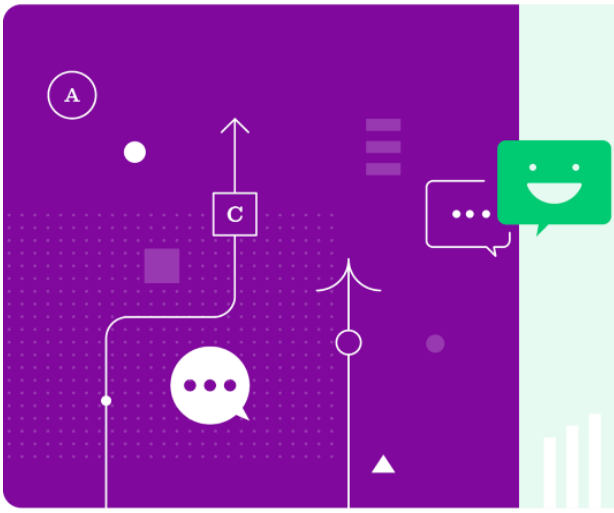
AMIRA
LEARNING

Using Amira to Increase Performance on DIBELS Reading Outcomes

Amira Learning | Every Child Deserves the Chance to Become a Reader
5214F Diamond Heights Blvd, #3255, San Francisco, CA 94131
wwww.AmiraLearning.com | 866-883-7323 | info@amiralearning.com



THE INTELLIGENT
Growth Engine

Using Amira to Increase Performance on DIBELS Reading Outcomes

PREPARED BY
Matthew Jeans, PhD

DATE
July 2025

Executive Summary

Amira is an advanced AI-powered literacy assessment and tutoring platform that uses voice recognition and machine learning to deliver personalized reading practice, real-time feedback, and targeted skill development. Early literacy is a key predictor of future academic success, and interventions in the earliest grades can have lasting effects. This evaluation examined the effectiveness of Amira Tutor in improving reading performance, as measured by Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 8th Edition, among early elementary students in a large, urban Texas district during the 2024–25 school year.

Using a rigorous matched comparison group design with over 15,000 students from 174 schools equivalent at baseline on BOY DIBELS percentile scores, demographic characteristics, and other covariates, hierarchical linear models (HLM) accounted for clustering at the school level. Attrition rates were low and balanced between groups.

- Kindergarten students using Amira Tutor scored approximately **8 percentile points higher** than matched control students on the EOY DIBELS assessment, resulting in an **effect size of 0.26**.
- First grade students using Amira Tutor scored approximately **2 percentile points** higher than matched control students on the EOY DIBELS assessment, resulting in an **effect size of 0.06**.

This quasi-experimental study meets the criteria for Tier 2 moderate evidence under the Every Student Succeeds Act (ESSA), strengthening the case for Amira Tutor as an evidence-based intervention for early literacy. The results demonstrate that the usage of Amira Tutor, regardless of dosage, positively impacts DIBELS performance, particularly in diverse, economically disadvantaged early elementary students. These findings align with prior research validating AI-driven literacy interventions as transformative tools in education.

Table of Contents

Introduction

Early identification and targeted support are foundational to literacy development, especially in the earliest years of schooling. Kindergarten and first grade represent critical windows where timely intervention can significantly alter students' reading trajectories. To meet this need, Amira Learning provides a suite of AI-powered tools designed to deliver both high-quality assessment and adaptive instructional support. This study evaluates the effectiveness of Amira's platform in accelerating literacy outcomes for young learners, with a focus on kindergarten and first grade students in a large urban school district in Texas.

Amira's platform consists of three integrated components: ISIP Assess, Tutor, and Instruct. ISIP Assess is an adaptive, oral reading benchmark that is administered three times per year. It provides percentile scores, diagnostic flags, and reading level estimates to guide instructional decisions (Amira Learning, 2023). Tutor offers individualized oral reading coaching, using AI to deliver real-time modeling, feedback, and support as students read aloud. This component is designed to improve fluency through immediate corrective feedback and scaffolded practice (Amira Learning, 2024b). Instruct is Amira's adaptive instructional engine, which personalizes a path of digital learning activities based on each student's assessment results and reading needs. Drawing on evidence-based instruction, Instruct delivers systematic practice in phonics, phonemic awareness, comprehension, and vocabulary, aligned with early literacy standards.

