AMIRA
LEARNING

Authored in: May 2026

PROVEN TECHNOLOGY

Not Just Screen Time

Why the EdTech Reckoning Demands a New Framework—and How Active AI Tutoring Meets the Moment

Kelly Boden, PhD

EXECUTIVE SUMMARY

A neuroscientist's Senate testimony goes viral. Sixteen states introduce legislation to restrict classroom screens. The nation's second-largest school district votes to limit device time for its youngest students. The $30 billion EdTech industry faces a reckoning and the question is no longer whether technology belongs in schools, but whether we have the frameworks to distinguish technology that drives learning from technology that merely occupies attention.

This white paper argues that the current debate framing as "screen time" versus "no screen time" is the wrong framework. What matters is not the presence of a screen, but the nature of the interaction happening on it and the underlying theoretical framework and instructional design. We propose a distinction between **active** and **passive** digital instruction, a distinction between those products grounded in the cognitive science of learning and those not, and demonstrate how Amira Learning's AI-powered reading tutor exemplifies the active model grounded in cognitive science of learning: students perform the target skill (reading aloud), the platform assesses performance in real time, and instruction adapts at the point of need.

As Horvath himself writes in his Senate testimony: "This is not a debate about rejecting technology. It is a question of aligning educational tools with how human learning actually works." (Horvath, 2026a). Horvath's meta-analytic synthesis identifies Intelligent Tutoring Systems and Learning Disorder Interventions as the two categories that consistently show the strongest performance — with ITS showing a raw effect size of +0.52, well above the +0.40 to 0.50 threshold Horvath identifies as meaningful, and exceed the impact of typical classroom instruction (Horvath, 2026b; Hattie, 2023). By every criterion Horvath uses to define that exception—defined skill domain, adaptive practice, real-time corrective feedback—Amira qualifies.

We present evidence from multiple independent and large-scale studies to demonstrate that Amira's approach produces measurable, replicable learning gains, with relatively small amounts of screen time. These include two studies meeting ESSA Level II evidence standards (a third-party evaluation across nearly 80,000 Louisiana students and a matched comparison study of over 15,000 students on DIBELS in Texas), independent state evaluations by the University of Utah Education Policy Center and the Consortium for Policy Research in Education at Columbia University, and an upcoming randomized controlled trial with Boston University.

The students who benefit most from this distinction are those who can least afford to lose access to it: in the Louisiana study, 75% of the 79,000 students were economically disadvantaged, and the gains were largest among the students with the highest usage.

We conclude with recommendations for policymakers, district leaders, and the EdTech industry: blanket screen-time caps, while well-intentioned, risk eliminating proven interventions alongside ineffective ones. The right standard is not less technology, it is proven technology, evaluated against rigorous evidence standards, with transparent reporting of what students actually do during instructional time.

1. THE POLICY LANDSCAPE

1.1 The Viral Moment

In January 2026, cognitive neuroscientist Jared Cooney Horvath testified before the U.S. Senate Committee on Commerce, Science, and Transportation, arguing that the $30 billion spent putting laptops and tablets in schools has produced a generation less cognitively capable than their parents. His testimony has been viewed over 2 million times. His book, The Digital Delusion, is being carried into school board meetings across the country.

1.2 The Legislative Response

The policy response has been swift and broad. Alabama and Utah became the first states to enact classroom screen-time restrictions for early elementary students. As of the close of the 2026 legislative session, six states have enacted classroom screen time restrictions — Alabama, Tennessee, Utah, and Virginia among them — and at least 16 states introduced legislation this session. The movement is bipartisan: Missouri's House passed its bill with strong bipartisan support before the session ended (NPR, 2026; GovTech, 2026; Whiteboard Advisors, 2026). At the district level, the nation's second-largest Los Angeles Unified School District voted unanimously in April 2026 to ban district-issued devices for students through first grade, restrict one-to-one device use in elementary schools, and set maximum daily screen-time limits by grade level for the 2026–27 school year.

1.3 The Industry Response—and Its Limits

EdTech companies are responding to this moment by repositioning. Competitors are publishing blog posts framing themselves as "intentional" and "complementary to instruction." They recommend focused usage windows and emphasize that technology should support teachers, not replace them.

