EDUCATION

THE UTAH STATE BOARD OF EDUCATION

Report to the Education Interim Committee

Early Interactive Reading Software Program Report (EISP)

10/14/2024

Amber Wright
Digital and Instructional Materials Specialist, USBE
Amber.Wright@schools.utah.gov

David MacKay
Research Consultant, USBE
David.MacKay@schools.utah.gov

Stephanie Su
Research Consultant, USBE
Stephanie.Su@schools.utah.gov

STATUTORY REQUIREMENT

U.C.A. Section 53A-17a-167 requires the Utah State Board of Education (USBE) to submit a report on the final testing data regarding an interactive computer software program including student learning gains as a result of the interactive computer software program.

Early Interactive Software Program Report

EXECUTIVE SUMMARY

The Early Interactive Software Program (EISP) was designed to increase the literacy skills of all students in K-3 through adaptive computer-based literacy software. Eleven software vendors were selected by schools and qualified by the state to provide reading software programs through the EISP in 2023-2024. The vendors used during the school year were (in alphabetical order): 95 Reading Achievement Program ("95 RAP"), Age of Learning, Amira Learning, Curriculum Associates ("i-Ready"), Dreambox Reading Park ("Dreambox"), Imagine Language & Literacy, Lexia® ("Core5®"), Moby Max, Read Naturally Live, Reading Horizons, and Waterford Reading Academy. The Evaluation and Training Institute (ETI), the EISP program evaluator, studied two core aspects of the program: 1) students' use of the program during the school year (program enrollment and implementation); and 2) the effects of the program on increasing students' literacy achievement (program impacts). The current evaluation examined the program as a whole (refer to as "program-wide" throughout this report), encompassing all eleven vendors, as well as the reading outcomes for each individual provider. This report summarizes the overall results across the program. We conclude by presenting our longitudinal findings on the implementation and impact of the EISP over the last six school years. The vendor-specific findings can be found in separate, supplemental memos submitted along with this report.

Utah's Early Intervention Reading Software Program

2023-2024 Program Evaluation Findings

Submitted to the Utah State Board of Education
October 14, 2024

EVALUATION AND TRAINING INSTITUTE *A non-profit center for research & evaluation*
12240 Venice Blvd, Suite 27, Los Angeles, California 90066 || 310-473-8367

ABOUT EVALUATION AND TRAINING INSTITUTE

Founded in 1974, the Evaluation & Training Institute (ETI) is a non-profit consulting firm, headquartered in Los Angeles, dedicated to working with schools, post-secondary institutions, public agencies, private foundations, community-based organizations, and professional organizations. We specialize in third-party program evaluations covering many fields, including education, literacy, STEM, social services, health, and prevention. Many of our evaluations have been instrumental in the development of public policy as well as state and federal legislation. Throughout, our focus is on helping clients improve their programs as well as maintain accountability to funders and oversight committees.

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

The Evaluation and Training Institute (ETI) thanks Amber Wright (Software Initiatives Specialist, Digital Teaching and Learning) and Melanie Valentine (Specialist, Digital Teaching and Learning) and from the Utah State Board of Education (USBE) for their ongoing collaboration throughout this evaluation project. We also acknowledge David Mackay and Stephanie Su, research consultants, at USBE for all their efforts preparing and transferring the student data used for the analyses. Finally, the software vendor representatives played a key role in helping us understand their software programs, sharing their data, and working with us to prepare the data in a consistent and streamlined format. In particular, we give special thanks to: Justin Carlson and Alex Kent from 95 Reading Achievement Program, Krislyn Mimnaugh and Rhiann Eddy from Amira, Ann Minckler and Erin Barnett from Imagine Learning, Kelley Pasatta, Nikki Osborn, and Diane Stromberg from Lexia, Tracey Jackson and Mary Molusky from Curriculum Associates, Vicky Robl, Maddie Dreffer, and Nick Underwood from Dreambox Reading Park, Connor Todd from Moby Max, Dan Evans and Jenna Krautbauer from Read Naturally Live, Marci Houseman, Julia Romrell, and Andrea Colon from Age of Learning, John Rajikannu and Steffani Clark from Reading Horizons, and Haya Shamir, Steven Kiisel, Robin Van Dusen, and Erik Yoder from Waterford. Each of these individuals provided necessary data from their products that were used to complete the evaluation project.

EXECUTIVE SUMMARY

Evaluation Purpose

The Early Intervention Software Program (EISP) was designed to increase the literacy skills of all students in K-3 through adaptive computer-based literacy software. Eleven software vendors were selected by schools and qualified by the state to provide reading software programs through the EISP in 2023-2024. The vendors used during the school year were (in alphabetical order): 95 Reading Achievement Program ("95 RAP"), Age of Learning, Amira Learning, Curriculum Associates ("i-Ready"), Dreambox Reading Park ("Dreambox"), Imagine Language & Literacy, Lexia® ("Core5®"), Moby Max, Read Naturally Live, Reading Horizons, and Waterford Reading Academy. The Evaluation and Training Institute (ETI), the EISP program evaluator, studied two core aspects of the program: 1) students' use of the program during the school year (program enrollment and implementation); and 2) the effects of the program on increasing students' literacy achievement (program impacts). The current evaluation examined the program as a whole (refer to as "program-wide" throughout this report), encompassing all eleven vendors, as well as the reading outcomes for each individual provider. This report summarizes the overall results across the program. We conclude by presenting our longitudinal findings on the implementation and impact of the EISP over the last six school years. The vendor-specific findings can be found in separate, supplemental memos submitted along with this report.

Program Enrollment and Implementation

In 2023-2024, eleven EISP software vendors were used in a total of 140 LEAs, in 692 schools and by 161,602 students. Core5 was the most widespread program in the state relative to other EISP providers, reaching 86 LEAs, 456 schools, and 116,268 students, followed by i-

Ready (23,441), Amira (13,075), Imagine Language & Literacy (4,237), Waterford (3,347), with all other providers enrolling fewer than 500 students.

Our implementation study revealed that many students struggled to meet the recommended minimum usage levels set by the software providers. While statewide program implementation offers a significant opportunity for numerous students to benefit, it is essential that they engage with the program for the recommended duration to effectively enhance their literacy achievement.

Program-Wide Impact on Acadience Achievement Scores

Literacy achievement was assessed using state-provided Acadience Reading scores. Our findings indicated that predicted end-of-year Acadience scores for students in grades K, 1, and 3 were higher than those of their control counterparts, with second grade the exception, which we discuss in more detail in the report. Additionally, we observed that the treatment effects were most pronounced for students who utilized the program as intended. Notably, kindergarten and first grade students experienced the most significant benefits from their participation in EISP.

EISP and Different Student Populations. We also analyzed the program's impact on specific demographic subgroups, including English Language Learners, low-income students, and those with special education designations. Our findings showed that in every subgroup, students participating in the EISP who met the vendors' recommended usage criteria outperformed their non-program peers, with the exception of second grade.

Multi Year Analysis

Finally, we examined program enrollment, implementation, and impact over the past six years of the EISP. The EISP has experienced significant growth over the past six years, expanding from four providers to eleven and expanding its reach to more students, schools and Local Education Agencies (LEAs). The program has had a positive impact on student literacy, particularly in early grades, with strong results in kindergarten and first grade. Excluding the atypical patterns of usage during and after the pandemic, the percentage of students meeting the usage recommendations has declined in the earliest grades (K and 1st) and remained more stable in the older primaries (2nd and 3rd).

Recommendations

The current evaluation revealed positive literacy achievement outcomes, particularly for kindergarten students who met the vendors' usage recommendations. Our findings underscore the importance of meeting minimum usage thresholds and sustaining consistent program use on a weekly basis. Below, we outline several key recommendations to enhance this engagement:

- **Encourage Use Consistent with Vendors' Recommendations:** A considerable number of EISP students did not meet the minimum usage recommendations set by the software vendors. We recommend that the state promote regular usage and hold Local Education Agencies (LEAs) accountable for adhering to these guidelines, ensuring students have the best opportunity to enhance their literacy skills.
- **Extend Kindergarten Impact:** Given the substantial benefits observed consistently for kindergarteners, we recommend that the state further explore how program participation can enhance more advanced literacy skills for students in subsequent grades.

- **Promote Awareness of Program Benefits for Diverse Learners**

  Raise awareness among educators and stakeholders about the program's demonstrated benefits for English Language Learners (ELL), special education, and low-income students. Highlighting these outcomes can help motivate schools to prioritize the program more effectively for these populations.

INTRODUCTION

The Early Intervention Software Program (EISP) was designed to increase the literacy skills of all students in K-3 through adaptive computer-based literacy programs. Over the last decade, the program has evolved to serve over 150,000 Utah students, and a yearly evaluation mandated through enacting legislation has been used to track the program's impact on student achievement as well as program use across the state. Historically, the Utah State Board of Education (USBE) offered local education agencies (LEAs) a curated list of approved literacy software providers to select from for the academic year. Utah Senate Bill 44 substantially changed the EISP program implementation model starting in SY 2023-2024: Local Education Agencies (LEAs) were able to autonomously select and purchase an interactive reading software for their students enrolled in kindergarten through third grade. In order to participate in the EISP, LEAs submit applications to the Utah State Board of Education (USBE) requesting reimbursement for the use of specific reading software programs prior to the start of each school year. Selected vendors must meet specific state-mandated criteria to qualify for EISP reimbursement, including tracking student-level usage and adhering to external evaluation requirements. In the end, eleven software vendors were selected by schools and qualified by the state to provide reading software programs through the EISP in 2023-2024. The number of providers has more than doubled compared to the previous program year. The vendors used during the school year were (in alphabetical order): 95 Reading Achievement Program ("95 RAP"), Age of Learning, Amira Learning, Curriculum Associates ("i-Ready"), Dreambox Reading Park ("Dreambox"), Imagine Language & Literacy, Lexia® ("Core5®"), Moby Max, Read Naturally Live, Reading Horizons, and Waterford Reading Academy.

