results using this type of approach. It serves as a baseline for comparing bias.

We know Model 1 will be biased upwards since it does not account for factors like higher implementation teacher could be more effective and based upon prior work (Chetty et al., 2014) that a more flexible baseline test scores is an important modeling decision. As such we estimate Model 2, which we label the **Teacher Fixed Effect Model**

$$Score_{ic} = \beta_0 + \sum_{j=1}^{3} \beta_j Digital\ Act\ Minutes_{jic} + f(Score(t-1)_{ic}) + \sum_k X_{kic} + \gamma_c + \varepsilon_{ic} \quad (2)$$

where $Score\ (t-1)_{ic}$ is now a cubic function of both baseline DIBELS and iReady, we add X_{ic} that are additional student-level covariates including gender, race, whether the students received an individual education program, enrolled in a limited English proficiency program, and came from a low, medium or high socioeconomic neighborhood based upon census data. This model also adds γ_c a classroom fixed effect as well. The fixed effect accounts for the teacher's average effect on student achievement, including whether they taught MORE. This model isolates within-classroom variation—comparing high vs. low activity users taught by the same teacher—which reduces bias from teacher quality differences.

For the two models above, we cluster our standard errors at the classroom-level.

Finally, we run a third model with two specifications that includes both student-level and period fixed effect that accounts for all unobserved, time invariant characteristics within students, teachers, and schools, which we label the **Student Fixed Effect Model:**

$$Score_{it} = \beta_0 + \sum_{j=1}^{2} \beta_j Digital\ Act\ Minutes_{it} + \alpha_i + \omega_t + \varepsilon_{it} \quad (4)$$

This model replaces student-level characteristics with student fixed effects (α_i) and adds period fixed effects (ω_t) for seasonal assessment differences. Student fixed effects eliminate all time-invariant unobservables (motivation, ability, family background). We can identify effects because students have multiple assessment periods, but in cases where there is no variation (e.g., student did not use the digital activities), we cannot estimate values. Thus in these models it changes our reference category to up to 10 minutes rather than 0 minutes. Because of observable difference between Light (e.g., up to 10 minutes) and full recommended usage, we also run student fixed effects models among moderate and full recommended usage as well because the pre-intervention gains are not statistically different from one another. Based upon the implementation design, students use the digital activities 0 minutes during the fall and winter assessment periods. We cluster our standard errors at the student-level.

In addition to average effects, we examine whether digital activity effects vary by baseline performance. Specifically, we interact the dosage categories by whether students performed below grade-level at baseline using the teacher **Teacher Fixed Effect Model**, because baseline test scores do not vary across time, as is required for the **Student Fixed Effect Model**. We hypothesize that students below grade-level may benefit more.

Results

Table 2 presents the results across all four models. Model 1 (naïve approach) shows that spending more than 60 minutes on digital activities increased end-of-year DIBELS scores by 0.113 standard deviations relative to non-users. However, this estimate is likely biased upward because higher performing students are more likely to engage independently with the digital activities. Once we add student covariates, a non-linear form of baseline test scores, and classroom fixed effects (Model 2), the effect drops to 0.085 SD—a 25% reduction. Next we run the two fixed effects models where the omitted category is light (Model 3) and moderate (Model 4) usage. The student fixed effect models (Model 3) and (Model 4) yield 0.061 and .026 SD, respectively. These estimates are similar to Model 2 given the changing omitted category and suggests that unobserved time-invariant characteristics (motivation, ability) are sufficiently addressed by Model 2. Students using digital activities for the full recommended dosage (>60 minutes/year) improve approximately 0.061 SD on foundational reading skills based upon Model 3. It is worth noting the dosage response effect for the moderate usage is statistically significant, but smaller in magnitude on DIBELS, as we would expect.

Table 3 presents results using iReady. As expected—since 94% of the MORE program time (e.g., lessons & assessment) emphasize comprehension—we observe only one, likely spurious, statistically significant effect across all models. Point estimates hover near zero.