These are reasonable principles. But they sidestep the harder question: it is not enough to recommend 30 minutes a week. **What matters is what the student is doing during those 30 minutes, whether the company can demonstrate it is working, and by what mechanism the interaction produces learning.**

2. ACTIVE VS. PASSIVE DIGITAL INSTRUCTION: A FRAMEWORK

We propose that the EdTech field adopt a clear distinction between two fundamentally different types of digital instructional interaction - active vs passive. This framework is grounded in decades of cognitive science research on learning, memory, and skill acquisition.

2.1 PASSIVE DIGITAL INSTRUCTION

In passive digital instruction, the student is primarily a receiver of content. The device delivers information; the student consumes it. The adaptive element, where it exists, adjusts what content to show but does not require the student to actively perform the target skill or provide real-time corrective feedback on that performance.

Examples include: watching instructional videos, clicking through adaptive content sequences, selecting answers in multiple-choice assessments, and dragging objects across a screen. These interactions may be well-designed and pedagogically sound, but they do not require the student to produce the complex cognitive skill that constitutes the learning target.

2.2 ACTIVE DIGITAL INSTRUCTION

In active digital instruction,t he device does not do the cognitive heavy lifting for the child. Rather, the student **performs the target skill**, the platform **assesses that performance in real time**, and **instruction adapts at the point of need.** The student is not just consuming content, they are producing the complex cognitive skill the technology is designed to develop.

This distinction maps directly onto established principles in the Science of Reading and cognitive psychology: productive struggle, desirable difficulties, retrieval practice, and immediate corrective feedback are all features of active instruction. Horvath's own framework, which identifies "cognitive offloading" as the primary risk of classroom technology, supports this distinction. Active digital instruction is the opposite of cognitive offloading: it demands that the student do the cognitive work. Horvath himself acknowledges that narrowly constrained tools that include adaptive drills for foundational skills and targeted remediation consistently approach meaningful gains. His critique targets undifferentiated screen deployment, not purpose-built instructional tools with evidence of impact.

2.3 WHY THIS DISTINCTION MATTERS FOR POLICY

In January 2026, the American Academy of Pediatrics released its first updated screen time guidance in ten years — abandoning fixed time limits entirely in favor of a framework centered on quality, context, and the difference between active and passive use. The AAP now recommends that families and policymakers ask not how many minutes a child spends on a screen, but what the screen displaces and whether the interaction is designed around

2026). Legislation that reinstates the time-limit framework the AAP just retired — without distinguishing between a student reading aloud to an AI tutor and a student watching YouTube — moves policy in the opposite direction from where the nation's leading pediatric authority is pointing.

Current and proposed legislation has yet to catch up to this distinction. A 45-minute daily screen-time cap treats a student watching YouTube on a Chromebook identically to a student reading aloud to an AI tutor that is measuring their fluency growth in real time.For the 75% economically disadvantaged students in the Louisiana study described below—students who may not have access to a reading tutor at school or at home—that distinction has real consequences.

3. AMIRA LEARNING: MECHANISM OF ACTION

Amira Learning is an AI-powered reading tutor built from the ground up on established theoretical frameworks in neuroscience and the Science of Reading. Its technology traces back more than two decades to Carnegie Mellon University's Project LISTEN, one of the earliest and most extensively studied automated reading tutors in the research literature. These frameworks shape both the assessment and tutoring systems and were developed through active, ongoing collaborations with leading researchers in the field. This section details those foundational principles and then describes how they are operationalized in the platform's core technologies.

3.1 THEORETICAL FOUNDATIONS OF AMIRA ASSESS

Amira's assessment is grounded in three synergistic research frameworks.

The IDA Pragmatic Framework. In collaboration with Dr. Jack Fletcher, a consulting psychometrician to the International Dyslexia Association, Amira's screening constructs are aligned with the IDA's recommended protocols at each grade level. The constructs and items were created or reviewed by Fletcher's team and subjected to rigorous statistical and qualitative review for conformance to IDA guidelines. Amira's screener incorporates every mode of identifying reading risk recommended by the IDA for each grade band.

The Multiple Deficit Model. Drawing on the work of Drs. David Francis and Bruce Pennington, Amira incorporates the Multiple Deficit Model (MDM) as a neuroscience-based framework for understanding the origins of reading difficulty. Unlike earlier models that attributed dyslexia primarily to phonological deficits, MDM posits that multiple cognitive, genetic, and environmental factors interact to produce reading difficulties. Amira's assessment is organized to deliver constructs across all seven MDM dimensions at each grade level: phonological processing, rapid automatized naming, orthographic processing, working memory, processing speed, language skills, and reading fluency (Pennington, 2006; Francis et al.).