The Evaluation and Training Institute (ETI) contracted with the USBE to study how the reading software programs were used by schools and the impact they had on students' literacy development. The evaluation included the results for both the combined impact of all the software programs used in Utah schools (referred to as "program-wide" throughout this report) as well as the individual outcomes on literacy achievement for each of the software providers (referred to as "vendor-specific"). This report highlights the program-wide findings only. The vendor-specific results can be found in supplemental memos provided to USBE separate from this report.

The current evaluation incorporates findings from the 2023-2024 academic year, marking the eleventh year of the EISP's implementation. As the final evaluation year of the current contract, we also provide a longitudinal analysis of trends observed over the past several years of the program in the state. These insights aim to assist the USBE and Local Education Agencies (LEAs) in assessing the program's effectiveness, identifying potential areas for improvement, and making evidence-based decisions for future iterations of the program. The following research questions were used to guide our program-wide evaluation:

1. **To what extent did students use the software program as intended?**
2. **How did the EISP impact students' Acadience scores across all vendors?**
3. **How did different program usage levels influence Acadience outcome scores?**
4. **What impact did EISP have on specific student populations?**
5. **What trends in the program have emerged over the past six years?**

The sections of this report include this year's program enrollment numbers across grade and vendor, program implementation findings including vendor recommendations and participants' ability to meet them, the impact that the EISP had on literacy achievement

including mean differences and effects sizes[1], and the impact that different amounts of program use have on literacy outcomes. The report also shows the impact that the EISP has on specific populations of students including English Language Learners, those classified as low-income, or special education. We then present the findings from our longitudinal analysis of the annual data collected over time. We summarize the key findings and study limitations in the final sections. A detailed summary of our research methods is included in **Appendix A**.

FINDINGS

Program Enrollment Analysis

In 2023-2024, 11 EISP software vendors were used in a total of 140 LEAs, in 692 schools and by 161,602 students (**Table 1**).

Table 1. 2023-2024 Program Enrollment Overview

Program	LEAs	Schools	Students (K-3)
95 Reading Achievement Program	1	1	25
Age of Learning	1	5	487
Amira	7	63	13,075
Core5	86	456	116,268
Dreambox Reading Park	1	1	101
Imagine Language & Literacy	10	25	4,237
i-Ready	21	90	23,441
Moby Max	2	4	379
Read Naturally Live	1	2	111

[1] ETI calculated effect sizes using the standardized mean difference calculation known as "Hedges' g" based on What Works Clearinghouse recommendations (WWC, 2020). For group design studies, this effect size is defined as the difference between the mean outcome for the intervention group and the mean outcome for the comparison group.

Program	LEAs	Schools	Students (K-3)
Reading Horizons	2	2	131
Waterford	8	43	3,347
Total	140	692	161,602
Data source: software vendor data, some LEAs and schools use more than one software vendor			

Core5 was the most widespread program in the state relative to other EISP providers, reaching 86 LEAs, 456 schools, and 116,268 students (refer to **Table 1** for enrollment numbers for all vendors).

Overall student enrollment was consistent across all grades, with similar numbers of students in kindergarten, first, second, and third grades. **Table 2** further details the breakdown by grade for each vendor.

Table 2. 2023-2024 Program Enrollment by Grade

Program	Kinder	1st	2nd	3rd
95 Reading Achievement Program	1	9	8	7
Age of Learning	180	120	187	-
Amira	2,435	3,235	3,828	3,577
Core5	25,360	29,488	30,983	30,437
Dreambox Reading Park	26	27	21	27
Imagine Language & Literacy	1,097	1,204	1,153	783
i-Ready	4,896	5,778	6,365	6,402
Moby Max	24	102	56	197
Read Naturally Live	-	-	95	46
Reading Horizons	38	42	34	17

Program	Kinder	1st	2nd	3rd
Waterford	2,716	455	120	56
Total	**36,773**	**40,464**	**42,850**	**41,549**

Data source: software vendor data in K-3

Program Implementation Analysis

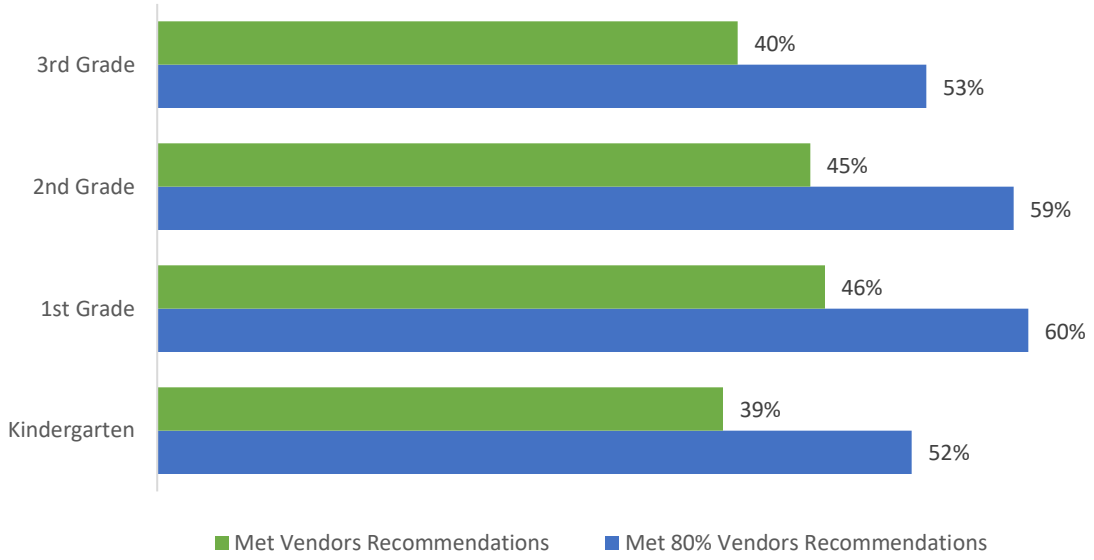
Studying program implementation prior to measuring the program impact provided a better understanding of the way the program was ultimately used by students. Namely, students must use the program long enough to influence the outcomes under study. Critical to successful EISP implementation was the amount of time and how consistently a student used the program during the school year.

Research Question: ***To what extent did students use the software program as intended?***

Key Takeaway: **EISP vendors have varying usage recommendations for students. Across all grades, fewer than half of EISP students met the minimum usage requirements set by each software provider, which included both weekly usage and the total number of weeks engaged with the program.**

This evaluation used two definitions of program use to capture students' EISP participation. Our goal was to align closely with each vendor's specified usage criteria, which varied among providers (refer to **Table 3**). This necessitated an initial step in our analysis to identify which students met the minimum requirement of the program, regardless of the provider that they used. Once the students were identified, we calculated the percentage of students in each grade who met the total weeks as recommended by the vendor *AND* whose average weekly minutes (for those weeks) was at or above the recommended minimum. Throughout this report we refer to this group of students as "met vendors recommendation." We found that participation was consistent among grades. That is, just about 40-46% of all students were able to adhere to the recommended weeks AND average weekly minutes (**Figure 1**; green bars).

Figure 1. Percentage of Students Meeting EISP Recommendations for Use



Note: Met Vendors Recommendations reflects 'Met minimum weeks and *average* weekly minutes'
Met 80% of Vendors Recommendations reflects 'Met 80% of weeks and 80% of *average* weekly minutes'

Next, we relaxed the vendors' minimum usage criteria, in order to create larger analytic samples. Specifically, we calculated the percent of students who met at least 80% of the vendors' total week recommendation AND met at least 80% of the average weekly minutes' recommendation. We refer to this group of students as "met 80% of vendors recommendation." This adjustment increased the representativeness of the children we studied and provided a larger sample of students who engaged with the program. As illustrated in **Figure 1** (blue bars), this adjustment increased the overall percentage of program students by nearly 15% across all grades.

As mentioned, each vendor provided recommendations for the amount of time that students should use the software program during the year, to have an impact on literacy achievement. As shown in **Table 3**, these recommendations differed by grade and by vendor.

Table 3. Vendor 2023-2024 Minimum Use Recommendations

Program	Kindergarten	First Grade	Second Grade	Third Grade	Suggested Minimum Weeks
95 Reading Achievement Program	120 min/week	120 min/week	120 min/week	120 min/week	30 weeks
Age of Learning	45 min/week	45 min/week	45 min/week	45 min/week	16 weeks
Amira	25 min/week	25 min/week	25 min/week	25 min/week	30 weeks
Core5	20 minutes to 60 min/week*	20 minutes to 60 min/week*	20 minutes to 60 min/week*	20 minutes to 60 min/week*	20 weeks
Dreambox Reading Park	30 min/week	30 min/week	30 min/week	30 min/week	30 weeks
Imagine Language & Literacy	40 min/week	50 min/week	50 min/week	50 min/week	18 weeks
i-Ready	30 min/week	30 min/week	30 min/week	30 min/week	20-25 weeks
Moby Max	30min/week	30min/week	30min/week	30min/week	36 weeks
Read Naturally Live**	1.5-2 stories/week	1.5-2 stories/week	1.5-2 stories/week	1.5-2 stories/week	24 stories
Reading Horizons	60 min/ week	60 min/ week	60 min/ week	60 min/ week	8 weeks
Waterford	60 min/week	75 min/week	75 min/week	75 min/week	28 weeks

* Core5 usage recommendations are automatically adjusted based on student need. Students working below grade level are assigned usage recommendations greater than those working at or above grade level.
**Read Naturally Live recorded usage in number of stories completed and does not have a set recommendation for weeks

Each software provider communicated both a range of minutes per week, and a minimum number of weeks for students to use the program. Across vendors, recommended weekly use ranged from 20 minutes to 120 minutes per week and total weeks ranged from 8 to 36 weeks.

