Implementation and Outcome Evaluation of the Early Interactive Reading Software Program (EISP): Amira Report

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

2025

Implementation and Outcome Evaluation of the *Early Interactive Reading Software Program (EISP): Amira Report*

Prepared for
the Utah State Board of Education

THE UNIVERSITY OF UTAH
UTAH EDUCATION POLICY CENTER

THE UNIVERSITY OF UTAH
UTAH EDUCATION POLICY CENTER

Bridging Research, Policy, & Practice

The Utah Education Policy Center (UEPC) is an independent, non-partisan, not-for-profit research-based center at the University of Utah founded in the Department of Educational Leadership and Policy in 1990 and administered through the College of Education since 2007. The UEPC's mission is to bridge research, policy, and practice by conducting rigorous and comprehensive research and evaluations, and providing expert and research-informed technical assistance and professional learning. We empower educators, policymakers, and leaders to make research actionable and impactful to transform education across early childhood education, K-12 schools, and higher education.

We are committed to supporting the understanding of whether educationalpolicies, programs, and practices are being implemented as intended, whether they are effective and impactful, and how they may be improved and scaled-up and become sustainable.

Please visit our website for more information about the UEPC:
http://uepc.utah.edu

Andrea K. Rorrer, Ph.D., Director
andrea.rorrer@utah.edu

Cori Groth, Ph.D., Associate Director
cori.groth@utah.edu

Ellen Altermatt, Ph.D., Assistant Director for Research and Evaluation
ellen.altermatt@utah.edu

T. W. Altermatt, Ph.D., Assistant Director and Lead Data Scientist
bill.altermatt@utah.edu

Citation: 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.

Introduction

This report presents data from Amira from the Utah Education Policy Center's (UEPC's) 2024-2025 evaluation of the Early Interactive Software Program (EISP), a state-supported initiative to strengthen literacy skills among students in Grades K through 3. The evaluation focused on both implementation (number of students, amount of usage) and impact (change in reading proficiency as a result of using early literacy software). Comparative data on other vendors is available in the statewide report (available at https://uepc.utah.edu/_resources/documents/eisp-implementation-and-outcomes-report-2025.pdf. Although the implementation data we review below are available in the full report, we summarized that data here with some added interpretation of Amira's data, specifically. Vendor-specific information about impact is unique to this report and does not appear in the full report.

Methods

Using data sharing agreements between vendors and the Utah State Board of Education (USBE) and a Master Data Sharing Agreement between the UEPC and USBE, the UEPC connected data provided by vendors on weekly student use of early literacy software with data provided by the USBE on student demographics, school information, and scores on the Acadience Reading standardized assessment. Implementation was evaluated by tabulating the number of students, schools, and LEAs served by each vendor as well as the mean level of student engagement with the software per week and the mean number of weeks per year that the software was used. These were compared to vendor-supplied cutoffs for the minimum recommended use, and reported as the percentage of students who met 80%, 100%, 200%, and 300% of vendor recommendations. When the data permitted, impact was evaluated using a method designed to reduce the correlation between student characteristics and early literacy software use: covariate balancing propensity score weighting. Like matching and random assignment, weighting increases confidence in cause-and-effect conclusions between early literacy software use and learning gains by controlling for other variables that might systematically co-vary with reading software use. The relationship between "dose" (level of early literacy software use) and "response" (learning gains) was modeled using statistical regression tailored to the weighting process. We then calculated two standardized effect sizes (i.e., Cohen's *d*) (1) comparing the predicted Acadience reading score at Amira's median level of usage with the predicted Acadience reading score at 0 usage, and (2) comparing the predicted Acadience reading score at the 90th percentile of Amira's usage (i.e., among Amira's users, the level of usage that was greater than 90% of all users) and the predicted Acadience reading score at 0 usage. When the data were too sparse to permit weighting and regression, the relationship between student usage and Acadience Reading performance was explored using data visualization.

Implementation

Enrollment

Among vendors, Amira has the third largest user base in the state (See Table 1 for Amira's users and see Table 1 in the statewide report for the number of students using platforms from other vendors).