Duke's Active View Framework. In collaboration with Dr. Nell Duke, Amira employs the Active View as the theoretical basis for ensuring comprehensive coverage of reading mastery. The Active View extends earlier models — the five pillars of reading, the Simple View of Reading, and Scarborough's Reading Rope — by emphasizing that successful reading comprehension involves active self-regulation, word recognition, bridging processes, and language comprehension. Amira evaluates students within each of these four major domains, providing a more holistic assessment than screeners focused solely on word recognition.

A Developmental Approach. Amira's constructs vary across grade levels, reflecting the recognition that reading skills develop along a continuum. Kindergarten and early first-grade assessments focus on foundational skills such as phonological awareness, letter knowledge, and early decoding. Later grades incorporate advanced decoding, fluency, and comprehension measures. This developmental progression reflects established models of reading acquisition (Chall, 1983; National Reading Panel, 2000) and ensures the assessment is both appropriate and predictive at each stage.

3.2 THEORETICAL FOUNDATIONS OF AMIRA TUTOR

The tutoring system is grounded in the Science of Reading and organized around several core principles that Amira identifies as central to its instructional design.

Scarborough's Reading Rope. Amira Tutor's instructional interactions are organized around Scarborough's Reading Rope, which identifies the interwoven strands of word recognition (decoding, phonological awareness, sight recognition) and language comprehension (vocabulary, background knowledge, verbal reasoning) that together produce skilled reading. The platform tracks student mastery across each strand and uses this profile to inform text selection, micro-intervention targeting, and instructional focus.

Guided Oral Reading with Feedback. The National Reading Panel (2000) identified guided oral reading with feedback as one of the most effective instructional approaches for developing fluency. Amira operationalizes this at scale: the student reads aloud, the AI detects errors in real time, and corrective feedback is delivered at the point of need. This approach is consistent with broader learning science research on the value of immediate corrective feedback and effortful retrieval in strengthening skill acquisition.

Adaptive Practice and Explicit Phonics. Amira's micro-intervention system selects the highest-leverage errors to address based on the student's skill profile and the linguistic context of the error, rather than intervening on every miscue. Text complexity is adjusted dynamically based on session performance. The platform's instructional interactions incorporate explicit phonics instruction, decoding support, and systematic skill development aligned with Science of Reading principles.

Knowledge Building Through Core Knowledge-Aligned Text Sets. Amira's text sets are aligned to Core Knowledge topics and district curriculum scope and sequence. Rather than assigning isolated passages, Amira sequences texts around coherent knowledge domains, enabling students to build the content knowledge that supports deeper comprehension. This approach reflects a growing body of research emphasizing the role of background knowledge in reading comprehension.

3.3 CORE AI TECHNOLOGIES

Reading Error Detection (RED) Model

When a student reads aloud, Amira's RED model analyzes their speech patterns to identify specific reading challenges. Unlike generic speech recognition, the RED model is specifically designed to evaluate reading skill, distinguishing between genuine reading errors and natural speech variation due to developmental factors, accent, dialect, or other factors unrelated to reading ability. The system was developed over five years in collaboration with experts from Carnegie Mellon and Johns Hopkins Universities, achieving 96% agreement with human expert judgment which is equivalent to the inter-rater agreement between two expertly trained teachers scoring the same reading.

Micro-Intervention Selection Model

Working in tandem with the RED model, this system analyzes detected errors in real time, evaluates the context of the student's struggle, considers the student's historical performance and skill mastery profile, examines linguistic properties of the text, and selects the most appropriate micro-intervention based on reading science principles. Amira does not intervene on every miscue; rather it selects the highest-leverage errors to address, mirroring the judgment of an expert reading tutor.

Comprehension Conversations

After completing a passage, Amira engages students in dynamic, AI-led Socratic dialogues that prompt critical thinking, inference-making, summarization, and evidence citation. These conversations mirror authentic classroom discussion and are aligned to research-based comprehension strategies.

