



Waterford Research Institute, LLC  
ESSA Evidence-Based Research

*An Alignment of Waterford Studies to ESSA's  
Four Tiers of Evidence*

By: Dr. Haya Shamir

## Table of Contents

<b>List of Tables .....</b>	<b>4</b>
<b>List of Figures .....</b>	<b>6</b>
<b>List of Acronyms .....</b>	<b>9</b>
<b>Tier I: Strong Evidence .....</b>	<b>10</b>
13 Rural UPSTART Preschool Study .....	10
13 WACS Randomized Control Trial .....	12
Randomized Controlled Trial Evaluation of Waterford Early Learning in Indiana .....	14
<b>Tier II: Moderate Evidence .....</b>	<b>18</b>
Evaluation of the Waterford Early Reading Program in Ohio .....	18
Effectiveness of UPSTART .....	20
Evaluation of the UPSTART Elementary Program .....	30
High-Quality School Readiness Expansion in Utah, 2017-2018.....	31
<b>Tier III: Promising Evidence .....</b>	<b>33</b>
Evaluation of a Tucson Unified School District.....	33
Evaluation of a School District in Florida, 2014-2015 .....	35
Evaluation of the Waterford Early Reading Program in California.....	39
Evaluation of the Waterford Early Math and Science Program in Indiana, 2015-2016.....	40
Effectiveness of UPSTART .....	48
Evaluation of a School District in South Carolina, 2016-2017 .....	59
Evaluation of Waterford Early Learning in Maryland, 2016-2017 .....	60
Evaluation of Waterford Early Learning in Tennessee.....	65
Evaluation of Waterford Early Learning in Maryland, 2017-2018.....	68
<b>Tier 4: Demonstrated Rationale .....</b>	<b>70</b>
Early Studies .....	70
Case Studies.....	70
Later Effectiveness Studies.....	71
Evaluation of the Waterford Early Reading Program in Alabama.....	73
Evaluation of the Waterford Early Reading Program in Colorado .....	75
Evaluation of the Waterford Early Reading Program in South Carolina.....	76
Evaluation of the Waterford Early Reading Program in Florida.....	78
Evaluation of the Waterford Early Reading Program in Indiana, 2015-2016.....	79

---

Evaluation across Two School Districts in Texas and Florida, 2015-2016 .....	83
Evaluation across Two School Districts in South Carolina and Indiana, 2015-2016 .....	85
Evaluation across Two School Districts in Texas and Indiana, 2015-2016 .....	86
Evaluation of a School District in Texas, 2015-2016 .....	88
Evaluation of Waterford Early Learning in Idaho, 2017-2018 .....	92
Evaluation of Waterford Early Learning in Three School Districts .....	94
Longitudinal Evaluation of Waterford Early Learning in Florida, 2014-2016 .....	96
Longitudinal Evaluation of Waterford Early Learning in South Carolina, 2015-2017 .....	99
Longitudinal Evaluation of a School District in Maryland, 2015-2018.....	100
Effectiveness of UPSTART .....	104
<i>References</i> .....	<i>115</i>

# List of Tables

Table 1: Brigance Subscale Posttest Means of Treatment-Control Groups .....	10
Table 2: Brigance Subscale Posttest Effect Size Estimates .....	11
Table 3: PELI Subscale Posttest Means of Treatment-Control Groups .....	11
Table 4: PELI Subscale Posttest Effect Size Estimates .....	11
Table 5: WACS End of Year Scores Controlling for Beginning of Year Scores.....	12
Table 6: Overall WACS End of Year Scores by Demographics .....	13
Table 7: Blending WACS End of Year Scores by Demographics.....	13
Table 8: Letter Sound WACS End of Year Scores by Demographics .....	13
Table 9: Letter Recognition WACS End of Year Scores by Demographics .....	14
Table 10: Descriptive Statistics for All Tasks .....	19
Table 11: Brigance Growth Rate Comparisons.....	24
Table 12: Bader Growth Rate Comparisons .....	24
Table 13: Effect Size Estimates by Literacy Domain .....	32
Table 14: ELL & Non-ELL Students on DIBELS Total Reading Score .....	34
Table 15: ANCOVA & Effect Sizes on All Outcome Measures (ERP Students With 1100 or More Usage Minutes).....	34
Table 16: Spring VPK Sub-Strand Scores, Covarying for Fall.....	35
Table 17: Effects of Demographics on Scores for Each Sub-Strand.....	36
Table 18: Brigance Posttest Analysis of Treatment-Control Group Differences .....	49
Table 19: Bader Posttest Analysis of Treatment-Control Group Differences.....	49
Table 20: Treatment-Control Group Differences in Growth Rates on the Brigance .....	50
Table 21: Treatment-Control Group Differences in Growth Rates on the Bader.....	50
Table 22: Brigance Posttest Analysis of Treatment-Control Group Differences .....	51
Table 23: Bader Posttest Analysis of Treatment-Control Group Differences.....	52
Table 24: Treatment-Control Group Differences in Growth Rates on the Brigance .....	53
Table 25: Treatment-Control Group Differences in Growth Rates on the Bader.....	54
Table 26: Brigance Posttest Analysis of Treatment-Control Group Differences .....	54
Table 27: Brigance Effect Size Estimates.....	55
Table 28: Bader Posttest Analysis of Treatment-Control Group Differences.....	55
Table 29: Bader Effect Size Estimates .....	56
Table 30: Treatment-Control Group Differences in Growth Rates on the Brigance .....	56
Table 31: Treatment-Control Group Differences in Growth Rates on the Bader.....	56
Table 32: Brigance Posttest Analysis of Treatment-Control Group Differences .....	57
Table 33: Brigance Effect Size Estimates.....	58
Table 34: Bader Posttest Analysis of Treatment-Control Group Differences.....	58
Table 35: Bader Effect Size Estimates .....	58
Table 36: Treatment-Control Group Differences in Growth Rates on the Brigance .....	59
Table 37: Treatment-Control Group Differences in Growth Rates on the Bader.....	59
Table 38: Kindergarten DRA Gains & End of Year Scores Covarying For Middle of Year Scores .....	60
Table 39: First Grade DRA Gains & End of Year Scores Covarying For Middle of Year Scores .....	60
Table 40: Performance Measure Means at the Beginning, Middle, & End of the Year .....	71
Table 41: Beginning of Year First Grade & Second Grade Terra Nova Reading Performance Group Means .....	72
Table 42: District 1 Overall Kindergarten TPRI Gains by Substrand .....	83
Table 43: District 1 Overall Kindergarten TPRI Gains by ELL Students .....	84