Table 4 presents a comprehensive summary of the EISP students' average usage for each vendor and grade. These numbers represent the overall average of all students in their respective grade, and include average weekly minutes of use, average total minutes of use, and average number of weeks of use through the end of the school year.

Table 4. 2023-2024 Program Use by Vendor and Grade

Program	Grade	N	Ave Weekly Min	Ave Total Min	Ave Wks. Of Use
95 Reading Achievement Program	K	1	29	294	10
	1	9	30	156	7
	2	8	38	246	7
	3	7	28	187	7
	Total	25	32	199	7
Age of Learning	K	180	41	483	11
	1	120	46	557	12
	2	187	30	329	10
	Total	487	38	442	11
Amira	K	2,435	18	335	17
	1	3,235	21	441	18
	2	3,828	22	471	19
	3	3,577	19	377	17
	Total	13,075	20	412	18
Core5	K	25,360	48	1291	25
	1	29,488	55	1580	27
	2	30,983	51	1436	27
	3	30,437	48	1298	25
	Total	116,268	50	1405	26
Dreambox	K	26	32	1416	29
	1	27	39	1753	31
	2	21	21	944	25
	3	27	19	896	25
	Total	101	28	1269	27
Imagine Learning	K	1,097	44	1138	23
	1	1,204	48	1199	23
	2	1,153	39	928	19
	3	783	43	1175	23
	Total	4,237	44	1105	22
i-Ready	K	4,896	32	699	21
	1	5,778	39	960	23
	2	6,365	38	958	24
	3	6,402	43	1004	23
	Total	23,441	38	917	23

Program	Grade	N	Ave Weekly Min	Ave Total Min	Ave Wks. Of Use
Moby Max	K	24	19	184	7
	1	102	16	237	12
	2	56	20	197	9
	3	197	16	153	8
	Total	379	17	184	9
Read Naturally Live*	2	65	2.4	19.2	8
	3	46	1.6	7.3	5
	Total	111	2.1	14.3	7
Reading Horizons	K	38	15	90	5
	1	42	10	120	10
	2	34	6	28	4
	3	17	4	20	3
	Total	131	10	75	6
Waterford Reading Academy	K	2,716	52	1368	25
	1	455	60	1663	27
	2	120	51	1543	29
	3	56	39	482	13
	Total	3,347	53	1400	25

***Read Naturally Live tracks usage in story time. Instead of average minutes, we present average weekly stories completed, average total story time, and average weeks of use.

It warrants reiterating that just under half of the EISP student population achieved the levels of engagement put forth by the vendors. For the purposes of our impact evaluation, we analyzed 3 usage levels, (1) “met recommendations”, (2) “met 80% of recommendations” groups, and (3) “intent to treat” which included all students using the program in any amount.

Program Impact Analysis

We analyzed the program's impact on Acadience test scores by comparing students who used the program with students who did not. We have included a detailed methods section for technical reviewers in **Appendix A**[2].

Research Questions*: *How did the EISP impact students' Acadience scores? And how did different program usage levels influence Acadience outcome scores?

***Key Takeaway:* EISP Students in grades K, 1, and 3 achieved higher predicted literacy mean scores at the end-of-year compared to students not participating in the program, with the exception of second grade. Large substantive treatment effect sizes were found in kindergarten and first grade. Additionally, Acadience scores were highest among those using the program as recommended.**

Table 5 displays the differences in predicted mean scores derived from ordinary least squares (OLS) regression models for the treatment and control groups across all three usage levels, organized by grade. OLS regression modeling increases our confidence that any differences observed in literacy outcomes are due to participation in the EISP program while controlling for other influential variables (see Appendix A for more details).

As shown, generally the highest predicted Acadience scores are among the EISP students who used the program as recommended by the software vendors. With the exception of second grade, students in all other grades who participated in the program significantly exceeded their control group counterparts in predicted literacy outcome scores. This may be attributed to unique aspects of the Acadience test used in second grade (addressed in greater detail in the Discussion section).

[2] Please refer to the individual supplemental memos for vendor specific results.

Table 5. Acadience EOY Predicted Mean Scores by Usage and Grade

Grade	Condition	Intent to Treat	Met 80% of Rec.	Met Rec.
		End-of-Year Predicted Mean Scores		
K	Treatment	164.41	173.87	178.42
	Control	156.48	160.14	162.53
(diff)		*7.93*	*13.73*	*15.89*
1	Treatment	86.65	91.45	94.59
	Control	80.89	83.77	85.97
(diff)		*5.76*	*7.68*	*8.62*
2	Treatment	272.74	287.42	N/S
	Control	277.19	289.32	
(diff)		*-4.45*	*-1.9*	
3	Treatment	N/S	405.67	416.67
	Control		400.48	410.08
(diff)			*5.19*	*6.59*

Data source: Matched K-3 ITT, MRU80, MRU samples. All mean comparisons displayed between treatment and control were statistically significant at $p \leq .05$.

Effect sizes describe the magnitude of the difference between two groups on an outcome measure. We adapted a set of effect size benchmarks based on categories from Kraft (2020) that were adjusted for early literacy outcome measures: less than 0.10 is *small*, 0.10 to less than .30 is *medium* and .30 or greater is *large* (M. Kraft, personal communication, October 13, 2023). **Table 6** shows the effect sizes where the most meaningful program impact was on kindergarten students in the highest two usage groups, those who were able to meet the vendors recommendations for use (g = 0.40) and for those who met 80% of the vendors recommendations (g= 0.35). Effect sizes were also strong for first graders who used the program as intended (g=.32). All other grades and usage levels had effect sizes that reflected medium or small treatment effects.

Table 6. Effect Sizes by Grade and Usage Level

Grade	Intent to Treat	Met 80% of Rec.	Met Rec.
K	*0.197*	**0.346**	**0.400**
1	*0.209*	*0.279*	**0.316**
2	-0.079	-0.034	N/S
3	N/S	0.081	*0.104*

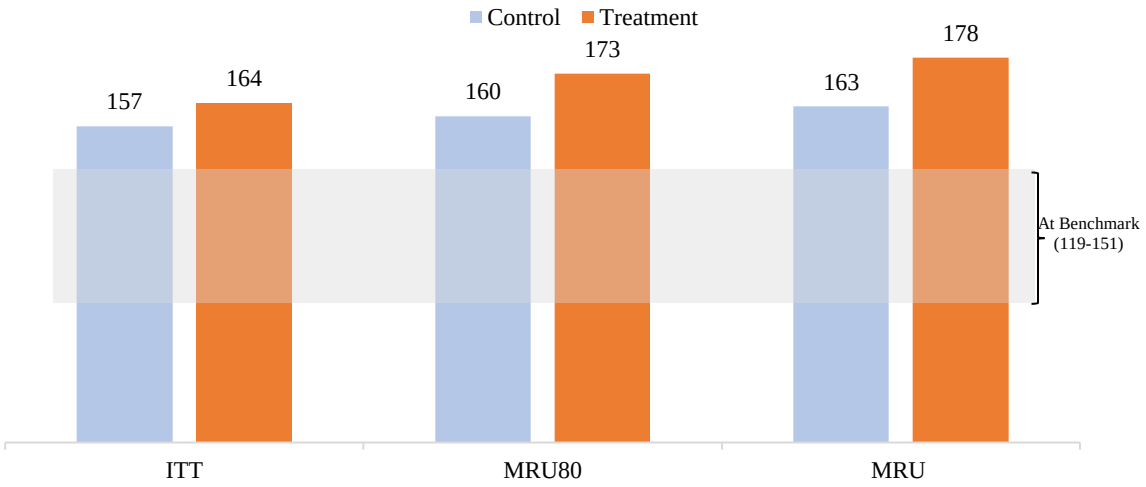
Data source: Matched K-3 ITT, MRU80, MRU samples[3]. All effect sizes displayed represent statistically significant mean differences at p≤ .05. Hedges' g effect size benchmarks are indicated in the table as follows: Small: 0 to < .10; *Medium*, italicized text: .10 < .30, **Large**: bold and underlined text: .30 or greater.

Acadience Grade Level Benchmark Analysis

It is also important to understand how the EISP impacted students' progress relative to grade level expectations. The following graphs depict not only the elevated performance of the EISP students, but also provide evidence that all students generally performed as expected for grade level regardless of program participation.

[3] Kindergarten sample size –ITT ctrl=4,249.426 , tr= 26,287; MRU80 ctrl= 4,007.232, tr= 14,798; MRU ctrl= 4,012.48, tr= 11,415; First Grade- ITT - ctrl= 4,983.299, tr= 30473; MRU80 - ctrl= 5321.6, tr=19,388; MRU- ctrl= 5,414.172, tr= 15,194; Second Grade sample size - ITT ctrl= 5,493.361, tr= 33,982; MRU80 ctrl= 5,721.207, tr=21,131; MRU ctrl= 5,731.58, tr= 16,308; Third Grade sample size – ITT ctrl= 5,575.32, tr=34,489; MRU80 ctrl=5,184.922, tr=19,147; MRU ctrl= 5000.573, tr=14,226.