3.4 WHAT A TUTORING SESSION LOOKS LIKE

Every session is an active performance of the skill itself. A student reads aloud, the AI listens in real time, evaluates fluency, accuracy, and comprehension, identifies specific skill gaps, and adjusts the next passage accordingly. The student is doing the thing the technology is designed to teach, reading connected text, and the platform is measuring and responding to that performance in the moment.

Amira sits in the space between a human reading tutor and a standardized assessment, capturing the kind of data that used to require a trained specialist sitting one-on-one with a child, running records in hand. It does this automatically and at scale.

3.5 ASSESSMENT: FORMAL SCREENING AND CONTINUOUS PROGRESS MONITORING

The same core principle that defines Amira's tutoring—the student reads aloud, the AI listens and evaluates in real time—also powers a comprehensive assessment system. Amira's assessment operates at two levels, both grounded in active oral reading performance rather than passive multiple-choice selection:

Amira Assess is a formal screening and benchmarking tool administered within defined assessment windows (beginning, middle, and end of year). When a student reads aloud during a benchmark assessment, the platform produces a psychometrically validated Amira Reading Mastery (ARM) score, oral reading fluency measures (WCPM, accuracy, prosody), a dyslexia risk indicator, a skills diagnostic mapped to Scarborough's Reading Rope, and standards mastery scores aligned to state-specific grade-level expectations. These are not proxy metrics derived from click patterns. They are direct measures of reading ability captured from the student's reading performance itself.

Continuous progress monitoring operates at two levels. Each time a student reads with Amira Tutor, the platform updates the student's estimated mastery profile, tracks skill development across the components of the Reading Rope, and feeds data back into the instructional model. In addition, teachers can assign formal progress monitoring assessments at any point during the year to capture a standardized snapshot of a student's reading growth between benchmark windows. Together, these two layers mean districts receive ongoing, assessment-grade data on every student's reading trajectory without sacrificing instructional time for standalone testing.

Both layers use the same underlying AI—the Reading Error Detection model and speech recognition engine developed at Carnegie Mellon. And both require the same active interaction: a student performing the skill of reading, not selecting answers on a screen. This is assessment as it should work in a Science of Reading framework, listening to a child read and drawing precise, actionable conclusions from that performance.

3.6 BILINGUAL CAPABILITIES

Amira supports students in both English and Spanish, offering assessment, instruction, and tutoring in both languages. Students can read in English while receiving scaffolding and directions in Spanish, supporting English Learners' access to literacy practice with appropriate linguistic support.

3.7 MTSS ALIGNMENT

Amira is designed to operate across all three tiers of a Multi-Tiered System of Supports: universal screening and supplemental support at Tier 1, targeted intervention with increased precision at Tier 2, and individualized intensive support at Tier 3. This alignment matters for funding: Amira's uses are allowable under Title I, Title II, <rectangle>Title III, Title IV, and IDEA.

4. EVIDENCE OF IMPACT

Amira's evidence base spans more than two decades and includes peer-reviewed publications, independent third-party evaluations, state department of education analyses, and large-scale quasi-experimental studies meeting federal evidence standards.

Interpreting the Evidence: The Hattie Benchmark

Effect sizes require context to be meaningful. John Hattie's large-scale synthesis of education research — the same body of evidence Horvath cites in his Senate testimony to dismiss most EdTech — establishes d = 0.40 as the hinge point above which an intervention meaningfully exceeds what typical classroom instruction already delivers (Hattie, 2023). Horvath uses this threshold to argue that general-use EdTech underperforms. As the studies below demonstrate, Amira's evidence base tells a different story — with effect sizes in independent evaluations that meet and, at high usage levels, exceed that threshold. The same benchmark used to indict general EdTech is the benchmark that distinguishes active AI tutoring from it.

4.1 LOUISIANA ESSA LEVEL II STUDY (2023–24)

Evaluator: Instructure (third-party)

Sample: 79,084 students (39,542 treatment, 39,542 matched comparison) across 12 Louisiana school districts

Design: Quasi-experimental matched-comparison, aligned to ESSA Level II (Moderate Evidence) standards

Outcomes: DIBELS composite (K–3) and LEAP ELA (Grades 4–5)

Methods: Multilevel models with district-level random effects and student-level covariates (baseline achievement, sex, race/ethnicity, SES, ELL status, special education designation). Baseline equivalence met WWC Version 5.0 standards at all grade levels.