Table 44: District 1 Overall Second Grade TPRI Gains by Substrand.....	84
Table 45: District 1 Overall Second Grade TPRI Gains by ELL Students.....	84
Table 46: District 2 Overall VLT End of Year Scores.....	84
Table 47: District 2 VLT End of Year Scores by ELL Students.....	84
Table 48: District 1 Kindergarten DRA End of Year Scores.....	85
Table 49: District 1 First Grade DRA Scores.....	85
Table 50: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores.....	85
Table 51: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores by Low-SES ...	85
Table 52: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores.....	85
Table 53: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Low-SES	86
Table 54: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores.....	86
Table 55: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores by Low-SES ..	86
Table 56: Kindergarten TPRI Gains by Special Education Status.....	90
Table 57: Kindergarten TPRI Gains by EL Status.....	90
Table 58: Kindergarten TPRI Gains by Ethnicity.....	90
Table 59: First Grade TPRI Gains by Special Education Status.....	91
Table 60: First Grade TPRI Gains by EL Status.....	91
Table 61: First Grade TPRI Gains by Ethnicity.....	92
Table 62: District 1 – Kindergarten TPRI Gains by Ethnicity.....	94
Table 63: District 1 – First Grade TPRI Gains by Ethnicity.....	95
Table 64: District 2 – Second Grade DRA Percent Gains by Ethnicity.....	95
Table 65: District 3 – Kindergarten VLT End of Year Scores by Ethnicity.....	95
Table 66: District 3 – First Grade VLT End of Year Scores by Ethnicity.....	95
Table 67: End of Kindergarten Scores by Strand & Demographics.....	98
Table 68: End of First Grade DRA Scores.....	99
Table 69: End of Second Grade DRA Scores.....	99
Table 70: End of First Grade High Usage DRA Scores.....	100
Table 71: End of Second Grade High Usage DRA Scores.....	100
Table 72: WACS Question Difficulty Ranges by Grade.....	104
Table 73: WACS Year 5 Posttests.....	106
Table 74: WACS Year 6 Posttests.....	109

# List of Figures

Figure 1: MAP End of Year Scores by Strand .....	15
Figure 2: MAP RIT Scores by Demographics.....	15
Figure 3: MAP Reading Foundations Scores by Demographics .....	16
Figure 4: MAP End of Year Scores - African American Students .....	16
Figure 5: TRC End of Year Lexile Levels .....	17
Figure 6: TRC End of Year Lexile Levels by Demographics.....	17
Figure 7: TRC End of Year Proficiency Benchmarks.....	18
Figure 8: Brigance & Bader Posttest Analysis of Composite Scores.....	21
Figure 9: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests.....	21
Figure 10: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores .....	22
Figure 11: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores .....	22
Figure 12: Pre-Literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores.....	23
Figure 13: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores.....	23
Figure 14: Brigance & Bader Posttest Analysis of Composite Scores.....	25
Figure 15: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests.....	25
Figure 16: Phonological Awareness: Treatment and Control Group Pretest & Posttest Mean Scores.....	26
Figure 17: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores .....	26
Figure 18: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores.....	27
Figure 19: Brigance & Bader Posttest Analysis of Composite Scores.....	28
Figure 20: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests.....	28
Figure 21: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores .....	29
Figure 22: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores .....	29
Figure 23: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores.....	30
Figure 24: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores.....	30
Figure 25: Listening Comprehension Benchmarks at Posttest.....	32
Figure 26: Post-Program Social-Emotional Development Skills by Program Group.....	33
Figure 27: Spring VPK Sub-Strand Scores, Covarying for Fall .....	36
Figure 28: Oral Language Vocabulary Spring Scores by Demographics .....	37
Figure 29: Phonological Awareness Spring Scores by Demographics .....	37
Figure 30: Print Knowledge Spring Scores by Demographics.....	38
Figure 31: Math Spring Scores by Demographics.....	38
Figure 32: Mean Scores by AIMSweb Sub-Strands for Kindergarten Students.....	39
Figure 33: Mean Scores by AIMSweb Sub-Strand NWF for First Grade Students.....	39
Figure 34: Kindergarten mCLASS: Math End of Year Scores by Sub-Strand .....	41
Figure 35: Kindergarten mCLASS: Math Counting End of Year Scores .....	41
Figure 36: Number Identification End of Year Scores by Demographics.....	42
Figure 37: Quantity Discrimination End of Year Scores by Demographics.....	43
Figure 38: First Grade mCLASS: Math End of Year Scores by Sub-Strand .....	44
Figure 39: First Grade mCLASS: Math Counting End of Year Scores .....	44
Figure 40: First Grade Number Facts End of Year Scores by Demographics .....	45
Figure 41: First Grade Quantity Discrimination End of Year Scores by Demographics .....	46
Figure 42: First Grade Missing Number End of Year Scores by Demographics .....	47
Figure 43: First Grade Quantity Discrimination End of Year Scores by Demographics .....	48
Figure 44: UPSTART’s Impact on the Brigance in Effect Size Units.....	52
Figure 45: UPSTART’s Impact on the Bader in Effect Size Units.....	53

Figure 46: Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Strand..... 61

Figure 47: Kindergarten End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics ..... 61

Figure 48: Kindergarten End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics ..... 61

Figure 49: Kindergarten End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics ..... 62

Figure 50: Kindergarten End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics ..... 62

Figure 51: First Grade End of Year Scores Covarying for Beginning of Year Scores by Strand ..... 62

Figure 52: First Grade End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics ..... 63

Figure 53: First Grade End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics ..... 63

Figure 54: First Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics ..... 63

Figure 55: First Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics ..... 64

Figure 56: Second Grade End of Year Scores Covarying for Beginning of Year Scores by Strand ..... 64

Figure 57: Second Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics ..... 64

Figure 58: Second Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics ..... 65

Figure 59: Kindergarten Beginning & End of Year Overall RIT Scores ..... 65

Figure 60: Kindergarten Beginning & End of Year RIT Scores by Strand ..... 66

Figure 61: Kindergarten End of Year RIT Scores while Covarying for Beginning of Year Scores ..... 66

Figure 62: First Grade Beginning & End of Year RIT Scores..... 67

Figure 63: First Grade Beginning & End of Year RIT Scores by Strand..... 67

Figure 64: First Grade End of Year RIT Scores while Covarying for Beginning of Year Scores..... 68

Figure 65: Second Grade MAP End of Year Scores by Strand..... 68

Figure 66: Second Grade MAP End of Year RIT to Reading Scores..... 69

Figure 67: Second Grade MAP End of Year Scores Controlling for Beginning of Year Scores by Strand.... 69

Figure 68: Second Grade MAP End of Year RIT to Reading Scores Controlling for Beginning of Year Scores ..... 69

Figure 69: Kindergarten Students’ Phonological Awareness Growth ..... 72

Figure 70: Reading Gains for Students with Low, Moderate, & High Reading Skills at First Grade ..... 73

Figure 71: Mean Gains by STAR Sub-Strands for First Sample Group ..... 74

Figure 72: Mean Gains by STAR Sub-Strands for Second Sample Group ..... 74

Figure 73: Mean Posttest Scores for DIBELS Sub-Strands for Third Sample Group ..... 75