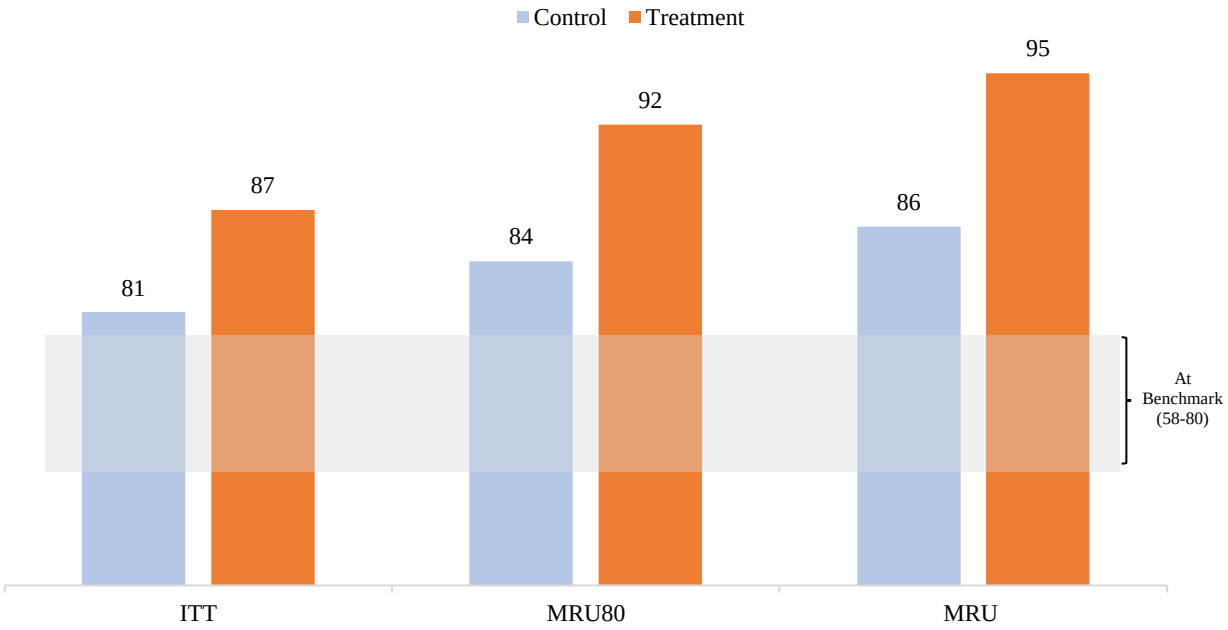
Figure 2. Kindergarten Predicted Mean Scores by Usage Level and Matched Sample



*Note: Students scoring **At Benchmark** (119-151), or **Above Benchmark** goal (152 or greater) have the odds in their favor (approximately 80% to 90% overall) of achieving later important reading outcomes. Data source: Matched K-3 ITT, MRU80 and MRU samples. All mean comparisons displayed in the figure were statistically significant at p≤.05.*

Figure 2 presents the predicted end-of-year mean scores for kindergarten students who used the EISP at different levels, along their matched control counterparts. Students in the two highest usage subgroups (those that met vendors recommendations and those that met 80% of the recommendations) had the highest end-of-year mean score (173 and 178, respectively), putting them in the "above benchmark" score range. Though the matched control students for the higher usage groups had predicted mean scores in the "above benchmark" range, treatment students had end-of-year mean scores 13-15 points higher than the control students. These findings further support that when the program is used consistently in kindergarten, students receive the highest program benefits.

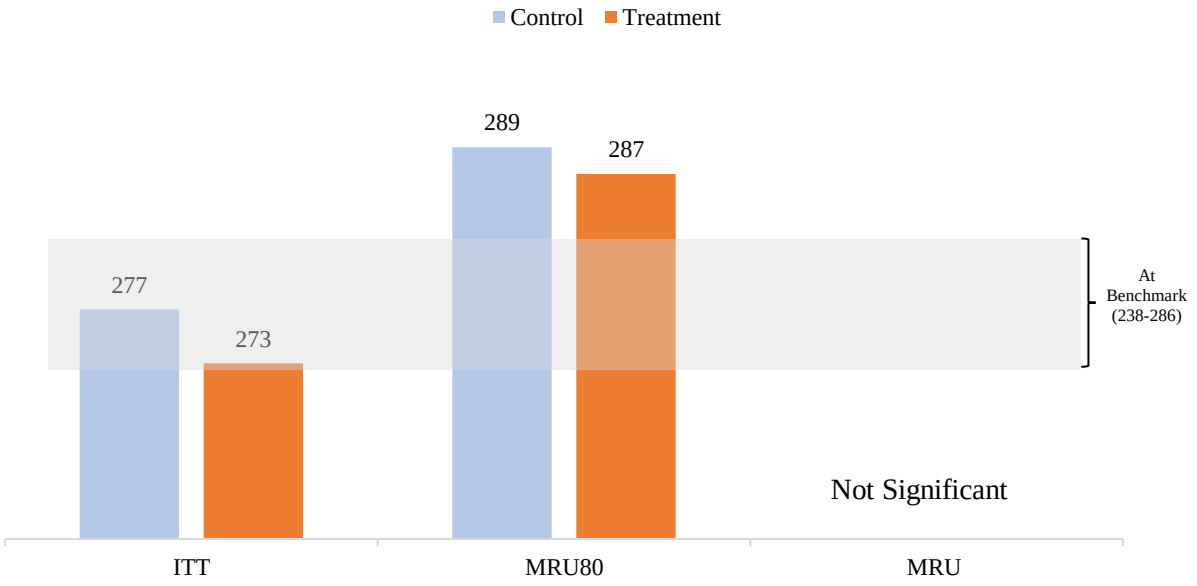
Figure 3. First Grade Predicted Mean Scores by Usage Level and Matched Sample



Note: First grade end-of-year predicted outcomes were measured with the Nonsense Word Fluency- Correct Letter Sounds scale and has a different range than the reading composite scale. Students scoring ***At Benchmark*** *(58-80), or* ***Above Benchmark*** *goal (81 or greater) have the odds in their favor (approximately 80% to 90% overall) of achieving later important reading outcomes. Data source: Matched K-3 ITT, MRU80 and MRU samples. All mean comparisons displayed in the figure were statistically significant at p≤ .05.*

Figure 3 shows the predicted end-of-year mean scores for first grade students who used the EISP at different levels, along with their matched control counterparts. Similar to kindergarten, first graders who used the program closest to the vendors' intention, had the highest end-of-year mean score (95 and 92, respectively). First grade students using the software in any amount (ITT) also had predicted end-of-year mean scores higher than the comparison students.

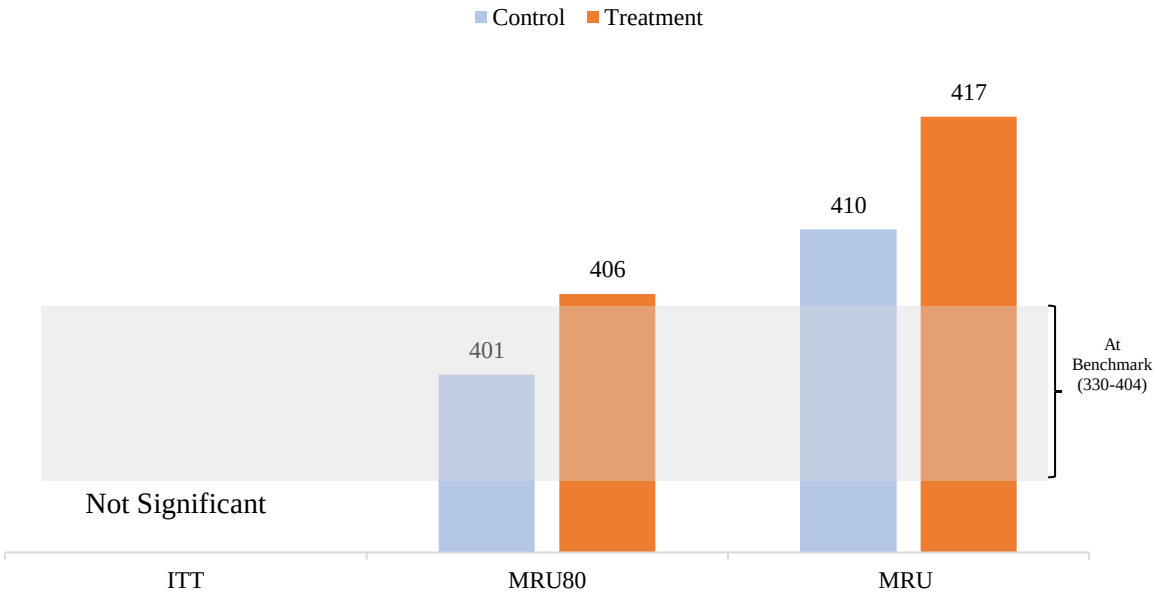
Figure 4. Second Grade Predicted Mean Scores by Usage Level and Matched Sample



*Note: Students scoring **At Benchmark** (238-286), or **Above Benchmark** goal (287 or greater) have the odds in their favor (approximately 80% to 90% overall) of achieving later important reading outcomes. Data source: Matched K-3 ITT, MRU80 and MRU samples. All mean comparisons displayed in the table were statistically significant at p≤ .05.*

Figure 4 illustrates the predicted end-of-year mean scores for second grade students who used the EISP at different levels. The program did not have statistically significant results for those who used the program in the highest usage group - those who met the recommendations. In the lowest usage groups (ITT, MRU80) treatment students had predicted mean scores that were lower than the matched control students, with differences of 4 or 2 points, respectively. Both treatment and control students had predicted end of year scores that were above benchmark. Again, a possible explanation for the second-grade results is provided in the discussion section below.

Figure 5. Third Grade Predicted Mean Scores by Usage Level and Matched Sample



*Note: Students scoring **At Benchmark** (330-404), or **Above Benchmark** goal (405 or greater) have the odds in their favor (approximately 80% to 90% overall) of achieving later important reading outcomes. Data source: Matched K-3 ITT, MRU80 and MRU samples. All mean comparisons displayed in the table were statistically significant at p≤ .05.*

Figure 5 presents the predicted end-of-year mean scores for third grade students. The highest achievement scores were aligned to the students who used the program as the vendor intended (417). The program did not have statistically significant results for those who used the program in the lowest usage group.