Extensive research has established Amira Tutor's efficacy across multiple literacy domains and student populations. Studies have shown that Amira significantly improves oral reading fluency and comprehension, with users demonstrating nearly twice the rate of fluency growth compared to non-users in New York and Nebraska (Amira Learning, 2022). The AI-guided practice provided by Amira has also been found to outperform independent reading in developing comprehension and word recognition (Mostow et al., 2013).

Amira has demonstrated positive impacts among diverse learners, including English Language Learners (ELLs), bilingual students, and students from under-resourced communities (Amira Learning, 2024a; Korsah et al., 2010; Poulsen & Wiemer-Hastings, 2007). For example, bilingual students in Canada showed the greatest gains among all language groups, and ELLs in the U.S. have benefitted from Amira's real-time adaptive support in English, Spanish, or bilingual modes (Amira Learning, 2024a; Reeder et al., 2015). At-risk students in Utah and low-income students in Ghana have also shown

measurable improvements with Amira use, highlighting its utility across cultural and economic contexts (USBE, 2023; Mills-Tettey et al., 2005).

The present study builds on the foundation of research by examining the impact of Amira Tutor in a real-world district implementation, using a rigorous design and multilevel modeling to evaluate its impact on early literacy growth.

Methodology

Study Design

This study meets ESSA Tier 2 evidence standards for moderate evidence, based on its matched comparison design, rigorous control of confounding variables, and statistically significant positive impacts on a standardized outcome measure. Students were matched prior to the administration of outcome measures using a one-to-one, nearest neighbor approach without replacement. Matching was conducted within grade level to ensure comparability across developmental stages. Students in the treatment group engaged with Amira's literacy platform during the 2024–2025 school year, while those in the control group did not receive exposure to the platform.

Matching was performed on a comprehensive set of covariates selected to address potential confounding factors related to literacy development. These included beginning-of-year (BOY) DIBELS reading percentile scores, gender, race/ethnicity, economic disadvantage status, special education (SPED) status, ELL status, chronic absenteeism, and campus type, specifically whether the school was part of the district's New Education System (NES). Baseline equivalence between treatment and control groups was assessed following matching to verify balance on key characteristics.

Setting

As aforementioned, students were matched on campus type (i.e., NES vs. non-NES). NES campuses implement a comprehensive reform model that includes centralized instructional support, expanded enrichment, and intensive coaching for educators. A key difference between NES and non-NES campuses is that Amira usage is systematically embedded into the school day on NES campuses, where usage is more prescribed as part of the instructional model. Students in both treatment and control groups attended NES and non-NES schools. While campus type was used in matching and initially tested as a covariate, it was excluded from final models due to a lack of statistical significance.

Measures

DIBELS

The outcome of interest was the EOY percentile score from DIBELS 8th Edition, a standardized assessment system designed to measure the acquisition of early literacy skills from kindergarten through sixth grade. The assessment includes brief, individually administered subtests that evaluate key components of reading development, including phonemic awareness, phonics, fluency, vocabulary, and comprehension.

In this study, students' performance was summarized using national norm-referenced percentile scores, which indicate each student's standing relative to a nationally representative peer group.

Amira Tutor

Amira Tutor is a research-based, AI-driven instructional program designed to improve literacy outcomes by providing personalized reading practice and support. Unlike traditional assessments, Amira Tutor serves as a one-on-one digital tutor, using advanced artificial intelligence to deliver targeted instruction in real-time. The system engages students in oral reading activities, providing immediate feedback on fluency, pronunciation, and comprehension while tracking progress over time.

The tutor incorporates tasks aligned with foundational literacy skills, including phonological awareness, phonics, decoding, oral reading fluency, and vocabulary acquisition (Amira Technical Guide, 2024b). By leveraging voice recognition and machine learning, Amira Tutor listens to students read aloud, identifies errors, and provides corrective feedback tailored to their specific needs. These interactions are informed by principles from the science of reading and are further enriched with diagnostic reports that offer actionable insights for educators and parents.

Key features of Amira Tutor include:

- **Immediate Feedback**: Amira analyzes students' reading in real time, helping them self-correct errors and improve fluency.
- **Adaptive Challenges**: The platform adjusts the difficulty of reading passages and exercises based on student performance.
- **Progress Monitoring**: Detailed, automatically generated reports highlight student strengths and areas for improvement, empowering educators to make data-driven instructional decisions.