Do struggling students benefit more from code-focused digital practice? We tested this hypothesis by interacting our predictor of interest in Model 2 (our teacher fixed effects model) with baseline performance. As noted in the methods section, we cannot estimate the student fixed effect model because there is no variation in baseline performance over time, but because Model 2 achieves similar bias reduction, using that model is justified. Students below grade-level at baseline showed effects similar to their grade-level peers. This suggests that digital activities provide equitable benefits across proficiency levels, rather than disproportionately helping weaker readers, which was unexpected.

To assess cost-effectiveness of integrated CAL, we estimated the per-student cost of the digital activities component in Table 5. MORE is sold as a comprehensive package ($380 per teacher annually) including classroom lessons, digital activities, formative assessment, and professional learning. Based on time allocation, the digital activities account for approximately 6% of total program time. Direct costs are thus $23.75 per teacher for the digital activity materials. Professional development for the digital activities occurs in one 30-minute professional learning community session. The Bureau of Labor Statistics indicated the median national salary for Kindergarten and Elementary School teachers is $63,000 (BLS, 2024). Assuming a 36-week school year and 40 hours a week, teacher salary is approximately $43.75 per hour. Thus, the opportunity cost of the teacher's salary for 30 mins is $21.88. If we assume there are 20 students per classroom, the total per-student cost is approximately $2.28 for one year of implementation.

For context, the 0.075 SD effect on code-based reading skills is modest but meaningful. A recent study of an early literacy hybrid tutoring program found effects of 0.23 standard deviations. However, the program cost between $375-$450 per student and required 15 minutes each day of independent practice as well as the 5-10 minute tutoring sessions (Cortes et al., 2025), a much higher dosage and cost per student.

Discussion

Our paper is the first, to our knowledge, to assess the incremental benefit of a CAL integrated with whole class Tier I curriculum. The CAL complements classroom instruction targeting metalinguistic skills for words (e.g., nodes) in vocabulary networks. Despite low-cost implementation and minimal time investment, effects on standardized formative assessments are meaningful. Classroom components account for approximately 94% of instructional time, while CAL delivers outsized returns through targeted practice on specific elements directly connected to core content. This could provide a model of how CAL could be better integrated into core instruction: teachers deliver conceptual content and reasoning, while CAL provides specific just-in-time, focused and varied practice on content directly connected to the classroom instruction.

Explicitly aligning CAL to core content could reduce implementation barriers and increase adoption. Our study achieved 69% utilization in a large urban district, while most CALs suffer from much lower adoption rates and usage (Holt, 2024). Research on instructional coherence has emphasized aligning professional development, curriculum, and assessment (Wang et al., 2024) and more recently whole-class and small group tutoring (Jackson & Shakeel, 2025). We are unaware of work examining whether this principle extends to digital supplements—such as CAL—aligned to concurrent whole-class instruction. Unlike prior CAL research focused primarily on outcomes, alignment to core content might also address some adoption challenges and requires minimal additional class time, potentially offering a cost-effective alternative to hiring dedicated coordinators.

Finally, our results highlight a critical issue: potential bias in educational technology company-reported results. Since the COVID-19 pandemic, software adoption has surged in districts, yet the public conversation has been distorted by analyses showing dramatic student gains. Our analysis is among the first to quantify this overstatement. Prior work (Kraft, 2020) shows that researcher-developed measures produce effects two to four times larger than standardized assessments, as they typically align closely with intervention content. Combining the differential effects between our naïve and fixed effect estimate with the fact that most company-reported effects do not use standardized assessments suggests that simple estimates could be more than four times larger than what companies promise districts. More importantly, if we put confidence intervals on these estimates, it suggests that many of the eye-popping effect sizes could be null. This discrepancy may explain why CAL adoption and favorable meta-analyses coexist with declining NAEP scores. Districts should interpret company- or researcher-developed gains skeptically and prioritize results from rigorous evaluations using standardized assessments.

Limitations

Our study has four key limitations. First, students were not randomly assigned to use the digital activities. Although our student-by-period fixed-effects model reduces bias by comparing each student to themselves across time periods and controlling for seasonal effects, it cannot address all sources of bias. While our short evaluation window limits unobserved student-by-time interactions, potential differential learning gain between Fall and Winter and Winter and Spring could influence our results, though we believe it is reasonable to assume the associated bias to be small. We therefore interpret our findings as credible approximations of causal effects, while acknowledging this limitation. Randomized assignment in future studies would provide stronger causal evidence.