Subgroup Impact Analysis

We were also interested in studying how the program may benefit students in specific demographic subgroups.

Research question: ***What impact did EISP have on specific student populations?***

We conducted a separate analysis of program impacts on students identified as English Language Learners, low-income, and special education designation status. **Table 7** presents the predicted mean scores for the Acadience Reading composite.

Table 7. Subgroup Analysis of Predicted End-of-Year Acadience Mean Scores

		Kindergarten	First Grade	Second Grade	Third Grade
Special Education	Treatment	167.66	87.83	N/S	393.57
	Control	151.76	79.22		386.98
ELL	Treatment	178.46	90.91	N/S	404.31
	Control	162.57	82.30		397.72
Low-Income	Treatment	177.67	91.71	N/S	412.36
	Control	161.77	83.10		405.77
	Data source: Matched K-3 MRU sample. All data points displayed in figure were statistically significant at p≤ .05.				

Across all demographic subgroups in kindergarten, first and third grade, students in the EISP who were able to meet the vendors' recommended use criteria outperformed their non-program counterparts. The differential treatment effects were most pronounced in kindergarten, but still show positive impacts in end-of-year literacy scores for first and third grade students.

Analysis of Program Use on Impact

As shown in the analysis sections above, our evaluation sought to show differences *between* treatment and control students, but equally important was understanding how different levels of program participation specifically among EISP students impacted literacy outcomes. **Table 8** shows a side-by-side view of each grade and the three defined usage levels among treatment students who (1) met the recommendation for weeks and average minutes, (2) met 80% of the recommendation, and (3) who had any use, ITT. The data suggest that as usage of the program increased within each grade (i.e., more adherence to the way program use was intended), predicted end-of-year mean scores also increased. This finding is especially pronounced in 2nd and 3rd grade.

Table 8. EISP Students' Predicted Mean Scores by Grade and Usage Level

Grade	Intent to Treat	Met 80% of Rec.	Met Rec.	Diff ITT to MRU
K	164	174	178	**+14**
1	87	92	95	**+8**
2	273	287	297	**+24**
3	384	406	417	**+33**

Note: ITT: Intent to Treat; MRU80: Met 80% of recommendation; MRU: Met recommendation. Kindergarten, second and third grade students predicted means were measured with the reading composite scale and first grade end-of-year predicted outcomes were measured with the Nonsense Word Fluency- Correct Letter Sounds scale, which has a different range than the reading composite scale.

Like in previous school years, the greatest benefits of consistent program use are seen among the 2nd and 3rd grade students. As seen in **Table 8**, the point difference in 2nd and 3rd literacy outcomes was 24 and 33, respectively, when comparing students engaged in casual program use to those engaged in vendor-recommended use. Results also suggest that as more advanced reading skills are practiced and acquired, adequate use of supplemental literacy interventions provide beneficial support within the classroom.

MULTI YEAR FINDINGS

In this section, we explored the key trends in program implementation, focusing on enrollment and student usage, and literacy outcomes over the past several years of the EISP.

Research question*: *What trends have emerged in the program over the last six years, including how the EISP has influenced student literacy achievement?

***Key Takeaway.* Year after year, the EISP has reached a substantial number of students across Utah. For those who engaged with the program as recommended by the software providers, the data indicate a significant positive trend in literacy achievement, particularly among kindergarten students.**

Multi Year Enrollment and Implementation Analysis

Table 9 depicts program enrollment of Local Education Agencies, schools and students over the past six years of the EISP. Program enrollment has steadily increased over the last six school years, serving 124,000 students in 2018-2019 and 161,000 students in 2023-2024. During the last two school years, the program grew from four participating vendors to eleven vendors. The number of LEAs in the program grew significantly from 2018-2019 to 2019-2020, serving an additional 50 LEAs, but has since stabilized, serving roughly 130-140 LEAs annually.

Table 9. Program Enrollment from 2018- 2024

	School Year					
	2018-2019	**2019-2020**	**2020-2021**	**2021-2022**	**2022-2023**	**2023-2024**
LEAs	88	139	142	133	140	140
Schools	438	573	605	565	692	692
K-3 Students	124,378	150,169	158,695	155,222	166,468	161,602

We also explored how students used the software throughout the last six years of implementation, specifically the percentage of students who met the vendors' recommendations. In **Table 10**, we show the percentage of students who met the recommendations for both minutes and weeks of use. Overall, we see a significant dip in the percentages during the 2019-2020 school year, as a result of school closures related to the Covid-19 pandemic[4]. That said, there was a large boost in usage in the following school year,

[4] During the 2019-2020 school year the evaluation methods were adjusted to accommodate school closures in March 2020 related to the Covid-19 pandemic.

as students returned to in-person instruction. Excluding the atypical patterns of usage during and after the pandemic, the percentage of students meeting the usage recommendations in kindergarten and first grade declined over time, while second and third grade remained more stable. It is important to note that in some cases, vendors adjust their recommendations from year to year, which can impact the percentage of students meeting the recommendations. Further, as discussed, the number of software providers grew from 5 vendors to 11 vendors in 2023-2024.

Table 10. Multi-year Trends in the Percent of Students who met recommendations

	School Year					
Grade	**2018-2019**	**2019-2020***	**2020-2021**	**2021-2022**	**2022-2023**	**2023-2024**
Kindergarten	46%	25%	51%	48%	35%	39%
First Grade	54%	39%	60%	55%	43%	46%
Second Grade	46%	42%	58%	57%	44%	45%
Third Grade	41%	36%	51%	45%	43%	40%

*Note-The definition of met recommended usage was adjusted in 2020, given the school closures and potentially shortened program exposure. Percentages in that year reflect the total percentage of weeks that students used the software through March 2020.

Multi Year Impact Analysis

Finally, we studied the extent to which the program had an impact on students' literacy over the last six school years. For this year-over-year analysis, we focused on the students who used the software as intended by the vendors. Results strongly support that the EISP has been effective in impacting the literacy performance of kindergarten students over time. Since 2019, we see that the strongest program effect sizes were reported at this early grade level (**Table 11**).

Table 11. Trends in Effect Sizes

	School Year					
Grade	**2018-2019***	**2019-2020***	**2020-2021**	**2021-2022**	**2022-2023**	**2023-2024**
Kindergarten	*0.13*	**0.33**	**0.34**	*0.27*	**0.37**	**0.40**
First Grade	0.04	*0.13*	*0.23*	*0.10*	*0.18*	**0.32**
Second Grade	N/S	0.07	*0.13*	-0.03	0.03	N/S
Third Grade	N/S	*0.17*	*0.17*	0.10	*0.13*	*0.10*

All effect sizes displayed represent statistically significant mean differences at p≤ .05. Hedges' g effect size benchmarks are indicated in the table as follows: Small: 0 to < .10; *Medium*, italicized text: .10 < .30, **Large**: bold and underlined text: .30 or greater. *Cohen's d effect sizes were calculated during these school years. Note that 2020 effect sizes were calculated on middle of year composite scores due to Covid-19 pandemic school closures.

Over the years, we have observed medium treatment effects in first grade, with the last three school years showing an upward trend, culminating in a *large* effect size of 0.32 in 2023-2024. In contrast, the effect sizes among second and third grade students have exhibited greater variation over time. This variability is likely linked to the skills needed in upper primary grades and the structure of the Acadience assessment, a trend that our evaluation has consistently highlighted in annual reports.

DISCUSSION, LIMITATIONS, AND RECOMMENDATIONS

There were two primary goals for the 2023-2024 EISP evaluation: (1) to study program implementation, and (2) to determine the program's impact on Acadience literacy scores. In this section, we summarize those findings, and present the known limitations, as well as our recommendations for program improvement.

Implementation

On average, only 43% of all EISP students across grades K-3 met the recommended minimum usage levels established by program vendors. These usage thresholds are communicated to Local Education Agencies (LEAs) each year as benchmarks to encourage the necessary levels of engagement that can effectively enhance literacy achievement. Given that less than half of the students are able to meet these requirements, expectations for literacy gains linked to the software should be approached with caution.

This pattern raises important questions about how schools prioritize the software in their classrooms. It seems possible that while schools view the software as a valuable tool, they may not fully commit to leveraging it to its maximum potential. The responsibility for successful implementation and adherence to the recommended usage guidelines ultimately lies with the schools and teachers using the software.

Impacts

Significant treatment effects were observed in kindergarten and first grade among students who met the vendors' usage requirements, with effect sizes of g=0.40 for kindergarten and g=0.32 for first grade. Additionally, kindergarten students who adhered to 80% of the recommended usage showed an effect size of g=0.35. In the highest usage group, third grade students also demonstrated improved predicted literacy scores by the end of the year compared

to their peers not participating in the program; however, the effect sizes in this group were medium to small.

To better inform stakeholders about the impact of varying levels of program usage on student outcomes, we created samples of students based on usage in our analysis. Overall, EISP students who engaged with the program as intended consistently outperformed their control counterparts on predicted end-of-year Acadience outcomes across nearly all grades (K, 1, and 3). We also found that higher program use was associated with stronger effects, although this relationship was less pronounced in second grade. This discrepancy may be linked to significant changes in the Acadience assessment specifically for that grade level.

Second Grade Acadience. The relationship between program participation and outcomes was less clear for second grade, a finding that was also observed in previous evaluation results. This may be attributed to unique aspects of the Acadience test used in second grade. In consultation with the USBE, we found that the skills targeted in second grade differ from those in other grades. The second-grade assessment places a stronger emphasis on identifying correct words, as measured by the Oral Reading Fluency- Words Correct scale, compared to first and third grades. Additionally, the scoring standards for this specific skill peak at this level, with second graders now required to achieve 104 correct words—an increase from 67 in first grade—before decreasing to 90 words in third grade. The change at second grade could mean that the skills developed through a specific vendor's curriculum may not be adequately captured by the Acadience composite measure. As a result, differences between treatment and control students might be difficult to discern, leading to challenges in interpreting the results.

