

Early Intervention Program

Report of FY 2014



Prepared by the

Utah State Office of Education

November 19, 2014

Diana Suddreth, Interim Teaching and Learning Director
diana.suddreth@schools.utah.gov

Jennifer Throndsen, Literacy Coordinator
jennifer.throndsen@schools.utah.gov

Sheri Ebert, K-3 Literacy
sheri.ebert@schools.utah.gov

Early Intervention Program

In Compliance with Intent Language of HB513 of the 2012 Legislative General Session

Minimum School Program Title:	Early Intervention Program
USOE Section Reporting:	Teaching and Learning
FY14 Allocation:	\$4,600,000 and \$132,200/year for evaluation- ETI (Evaluation Technology Institute)
Authorization:	53A-17a-167

Program Description

The Utah State Legislature funded the Early Intervention Program (HB 513) to encourage literacy growth and achievement in students grades K-3. In its second year of implementation statewide, the Early Intervention Program offers, through a competitive process, local education agencies (LEAs) the option to select from five interactive computer software programs designed to improve reading. LEAs participating in year one were given first priority to receive equivalent licenses during year two. An external evaluator, Evaluation and Training Institute (ETI) compiled and disaggregated data for this report.

Early Intervention Program and Five Interactive Software Vendors

The Early Intervention Program FY14 allocation was \$4,600,000. Through an RFP process, five software vendors were awarded contracts to supply interactive software licenses, staff training and curricular and technical support to participating Local Education Agencies and to individual school sites. **Table 1** shows the breakdown of schools and students by vendor.

Table 1. Software Vendors Used by Program Participants

Program	# Schools	# Students	% Students of Total Sample
Imagine Learning	131	18,455	48%
i-Ready	49	8,102	21%
SuccessMaker	22	4,357	11%
Waterford	33	4,236	11%
IStation	19	3,403	9%
Totals	254	38,553	100%

Source: USOE staff communication

Evaluation of Student Learning Gains

Method

A mixed-method evaluation design was used, which included student demographics including risk factors, program usage fidelity, student data resulting from software usage, and student testing results from school-based Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next benchmarks.

Student Demographics and Risk Indicators

The Early Intervention Program targets K-3 students with at-risk indicators. **Table 2** shows the demographic information for all the students who participated in the program. Specifically, 45 percent of student participants were from low-income households (as measured by eligibility for free or reduced lunch). Additionally, 56 percent of program participants attended a Title 1 school.

Table 2. Demographic Characteristics of Program Participants

	N	% Total Sample
Gender		
Female	17,250	45%
Male	18,962	49%
Unknown	2,341	6%
Race/Ethnicity		
White	27,493	71%
Hispanic	5,694	15%
Other	3,025	8%
Unknown	2,341	6%
Title 1 School Status		
School wide	16,072	42%
Targeted Assistance	5,187	14%
Low Income Family	17,225	45%
English Language Learner	3,162	8%
Special Education	4,923	13%
Migrant Family	61	0%

Source: USOE Demographic Data, program vendor data

DIBELS Next: Beginning of Year Literacy Benchmark Indicating Student Risk

Table 3 shows that approximately 42 percent of students participating in the program scored below benchmark on the beginning of year DIBELS literacy benchmarks.

Table 3. DIBELS Next BOY Literacy Benchmarks

DIBELS Literacy Benchmarks: (Beginning of Year)	Program Participants
At or Above Benchmark	8,965 53%
Below Benchmark	2,476 15%
Well Below Benchmark	4,614 27%
N/A	854 5%
Total	16,909 100%

Program Fidelity

The Utah State Office of Education (USOE) defines fidelity according to the following criteria: *“Schools will be considered to be implementing appropriately when 90 percent of the enrolled students complete 90 percent of the average number of minutes calculated per week from the start date (initial assessment) to the end of the student’s program period.”*

Results from the fidelity implementation analysis are shown below for schools and Local Education Agencies (LEAs). **Table 4** is a summary of program fidelity by school with 11 percent of schools meeting the fidelity requirement set by USOE. **Table 5** summarizes fidelity of usage by district with 3 percent of LEAs meeting the fidelity requirement set by USOE.

Table 4. Summary of Program Fidelity by School

Indicator	Count	%
Met fidelity requirements	16	11%
Did not meet fidelity requirements	129	89%
Total Schools	145	100%

Table 5. Summary of Program Fidelity by District

Indicator	Count	%
Met fidelity requirements	1	3%
Did not meet fidelity requirements	31	97%
Total Districts	32	100%

Student Learning Outcomes

Evaluation of student learning outcomes was measured from two sources of data:

- Vendors’ proprietary testing outcomes (recorded in the software program); and,
- DIBELS Next literacy benchmark categories and test scores (conducted in schools).