Figure 74: Mean Reading Gains by DRA Overall Kindergarten Scores ..... 75

Figure 75: Kindergarten DRA End of Year Scores ..... 76

Figure 76: Kindergarten DRA End of Year Scores by Demographics ..... 77

Figure 77: First Grade DRA End of Year Scores ..... 77

Figure 78: First Grade DRA End of Year Scores by Demographics..... 78

Figure 79: End of Year VLT Scores by Grade..... 79

Figure 80: First Grade DRA Percent Gains..... 80

Figure 81: First Grade DRA Percent Gains by Demographics ..... 81

Figure 82: Second Grade DRA Percent Gains .....	81
Figure 83: Second Grade DRA Percent Gains by Demographics .....	83
Figure 84: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Whole Words Read .....	87
Figure 85: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Letter Naming Fluency & Correct Letter Sounds.....	87
Figure 86: mCLASS DIBELS Next and Math Scores by Strand by Special Education Control—Number Identification & Quantity Discrimination .....	87
Figure 87: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Whole Words Read.....	87
Figure 88: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Letter Naming Fluency & Correct Letter Sounds .....	87
Figure 89: mCLASS DIBELS Next and Math Scores by Strand by Non-Special Education Control—Number Identification & Quantity Discrimination .....	87
Figure 90: TPRI Gain Scores by Strand by Special Education Control.....	88
Figure 91: TPRI Gain Scores by Strand by Non-Special Education Control.....	88
Figure 92: Kindergarten TPRI Gains by Substrand.....	89
Figure 93: Kindergarten TPRI Gains by Substrand.....	89
Figure 94: First Grade TPRI Gains by Substrand .....	91
Figure 95: STAR Early Literacy Scaled Score Gains from Beginning of Year to End of Year.....	93
Figure 96: STAR Early Literacy Gains from Beginning of Year to End of Year by Strand.....	93
Figure 97: IRI Gains from Beginning of Year to End of Year by Strand.....	94
Figure 98: KRT Overall End of Year Scores by Demographics.....	96
Figure 99: KRT Letter ID & Sounds End of Year Scores by Demographics.....	97
Figure 100: KRT Concepts of Print End of Year Scores by Demographics .....	97
Figure 101: KRT Phonemic Awareness End of Year Scores by Demographics.....	98
Figure 102: Three Years of Usage vs. No Usage End of Year Scores by Strand .....	101
Figure 103: Three Years of Usage vs. No Usage RIT to Reading End of Year Scores .....	101
Figure 104: Two Years of Usage vs. No Usage End of Year Scores by Strand.....	102
Figure 105: Two Years of Usage vs. No Usage RIT to Reading End of Year Scores.....	102
Figure 106: Three Years of Usage vs. Two Years of Usage End of Year Scores by Strand .....	103
Figure 107: Three Years of Usage vs. Two Years of Usage RIT to Reading End of Year Scores .....	103
Figure 108: Year 1-6 WACS Overall Scores.....	105
Figure 109: Year 5 WACS Reading Scores, Grouped by Subtests .....	106
Figure 110: WACS Reading Score Gains Grouped by Ethnicity .....	107
Figure 111: WACS Reading Score Gains Grouped by Other Preschool Attendance.....	107
Figure 112: WACS Reading Score Gains Grouped by SES.....	108
Figure 113: Year 6 WACS Reading Scores, Grouped by Subtests .....	108
Figure 114: WACS Reading Score Gains Grouped by Ethnicity .....	109
Figure 115: WACS Reading Score Gains Grouped by Other Preschool Attendance.....	110
Figure 116: WACS Reading Score Gains Grouped by SES.....	110
Figure 117: WACS Reading Score Gains Grouped by SES.....	111
Figure 118: UPSTART Students & State Average DIBELS Scores, Grades 1 through 3.....	111
Figure 119: UPSTART Students & State Average SAGE Overall Scores, Grades 3 & 4.....	111
Figure 120: UPSTART Students & State Average SAGE SPED Scores, Grades 3 & 4.....	112
Figure 121: UPSTART Students & State Average SAGE Minority Scores, Grades 3 & 4.....	113
Figure 122: UPSTART Students & State Average SAGE Low Income Scores, Grades 3 & 4.....	113



Figure 123: UPSTART Students & State Average SAGE EL Scores, Grades 3 & 4..... 114

## List of Acronyms

ACRONYM	MEANING
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DRA	Developmental Reading Assessment
ELL	English Language Learners
EMS	Waterford Early Math and Science Program
ERP	Waterford Early Reading Program
ES	Effect Size
ILS	Integrated Learning System
KRT	Kindergarten Readiness Test
LEP	Limited English Proficiency
LNF	Letter Name Fluency
LSF	Letter Sound Fluency
MAP	Measures of Academic Progress
NWEA	Northwest Evaluation Association
NWF	Nonsense Word Fluency
PAT	Phonological Abilities Test
PELI	Preschool Early Literacy Indicators
RCT	Randomized Control Study
RIT	Rasch Unit (scale used in the NWEA MAP Growth test)
SAGE	Student Assessment of Growth and Excellence
SES	Socio-Economic Status
TPRI	Texas Primary Reading Inventory
TRC	Text Reading Comprehension
VPK	Voluntary Pre-Kindergarten assessment
WACS	Waterford Assessment of Core Skills
WEL	Waterford Early Learning
WERI	Waterford Early Reading Instrument
WWC	What Works Clearinghouse

## Waterford Reading Academy

Waterford Early Learning: Reading (formerly Waterford Early Reading Program) is the PreK-2 sequence of Waterford Reading Academy.

Waterford Early Learning has been formally assessed in a variety of schools and districts of varying size, location, and socioeconomic status, and results are consistent in supporting the software’s considerable effectiveness. The consistency of the research results, both within and between studies, is striking. In each of the studies, students using Waterford programs outperformed comparison-groups in most, if not all, of the examined assessment measures. In no case did the comparison group outperform Waterford students. Waterford has always focused its development and iteration on research; because of this, its software has demonstrated remarkable strength, robustness, and adaptability. Results have been consistent in a wide variety of early-education contexts, regardless of which assessments have been used. Waterford’s software has been proven to be a flexible tool for helping children reach their full potential.

## Tier I: Strong Evidence

For a study to be considered strong evidence, it must meet the What Works Clearinghouse (WWC) Evidence Standards without reservations. This includes having a randomized control design, with a level of attrition below the required threshold. Additionally, the intervention in the study must demonstrate a statistically significant and positive effect on a student outcome, without having statistically significant and negative outcomes in other studies that meet WWC Evidence Standards with or without reservations. The sample analyzed must be large, include multiple sites, and be representative of the population of interest.

### 13 Rural UPSTART Preschool Study

The Rural UPSTART program prepares children for success upon entering kindergarten by providing computer-adaptive reading curriculum to pre-kindergarteners (Hobbs & Overby, 2019). This Randomized Control Study (RCT) study sampled 491 preschoolers from 13 of the most rural school districts in Utah during the 2014-2015 school year. Students were randomly assigned to receive either the UPSTART Reading program (the treatment group) or the UPSTART Math/Science program (the control group). The treatment group significantly outperformed the control group on six of the eight subtests of the Brigance and the initial word sounds subtest of the PELI (identifying uppercase letters, reciting the alphabet, phonological awareness, phoneme manipulation, word recognition, and reading words from common signs). The UPSTART Reading program improved foundational literacy skills in treatment students, with meaningful effect sizes for phonological awareness ( $d = 0.30$  to  $0.32$ ), letter knowledge ( $d = 0.21$  to  $0.51$ ), and decoding ( $d = 0.22$  to  $0.49$ ).