Furthermore, the EISP demonstrated substantial benefits for students classified as English Language Learners (ELL), those in special education, and low-income students, compared to matched peers who did not participate in the program.

Limitations

While we strive to account for all potential influences on student reading outcomes through our sampling methods and statistical techniques, research conducted in real educational environments is inherently subject to factors beyond the specific program being studied.

Expanding Nature of EISP Providers. As more software vendors join the program, the variability in its implementation and usage across classrooms is likely to increase. This diversity presents challenges for our research and evaluation efforts as we aim to maintain consistency in studying the success of implementation and its impact on student literacy. Given the range of programs being implemented across various school districts, we attempt to control for these differences in our research methods; however, the dynamic nature of live educational settings means that our research can still be influenced by external factors.

Differences in Teacher Prioritization. The variability in how teachers implement the program significantly affects our ability to assess program-wide impacts. With over a hundred thousand students across thousands of classrooms, we cannot fully control for the extent to which different teachers engage students with the software. Gaining deeper insights into teachers' prioritization and implementation of the intervention could enhance our understanding of the usage data and its corresponding impacts.

Control Group Contamination. The rise of digital technology in educational interventions is notable in Utah, leading to an increasing number of students utilizing various software programs each year. Our control group consists of students not participating in the EISP, but the growing prevalence of educational technology raises the possibility that some control students may have been exposed to non-EISP reading interventions. Future evaluations would benefit from the collaboration between the USBE and program vendors to track and share this information, allowing for a more comprehensive analysis.

Recommendations

The results of the evaluation underscore the importance of supporting students' literacy development and creating opportunities for our youngest learners. Generally speaking, students served by the EISP outperformed the students who were not. Further, the students who were able to engage with the software as it was intended by the vendors also showed greater end-of-year literacy scores relative to those participating more casually in the program.

Based on the findings from the 2023-2024 EISP evaluation, we propose the following recommendations to enhance program implementation and improve student literacy outcomes:

- **Expand Collaboration, Enhance Training and Support for Educators**
  Enhance collaboration between program vendors, LEAs, and the Utah State Board of Education (USBE) to share data on usage and educational interventions. Develop targeted professional development programs that equip teachers with effective strategies for integrating the EISP software into their classrooms. Providing ongoing support and resources can help ensure that teachers prioritize the program and maximize its potential.
- **Reevaluate Program Prioritization**
  Encourage schools to assess how they prioritize the EISP software in their curricula. This could involve discussions on its role as a core instructional tool versus a supplemental resource, promoting a shared understanding of its importance in supporting literacy development.
- **Foster Consistent Student Engagement**
  Encourage strategies that promote consistent student engagement with the program. This could involve integrating the software into daily routines, assigning regular usage tasks, or creating incentives for students to use the program more frequently.
- **Extend Kindergarten Impact:** Given the substantial benefits observed for Kindergarteners, we recommend that the state further explore how program participation can enhance more advanced literacy skills for students in subsequent grades.
- **Promote Awareness of Program Benefits for Diverse Learners**
  Raise awareness among educators and stakeholders about the program's demonstrated benefits for English Language Learners (ELL), special education, and low-income students.

Highlighting these outcomes can help motivate schools to prioritize the program more effectively for these populations.

By considering these recommendations, the state may improve the effectiveness of the EISP and further support the development of student literacy outcomes across participating schools. With a focused commitment to accountability, consistency in program usage, and the integration of various literacy-focused approaches, an increasing number of students may benefit from the *Early Intervention Software Program*.

REFERENCES

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dynamic Measurement Group, Inc. (2016, September). *Acadience Reading Benchmark Goals and Composite Score.* https://Acadience.org/papers/AcadienceNextBenchmarkGoals.pdf.

Evaluation and Training Institute. (2014-2020, October). *Early Intervention Software Program Evaluation: Results.* Culver City, CA

Hill, C. J., Bloom, H. S., Black, A. R. and Lipsey, M. W. (2008), *Empirical Benchmarks for Interpreting Effect Sizes in Research.* Child Development Perspectives, 2: 172–177. doi: 10.1111/j.1750-8606.2008.00061

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2008. *Matching for Causal Inference without Balance Checking*. http://gking.harvard.edu/files/abs/cem-abs.shtml.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189x20912798

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington DC: Institute of Education Sciences.

Powell-Smith, K., Good, R.H., III, & Dewey, E.N., & Latimer, R.J. (2014). *Assessing the Readability of Acadience AD Oral Reading Fluency and Daze.* (Technical Report No.16). Eugene, OR: Dynamic Measurement Group.

APPENDIX A. EVALUATION METHODS

The following is an overview of our research methods, samples and data sources that were used to answer each research question. The methods are described for the two studies, the impact study of students' achievement outcomes and the implementation study of students' program use, that were used to inform the program evaluation. **Appendices A-C** provide additional details on our methods, data processing procedures and samples.

Program Participants

Implementation Study Evaluation Participant Samples

The goal of the implementation study was to examine the extent to which students used the software as intended by each program vendor. All students captured in the vendors' usage data were included in our implementation study. Our goal was to provide the most accurate depiction of students' program use, regardless of how much students engaged with the program. To do so, for K-3 students we used the vendor data and did not remove students with incomplete Acadience data.

Impact Study Evaluation Participant Samples

To study program impact, we created three different groups of treatment students based on their level of program usage, (1) those who used the software in any amount (Intent to Treat or "ITT"), (2) students who used the software for at least 80% of the minimum recommended amount, and (3) students who used the software as intended by the vendors including weekly minutes and total weeks. To be included in our analytic samples, students needed to have accurate state student SSIDs (unique identification numbers used by the state to track students

in K-12) and complete Acadience test score data. We removed students with a beginning of year composite score of three or less. These low scoring cases indicated that the measurement was not appropriate for their skill level. Further, we excluded students who may have used multiple software programs during the year to reduce "treatment cross-program contamination" effects.

Control Student Matching Process

Our impact study compared Acadience literacy test scores between EISP program students (the treatment group) to a group of non-program students (the control group). Since we were not able to randomly assign students to treatment or control groups, we matched preexisting program to control students using Coarsened Exact Matching (CEM; Iacus et al., 2008). The students were matched on data from the beginning of the school year, and across several important characteristics (covariates used included: grade, beginning-of-year achievement scores, gender, race, English Language Learner status, and poverty status).

We employed a CEM approach designed to retain as many treatment cases as possible. There were fewer control students than treatment students, which resulted in slight pretest imbalances between our matched treatment and control groups (these imbalances were statistically corrected by using weighting to balance the differences in mean values of the covariates between groups; see the below description about linear regression models). Despite these slight differences, our approach led to a well-balanced analytic samples, as indicated by the following three L1 scores,[5] ITT; 0.00000000000002186; MRU80; 0.000000000000001298, and MRU;

[5] The L1 statistic is a comprehensive measure of global imbalance (Iacus, King and Porro, 2008). It is based on

0.000000000000002782. Lower values indicate less imbalance, and the closer to zero the better the two samples were balanced across covariates.

To summarize, we created and matched three treatment and control samples based on three different levels of usage. The EISP students were categorized into 3 subgroups (1) those who used the software in any amount (Intent to Treat or "ITT"), (2) students who used the software for at least 80% of the minimum recommended amount, and (3) students who used the software as intended by the vendors including weekly minutes and total weeks. Each of these groups had matched control counterparts.

What sources of data were used in our analyses?

We collected data from ten different sources to create our master dataset for the EISP analyses. The data sources included: eleven program vendors, who provided us with usage information for each student who used their programs; state Acadience Learning (Acadience Reading) testing data; and student information system (SIS) demographic data provided by the Utah State Board of Education (USBE). See **Appendix D** for details on how we created our master dataset.

Which instruments did we use to measure literacy achievement?

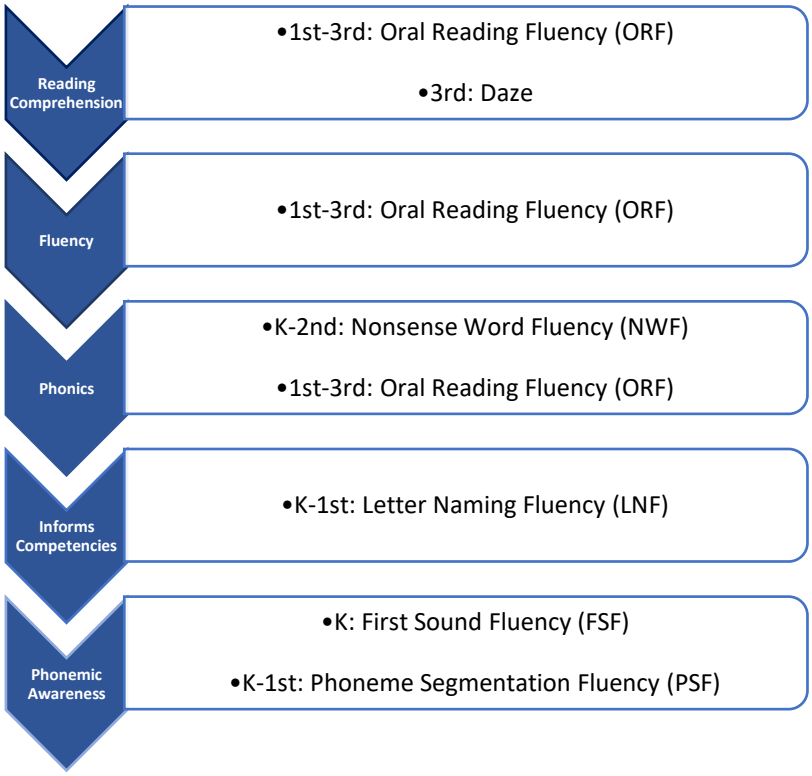
We measured literacy achievement using Acadience Reading, which was administered in schools throughout the state in Grades K-3. The Acadience Reading measures were used

---

the L1 difference between the multidimensional histogram of all pretreatment covariates in the treated group and that in the control group.

throughout Utah and are strong predictors of future reading achievement. Acadience Reading is comprised of six measures that function as indicators of critical skills students must master to become proficient readers, including: First Sound Fluency (FSF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF), and reading comprehension (DAZE). In addition to scores for the six subscale measures described above, we used reading composite scores and benchmark levels, or criterion-reference target scores that represent adequate reading progress. See **Appendix E** for additional detail on the Acadience Reading measures.