Software Vendor Testing Outcomes

Q. Did the computer-based program improve literacy skills as measured by the vendor?

A: Analyses of vendor reported student outcomes show a positive relationship between time spent on the program and an increase in students' scores. This held true for the following vendors: i-Ready, iStation, SuccessMaker, and Waterford ERP.

A: Imagine Learning: the results indicate that a majority of students met (or almost met) Imagine Learning's recommendation for proficiency in a skill.

iReady

Curriculum Associates provided ETI with student outcome measures that included: first and final diagnostic overall scaled scores (300-800), the grade level equivalency for each literacy strand (early, mid, late plus the grade) and the grade level at first and final sessions. Filters were added to exclude students whose initial or final diagnostic scores were below 300, as these scores were considered to be a result of measurement error.

- Across all grade levels, the average final test score was significantly higher than the pre-test score (473 versus 440). This was especially true for younger students, with kindergarten and first grade students scoring 38.84 and 42.61 points higher on the final test, respectively.
- Linear regression results indicate a positive, significant relationship between final test scores and the amount of time students' spent on the program across all grade levels.
- For students with the same pre-test score, those who spent more time on the program had higher second test results.
- Program usage had a slightly stronger effect on test scores for kindergarten students, followed by students in the first grade.
- For every hour spent on the software, the final test scores went up by approximately eight-tenths of a point.

iStation

iStation provided ETI with the following student outcome measures: overall ability scores at first and final sessions; ability scores at first and final session for each literacy strand (eight domains); and students percentile rank at first and final session.

- ETI regressed total time on the software on final test scores, holding first test scores constant.
- The results of the linear regression show a slight but statistically significant, positive relationship between program usage and final test scores for 1st- 3rd grade students. This effect was the most pronounced in 1st grade students.

- In contrast, time spent on the software was not statistically significant for kindergarten students. However, after running regressions for final test scores for each literacy strand, we found a statistically significant, positive relationship between program use and two literacy strands: listening comprehension and letter knowledge.

SuccessMaker

SuccessMaker provided ETI with a score for students starting and ending course level (from 0.00 – 4.00) and the percentage of skills and exercises correct.

- Regression analysis indicated a positive and significant relationship between time spent on the program and students ending course level (final scores), regardless of students beginning course level (starting score).
- The relationship between program use and final scores was moderate to strong. This relationship was most evident for 3rd grade students, in which 1 hour spent on the software was associated with an increase of one course level.

Waterford

Waterford provided difficulty level scores at first and final session for each literacy strand (5 domains). Scores ranged from 100 to 3000 (Level 1 difficulty=100-1000; Level 2 difficulty=1001-2000; Level 3 difficulty=2001-3000).

- ETI computed a growth score (Level at first session – Level at final session) to use as the outcome variable in our regression models.
- ETI regressed the growth scores for each literacy strand on the total time spent on the software. The results suggest that students growth scores increased as students spent more time on the software.
- Total time on the software was statistically significant across all 5 literacy strands: Phonological Awareness; Phonics; Fluency; Comprehension and Vocabulary; and Language Concepts.
- This relationship was strongest for “Comprehension and Vocabulary” and “Phonics”, for which scores increased by approximately 19 points for every hour on the software, respectively.

Imagine Learning

Imagine Learning provided ETI with a final proficiency score for each literacy strand (eight domains), students starting and ending points in the curriculum (percent of the curriculum where the students started and ended), and the percent gain (difference between starting and ending points in the curriculum).

There were several limitations to using regression to measure student performance with the Imagine Learning data:

- First, ETI did not have a pre-test variable to use as a control, which limited the ability to determine how time on the software affected proficiency.
- Second, the program ends once a student progresses all the way through the curriculum, which caused students with higher starting scores to experience low growth.
 - As a result of these constraints, ETI looked for other ways to examine student performance.
 - ETI generated frequencies for students who scored over and under 80 percent proficiency (80 percent is considered proficient by Imagine Learning) for each subscale and grade.
 - A majority of students scored above 80 percent proficiency for 4 out of 8 literacy strands: Phonological Awareness; Comprehension, Vocabulary, Fluency; Listening Comprehension; and Basic Vocabulary across all grade levels.
 - Slightly more than half of the students scored less than 80 percent proficiency for the remaining four strands, with the exception of kindergarten students.