- **Evidence-Based Design**: Developed in collaboration with leading institutions such as Carnegie Mellon University and Johns Hopkins University, Amira is rooted in validated literacy research and predictive accuracy for learning outcomes.

Analytical Approach

HLM was used to account for the nested structure of the data, with students clustered within schools. Separate models were estimated for kindergarten and first grade cohorts. Three models were tested in each grade:

1. Model 1 (Null Model): Estimated the intraclass correlation coefficient (ICC) to quantify between-school variance.
2. Model 2 (Fixed Effects Model): Included treatment status and all matched covariates.
3. Model 3 (Random Slopes Model): Added a random slope for the treatment variable (Amira usage) to account for variation in treatment effects across schools.

All models were estimated using restricted maximum likelihood (REML) with Satterthwaite approximations for degrees of freedom. The primary coefficient of interest in each model was the fixed effect treatment (Amira), representing the average difference in EOY percentile ranks between treatment and control students after adjusting for covariates.

Effect sizes were calculated by dividing the estimated treatment coefficient by the standard deviation of the EOY DIBELS percentile ranks in the analytic sample.

Results

Sample and Attrition

The analytic sample from 174 schools included 15,424 students, with 6,694 in kindergarten and 8,730 in first grade. After matching and applying inclusion criteria (i.e., having EOY DIBELS scores), the treatment group included 7,691 students, and the control group included 7,733 students. Attrition was low and balanced across groups: 4.2% (treatment) vs. 4.0% (control) in kindergarten, and 4.0% (treatment) vs. 3.2% (control) in first grade, as shown in Table 1.

Table 1. Pretest and Posttest Sample Sizes and Attrition by Grade Level

Grade	Group	Pretest Sample	Posttest Sample	Attrition N	Attrition %
K	Treatment (Amira)	3,489	3,344	145	4.2%
	Control	3,489	3,350	139	4.0%
	Total	6,978	6,694	284	4.1%
1	Treatment (Amira)	4,528	4,347	181	4.0%
	Control	4,528	4,382	146	3.2%
	Total	9,056	8,729	327	3.6%

Baseline Equivalence and Descriptive Statistics

Post-matching comparison confirmed baseline equivalence on pretest measures (i.e., DIBELS BOY percentile), as shown in Table 2. No statistically significant differences were observed for BOY scores between treatment and control for either grade level. However, in kindergarten students, there was a significantly higher mean EOY percentile rank observed for the treatment group compared to the control group (54.66 vs. 47.62, respectively).

Table 2. BOY and EOY DIBELS Percentile Ranks by Treatment Group and Grade Level, Mean (SD)

Grade	BOY Percentile		EOY Percentile	
	Treatment	Control	Treatment	Control
K	43.15 (29.92)	42.67 (29.73)	54.66* (29.56)	47.62 (30.69)
1	41.84 (28.83)	41.27 (29.34)	48.61 (29.73)	47.64 (29.75)

* $p < 0.05$

Furthermore, demographic characteristics of the sample were compared, as shown in Table 3, and no statistically significant differences were observed between treatment groups for gender, SPED status, or economic disadvantage in either grade level.

However, statistically significant differences, analyzed using ANOVA, were observed for race/ethnicity and campus type for both grade levels. Additionally, chronic absenteeism was significantly different in kindergarten students, and ELL status was significantly different in first grade students. Students in the control group had higher proportions of Hispanic students, students classified as "Other" race or ethnicity, and ELL students. Generally, the treatment group had a higher proportion of Black or African American students, chronically absent students, and students in NES campuses.

Given that over 75% of students in the treatment group attended NEW campuses (compared to about 20% in the control group), this imbalance likely reflects

implementation priorities or targeting within the district. While this was not the primary focus of the study, it is notable that NES participation may represent a co-occurring reform effort that could interact with or enhance the effectiveness of Amira implementation.