Figure A1. Acadience Indicator & Literacy Skill Measures



How did we study program implementation?

Our program implementation findings focused on program usage in relationship to its intended use, as described through vendors' use recommendations. Program usage data included the

following: total minutes of software use, from log-in to logoff for each week the program was used during the school year; total weeks, and average weekly use. Program vendors supplied the usage data.

How did we study the program-wide impacts across all vendors?

Our study relied on statistical analyses to measure program impacts, which included linear regression modeling (OLS), and descriptive analyses of trends related to levels of program use and Acadience benchmark category outcomes.

Linear regression models

We studied the program impacts on students' Acadience test scores by comparing a sample of treatment group students drawn from all vendors to a matched sample of control students. We determined that using an ordinary least squares (OLS) regression model allowed us to study the differences in treatment and control group test scores, while controlling for other important predictors of reading achievement. We used OLS to regress student outcomes on our predictor variables. Our independent variable was treatment group status (1/0), and we included other predictor variables to control for their effects in our models, including: beginning-of-year (BOY) test scores, gender, special education status, economic disadvantaged status, and ethnicity to adjust for their influence on end-of-year reading scores. By accounting for these additional predictor variables, we increased our ability to show a causal link between program use and outcomes while holding other factors unrelated to the program constant.

In addition, we applied the use of weights to our regression analysis to balance the differences in mean values of the covariates between treatment and control groups. The control observations were given weights such that the joint distribution of the multidimensional analytic sample achieved balance. Sometimes, this meant the controls were given more weight and sometimes it means they were given less weight.

Treatment Outcome Descriptive Analyses

To present our findings in an intuitive and applicable context, we measured the differences in students' reading scores at the end-of-year based on different categories of program exposure, or use. Use categories ranged from any use (i.e., Intent to Treat) to the highest category of meeting vendors' minimum recommended use requirement. As a complement to our OLS regression (causal) analysis, we used the descriptive analysis to show the association between levels of program use and outcomes for all students in the program.

What statistics do we provide in our results?

Where appropriate, we provided predicted mean scores and mean score differences for our treatment and control groups, which are meaningful when comparing treatment and control groups from the same sample. Statistical significance testing allowed us to determine the likelihood that a finding was a result of chance, or due to the treatment effect. We also provided treatment effect sizes (ES; based on Hedges G) to help readers understand the magnitude of treatment effects. Presenting effect sizes enabled us to provide a standardized scale to compare results based on different samples and measure the relative strengths of program impacts.

There are multiple ways to interpret effect sizes, including the use of categories such as small, medium, or large (e.g., Cohen, 1988; Kraft, 2020), or using a minimum threshold (Hill 2008). Variations of both approaches are widely used and accepted, yet both require careful considerations of the research design and key study components (such as sample, measures, etc.) Our effect size interpretation approach uses a categorical range based on effect sizes for similar types of research, studying similar interventions (early literacy programs) and with similar populations (elementary students). We adapted a set of effect size benchmarks based on categories from Kraft (2020) that were adjusted for early literacy outcome measures: less than 0.10 is *small*, 0.10 to less than .30 is *medium* and .30 or greater is *large* (M. Kraft, personal communication, October 13, 2023).Specifically, the range used in the current study represents the benchmarks for early literacy found in a summary of meta-analyses of relevant and similar educational studies, as well as the direct recommendation from the author (Kraft, 2020; M. Kraft, personal communication, October 13, 2023).

Methods Summary

In order to study EISP's impact on Acadience literacy test scores, we needed two samples of students, those who participated in the program (Treatment group) and those who were matched to the treatment students across characteristics that influence learning, such as socio-economic status, demographic information, and beginning-of-year Acadience test scores, but who did not participate in the program (Control group). The students who made up our treatment and control groups, within each grade K-3, were considered our analytic samples (i.e., the samples we used in the analysis).

Among the overall treatment sample, we created three subgroups of students to account for different levels of program usage. These subgroups were created to evaluate how different levels of use influenced the program's impact on literacy achievement. We considered three main factors in creating the subgroups for EISP students: (1) students who met the minimum weeks and average weekly use recommendations as defined by each vendor (MRU), (2) students who met at least 80% of the recommended weeks and average weekly minutes (MRU80), and (3) the broadest use group, inclusive of those who used the program in any amount throughout the program year (Intent to Treat).

We then matched comparison (control) students who did not participate in the program to the three EISP usage groups using Coarsened Exact Matching (CEM). We used CEM to match students on grade, beginning-of-year achievement scores and benchmark levels[6], gender, race, English Language Learner (ELL) status, and poverty status. The baseline characteristics of the treatment and control samples can be found in **Appendix A and B**. The matched samples were statistically well-balanced as indicated by L1 coefficients. For more detail on our statistical matching process, please refer to **Appendix A**.

Statistical Modeling of Program Impacts on Acadience Test Scores. Ordinary least squares (OLS) regression models were computed for each analytic sample. The OLS models predicted the differences in treatment and control groups' end-of-year group mean scores, while controlling for students' beginning-of-year (BOY) reading scores and key demographics;

[6] Students in kindergarten, 2nd and 3rd grade were matched on reading composite scores (BOY Comp) and students in 1st grade were matched on nonsense word fluency, correct letter sounds (NWF-CLS) scores.

gender, race, ELL status, SPED designation, and poverty status. We examined treatment effects for each analytic sample based on their usage and grade. For kindergarten, 2nd and 3rd grade end of year group mean scores, we used the reading composite score to measure student outcomes and for 1st grade students, we used the nonsense word fluency, correct letter sounds as our outcome variable.

APPENDIX B. ANALYTIC SAMPLES

Tables B1 – B3 present the characteristics for the population sample, as well as the matched sample used in our analyses. We also present the L1 statistic for each covariate in the matches ample. Lower values indicate less imbalance, and the closer to zero the better the two samples were balanced across covariates.

Table B1. Matched Treatment ITT Sample Demographics

	Grade	N	Female	Caucasian	SPED	Low-Income	ELL	BOY Score
Total Treatment Sample	K	27,204	49%	72%	9%	27%	6%	39.24
	1	33,132	49%	70%	12%	33%	9%	40.47
	2	35,480	49%	72%	13%	32%	9%	184.01
	3	35,610	49%	70%	16%	34%	11%	258.87
Matched ITT Treatment Sample	K	26,287	49%	74%	9%	26%	5%	39.56
	1	30,473	49%	75%	11%	31%	7%	39.29
	2	33,982	49%	74%	12%	32%	8%	186.29
	3	34,489	49%	73%	15%	33%	9%	260.77

Note: The matched sample had a multivariate L1 score of 0.00000000000002186. Lower values indicated less imbalance, and the closer to zero the better the two samples are balanced across covariates. Additionally, all covariates in the matched sample were found to be balanced: Female (L1= 0.000000000000024), White (L1= 0.000000000000026), SPED (L1 = 0.000000000000018), Low-Income (L1= 0.000000000000024), and ELL (L1= 0.0000000000000089).

Table B2. Matched Treatment MRU80 Sample Demographics

	Grade	N	Female	Caucasian	SPED	Low-Income	ELL	BOY Score
Total Treatment Sample	K	15,255	48%	74%	8%	26%	5%	42.42
	1	21,012	49%	72%	11%	32%	8%	43.07
	2	22,011	49%	72%	11%	31%	9%	196.55
	3	19,777	50%	71%	13%	33%	10%	275.52
Matched MRU 80 Treatment Sample	K	14,798	48%	76%	8%	25%	4%	42.75
	1	19,388	49%	76%	10%	30%	6%	41.78
	2	21,132	50%	75%	11%	30%	7%	198.76
	3	19,147	50%	73%	13%	32%	9%	277.73

Note: The matched sample had a multivariate L1 score of 0.000000000000001298. Lower values indicated less imbalance, and the closer to zero the better the two samples are balanced across covariates. Additionally, all covariates in the matched sample were found to be balanced: Female (L1= 0.0000000000000053), White

(L1= 0.000000000000007), SPED (L1 = 0.0000000000000031), Low-Income (L1= 0.0000000000000077), and ELL (L1= 0.00000000000000037).

Table B3. Matched Treatment MRU Sample Demographics

	Grade	N	Female	Caucasian	SPED	Low-Income	ELL	BOY Score
Total Treatment Sample	K	11,731	48%	75%	8%	25%	5%	44.44
	1	16,470	49%	73%	10%	31%	7%	45.27
	2	16,942	49%	74%	10%	30%	8%	205.16
	3	14,687	50%	72%	12%	32%	10%	286.08
Matched MRU Treatment Sample	K	11,415	48%	77%	7%	24%	4%	44.75
	1	15,194	49%	77%	9%	29%	5%	43.89
	2	16,309	50%	76%	10%	29%	6%	207.28
	3	14,226	50%	73%	12%	31%	8%	288.25

Note: The matched sample had a multivariate L1 score of 0.000000000000002782. Lower values indicated less imbalance, and the closer to zero the better the two samples are balanced across covariates. Additionally, all covariates in the matched sample were found to be balanced: Female (L1= 0.0000000000000081), White (L1= 0.000000000000021), SPED (L1 = 0.000000000000013), Low-Income (L1= 0.00000000000002), and ELL (L1= 0.000000000000009).