Key Findings: Statistically significant positive associations between Amira usage and literacy outcomes at every grade level (K–5). High-use kindergarten students performed at the 58th percentile vs. 50th percentile for matched non-users ($g = 0.21$, $p < .001$). High-use Grade 1 students performed at the 59th percentile ($g = 0.22$, $p < .001$). Clear dose-response patterns across all grades.

Demographics: 75% economically disadvantaged, 38% African American, 18% Hispanic, 10% ELL, 14% special education.

4.2 TEXAS DIBELS ESSA LEVEL II STUDY (2024–25)

Evaluator: Amira Research Team, Reviewed by Johns Hopkins University

Sample: 15,424 students (K–1) across 174 schools in a large urban Texas districts

Design: Matched comparison group with one-to-one nearest-neighbor matching on BOY DIBELS percentile, gender, race/ethnicity, economic disadvantage, SPED, ELL, chronic absenteeism, and campus type. Meets ESSA Tier 2 moderate evidence standards.

Outcomes: DIBELS percentile ranks

Methods: Hierarchical linear models with school-level clustering and student-level covariates (BOY DIBELS percentile, gender, race/ethnicity, economic disadvantage, special education, ELL status, chronic absenteeism). Baseline equivalence met.

Key Findings: Kindergarten students using Amira scored approximately 8 percentile points higher than matched controls on EOY DIBELS (effect size = 0.26, $p < .05$). First grade students scored approximately 2 percentile points higher. The sample was 90% economically disadvantaged in kindergarten and 86% in first grade, with approximately two-thirds Hispanic and one-quarter Black or African American.

4.3 UTAH STATE EVALUATION (2024–25)

Evaluator: Utah Education Policy Center, University of Utah (independent, commissioned by the Utah State Board of Education)

Sample: 8,222 students in K–3 across 97 schools and 7 LEAs

Design: Covariate balancing propensity score weighting with dose-response modeling, controlling for BOY Acadience Reading scores, demographics, and school-level characteristics

Outcomes: EOY Acadience Reading composite scores

Key Findings: Statistically significant positive dose-response relationships between Amira usage and Acadience Reading performance in kindergarten, 2nd grade, and 3rd grade. Effect sizes at the 90th percentile of usage were as high as 0.56 (kindergarten) compared to zero usage.

4.4 SAVANNAH-CHATHAM COUNTY PUBLIC SCHOOLS (2020–21)

Evaluator: Consortium for Policy Research in Education (CPRE) at Teachers College, Columbia University (independent, third-party)

Sample: 2,305 first and second grade students across 29 schools, with additional sub-analyses for kindergarten and third grade

Design: ANCOVA with classroom fixed effects, controlling for baseline literacy scores, demographics, special education status, and minutes read per week

Outcomes: Oral reading fluency (WCPM), vocabulary size, sight recognition (ESRI), phonological awareness, and Lexile scores

Key Findings: High-usage students (10+ weeks) gained 0.25 to 0.45 SD more than low-usage peers across all five literacy outcomes during the fall-to-winter period. Each additional week of usage was associated with approximately 0.03 SD additional growth in WCPM. Dose-response patterns were consistent across all outcomes. The study was conducted during COVID-19, with implementation shifting from remote to in-person mid-year, and the majority of students received less than half the recommended dosage, suggesting these may be conservative estimates of Amira's potential.

4.5 ADDITIONAL EVIDENCE

Texas (2024-25): 36,123 students in Grades 3–5 across 130 Texas districts. Students with high usage scored 6 percentile points higher on Texas STAAR than low-usage peers (d = 0.32), with the strongest effects among economically disadvantaged and Black students.

North Dakota (2023–24): Students in the highest usage quartile scored up to 15 points higher on the NDSA ELA than lowest-usage peers.

Heritage Research: Amira's foundational technology (Project LISTEN, Carnegie Mellon University) has been studied since 2000, with peer-reviewed publications demonstrating effectiveness comparable to certified human tutors, producing effect sizes that range from 0.38 to 1.82 across various reading domains and student populations.