Table 3. Demographic Characteristics of the Samply by Treatment Group and Grade Level

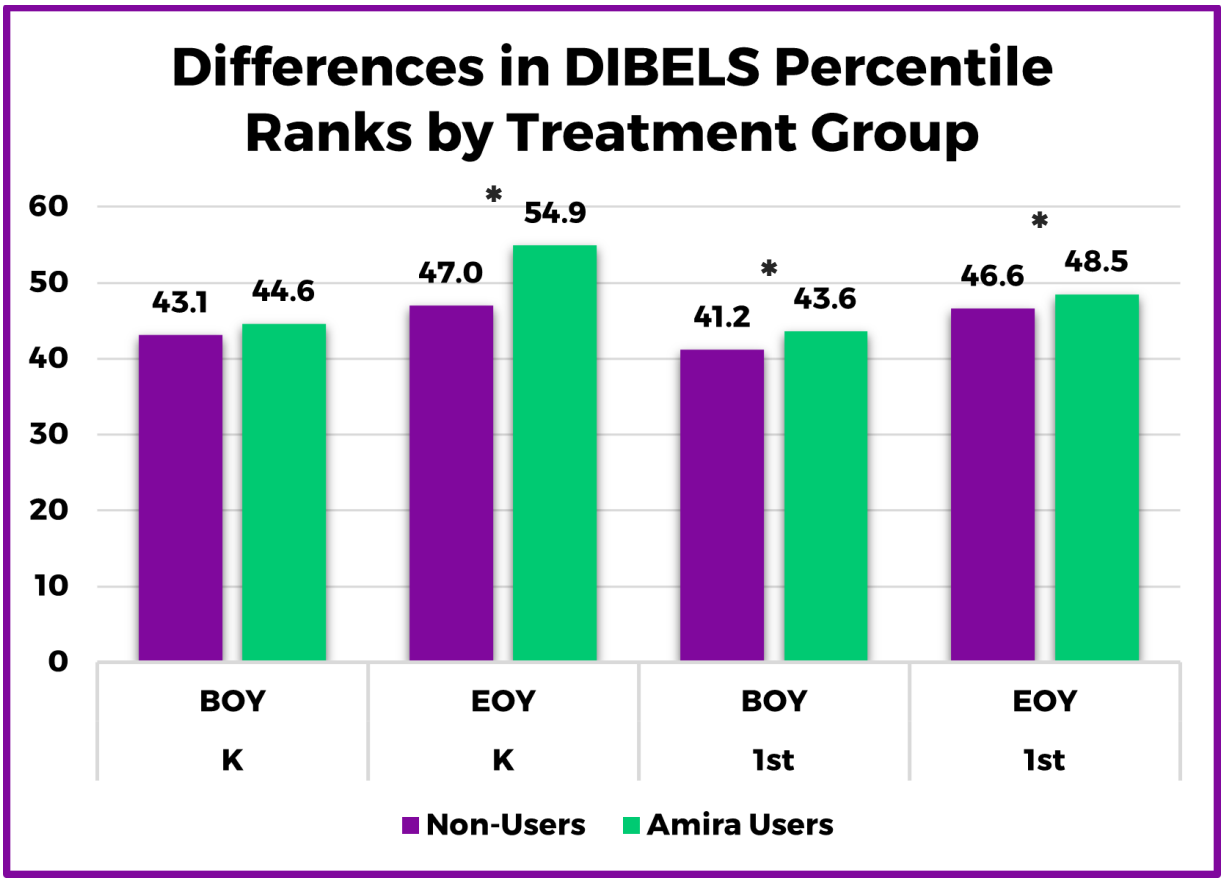
Grade	Group	Treatment (Amira)	Control
K	Gender (F)	51.7%	51.4%
	Race and Ethnicity		
	Hispanic	66.5%	72.2%*
	Black or African American	27.2%*	20.3%
	Other	6.3%	7.6%*
	Special Education	5.0%	5.2%
	Economically Disadvantaged	90.0%	90.2%
	ELL	41.4%	42.0%
	Chronically Absent	22.3%*	19.6%
	NES Campus	77.6%*	20.6%
1	Gender (M)	51.2%	51.5%
	Race and Ethnicity		
	Hispanic	66.1%	74.7%*
	Black or African American	26.2%*	17.1%
	Other Race and Ethnicity	7.7%	8.2%
	Special Education	6.3%	7.2%
	Economically Disadvantaged	86.4%	86.1%
	ELL	50.2%	53.7%*
	Chronically Absent	18.1%	16.6%
	NES Campus	77.0%*	16.7%

* $p < 0.05$

Amira Usage on DIBELS Reading Outcomes

Figure 1 shows the graphical representation of increases in DIBELS percentile ranks observed from the HLM analyses. While BOY scores in Figure 1 appear significantly different between treatment and control for first grade, the values shown are adjusted means controlling for BOY percentile (when EOY is the outcome) and demographic characteristics shown in Tables 1 and 2.