*Table 1: Brigance Subscale Posttest Means of Treatment-Control Groups*

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>
Visual Discrimination	Reading	252	16.59	3.204	.96
	Math	239	16.33	2.915	
Auditory Discrimination	Reading	252	7.79	2.693	.37
	Math	239	7.70	2.458	
Recites Alphabet	Reading	252	19.39	9.074	2.86**
	Math	239	16.91	10.132	
Identifies Uppercase Letters	Reading	252	20.53	7.106	6.09**
	Math	239	16.19	8.634	

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>
Phonological Awareness	Reading	252	6.57	2.444	2.25*
	Math	239	6.06	2.564	
Phoneme Manipulation	Reading	252	4.07	1.529	3.58**
	Math	239	3.56	1.631	
Word Recognition	Reading	252	3.38	3.997	5.67**
	Math	239	1.50	3.288	
Reads Words from Signs	Reading	252	2.25	1.951	2.85*
	Math	239	1.75	1.969	

*\*p < .05, \*\*p < .01*

*Table 2: Brigance Subscale Posttest Effect Size Estimates*

<i>Brigance Posttest</i>	<i>Effect Size</i>	<i>Significance</i>
Visual Discrimination	0.05	NS
Auditory Discrimination	0.02	NS
Recites Alphabet	0.21	*
Identifies Uppercase Letters	0.51	**
Phonological Awareness	0.14	NS
Phoneme Manipulation	0.32	**
Word Recognition	0.49	**
Reads Words from Signs	0.22	**

*\*p < .05, \*\*p < .01*

*Table 3: PELI Subscale Posttest Means of Treatment-Control Groups*

<i>PELI Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>
Initial Word Sounds	Reading	252	10.53	4.673	3.20**
	Math	239	9.19	4.592	
Vocabulary/Oral Language	Reading	252	22.48	6.197	-.37
	Math	239	22.68	6.043	
Listening Comprehension	Reading	252	18.36	3.642	.70
	Math	239	18.14	3.429	

*\*p < .05, \*\*p < .01*

*Table 4: PELI Subscale Posttest Effect Size Estimates*

<i>PELI Posttest</i>	<i>Effect Size</i>	<i>Significance</i>
Initial Word Sounds	.30	**
Vocabulary/Oral Language	-.03	NS
Listening Comprehension	.02	NS

*\*p < .05, \*\*p < .01*

### 13 WACS Randomized Control Trial

In a recent randomized control trial, 523 four-year old pre-kindergarten students were randomly assigned to use only the Waterford Early Reading program (ERP) or the Waterford Early Math and Science program (EMS) at home in 13 rural school districts in Utah (Shamir, Miner, Izzo, Feehan, Yoder, & Pocklington, 2019). The Waterford Assessment of Core Skills (WACS) was administered at the beginning and end of the program to assess students’ literacy skills across multiple strands. At the end of the program year, students who used the computer-adaptive reading program significantly outperformed their control counterparts on Overall WACS scores; furthermore, students who used ERP outperformed their control counterparts on literacy strands of the assessment. The improvement was seen across demographics, including socioeconomic status, whether students attended another preschool, ethnicity, and active special education status. These findings indicate that computer-assisted instruction improves students’ early literacy skills after one year in the program and prepares them for kindergarten.

Table 5: WACS End of Year Scores Controlling for Beginning of Year Scores

	Experimental			Control			p	g
	M	SD	N	M	SD	N		
Overall	2597.90	373.85	273	2456.53	372.44	250	.00**	0.42
Blending	2669.03	722.27	270	2478.63	684.54	249	.00**	0.27
Initial Sound	2537.51	267.00	270	2499.37	248.54	249	.09	0.16
Letter Sound	2529.91	558.38	272	2254.86	531.94	249	.00**	0.52
Letter Recognition	2016.84	220.07	273	1948.41	227.57	249	.00**	0.32
Listening Comprehension	2799.08	1118.57	173	2593.70	1199.91	191	.08	0.18
Vocabulary	2784.81	656.44	270	2737.86	666.79	248	.42	0.06

\*p < .05, \*\*p < .01

Table 6: Overall WACS End of Year Scores by Demographics

	Experimental			Control			p	g
	M	SD	N	M	SD	N		
<b>185 % Poverty</b>								
Above	2624.09	365.58	118	2515.95	376.15	126	.01*	0.22
Below	2578.38	376.17	155	2395.63	355.42	124	.00**	0.40
<b>Gender</b>								
Female	2629.05	355.70	141	2442.70	396.28	115	.00**	0.39
Male	2564.91	385.60	132	2468.03	351.98	135	.02*	0.21
<b>Other Preschool</b>								
No Other Preschool	2576.03	401.85	119	2449.28	372.74	114	.00**	0.25
Another Preschool	2614.80	351.29	154	2462.60	372.86	136	.00**	0.34
<b>Ethnicity</b>								
Caucasian	2624.48	358.80	243	2483.10	365.82	221	.00**	0.41
Hispanic	2401.61	424.90	19	2304.63	338.36	19	.36	0.09
<b>Special Education Status</b>								
No Active Special Education Status	2602.98	375.34	256	2468.22	363.34	236	.00**	0.39
Active Special Education Status	2526.71	327.69	17	2252.96	386.87	14	.02*	0.20

\* $p < .05$ , \*\* $p < .01$

Table 7: Blending WACS End of Year Scores by Demographics

	Experimental			Control			p	g
	M	SD	N	M	SD	N		
<b>185 % Poverty</b>								
Above	2729.99	684.40	118	2560.91	721.30	125	.06	0.17
Below	2622.32	747.02	152	2394.96	634.86	124	.01**	0.24
<b>Gender</b>								
Female	2685.59	704.55	141	2462.56	677.15	114	.01*	0.22
Male	2651.05	742.76	129	2492.10	692.78	135	.06	0.17
<b>Other Preschool</b>								
No Other Preschool	2670.88	715.62	118	2420.96	661.71	114	.01**	0.24
Another Preschool	2667.86	729.74	152	2527.04	699.83	135	.09	0.15
<b>Ethnicity</b>								
Caucasian	2694.42	702.92	240	2513.20	688.76	220	.01**	0.25
Hispanic	2473.85	944.77	19	2181.40	590.41	19	.19	0.11
<b>Special Education Status</b>								
No Active Special Education Status	2681.43	729.88	256	2504.63	689.90	236	.01**	0.25
Active Special Education Status	2439.37	525.27	14	2009.96	391.85	13	.11	0.14

\* $p < .05$ , \*\* $p < .01$

Table 8: Letter Sound WACS End of Year Scores by Demographics

	Experimental			Control			p	g
	M	SD	N	M	SD	N		
<b>185 % Poverty</b>								
Above	2575.46	543.65	118	2340.41	506.60	125	.00**	0.31
Below	2495.01	566.58	154	2168.61	542.42	124	.00**	0.45
<b>Gender</b>								
Female	2557.28	543.79	141	2273.01	551.91	114	.00**	0.37

	Experimental			Control			<i>p</i>	<i>g</i>
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>		
Male	2500.45	573.27	131	2239.53	515.85	135	.00**	0.35
Other Preschool								
No Other Preschool	2490.70	571.91	119	2241.77	529.11	114	.00**	0.31
Another Preschool	2560.41	547.51	153	2265.90	535.99	135	.00**	0.41
Ethnicity								
Caucasian	2560.42	551.22	242	2275.63	537.37	220	.00**	0.51
Hispanic	2282.17	585.34	19	2219.49	512.13	19	.71	0.03
Special Education Status								
No Active Special Education Status	2527.30	565.56	256	2262.73	530.32	236	.00**	0.49
Active Special Education Status	2571.75	437.85	16	2111.95	544.76	13	.02*	0.20