Student Achievement Outcomes: DIBELS Next Benchmark Scores

DIBELS Next measures are used statewide to screen for literacy risk factors and to measure growth in foundational literacy skills and reading fluency. ETI analyzed DIBELS benchmark data using two methods:

- A descriptive analyses of trends in DIBELS Next data compared across software program (“treatment”) students and non-program (“control students”; the descriptive analyses were conducted prior to implementing a full set of inferential analyses); and,
- Multi-level regression models were computed to compare treatment students to control students while controlling for other factors that could influence achievement.
 - Sample of students for these analyses included only those treatment students that met the vendor recommended minimum program use requirements (i.e. students that met program fidelity). In addition, treatment students were matched to ensure a balance across covariates.

DIBELS Next Benchmarks: Descriptive Findings

Q: Do students who use the software program have better learning outcomes than those who do not as measured by DIBELS Next test scores?

A: The findings show that two factors influence treatment outcomes: Grade and Dosage of treatment:

- *Kindergarten students had more treatment benefits than 1st and 2nd grade, with no positive effects observed in 3rd grade.*
- *The “Optimal Dosage” for treatment was between 120 minutes to 1,300 minutes of software use, and was associated with positive treatment effects.*
- *“High Dosage” of treatment, over 1,300 minutes, was not associated with positive treatment effects.*

DIBELS Next benchmark rankings were applied for treatment and control students from beginning of the year (BOY) to end of the year (EOY) benchmarks. The results are depicted in **tables 6 and 7**. For ease in reading, bold borders are imposed on the table cells showing movement from lower benchmarks at BOY (Below Bench or Well Below Bench) to higher levels at the EOY benchmark.

During the visual analyses of benchmark changes from BOY to EOY, it was identified that students with the “extended dosage” (minutes of software use) of 1300 minutes or more showed little to no improvement in the BOY to EOY literacy benchmark when compared to control students. **Table 6** shows findings related to the extended dosage treatment students when compared to the control students.

**Table 6: DIBELS Next Benchmark Levels at Beginning (BOY) and End (EOY)
HIGH DOSAGE (>1300 minutes) Treatment and Control Students**

<i>1st Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	90%	8%	3%	91%	6%	3%
Below Bench	63%	20%	17%	67%	17%	16%
Well Below Bench	35%	20%	45%	38%	19%	43%
<i>2nd Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	87%	10%	2%	90%	8%	2%
Below Bench	48%	29%	23%	49%	32%	20%
Well Below Bench	15%	19%	66%	13%	20%	68%
<i>3rd Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	92%	6%	1%	92%	6%	2%
Below Bench	62%	24%	15%	62%	24%	14%
Well Below Bench	15%	23%	62%	19%	18%	63%
<i>Total</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	89%	9%	2%	90%	8%	2%

Below Bench	61%	25%	14%	62%	23%	15%
Well Below Bench	34%	23%	43%	30%	20%	50%

**Table 7: DIBELS Next Benchmark Levels at Beginning (BOY) and End (EOY)
OPTIMAL DOSAGE (120 to 1300 minutes) Treatment and Control Students**

<i>Kindergarten</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	93%	6%	1%	84%	13%	3%
Below Bench	83%	17%	0%	68%	23%	10%
Well Below Bench	74%	15%	12%	53%	25%	22%
<i>1st Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	94%	4%	2%	91%	6%	3%
Below Bench	68%	20%	12%	67%	17%	16%
Well Below Bench	35%	26%	40%	38%	19%	43%
<i>2nd Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	92%	7%	1%	90%	8%	2%
Below Bench	61%	25%	14%	49%	32%	20%
Well Below Bench	13%	18%	69%	13%	20%	68%
<i>3rd Grade</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below
At or Above Bench	95%	4%	1%	92%	6%	2%
Below Bench	58%	25%	17%	62%	24%	14%
Well Below Bench	22%	15%	63%	19%	18%	63%
<i>All Grades</i>	Treatment			Control		
	(EOY)			(EOY)		
(BOY)	At or Above	Below	Well Below	At or Above	Below	Well Below

At or Above Bench	93%	6%	1%	90%	8%	2%
Below Bench	71%	21%	9%	62%	23%	15%
Well Below Bench	39%	19%	42%	30%	20%	50%

Table 7 shows comparisons of student outcomes by grade with an **Optimal Dosage** of between 120 to 1,300 minutes of use (gains shown in green). Important findings included:

- Across all grades (K-3):
 - 71 percent of treatment students moved from Below Bench at BOY to At or Above Bench at EOY vs. 62 percent of control students (a 9 percent difference).
 - 39 percent of treatment students moved from Well Below Bench to At or Above bench vs. 30 percent of control students (a 9 percent difference).
- Kindergarten:
 - 83 percent of treatment students moved from Below Bench to At or Above Bench vs. 49 percent in the control group (a 15 percent difference).
 - 74 percent of treatment students moved from Far Below Bench to At or Above Bench vs. 53 percent of control students (a 21 percent difference).
- 2nd Grade:
 - 61 percent of treatment students moved from Below Bench at BOY to At or Above Bench at EOY vs. 49 percent of control students (a 12 percent difference).

DIBELS Next EOY Regression Analyses

Q: Do students who use the software program have better learning outcomes than those who do not as measured by DIBELS Next test scores?

A: Kindergarten students who used the program for between 120 to 1,300 minutes had better scores than non-program students on the following End of Year test scores:

- *Composite literacy scores.*
- *Letter naming fluency*
- *Nonsense word fluency: correct letter sounds*
- *Nonsense Word Fluency: whole words read*

A: Large effect sizes were observed for treatment effects in Kindergarten.

A: Number of minutes using the software above 1,300 minutes is negatively associated with almost every subscale and grade level (except Kindergarten), but is significant in only a few cases.

Graphical Depiction of Learning Outcomes

Figures 1 - 4 (below) visually display the positive treatment effects in Kindergarten by comparing mean scores and a linear model between students who used the software to those who did not.

Figure 1. Kindergarten Program Effects on Nonsense Word Fluency: Correct Letter Sound

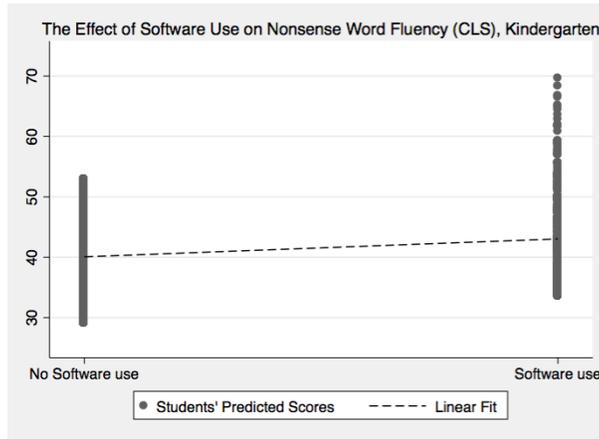


Figure 2. Kindergarten Program Effects on Composite Scores

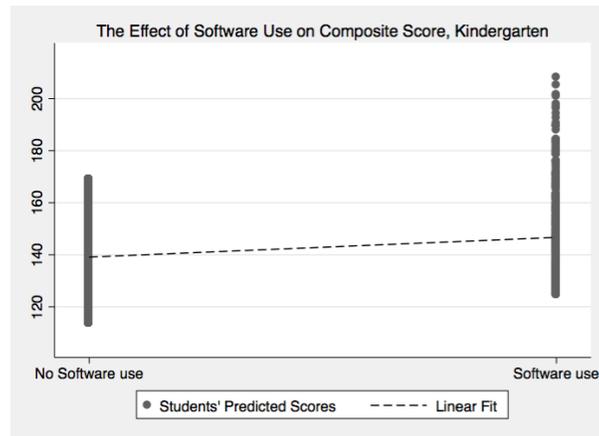


Figure 3. Kindergarten Program Effects on Letter Naming Fluency

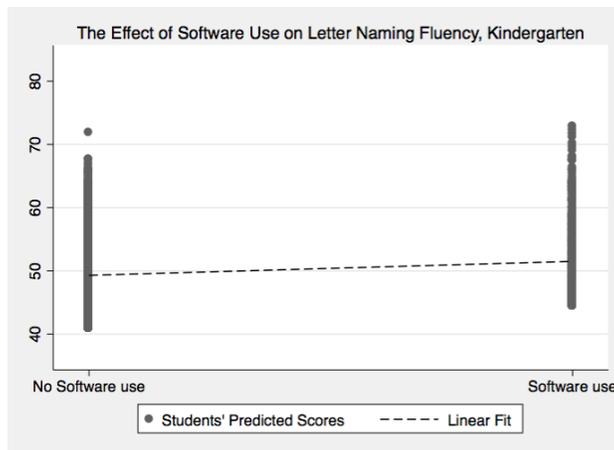


Figure 4. Kindergarten Program Effects on Nonsense Word Fluency: Whole Words Read