Figure 1. Differences in DIBELS Percentile Ranks by Treatment Group and Grade Level for BOY and EOY Benchmarks.



* $p < 0.05$

Figure shows adjusted means.

Usage Results in Kindergarten Students

Table 4 presents the results of the two-level HLM for kindergarten students (n = 6,694). In the final model, the treatment effect of Amira was statistically significant and positive. Students who used Amira outperformed their matched peers by an average of nearly 8 percentile points on the EOY DIBELS assessment, adjusting for BOY percentile, demographic variables, and school-level clustering. This translates to an effect size of 0.26, indicating a small-to-moderate positive effect of Amira in kindergarten students.

Table 4. Two-Level HLM for Kindergarten, Coefficients and Standard Errors (SE)

Fixed Effects	Model 1	Model 2	Model 3
Intercept	51.87* (0.86)	23.62* (1.45)	23.79* (1.44)
Treatment		8.21* (0.90)	7.93* (1.14)
BOY Percentile		0.61* (0.01)	0.61* (0.01)
SPED		-12.60* (1.29)	-12.62* (1.29)
Economically Disadvantaged		-3.16* (1.02)	-3.20* (1.02)
ELL		5.26* (0.70)	5.26* (0.70)
Chronically Absent		-6.21* (0.70)	-6.18* (0.70)
***Race and Ethnicity* (Hispanic Referent)**			
Black or African American		-2.69* (0.87)	-2.86* (0.87)
Other		4.53* (1.21)	4.41* (1.21)
Error Variance			
Level-1	815.56* (14.27)	511.34* (8.96)	507.18* (8.93)
Level-2 Intercept	98.77* (13.34)	87.19* (11.20)	77.95* (11.23)
Treatment			42.96* (15.29)
Model Fit			
AIC	64159.2	61085.7	61071.3
BIC	64179.6	61160.6	61153.0

* $p < 0.05$; ICC = .11

Values based on Stata 19.5 Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = REML; Satterthwaite degrees of freedom.

Usage Results in First Grade Students

Table 5 presents the results of the two-level HLM for first grade students (n = 8,729). In the final model, the treatment effect of Amira was statistically significant and positive. Students who used Amira outperformed their matched peers by an average of nearly 2 percentile points on the EOY DIBELS assessment, adjusting for BOY percentile, demographic variables, and school-level clustering. This translates to an effect size of 0.06. While modest, this effect is in the expected direction and reinforces the positive relationship between Amira use and reading outcomes in early elementary grades.

Table 5. Two-Level HLM for First Grade, Coefficients and Standard Errors (SE)

Fixed Effects	Model 1	Model 2	Model 3
Intercept	48.77* (0.80)	12.31* (0.95)	12.24* (0.95)
Treatment		2.08* (0.58)	1.92* (0.84)
BOY Percentile		0.81* (0.01)	0.81* (0.01)
SPED		-4.50* (0.75)	-4.60* (0.75)
Economically Disadvantaged		-0.85 (0.61)	-0.83 (0.61)
ELL		4.73* (0.47)	4.60* (0.47)
Chronically Absent		-2.41* (0.51)	-2.39* (0.50)
***Race and Ethnicity* (Hispanic Referent)**			
Black or African American		-2.21* (0.61)	-2.02* (0.62)
Other		-0.19 (0.76)	-0.13 (0.76)
Error Variance			
Level-1	799.79* (12.09)	296.14* (4.53)	291.56* (4.49)
Level-2 Intercept	90.87* (12.10)	46.87* (5.90)	43.62* (6.17)
Treatment			37.08* (9.25)
Model Fit			
AIC	83436.1	74823.0	74770.3
BIC	83457.3	74900.8	74855.2

* $p < 0.05$; ICC = .10

Values based on Stata 19.5 Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = REML; Satterthwaite degrees of freedom

Discussion

This study provides rigorous evidence that Amira, an AI-powered literacy platform, produces positive impacts on early reading outcomes when implemented in elementary classrooms. Using a matched sample design and HLM, students who used Amira outperformed their matched peers on the EOY DIBELS assessment, with a statistically significant and educationally meaningful effect in kindergarten and a smaller but significant effect in first grade. The kindergarten effect size of 0.26 suggests that the platform can contribute to accelerated reading growth during a critical period in literacy development.