Second, prior meta-analyses suggest that CAL programs focused on more constrained skills produce larger gains than language-focused CAL (Silverman et al., 2025; Gilbert & Kim, 2026). In our study, all students received intensive language comprehension instruction during whole-class lessons, which may explain why we only observed effects on code-based measures—the CAL provides incremental practice on skills less covered in core instruction. Alternatively, code-based skills may be more amenable to improvement through digital practice because these skills are constrained – they have a ceiling. To distinguish between these explanations, future research should try to disentangle these effects. For example, studies could test whether increased CAL dosage on language skills produces gains when baseline classroom exposure is lower.

Third, although Table 1 shows the district serves a diverse student population, this study was conducted in the context of a long-standing research partnership that allowed for the alignment with core instruction and state standards. Replication across different districts, grade levels, and demographics would strengthen confidence in the generalizability of these findings.

Finally, this CAL was designed for concentrated, short-term use. Our analysis shows increasing dose-response effects: larger CAL dosages produce larger gains. However, we suspect these gains plateau or reverse at higher dosages. Because our implementation concentrated on a limited time window, we lack data to identify this inflection point. Future research should test whether higher-dosage use produces diminishing returns or harm.

Implications for Policy and Practice
Our findings support decades of research showing that students learn best when concepts and vocabulary are networked and reinforced across contexts. However, the implication is not simply that "aligning CAL to daily lessons helps students." Given the magnitude of the effects for this relatively small dosage, we believe this work complements a large body of work by suggesting targeted digital practice on densely networked, semantically related vocabulary—carefully selected by content experts and tied to content and literacy state standards—appears to strengthen students' understanding of untaught related words. While it might be tempting to conclude our results only apply to ELA-focused CALs, recent studies have shown how effective elementary-level literacy interventions translate into improvements in mathematics (Gilbert & Kim, 2026; Spencer, 2024; Kim et al. 2024). Thus, students benefit not because they happened to encounter an assessed word on the CAL on their standardized assessment, but because the CAL reinforced a cluster of interconnected concepts.

Two practical takeaways emerge: First, for elementary literacy instruction, districts can apply this principle beyond any single program by ensuring supplementary digital activities target conceptually related vocabulary within their existing curriculum. Second, emerging technologies like GraphRAG (Edge et al., 2025), which combines text extraction, network analysis and summarization, could automate the identification of semantic networks, making it feasible for educators and developers to design CALs this way at scale.

In addition, the field debates whether "screens" are beneficial or harmful, but this framing misses the point. Students use technology for vastly different purposes—writing, research, communication, or passive consumption—some of which are productive while others are likely detrimental. Rather than asking whether screen time is good or bad, districts should ask: What is CAL's comparative advantage that can complement classroom instruction? For example, a meta-analysis of a single CAL assessed across countries finds that adult interaction is a key moderator in effectiveness, even when on average the effects of the program were null (McTigue et al., 2020).

Our study suggests another answer for improving literacy outcomes in particular: targeted, repeated practice on narrowly defined skills. Skilled classroom teachers excel at building foundational comprehension, modeling reasoning, and fostering discussion—tasks where there is little evidence that digital delivery improves student achievement at scale. The productive question is not whether to use CAL, but how to allocate instructional time across teacher-led and technology-supported activities to maximize learning. This complementarity lens—largely absent in both research and policy—deserves greater attention, particularly because schools face persistent time constraints during the school day.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, *32*(8), 10–14. https://doi.org/10.3102/0013189X032008010

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). Longman.

Bureau of Labor Statistics, U.S. Department of Labor, *Occupational Outlook Handbook*, Kindergarten and Elementary School Teachers,
at
https://www.bls.gov/ooh/education-training-and-library/kindergarten-and-elementary-school-teachers.htm (visited *July 02, 2026*).

Cortes, K. E., Kortecamp, K., Loeb, S., & Robinson, C. D. (2025). A scalable approach to high-impact tutoring for young readers. *Learning and Instruction*, *95*, 102021.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American economic review*, *104*(9), 2593-2632.