\**p* < .05, \*\**p* < .01

Table 9: Letter Recognition WACS End of Year Scores by Demographics

	Experimental			Control			<i>p</i>	<i>g</i>
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>		
185 % Poverty								
Above	2026.35	183.12	118	1982.51	205.31	125	.10	0.14
Below	2009.77	244.46	155	1913.83	242.32	124	.00**	0.33
Gender								
Female	2041.25	193.44	141	1963.75	201.38	114	.00**	0.25
Male	1990.92	242.82	132	1935.31	246.59	135	.03*	0.19
Other Preschool								
No Other Preschool	2028.93	250.15	119	1939.28	240.86	114	.00**	0.28
Another Preschool	2007.45	194.47	154	1956.18	216.34	135	.04*	0.18
Ethnicity								
Caucasian	2021.44	209.73	243	1955.44	224.57	220	.00**	0.30
Hispanic	1964.15	331.11	19	1915.54	252.05	19	.48	0.06
Special Education Status								
No Active Special Education Status	2016.96	223.44	256	1957.05	224.55	236	.00**	0.28
Active Special Education Status	2016.87	166.22	17	1789.08	201.75	13	.00**	0.25

\**p* < .05, \*\**p* < .01

## Randomized Controlled Trial Evaluation of Waterford Early Learning in Indiana

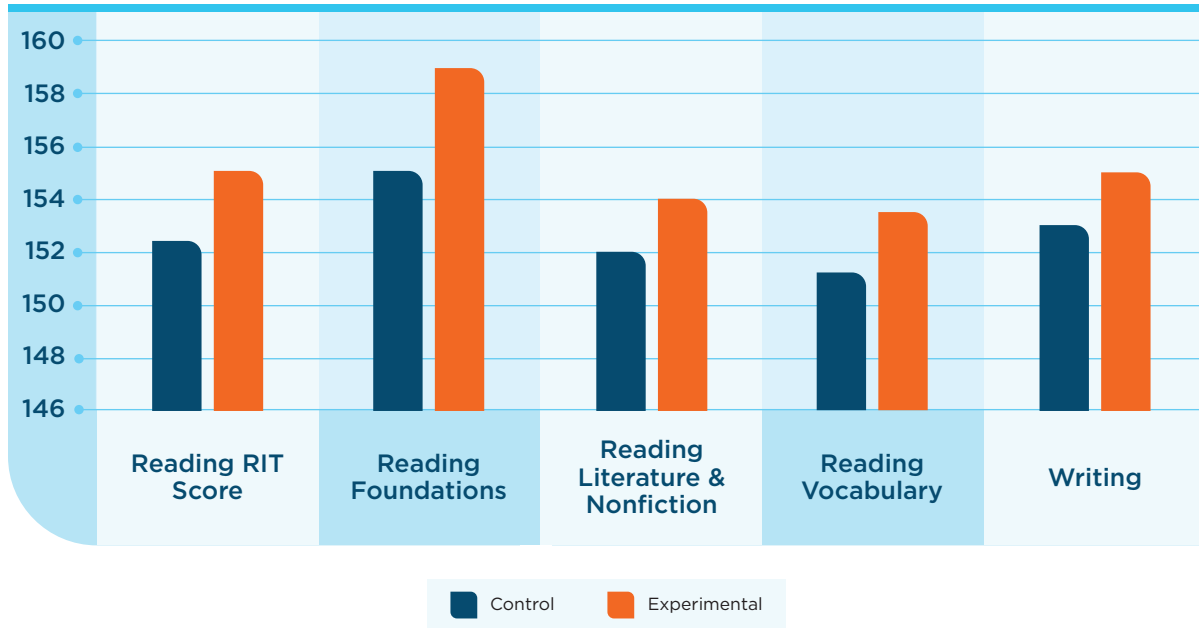
The following randomized controlled trial (Shamir, Yoder, Feehan, & Pocklington, 2019) assessed the efficacy of Waterford Early Learning (WEL). Eleven kindergarten classes (*n* = 273) were randomly assigned to the experimental condition: Students in these classes were expected to use WEL during the 2017-2018 school year for 15 minutes per day, five days per week. Eleven kindergarten classes (*n* = 263) were assigned to the control condition: Students received traditional literacy instruction for the same amount of time that the experimental group received WEL. Thus, overall exposure to literacy instruction was the same for both groups. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) and the Text Reading Comprehension (TRC) literacy assessments were administered at the end of the year. Independent samples *t*-tests showed that experimental students consistently

outperformed their control group counterparts on all strands, and ANOVAs showed that experimental students across all ethnicities and types of lunch status analyzed outperformed their control group counterparts.

## MAP

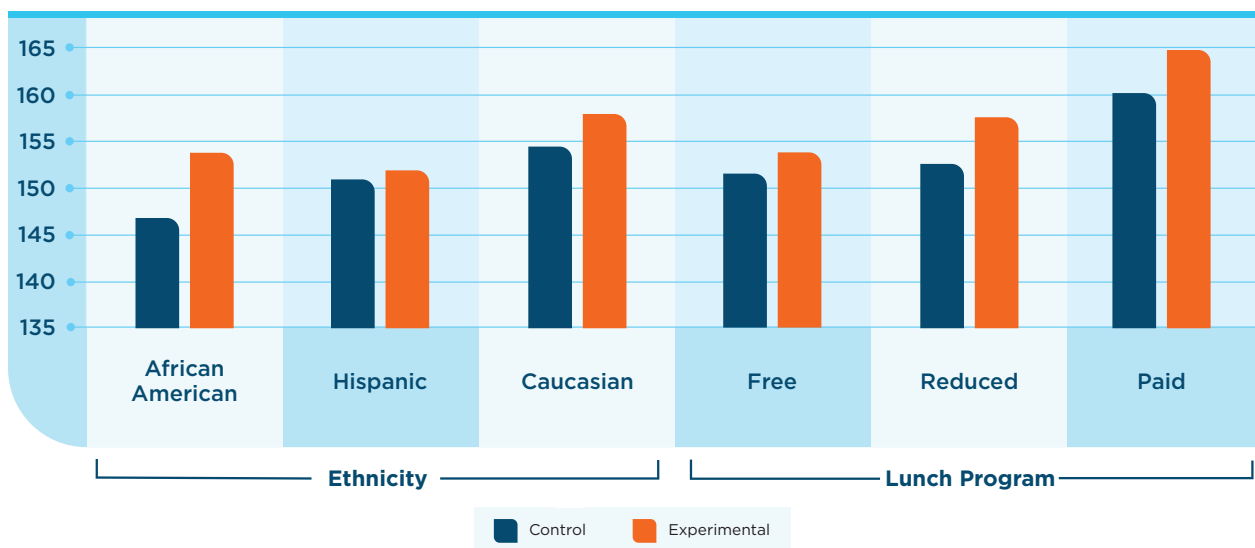
### GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS

Figure 1: MAP End of Year Scores by Strand



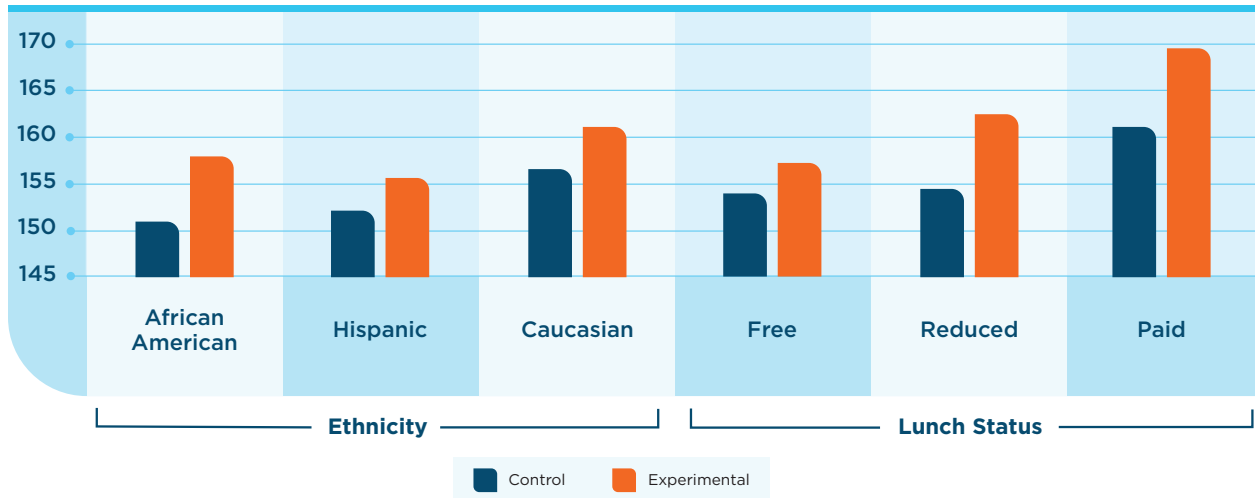
### GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS – RIT SCORE

Figure 2: MAP RIT Scores by Demographics



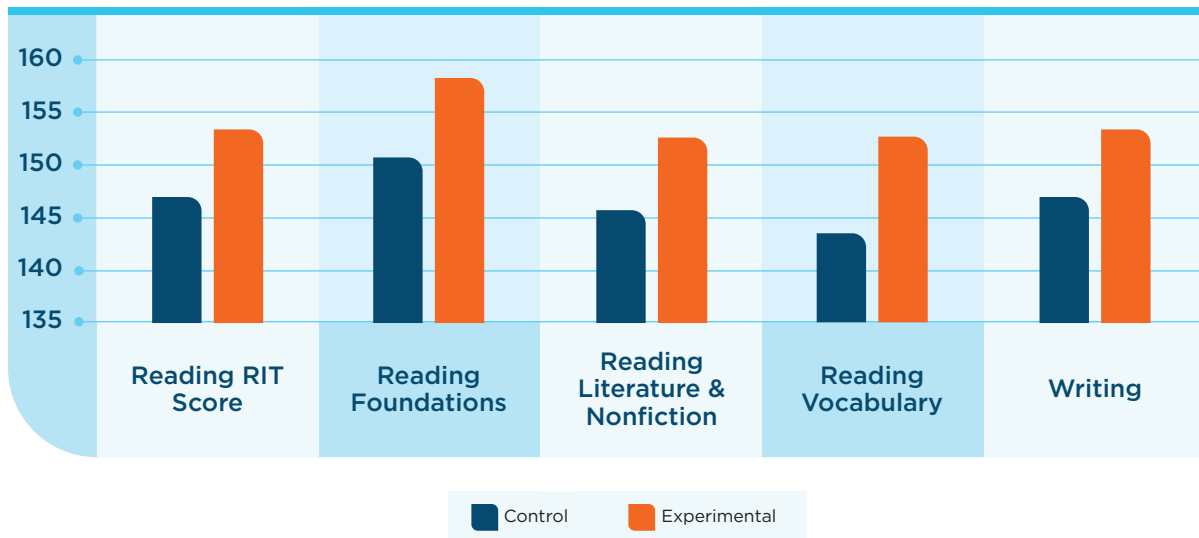
### GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS – READING FOUNDATIONS