4.6 UPCOMING: BOSTON UNIVERSITY RANDOMIZED CONTROLLED TRIAL

In March 2026, Boston University's Evidence-Based AI in Learning (EVAL) Industry Collaborative selected Amira Learning as the winner of its inaugural EdTech Evaluation Challenge—a national competition that drew 32 applicants. Amira was selected for its grounding in learning science and readiness for large-scale classroom evaluation. The resulting study will be a rigorous randomized controlled trial designed by a cross-disciplinary BU research team spanning education, neuroscience, computer science, economics, and public health. Critically, the research will be fully independent, transparent, and published regardless of outcome.

This represents the next step in Amira's evidence trajectory: moving from quasi-experimental evidence meeting ESSA Level II standards toward gold-standard RCT evidence at ESSA Level I.

4.7 NORTHERN IRELAND

Amira's independent evaluation portfolio now extends beyond U.S. borders. In June 2025, the Northern Ireland Department of Education announced it will fund a national matched-pair research study evaluating the impact of AI-powered literacy intervention on reading outcomes — selecting Amira as the tool and commissioning Oxford Brookes University as the independent evaluator. The study, conducted under Northern Ireland's RAISE initiative, focuses specifically on outcomes for disadvantaged pupils and students with Special Educational Needs and Disabilities. More than 15,000 students will participate, beginning with a 15-minute baseline literacy assessment and then engaging in weekly sessions with Amira over the course of the academic year. Results are expected to directly inform national literacy policy in Northern Ireland.

The significance of this partnership extends beyond a single study. A sitting national Department of Education choosing to fund independent research on an AI literacy tool — rather than relying on vendor-supplied evidence — reflects precisely the standard this paper advocates: government investment in rigorous, independent evaluation of tools before they are adopted at scale. That the department selected a matched-pair design, focused the study on the most vulnerable learners, and chose an independent university as evaluator makes it a model of the procurement accountability this white paper calls for.

5. PRIVACY, SAFETY, AND ETHICAL DESIGN

A common concern in the screen-time debate is student surveillance and data privacy. Amira's architecture addresses these concerns directly through Safety by Design and Privacy by Design frameworks embedded into the platform's core.

5.1 NOT SURVEILLANCE—STRUCTURED INSTRUCTIONAL INTERACTION

Amira integrates exclusively with school-managed digital environments. The platform only monitors when a student is actively engaged in an Amira session—there is no passive listening, no background data collection, and no data captured unless a student is logged in and in an active session.

5.2 PRIVACY COMPLIANCE

Amira maintains full compliance with FERPA, COPPA, and CIPA. Student data is never used for advertising, marketing, or any commercial purpose unrelated to the educational service. All data is encrypted in transit (TLS 1.2+) and at rest (AES-256), with logical segregation between districts and regular third-party SOC 2 audits.

5.3 ETHICAL AI AND BIAS MITIGATION

Amira's speech recognition and NLP models are designed to recognize and accept diverse accents and dialects, accommodate developmental speech patterns, and show equivalent accuracy across gender, racial, and ethnic groups. Regular evaluations confirm accuracy rates exceeding 95% for each demographic subgroup.

6. IMPLICATIONS FOR POLICY AND FUNDING

6.1 The Risk of Blunt Instruments

Blanket screen-time caps, applied uniformly across all technology regardless of purpose or evidence, risk eliminating proven interventions alongside ineffective ones. A policy that does not distinguish between a student watching YouTube on a Chromebook and a student reading aloud to an evidence-based AI tutor will disproportionately harm the students who benefit most from structured digital instruction—students in under-resourced schools who may not have access to a reading tutor at home.

These caps also threaten assessment. Over 40 states now mandate universal dyslexia screening in early grades, and many districts rely on digital tools to administer those screeners efficiently. Amira Assess satisfies these screening requirements through the same active oral reading interaction that powers the tutor—producing dyslexia risk indicators, fluency measures, and skills diagnostics from a single session in which a student reads aloud. A district that caps screen time without distinguishing between instructional technology and assessment technology may find itself unable to conduct the very screening that state law requires. The same legislation intended to protect students could prevent districts from identifying the students who need the most help.

6.2 FUNDING AT RISK

Amira's uses are allowable under Title I, Title II, Title III, Title IV, IDEA, and Comprehensive Literacy State Development grants. Legislation that restricts classroom technology without evidence-based exemptions could prevent districts from deploying federally funded literacy interventions that meet ESSA evidence standards—the very standards that Congress established to ensure educational investments are backed by rigorous research.