The effect of Amira Tutor was most pronounced in kindergarten, where students in the treatment group scored approximately 8 percentile points higher than their matched peers. While the statistical effect size is modest, this magnitude of gain is pedagogically meaningful, particularly given the importance of early reading

trajectories and the high concentration of historically underserved students in the sample. In contrast, the impact in first grade, though still statistically significant, was smaller. This difference may be attributable to several factors, including reduced time for intervention, differences in how Amira was implemented across grades, or ceiling effects among higher-performing first graders.

It is also important to consider that a substantial portion of the treatment group attended schools participating in the district's NES. As such, NES schools required structured Amira usage as part of instruction, which may have contributed to more consistent or higher-fidelity implementation. While students in the control group were also present in NES schools, they were far fewer in number, and usage expectations differed. Although campus type was included in the matching process and tested during model building, it did not significantly predict outcomes and was therefore excluded from the final HLM models. Nonetheless, the instructional environment and implementation requirements in NES schools may have amplified the impact of Amira, warranting future research into how policy-driven implementation models interact with digital learning tools to support reading growth.

Given that this study meets ESSA criteria for moderate evidence, these findings provide actionable guidance for school systems seeking federally aligned interventions to support early literacy growth. The results are particularly relevant for large, urban districts serving high proportions of economically disadvantaged and linguistically diverse students, as in the present sample. The consistent positive effects across grades suggest that Amira Tutor can be an effective and scalable intervention in similar educational contexts, even when implemented in conditions with variability in usage patterns.

The strength of evidence in this study is bolstered by the study's design features: matching on a comprehensive set of pre-treatment variables, including BOY percentile scores, student demographics, attendance, and school context; estimation of treatment effects with appropriate multilevel modeling; and use of a well-validated, norm-referenced outcome measure. The low and balanced attrition rates further strengthen internal validity and support the reliability of the findings, which contribute important new evidence that Amira can effectively support early reading development in real-world school settings. However, future research may further investigate how usage patterns, dosage, and implementation fidelity interact with student outcomes, and explore whether differential effects emerge across subgroups.

References

Amira Learning. (2022). *Meta-Analysis of Research on Amira Intelligent Tutoring's Impact*.

Amira Learning. (2023). *Amira Benchmark Assessment Technical Guide*.

Amira Learning. (2024a*). AI Tutoring Powers Reading Growth for MLLs/ELLs in English and Spanish*.

Amira Learning. (2024b). *Amira Teacher Manual*.

Korsah, G. A., Mostow, J., Dias, M. B., Sweet, T., Belousov, S. M., Dias, M. F., & Gong, H. (2010). Improving literacy in Africa: Experiments with an automated reading tutor. *Information Technologies & International Development*, 6(2), 1–19. Retrieved from https://www.cs.cmu.edu/~listen/pdfs/2010%20ITID%20Korsah%20517-1388-1-PB.pdf

Mills-Tettey, G. A., Mostow, J., Dias, M. B., Belousov, S., Dias, M. F., Sweet, T. M., & Gong, H. (2005). *Improving child literacy in Africa: Experiments with an automated reading tutor*. Carnegie Mellon University, Robotics Institute.

Mostow, J., Nelson, J., & Beck, J. E. (2013). Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research*, 49(2), 249–276. https://doi.org/10.2190/EC.49.2.g

Poulsen, R., Wiemer-Hastings, P., & Allbritton, D. (2007). Tutoring bilingual students with an automated reading tutor that listens. *Journal of Educational Computing Research*, 36(2), 191–221. https://doi.org/10.2190/A007-367T-5474-8383

Reeder, K., Shapiro, J., Wakefield, J., & D'Silva, R. (2015). Speech Recognition Software Contributes to Reading Development for Young Learners of English. *International Journal of Computer-Assisted Language Learning and Teaching*, 5(3), 60-74.

Utah State Board of Education. (2023). *Utah's Early Interactive Reading Software Program: Amira Learning 2022–2023 Evaluation Report*.