Figure 3: MAP Reading Foundations Scores by Demographics



### GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS: AFRICAN AMERICAN ANALYSIS

Figure 4: MAP End of Year Scores - African American Students

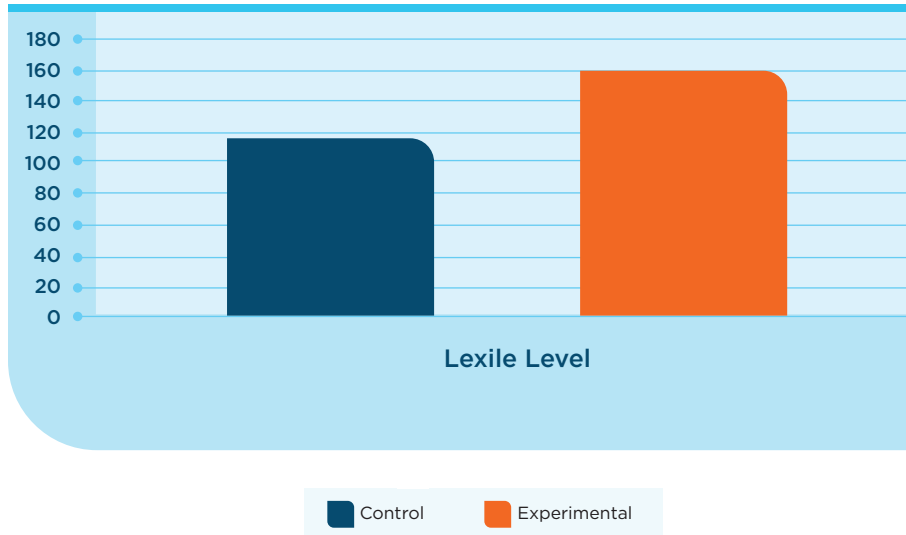




## TRC

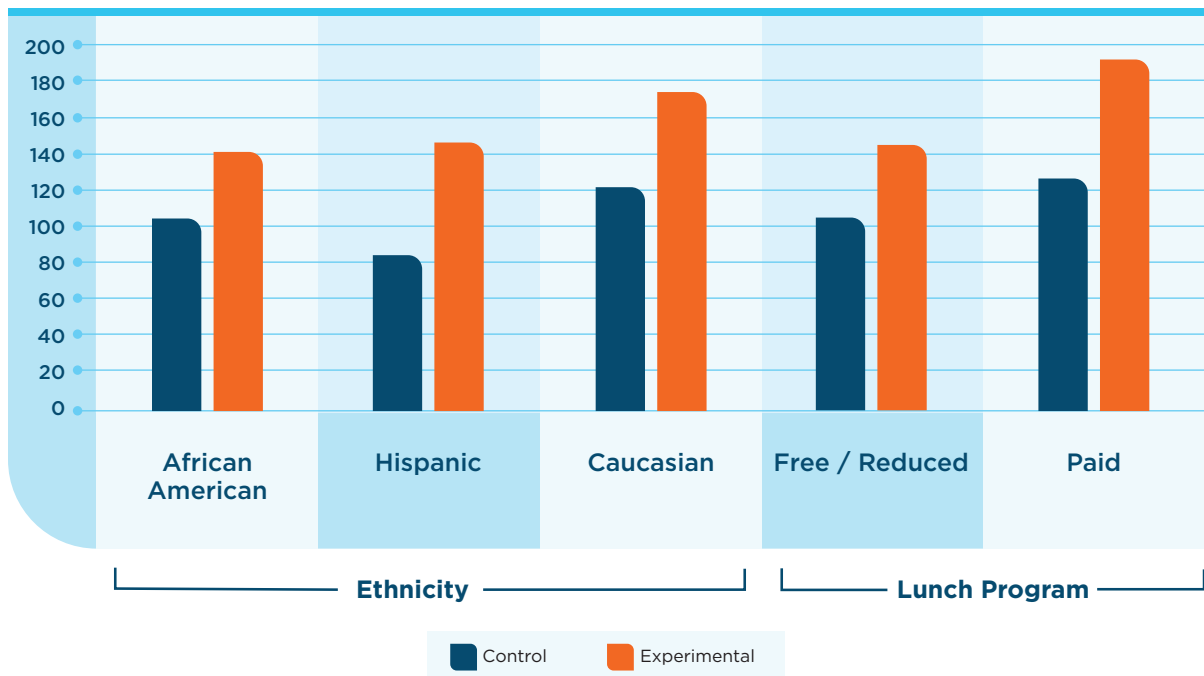
### GROUP DIFFERENCES USING AN INDEPENDENT SAMPLES T-TEST

Figure 5: TRC End of Year Lexile Levels



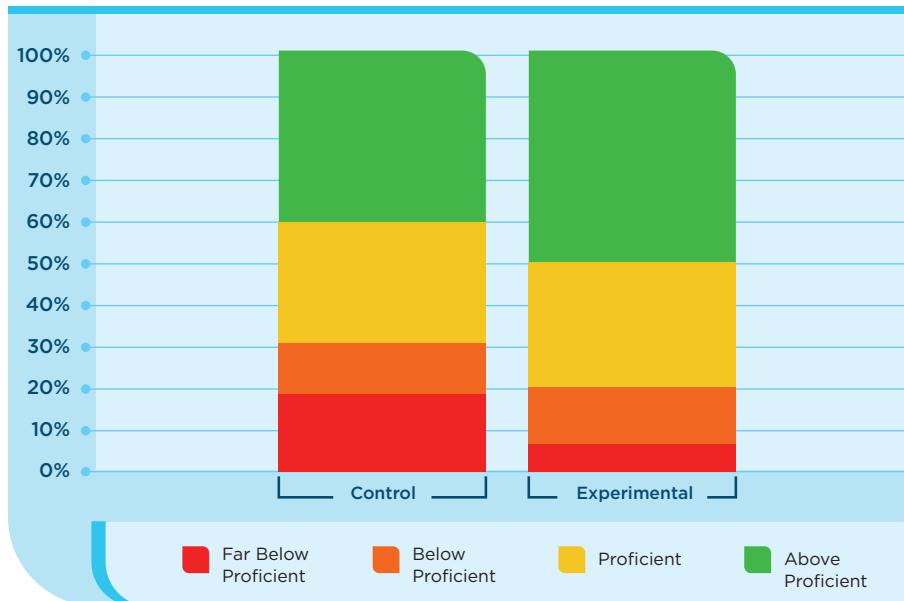
### GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS

Figure 6: TRC End of Year Lexile Levels by Demographics



## GROUP DIFFERENCES USING CHI-SQUARE

Figure 7: TRC End of Year Proficiency Benchmarks



## Tier II: Moderate Evidence

For a study to be considered moderate evidence, it needs to have a quasi-experimental design and be well-designed and well-implemented by meeting the WWC Evidence Standards with reservations. Additionally, the intervention in the study must demonstrate a statistically significant and positive effect on a student outcome, without having statistically significant and negative outcomes in other studies that meet the WWC Evidence Standards with or without reservations. The sample analyzed must be large, include multiple sites, and be representative of the population of interest.

### Evaluation of the Waterford Early Reading Program in Ohio

Hecht and Close (2002) investigated the use of the Waterford software by a group of disadvantaged kindergarteners, comparing their pre-literacy gains over the course of the year to a class that had not used the program. Assessments included a number of well-known standardized measures, among them the Wide Range Achievement Test, the Stanford-Binet, the Stones—Concepts About Print Test, the Woodcock-Johnson Tests of Achievement (Form B), and the Comprehensive Test of Phonological Processing. Analyses showed that the Waterford group significantly outperformed the comparison group in Phonemic Awareness tests (with effect sizes of 1.14 and 1.13 for the skills of Phonemic Segmenting and Blending, respectively) as well as Invented Spelling (effect size = 1.20) and word reading (effect size = 1.11). The researchers also noted that the amount of time children spent with the software was an important factor for its success: More time spent with the software uniquely contributed to performance in Phonemic Awareness, Invented Spelling, Letter Knowledge, and Print Concepts.

Table 10: Descriptive Statistics for All Tasks

Pretest Predictors	Training (n = 42)			Control (n = 34)		
	M	SD	Reliability	M	SD	Reliability
<b>Segmenting</b>						
Pre	1.76	3.31	.91	1.44	2.43	.88
Post	7.91	7.05	.95	1.53	2.84	.91
Gain	6.14***	6.09	—	.09	1.50	—
<b>Blending</b>						
Pre	4.98	5.01	.95	3.65	3.27	.85
Post	10.29	5.55	.93	4.24	5.08	.94
Gain	5.31***	5.94	—	.59	3.49	—
<b>Letter Name Knowledge</b>						
Pre	19.45	8.52	.97	16.82	9.82	.95
Post	24.21	4.43	.98	24.65	4.14	.93
Gain	4.76***	7.16	—	7.83***	8.07	—
<b>Letter Sound Knowledge</b>						
Pre	19.51	11.04	.96	13.79	11.45	.97
Post	24.81	8.87	.95	22.55	9.33	.92
Gain	5.30***	10.92	—	8.76***	9.06	—
<b>Letter Writing Knowledge</b>						
Pre	6.95	4.99	.94	5.41	4.41	.96
Post	12.00	2.35	.94	11.21	3.49	.98
Gain	5.05***	4.54	—	5.79***	4.46	—
<b>Word Reading</b>						
Pre	.40	.80	.55	.12	.41	.84
Post	3.72	3.43	.91	.77	1.16	.72
Gain	3.32***	3.20	—	.65**	1.27	—
<b>Invented Spelling</b>						
Pre	3.55	4.28	.96 <sup>a</sup>	2.41	3.84	.98 <sup>a</sup>
Post	26.71	19.67	.91 <sup>a</sup>	8.09	7.79	.94 <sup>a</sup>
Gain	23.16***	17.75	—	5.68***	5.77	—
<b>Vocabulary Knowledge</b>						
Pre	16.50	3.42	.88	15.60	6.07	.82
Post	17.41	3.66	.79	16.58	3.35	.95
Gain	.91	3.45	—	.58	2.34	—
<b>Print Concepts</b>						
Pre	7.83	2.92	.74	6.53	3.30	.79
Post	9.88	3.05	.84	9.01	4.57	.91
Gain	2.05***	2.79	—	2.48***	4.34	—
<b>Total time spent using the WERP-1</b>						
(hh:mm:ss)	21:25:49			7:17:38		

<sup>a</sup>Proportion of agreement between two independent raters.