Edge, D., Trinh, H., Newman Cheng, J. B., Chao, A., Mody, A., Truitt, S., ... & Larson, J. From local to global: A graph rag approach to query-focused summarization, 2025. *URL https://arxiv.org/abs/2404.16130*.

Escueta, M., Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, *58*(4), 897-996.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Gilbert, J. B., Domingue, B. W., & Kim, J. S. (2025). Estimating causal effects on psychological networks using item response theory. Psychological Methods. Advance online publication. https://dx.doi.org/10.1037/met0000764

Gilbert, J. B., & Kim J.S. (2026). Mapping the Mechanisms of Interdisciplinary Learning Transfer from Reading to Math Achievement: Evidence from a Large-Scale Randomized Controlled Trial. (EdWorkingPaper: 26-1376). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/nxqq-jc13

Gilbert, J. & Kim, J.S. (2026) Measures Still Matter: Replicating and Extending a Meta-Analysis of Educational Apps with Five Years of New Data. Paper presented at the Society for Research on Educational Effectiveness, Baltimore, MD

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In *Psychology of learning and motivation* (Vol. 16, pp. 59-109). Academic Press.

Holt, L. (2024). The 5 Percent Problem: Online mathematics programs may benefit most the kids who need it least. *Education Next*, Fall 2024.

Instructure. (2024). EdTech Top 40: A Look at K-12 EdTech Engagement During the 2023-24 School Year. https://www.instructure.com/resources/research-reports/edtech-top-40-look-k-12-edtech-engagement-during-2023-24-school-year

Instructure. (2026). New Instructure Data Shows K-12 Districts Are Demanding Evidence, Not Just Access to Edtech Tools. https://www.instructure.com/press-release/new-instructure-data-shows-k-12-districts-are-demanding-evidence-not-just-access.

Jackson, C. & Shakeel, A. (2025). Creating Coherence: Does Instructional Alignment Affect the Impact of Tutoring?. (EdWorkingPaper: 25-1332). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/en9j-xp53

Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2021). Improving reading comprehension, science domain knowledge, and reading engagement through a first grade content literacy intervention. Journal of Educational Psychology, 113(1), 3–26. https://doi-org /10.1037/edu0000465

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D., & McIntyre, J. (2023). A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension. Journal of Educational Psychology, 115(1), 73–98. https://doi.org/10.1037/edu0000751

Kim, J. S., Gilbert, J. B., Relyea, J. E., Rich, P., Scherer, E., Burkhauser, M. A., & Tvedt, J. N. (2024). Time to transfer: Long-term effects of a sustained and spiraled content literacy intervention in the elementary grades. Developmental Psychology. Advance online publication. https://doi.org/10.1037/dev0001710

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational researcher*, *49*(4), 241-253

Kraft, M. A., & Novicoff, S. (2024). Time in school: A conceptual framework, synthesis of the causal research, and empirical exploration. *American Educational Research Journal*, *61*(4), 724-766.

Marzano, R. J. (2004). A six-step process for teaching vocabulary. *Building background knowledge for academic achievement. Alexandria, VA: Association for Supervision and Curriculum Development*.

McKeown, M. G., & Beck, I. L. (2011). Making vocabulary interventions engaging and effective. In R. E. O'Connor & P. F. Vadasy (Eds.), *Handbook of reading interventions* (pp. 138–168). Guilford Press.

McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically reviewing GraphoGame across the world: Recommendations and cautions for research and implementation of computer‐assisted instruction for word‐reading acquisition. *Reading Research Quarterly*, *55*(1), 45-73.

Mosher, D. M., & Kim, J. S. (2024). Building a Science of Teaching Reading and Vocabulary: Experimental Effects of Structured Supplements for a Read Aloud Lesson on Third Graders' Domain-Specific Reading Comprehension. Scientific Studies of Reading, 1–25. https://doi.org/10.1080/10888438.2024.2368145

Oreopoulos, P., Keyes-Krysakowski, O., & Agarwal, D. (2026). *How In-School Supervised Ed-Tech Support Produces Massive Learning Gains: A Khan Academy Field Experiment in India* (No. w34683). National Bureau of Economic Research.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127-144.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of cognitive tutor geometry. *Journal of Research on Educational Effectiveness*, *3*(3), 254-281.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Robinson, C. D., Bisht, B., & Loeb, S. (2025). The inequity of opt-in educational resources and an intervention to increase equitable access. *Educational Researcher*, *54*(6), 328-338.