Table 1. 2024-2025 Program Enrollment Overview

Vendor	Reported Students	Matched SSID[a]	Zero Usage[b]	Shared Students[c]	Unique Schools	Unique LEAs
Amira	8,768	8,222	1	2,815	97	7

Note: [a]Number of students whose State Student Identifiers (SSIDs) could be matched to USBE student records. [b]Number of students whose total software usage for the year was zero and had an SSID that matched USBE records. [c]Number of students who appeared in the user lists of more than one vendor and had an SSID that matched USBE records.

Amira users are concentrated between 1st through 3rd grades, with fewer users in Kindergarten (see Table 2).

Table 2. 2024-2025 Program Enrollment by Grade

Vendor	K	1st	2nd	3rd
Amira	956	2,542	2,528	2,196

Note: Counts reflect the number of unique student users with non-zero usage for the year whose SSIDs could be matched to USBE records. Counts include students who appeared in multiple vendor lists.

Usage

Amira's usage recommendations for minutes per week are the lowest in the state, and its number of recommended weeks is the third highest among vendors.

Table 3. Vendor Use Recommendations

Vendor	K	1st	2nd	3rd	Weeks
Amira	20	20	20	20	25

Minutes per Week and Weeks per Year

Table 4 provides mean levels of usage (in minutes per week, total minutes, and number of weeks) by grade level. Amira users average 18.05 minutes per week (See Table 4). Comparing these numbers to Table 4 in the statewide report, this is on the lower end of all vendors.

Table 4. 2024-2025 Program Use by Vendor and Grade for Age of Learning

Vendor	Grade	*N*	Avg Weekly Minutes	Avg Total Minutes	Avg Weeks of Use
Amira	K	956	14.3	292.66	17.36
	1	2,542	16.84	332.31	17.34
	2	2,528	20.17	466.95	19.37
	3	2,196	18.65	454.45	19.29
	Total	8,222	18.05	401.72	18.49

Percent of Students Meeting Recommendations

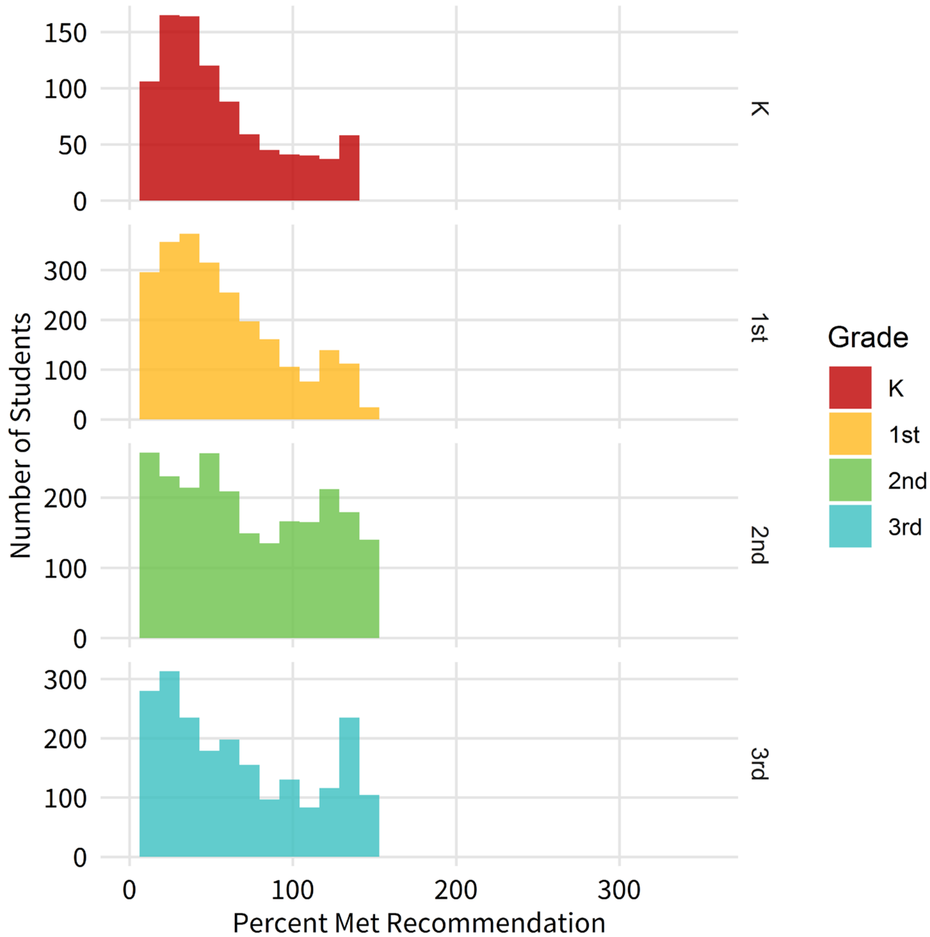
Because of the different standards and recommendations for weekly use across vendors, it is difficult to evaluate the average minutes of use with regard to whether students are using the software above or below expectations. To address this concern, we standardize student usage by considering it in the context of each vendor's level of recommended usage. The percentage of students meeting 80%, 100%, 200%, and 300% of vendor recommendations is presented in Table 5, and the distribution of student use is presented in Figure 1. Examining Figure 1 and Table 5, Amira has about an average rate of students who met at least 80% of vendor recommended use among vendors.