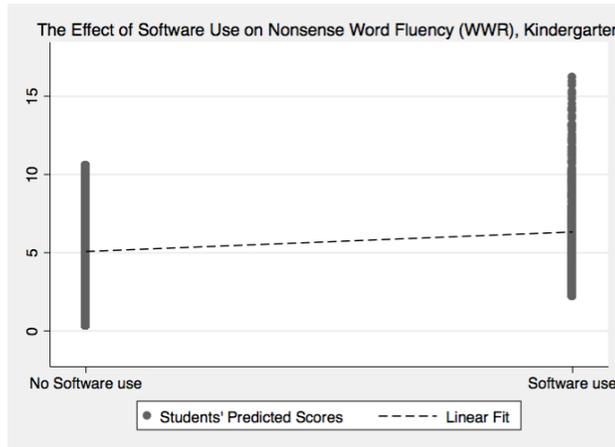
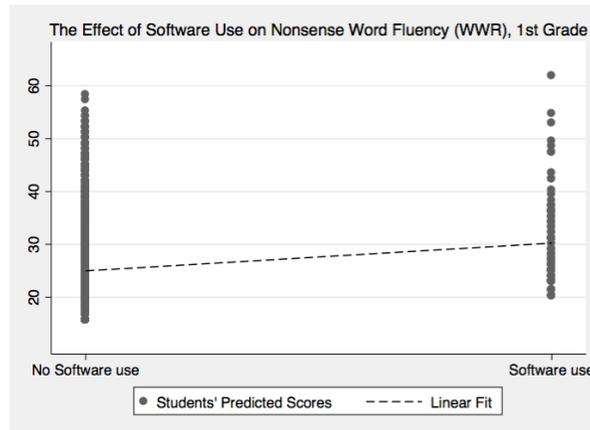


Figure 5. 1st Grade Program Effects on Nonsense Word Fluency: Whole Words Read



Effect Sizes

Any coefficient under .05 is too small to be considered meaningful, from .05 to .10 is small but meaningful, and .10 to .25 is a large effect. As can be seen in **Table 8**, all standardized coefficients computed for Kindergarten suggest large treatment effects for software users. The largest treatment effect is found in 1st grade, with the coefficient for Nonsense Word Fluency, whole words read (.30).

Table 8. Standardized Coefficients for the Influence of Program Results Predicting DIBELS Subscale Scores:
Optimal Treatment
N = 5,144 individuals from 107 schools

DIBELS Measure	Kindergarten	1 st Grade	2 nd Grade	3 rd Grade	Overall
Comp	0.27*	0.00	-0.05	-0.01	0.02
LNF	0.22*				.023**
PSF					
NWF-CLS	.22**	0.14			0.19
NWF-WWR	.23**	0.30**			.30**
DORF-Accuracy				0.04	.05*
DAZE				0.01	0.01

Summary Findings

Student Learning Gains

DIBELS Next literacy benchmark scores indicated strong evidence that the Early Intervention Program is associated with early literacy gains in Kindergarten and, to a lesser extent, 1st Grade. Students in the 1st grade who used the software over 1,300 minutes had some significant negative treatment effects.

In 2nd and 3rd Grade, students' DIBELS Next benchmark scores showed no evidence that participation in the program increased oral reading fluency; however, the program benefits were mediated by a dosage effect: students had benefits at an optimal level of software use from 120 to 1,300 minutes.

Data: Harvesting and Disaggregating

In year one of the Early Intervention Program, LEAs and schools did not use consistent student identification numbers. This lack of consistency, served as a barrier in the harvesting and disaggregating of the data for this report. In year two, the USOE's Teaching and Learning and Assessment Departments disseminated a memo to all LEAs requiring the use of a consistent student identification number statewide. This consistent student ID streamlined the data portions of this report.

Software Usage

Fidelity of software usage is defined by the software vendors and in some cases, may be for excessive time periods. Expecting young children to use software for long periods of time may have negative learning effects due to the following:

- Time on the software interferes with core instructional time in the classroom.
- Teachers are keeping students who struggle with reading on the software longer in hopes that eventually it will work; and,
- Software usage for extended periods of time is not developmentally appropriate as young students may become bored, and in turn, bars the learning process.

It may be beneficial for USOE to recommend fidelity of software usage, based on evidence, rather than simply relying on the recommendations of software vendors.