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA open*, *2*(4), 2332858416673968.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 335–360). Lawrence Erlbaum Associates.

Silverman, R. D., Keane, K., Darling-Hammond, E., & Khanna, S. (2025). The effects of educational technology interventions on literacy in elementary school: A meta-analysis. *Review of Educational Research*, *95*(5), 972-1012.

Spencer, N. (2024). Comprehensive early literacy policy and the "Mississippi Miracle". *Economics of Education Review*, *103*, 102598.

Tyner, A., & Kabourek, S. (2020). Social Studies Instruction and Reading Comprehension: Evidence from the Early Childhood Longitudinal Study. *Thomas B. Fordham Institute*.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, School Pulse Panel 2021–22, 2022–23, 2023–24, and 2024–25. https://ies.ed.gov/learn/press-release/more-half-public-school-leaders-say-cell-phones-hurt-academic-performance

Wang, Elaine Lin, Julia H. Kaufman, Sabrina Lee, Brian Kim, and V. Darleen Opfer, Instructional System Coherence: A Scoping Literature Review. Santa Monica, CA: RAND Corporation, 2024. https://www.rand.org/pubs/research_reports/RRA279-5.html.

Wright, T. S., Cervetti, G. N., Wise, C., & McClung, N. A. (2022). The impact of knowledge-building through conceptually-coherent read alouds on vocabulary and comprehension. *Reading Psychology*, *43*(1), 70-84.

Zipke, M. (2008) Teaching metalinguistic awareness and reading comprehension with riddles. The Reading Teacher, 62(2), 128-137.

Table 1: Student Demographics by Usage Categories

Usage	None	Light	Moderate	Full
Female	0.503	0.506	0.491	0.437
	(0.500)	(0.500)	(0.500)	(0.496)
Asian	0.054	0.069	0.087	0.098
	(0.226)	(0.254)	(0.281)	(0.298)
Black	0.358	0.268	0.271	0.214
	(0.479)	(0.443)	(0.445)	(0.411)
Hispanic	0.362	0.310	0.281	0.296
	(0.481)	(0.463)	(0.449)	(0.457)
White	0.038	0.046	0.051	0.052
	(0.192)	(0.210)	(0.220)	(0.222)
Other	0.188	0.307	0.311	0.339
	(0.391)	(0.461)	(0.463)	(0.474)
Students with Disabilities	0.142	0.145	0.112	0.095
	(0.350)	(0.352)	(0.315)	(0.294)
English Learner	0.295	0.262	0.224	0.238
	(0.456)	(0.440)	(0.417)	(0.426)
Low SES	0.432	0.344	0.341	0.298
	(0.495)	(0.475)	(0.474)	(0.458)
Medium SES	0.329	0.285	0.314	0.296
	(0.470)	(0.452)	(0.464)	(0.457)
High SES	0.231	0.359	0.334	0.398
	(0.422)	(0.480)	(0.472)	(0.490)
DIBELS Fall to Winter Percentile Change	0.409	0.689	0.973	1.152
	(10.977)	(10.030)	(10.938)	(10.453)
iReady Fall to Winter Percentile Change	-1.723	-2.181	-1.457	-1.465
	(12.030)	(12.411)	(12.309)	(12.081)
Number of Observations	4,737	1,130	2,842	1,366

Note: Table reports the means and standard deviations (in parentheses) for each student characteristic. Light usage corresponds to up to 10 minutes, moderate is 10 to 60 minutes, and full is more than 60 minutes per year. Students who used the CAL, but did not use the digital activities - ebook usage - are included in the zero usage category and represent approximately 16% of the no usage category. Both DIBELS and iReady report national percentile norms. The Fall to Winter percentile change represent the percentile point change from Fall to Winter. Sixty students took DIBELS, but not iReady in the sample. The summary statistics represent the students who took DIBELS.