Table 5. Percentage of Students Meeting Vendor Recommendations for Use

Vendor	Grade	*N*	% at 80% of Rec.[b]	% at 100% of Rec.[b]	% at 200% of Rec.[b]	% at 300% of Rec.[b]
Amira	K	956	23%	16%	0%	0%
	1	2,542	24%	16%	0%	0%
	2	2,528	39%	30%	0%	0%
	3	2,196	35%	27%	0%	0%
	Total	8,221	31%	23%	0%	0%

Note: [a]*N* is the count of unique student users, including those who had zero usage for the year and those who appear in multiple vendor lists but excluding those who could not be matched to USBE records by SSID. [b]Percentages are the number of students with usage at different percentages of vendor recommendations, divided by the number of students in the *N* column.

Figure 1. Distribution of usage

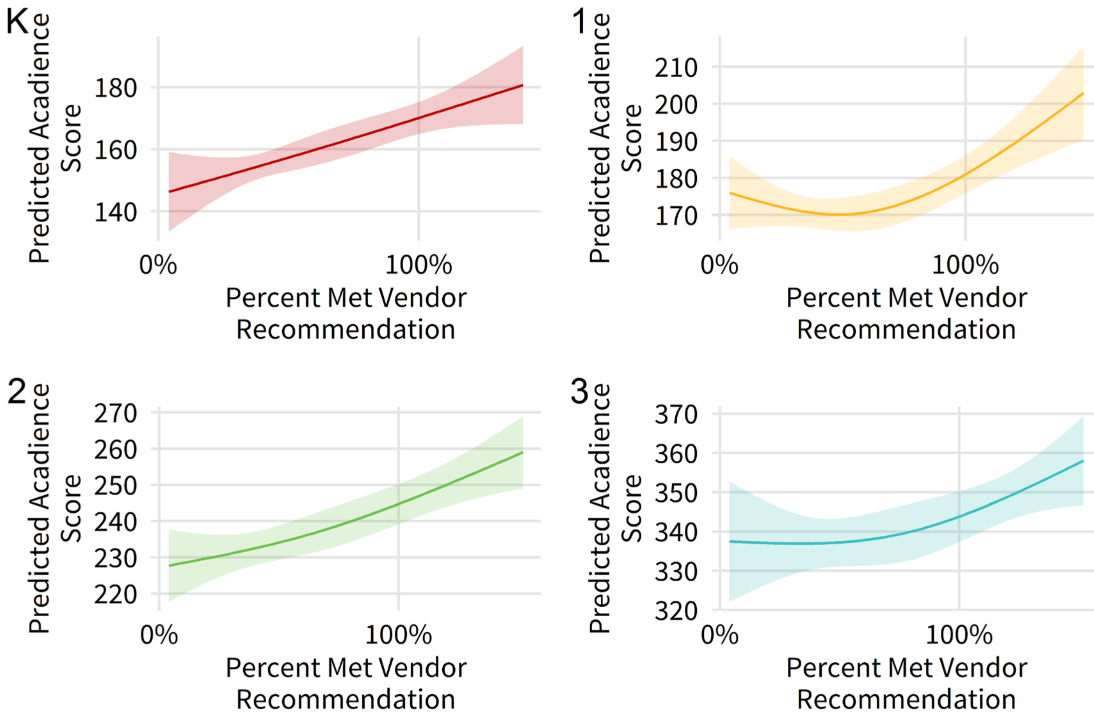


Early Literacy Software Usage and Reading Achievement

As our outcome measure, we used the end-of-year Acadience Reading composite score. As outlined by Good III and colleagues (2011), the Acadience Reading assessment is designed to measure early literacy and reading ability for students in kindergarten through 6th grade. See Appendix for a description of the statistical method used to estimate the relationship between student usage (expressed as a percentage of vendor recommendation) and Acadience Reading performance while controlling for possible confounding variables.