6.3 ACCOUNTABILITY BEYOND EVIDENCE: OUTCOMES-BASED CONTRACTING

A growing number of states and districts are adopting outcomes-based contracting (OBC) as a mechanism for holding EdTech vendors accountable for student learning. The Center for Outcomes Based Contracting, a national initiative of the Southern Education Foundation, defines OBC as a model "in which a substantial part of payment to a service provider is contingent on meeting agreed-upon student outcomes" — shifting procurement from a transactional license model to a mutual accountability partnership (Center for Outcomes Based Contracting, 2025).

Amira's contracting model aligns with the Center for OBC's Standards of Excellence for Outcomes Based Contracting™. Under this framework, contracts specify the research-based dosage required to achieve impact, define the student outcome measures against which performance will be assessed, and structure a portion of vendor compensation as contingent on those outcomes being met. The model requires that both parties — district and provider — commit to an implementation plan before the tool enters the classroom. Independent research on the first edtech cohort to use this model found dosage rates averaging 69%, nearly 14 times the 5% industry baseline, alongside deeper district-provider partnerships and meaningful shifts toward data-driven accountability. Southerneducation

Outcomes-based contracting aligns vendor incentives with student learning in a way that seat-license procurement cannot. When a portion of compensation is tied to demonstrated growth rather than adoption metrics, product development prioritizes efficacy over engagement, implementation support focuses on dosage compliance rather than renewal, and districts gain a partner whose financial interests are structurally aligned with their instructional goals. If policymakers are serious about ensuring technology serves learning, OBC should be part of the policy conversation — not just as a business model, but as a procurement standard that rewards results over promises.

6.4 RECOMMENDATIONS

For Policymakers

- Distinguish between active and passive digital instruction in legislation based on the
- nature of the student interaction, not merely the presence of a device.
- Require EdTech tools used in schools to meet recognized evidence standards (e.g., ESSA Level II or higher) within a reasonable implementation period.
- Exempt evidence-based instructional tools from blanket screen-time restrictions,
- provided they meet defined evidence and transparency thresholds.
- Encourage independent evaluation of educational tools, with results published regardless of outcome.

Several states have already demonstrated how to draft screen time restrictions that protect students without eliminating proven interventions. Tennessee's 2026 legislation exempts "targeted instructional support, intervention or remediation" — language broad enough to cover adaptive tools meeting ESSA evidence standards while excluding passive screen use. Alabama and Tennessee both carve out blanket protections for students with disability plans. Policymakers in states currently drafting legislation can build on these precedents. A functional framework includes three elements: (1) a definition of "active instructional use" tied to direct skill performance rather than content delivery; (2) an exemption for tools meeting ESSA Level II or higher evidence standards; and (3) a universal exemption for mandated assessments, including state-required dyslexia screening. Without the third element, states risk the specific contradiction of mandating screening by law while simultaneously preventing districts from conducting it.

For District Leaders

- Evaluate EdTech tools based on what the student is actually doing during instructional time, not time-on-task alone, but the nature and quality of the interaction.
- Demand effect sizes, comparison groups, and independent outcome measures from EdTech vendors, not self-reported engagement metrics or testimonials.
- Prioritize tools that generate assessment-grade data as a byproduct of instruction, reducing the need for separate testing windows and enabling continuous progress monitoring.
- Explore outcomes-based contracts that tie vendor compensation to measurable student growth, shifting the financial risk from the district to the provider.

For the EdTech Industry

- Submit products to independent, third-party evaluation and publish results regardless of outcome.
- Report transparently on what students do during product use, not just how long they use it, including whether the interaction requires active skill performance or passive content consumption.
- Align business models with evidence by offering outcomes-based contracts, demonstrating confidence in the product's ability to deliver measurable learning gains.
- Adopt transparent evidence reporting as a standard practice, including publication of study designs, effect sizes, and limitations alongside product claims

7. CONCLUSION

The EdTech reckoning is here, and it is overdue. For too long, the industry operated on good intentions rather than good evidence. The skepticism from parents, educators, and policymakers is earned.

But the path forward is not retreat from all technology. It is a higher standard, one that demands proof of impact, transparency about mechanism, and accountability for outcomes. As Horvath told the Senate: “This is not a debate about rejecting technology. It is a question of aligning educational tools with how human learning actually works.”