APPENDIX C. REGRESSION STATISTICS AND EFFECT SIZES BY SAMPLE

Table C1. ITT Regression Summary, by grade

	Grade	Condition	P-value	Marginal Mean	St. Error	Diff.	ES
Intent to Treat	K	Treatment	0.000	164.41	0.25	7.92	*0.197*
		Control		156.48	0.62		
	1	Treatment	0.000	86.65	0.16	5.76	*0.209*
		Control		80.89	0.39		
	2	Treatment	0.000	272.74	0.31	-4.45	-0.079
		Control		277.19	0.76		
	3	Treatment	N/S				
		Control					

Note. Hedges' g effect size (ES) benchmarks are indicated in the table as follows: Small: 0 to < .10; *Medium*, italicized text: .10 < .30, **Large**: bold and underlined text: .30 or greater. Data source: Matched K-3 ITT sample. Kindergarten sample size ctrl= 4,249.426, tr= 26,287; First Grade sample size ctrl= 4,983.299, tr= 30,473; Second Grade sample size ctrl= 5,493.361, tr= 33,982; Third Grade sample size ctrl= 5,575.32, tr= 34,489.

Table C2. MRU 80 Regression Summary, by grade

	Grade	Condition	P-value	Marginal Mean	St. Error	Diff.	ES
Met 80% of Recommended Use	K	Treatment	0.000	173.87	0.33	13.74	**0.346**
		Control		160.14	0.63		
	1	Treatment	0.000	91.45	0.20	7.68	*0.279*
		Control		83.77	0.38		
	2	Treatment	0.021	287.42	0.38	-1.90	-0.034
		Control		289.32	0.73		
	3	Treatment	0.000	400.67	0.46	5.18	0.081
		Control		400.48	0.89		

Note. Hedges' g effect size (ES) benchmarks are indicated in the table as follows: Small: 0 to < .10; *Medium*, italicized text: .10 < .30, **Large**: bold and underlined text: .30 or greater. Data source: Matched K-3 MRU80 sample. Kindergarten sample size ctrl= 4007.23, tr= 14,798, First Grade sample size ctrl= 5,321.60, tr= 19,388; Second Grade sample size ctrl= 5,721.207, tr= 21,131.97; Third Grade sample size ctrl= 5,184.922, tr= 19,147.

Table C3. MRU Regression Summary, by grade

	Grade	Condition	P-value	Marginal Mean	St. Error	Diff.	ES
Met Recommended Use	K	Treatment	0.000	178.42	0.37	15.89	**0.400**
		Control		162.53	0.63		
	1	Treatment	0.000	94.59	0.22	8.61	**0.316**
		Control		85.97	0.37		
	2	Treatment	N/S				
		Control					
	3	Treatment	0.000	416.67	0.53	6.59	0.104
		Control		410.08	0.90		

Note. Hedges' g effect size (ES) benchmarks are indicated in the table as follows: Small: 0 to < .10; *Medium*, italicized text: .10 < .30, **Large**: bold and underlined text: .30 or greater. Data source: Matched K-3 MRU sample. Kindergarten sample size ctrl= 4,012.48, tr= 11,415, First Grade sample size ctrl= 5,414.72, tr= 15,194, Second Grade sample size ctrl= 5,731.58, tr= 16,308.84; Third Grade sample size ctrl= 5000.573, tr=14,226.

APPENDIX D. DATA PROCESSING & MERGE SUMMARY

We reviewed and cleaned data from thirteen different sources in preparation of completing our analyses, including program usage data from four software program providers, student literacy achievement data, and demographic data (student information system, "SIS") data from the USBE. Throughout the different stages of data processing, a percentage of cases were dropped from each program vendor. In this Appendix, we show how our pool of treatment students shrank at each stage of the cleaning process and describe how we cleaned the different types of data in the creation of the final datasets used our analyses.

Software Program Data

Each software program provider provided student level data with the time students spent in the software for each week of school. To help vendors provide quality data and ensure consistency across software program providers, vendors received an example data file, a description of the correct format for each variable, and a checklist to conduct a final review of their data. Our cleaning process for the program vendor data files included making sure all program schools that received licenses were included in the data, identifying, and processing duplicate IDs within vendors' data, and formatting variables as needed, among other steps. We reviewed existing variables and created additional variables to use in our analyses, such as total weeks of use, average minutes of use, and other program fidelity measures.

When cleaning duplicate IDs within each vendors' data, we deleted cases that were the same student with different usage reported and kept any unique cases after removing exact replicas. We did not count weeks or include minutes, when there were fewer than five minutes recorded

in a given week. After removing these instances, we updated the usage variables, such as total minutes, to reflect the change in use, and then removed students who had fewer than five minutes of total use from the data. After we cleaned and processed the vendors data, the total count of students went from **164,537** to **161,602** students. We used this data to study program implementation.

To create the vendor data used in our outcome analyses, we identified any IDs that did not comply with the state student ID (SSID) 7-digit format (**529** cases). We also removed duplicate IDs (**4,426** cases), which indicated students used more than one software program, either because they moved to a different district, or because the LEA administered multiple programs to the same students. In either case, we did not include these students in order to report the individual impacts for each software provider. This left us with a file of **156,647** cases.

SIS Data

We were provided SIS data for all students in Grades K-3. We reviewed the SIS data provided by the USBE to ensure that all LEAs who were listed as 2023-2024 participants were included in the data. The SIS data file consisted of **203,052** cases, of which approximately three percent were duplicate records. After cleaning the data of duplicates, our SIS data consisted of **197,587** records.

Acadience Reading Data

In 2023-2024, the USBE prepared and transferred an Acadience Reading data file (n= **183,882**). After cleaning the IDs (e.g., deleting missing IDs and IDs that were not in a valid format),

removing duplicates and removing cases with missing outcome data, we were left with a master Acadience file containing **179,683** cases. This master file contained outcome data for our pool of treatment and control cases.

Master Merged Data File

We merged the SIS data from the USBE into our master Acadience Reading file and were left with **169,147** cases. Next, we merged our master vendor data into the Acadience and SIS data, which left us with **138,541** complete treatment cases and **30,606 control** cases.

Lastly, we identified (where possible) schools or students with program exposure, using one of the eleven program vendors through non-EISP funding. We removed these cases from our pool of potential controls[7]. This included excluding students who used Imagine Learning through a separate state-wide grant[8] prior to reporting the program impacts for similar reasons. After processing the data, our final, pre-matched dataset consisted of **160,274** cases, of which, **138,541** were treatment and **21,733** were potential controls.

Matched Data Files

Before we could run our analyses, the final step was to create our matched control groups. Control students were drawn from a group of children who were not exposed to an early intervention software program (EISP) in 2023-2024. We needed to create a comparison group

[7] We removed students from non-EISP funded schools who were using an EISP program based on information provided by vendors.

[8] We excluded these students from our analyses using the SSIDs provided by Imagine Learning to identify students who used their reading software through this separate state-wide initiative.

that matched the students in our treatment sample. We drew controls from a pool of non-program participants in the state of Utah, and in general, lost very few cases when creating our matched samples for individual vendors and the program-wide analyses which consisted of fewer students. However, for our largest sample of program students, the Intent to Treat (ITT) program-wide sample, there were more program students than control students. This automatically reduced the size of this particular sample.

APPENDIX E. ACADIENCE READING MEASURES

Acadience Reading is a statewide assessment used to measure students' acquisition of early literacy skills at the beginning, middle, and end of the academic year. According to a technical report produced by the Dynamic Measurement Group (Powell-Smith, et al., 2014), *"The Acadience measures map on to the critical early reading skills identified by the National Reading Panel (2002) and include indicators of phonemic awareness, Alphabetic principle, vocabulary and oral language development, accuracy and fluency with connected text, and comprehension."* **Table E1** provides a summary of the Acadience subscales used in our analyses.

Table E1. Acadience Reading Scales

Acadience Reading Scale	Description	Early Literacy Construct	Grade
Composite Score	Acadience Composite Score is a combination of multiple Acadience scores	Overall estimate of reading proficiency	K-6
First Sound Fluency (FSF)	A brief direct measure of a student's fluency in identifying initial sounds in words.	Phonemic Awareness	K
Letter Naming Fluency (LNF)	Assesses a student's ability to recognize individual letters and say their letter names.	Measure is an indicator of risk	K-1
Phoneme Segmentation Fluency (PSF)	Assesses the student's fluency in segmenting a spoken word into its component parts of sound segments.	Phonemic Awareness	K-1
Nonsense Word Fluency (NWF)	Assesses knowledge of basic letter sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant and vowel-consonant words. Designed to measure alphabetic principle and basic phonics.	Alphabetic Principle and Basic Phonics	K-2
Oral Reading Fluency (ORF)	Students are presented with grade-level passages and are asked to read aloud and retell the passage. Measures advanced phonics and word attack skills, accuracy, and fluency with connected text, reading comprehension.	Reading Comprehension Accurate and Fluent Reading of Connected Text	1-6
Maze (MAZE)	Students read a passage with every seventh word replaced by a box containing the correct word and two distractor words. Assesses student's ability to construct meaning from text using word recognition skills, background information and prior knowledge, and familiarity with linguistic properties (e.g., syntax, morphology).	Reading Comprehension	3-6

Evaluation and Training Institute
12240 Venice Blvd, Suite 27
Los Angeles, CA 90066
www.eticonsulting.org

For more information on the
Evaluation and Training Institute, contact ETI:

Jon Hobbs, Ph.D., President
Phone: 310-473 8367
jhobbs@eticonsulting.org