The dose-response curves for Amira are presented in Figure 2. These curves represent estimates for the relationship between usage and Acadience Reading that, to the degree possible, control for beginning-of-year score and other student- and school-level variables to provide a more accurate and unbiased understanding.

Figure 2. Relationship between usage and reading achievement



The slope demonstrating the relationship between usage and predicted Acadience Reading score was positive and significant for Kindergarten, 2nd grade, and 3rd grade (average slopes ranging from 0.14 to 0.25, *ps* < .05). For 1st grade, the average slope was positive but was not statistically significant (*p* = .09). It is important to understand for the dose-response curve for 1st grade, that while it seems to show a decline in the early percentages, there is a high degree of uncertainty about the slope of the line due to fewer students in this range (reflected in the wider confidence interval). The true relationship between student usage and Acadience Reading scores could be anywhere within the shaded region of the confidence interval, making it possible that there is actually a positive relationship or no relationship at low usage levels. Further research is recommended to investigate the relationship between lower levels of usage and Acadience Scores among 1st graders.

Table 6 displays the effect sizes of median Amira usage vs. 0% usage and 90th percentile Amira usage vs. 0%, by grade. Effects range from small to well above large. The effect sizes for 1st grade should be interpreted as descriptive and do not represent statistically significant effects.

Table 6. Effect sizes by grade

Comparison	Kindergarten	1st Grade	2nd Grade	3rd Grade
Median vs. 0%	0.23	-0.07	0.07	0.01
90th Percentile vs. 0%	0.56	0.16	0.24	0.11

Note: Median Amira usage was 52.03%, 52.00%, 60.00%, and 73.93% for Kindergarten through 3rd grade, respectively. 90th percentile Amira usage was 124.00%, 128.00%, 140.00%, and 140.00%, respectively.

Summary and Discussion

The data demonstrate that Amira is one of the largest vendors in EISP. Our impact results found that Amira is effective in Kindergarten, 2nd grade, and 3rd grade, but the current evidence supports its effectiveness only at higher rates of usage in 1st grade.

Appendix A: Method

Dose-response Curves

In contrast to the main report, for all individual vendor reports, we did not include students who did not have a vendor. Thus, all individual vendor results reflect data from students with non-zero usage using the vendor's software.

Our analysis of the relationship between usage and Acadience Reading used a method called "weighting." The goal of weighting is to minimize the correlation between the level of treatment received by a student (i.e., their level of software usage) and student and school characteristics. This is the same goal pursued by matching and by random assignment to conditions, the gold standard of research designs for causal inference. We used the WeightIt R package to conduct Covariate Balancing Propensity Score weighting to generate a weight for each student that reduced the correlation between the treatment variable (i.e., percent met vendor recommendation), and the following variables: free/reduced-price lunch status, student race, multilingual learner status, receipt of special education services status, beginning of the year Acadience Reading composite score, whether the beginning of the year Acadience Reading composite score was missing for that student, student gender, school-level percent multilingual learners, and school-level percent of students receiving special education services. Weighting and analysis were done separately by grade level. Missing values for beginning of the year Acadience Reading composite score were imputed with the median and accompanied by a dummy variable for missingness. After estimating the weights, we assessed weighting quality by evaluating the covariate-treatment correlations. After minimizing these correlations, we used a propensity score-weighted linear regression with natural splines (df = 2) to estimate the causal relationship between percent-met-vendor-recommendation and end-of-year Acadience Reading composite score. To estimate the average dose-response function, we used g-computation. Specifically, we predicted 100 evenly-spaced outcomes across the complete range of values of percent-met-vendor-recommendation. We tested whether the slope was significantly different from zero.