Table 2. Average Effect of Computer-assisted learning Minutes on Standardized DIBELS Composite

	Model 1	Model 2	Model 3	Model 4
Up to 10 minutes	0.0232	0.0233	NA	
	(0.0161)	(0.0159)		
10 to 60 minutes	0.0728***	0.0600***	0.0354**	NA
	(0.0116)	(0.0154)	(0.0171)	
More than 60 minutes	0.113***	0.0846***	0.0611***	0.0256*
	(0.0150)	(0.0199)	(0.0192)	(0.0153)
R2	0.811	0.904	0.963	0.962
N	10,075	10,075	16,014	12,624
Winter Standardized Score	X			
Teacher FE		X		
Student FE Light, Moderate, Full Usage Only			X	
Student FE Moderate & Full Usage Only				X

Note: The table shows a progression of Models. Model 1 only includes a linear baseline winter score term. Model 2 adds teachers fixed effects (FE), demographic controls, and a flexible baseline functional form. Model 3 is a student fixed effects model with the 5,338 whose usage varied between Fall, Winter, and Spring assessments and Model 4 is the same fixed effect model, but includes only moderate and full usage student (N=4,208). Standard errors for Model 1 and 2 are clustered at the classroom level and at the student-level for Model 3 and 4. * p<0.10 ** p<0.05 *** p<0.01

Table 3. Average Effect of Computer-assisted learning Minutes on Standardized iReady Scores

	Model 1	Model 2	Model 3	Model 4
Up to 10 minutes	0.0101	-0.00110	NA	
	(0.0146)	(0.0182)		
10 to 60 minutes	0.0185*	-0.00689	-0.0122	NA
	(0.0105)	(0.0186)	(0.0162)	
More than 60 minutes	0.0215	-0.00489	-0.0155	-0.0033
	(0.0136)	(0.0246)	(0.0185)	(0.0150)
R2	0.816	0.863	0.908	0.962
N	10015	10015	15987	12624
Winter Standardized Score	X			
Teacher FE		X		
Student FE Light, Moderate, Full Usage Only			X	
Student FE Moderate & Full Usage Only				X

Note: The table shows a progression of Models. Model 1 only includes a linear baseline winter score term. Model 2 adds teachers fixed effects (FE), demographic controls, and a flexible baseline functional form. Model 3 is a student fixed effects model with the 5,329 whose usage varied between Fall, Winter, and Spring assessments and Model 4 is the same fixed effect model, but includes only moderate and full usage student (N=4,208). Standard errors for Model 1 and 2 are clustered at the classroom level and at the student-level for Model 3 and 4. * p<0.10 ** p<0.05 *** p<0.01

Table 4. Average Effect of Computer-assisted Learning Minutes Moderated by Baseline Proficiency

	iReady	DIBELS
	Model 1	Model 2
Up to 10 minutes	0.00249	0.00873
	(0.0227)	(0.0204)
10 to 60 minutes	-0.00597	0.0467***
	(0.0211)	(0.0172)
More than 60 minutes	0.00807	0.0815***
	(0.0273)	(0.0234)
Below BOY Benchmark	-0.0367*	-0.133***
	(0.0193)	(0.0170)
Up to 10 minutesXBenchmark	-0.00806	0.0337
	(0.0293)	(0.0245)
10 to 60 minutesXBenchmark	-0.000553	0.0318
	(0.0219)	(0.0193)
More than 60 minutesXBenchmark	-0.0415	-0.00516
	(0.0286)	(0.0291)
R2	0.864	0.905
N	10015	10075
Teacher FE	X	X

Note: Model 1 includes teachers fixed effects (FE), demographic controls, and a flexible baseline functional form for iReady. Model 2 is the same model for DIBELS. Standard errors are clustered at the classroom level. * p<0.10 ** p<0.05 *** p<0.01

Table 5. Cost Calculation: Computer-assisted learning

	Total	Fraction Related to Integrated CAL
Cost per Teacher	$380	$23.75
Opportunity Cost of Professional Learning	$22	$22
Subtotal Per Teacher	$402	$45.63
Average Class Size		20
Cost per Student		$2.28

Table A1: Distribution of CAL Minutes

	Minutes
10th Percentile	0
25th Percentile	0
50th Percentile	2.4
75th Percentile	31.2
90th Percentile	75.4