Amira Learning was built for this standard. From its origins in university research to its current deployment serving thousands of school districts, the platform demonstrates what it looks like when an EdTech company does the thing the field is now demanding: requires students to actively perform the skill being taught, measures that performance with assessment-grade precision, adapts instruction in real time, proves its impact through independent evaluation, and stakes its revenue on results.

REFERENCES

Altermatt, T. W., Gallyer, J., Altermatt, E. R., Yildiz, M., & Rorrer, A. K. (2025). Implementation and Outcome Evaluation of the Early Interactive Reading Software Program (EISP): Amira Report. Salt Lake City, UT: Utah Education Policy Center, University of Utah.

American Academy of Pediatrics. (2026, January). Digital ecosystems, children, and adolescents [Policy statement and technical report]. **https://www.aap.org/en/patient-care/media-and-children/**

Amira Learning. (2025). Amira 24–25 STAAR Analysis. San Francisco, CA.

Boston University Human Intelligence + Computing Lab. (2026, March 3). Boston University's EVAL Collaborative names Amira Learning winner of inaugural EVAL EdTech Evaluation Challenge. **https://www.bu.edu/hic/2026/03/03/**

Center for Outcomes Based Contracting. (2025). Standards of Excellence for Outcomes Based Contracting™: EdTech instructional interventions. Southern Education Foundation. **https://obc.southerneducation.org/standards-of-excellence-edtech-intervention/**

Consortium for Policy Research in Education at Teachers College, Columbia University. (2021). Amira Learning in Savannah-Chatham County Public Schools. New York, NY.

Digital Promise. (2026, March). Exploring outcomes based contracting for EdTech [Report]. **https://digitalpromise.org/2026/03/16/new-report-exploring-outcomes-based-contracting-for-edtech/**

GovTech. (2026, March 19). Parents, states push back against screens in class. **https://www.govtech.com/education/k-12/parents-states-push-back-against-screens-in-class**

Hamilton Brookes / Oxford Brookes University. (2025). Reading with AI research project in Northern Ireland. **https://www.hamilton-trust.org.uk/blog/oxford-brookes-to-lead-major-research-study-on-reading-with-ai-in-northern-ireland/**

Hattie, J. (2023). Visible learning: The sequel. New York.

Horvath, J. C. (2026a, January 15). Written testimony before the U.S. Senate Committee on Commerce, Science, and Transportation. **https://www.commerce.senate.gov/wp-content/uploads/media/doc/Horvath_Written%20Testimony.pdf**

Horvath, J. C. (2026b). The Digital Delusion: How Classroom Technology Harms Our Kids' Learning—and How to Help Them Thrive Again. LME Global Press.

Horvath, J. C. (2026c, February 13). So, when is EdTech good for learning? The Digital Delusion (Substack). **https://thedigitaldelusion.substack.com/p/so-when-is-edtech-good-for-learning**

Instructure. (2025). Amira Learning in Louisiana public schools: ESSA Level II Study (2023–24). Prepared for Amira Learning.

Jeans, M. (2025). Using Amira to Increase Performance on DIBELS Reading Outcomes. Amira Learning.

Los Angeles Unified School District Board of Education. (2026, April 22). Resolution to establish screen time limits for district-issued devices. As reported by K-12 Dive: **https://www.k12dive.com/news/lausd-imposes-screen-time-limits-starting-in-2026-27/818224/**

Northern Ireland Department of Education. (2025, June 4). Northern Ireland Department of Education to launch research study to evaluate AI's role in advancing literacy for students [Press release]. **https://www.prnewswire.com/news-releases/northern-ireland-department-of-education-to-launch-research-study-to-evaluate-ais-role-in-advancing-literacy-for-students-302473722.html**

NPR. (2026, May 1). Several states — and the LA public schools — are setting limits on screen time. **https://www.npr.org/2026/05/01/nx-s1-5791657/states-schools-restricting-screen-time**

Whiteboard Advisors. (2026). Screen time legislation is moving fast. Here's what's actually enacted. **https://whiteboardadvisors.com/screen-time-legislation-is-moving-fast-heres-whats-actually-enacted/**

For more information about Amira Learning's research and evidence base, vist amiralearning.com/research or contact info@amiralearning.com.

amiralearning.com | info@amiralearning.com | 866-883-7323
490 Post St, Suite 500 PMB 2538, San Francisco, CA 94102