\*\* $p < .01$

\*\*\* $p < .001$

The What Works Clearinghouse reviewed the Ohio study (Hecht & Close, 2002) that included kindergarten students from six schools in Ohio and found evidence (with reservations) supporting the reading software's value for Alphabets and Comprehension. In evaluating its effectiveness, the Waterford Early Reading Program was found to have potentially positive effects on Alphabets—+19 percentile points—while the Comprehension improvement index was +4 percentile points (What Works Clearinghouse, 2007).

## Effectiveness of UPSTART

The Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) Program has provided an entirely new environment for measuring the effectiveness of Waterford software. Studying preschool children in the home is important because, although the software is designed to accommodate children of this age, all previous studies have been conducted with kindergarten or early primary students. Similarly, prior to UPSTART, the use of Waterford software had been confined to the school classroom.

UPSTART began in 2009 and was implemented to some degree in every public school district throughout Utah, including the state's most rural districts. Substantial efforts were made to reach low-income and minority students, and in the first year of the program 61% of UPSTART participants came from low-income homes, and 20% classified their ethnicities as Hispanic or non-Caucasian. The program provides state funding for the installation of computers and Internet access in homes that do not have them.

Developed by Waterford, UPSTART prepares pre-kindergarten students for academic success. The in-home program provides a research-based individualized math, reading, and science curriculum. The UPSTART program is in its eighth year of implementation: To date, pre-kindergarten students across demographics who have met the usage requirements of the UPSTART program for the school year score at the Kindergarten Advanced level on the Waterford Assessment of Core Skills™ (WACS) adaptive reading assessment. These findings indicate that, regardless of ethnicity, socio-economic status, and location in the United States, the children that participate in the UPSTART program begin kindergarten with abilities comparable to kindergarten students in the last three months of kindergarten nationwide. Results show how the UPSTART program has significantly closed the gap among pre-kindergarten students, across demographics, entering kindergarten.

## YEAR 6

Waterford Institute served 5,091 preschool children in its sixth year of operation during the 2014-15 school year (Evaluation and Training Institute, 2016). Combined posttest results showed that UPSTART participation using Waterford Early Learning had a large impact on students' early literacy skills. Large effect sizes (Bader = 0.95; Brigance = 0.81) were shown favoring UPSTART students as measured by the total Bader and Brigance composite scores. Favoring the UPSTART treatment group, differences in growth rates between the UPSTART treatment and control group were significantly different for the overall Brigance and for five of the Brigance subtests and for the Total Bader and all three Bader subtests.

Children participating in UPSTART demonstrated improvement in word decoding and phonological awareness skills. Medium effect sizes were observed for Survival Sight Words (0.45) and Rhyme Recognition (0.44). Large effect sizes were found for Pre-Primer Vocabulary (1.10), Phonemic Blending (0.99), and Phoneme Segmenting (0.85). Children participating in UPSTART also demonstrated greater

gains on Pre-Primer Vocabulary, Survival Sight Words, and all Phonological Awareness subscales than control students. Participation in UPSTART was associated with improvement in all phonological awareness strands of the Bader, including Rhyme Recognition, Phoneme Blending, and Phoneme Segmenting. Children participating in UPSTART had a 36-point advantage on Brigance posttest scores compared to non-participating children.

## BRIGANCE AND BADER RESULTS

Figure 8: Brigance & Bader Posttest Analysis of Composite Scores

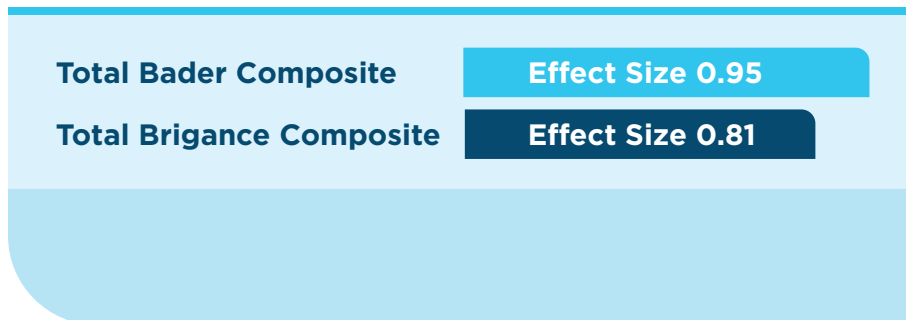
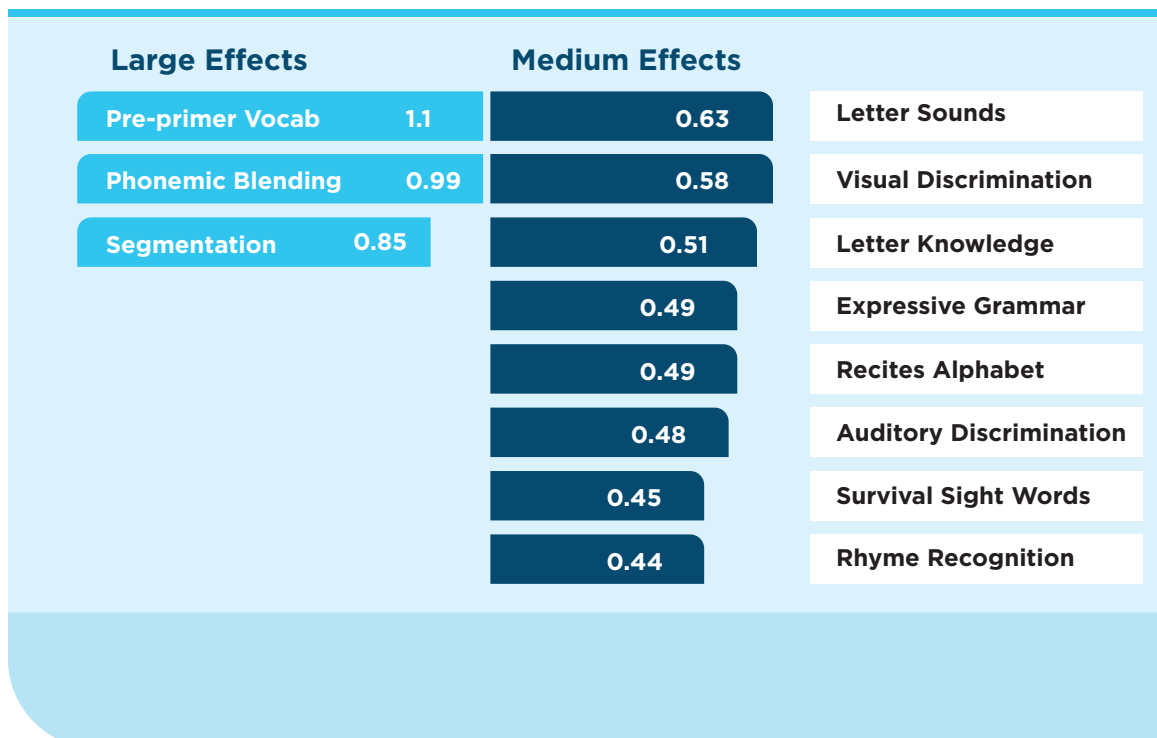
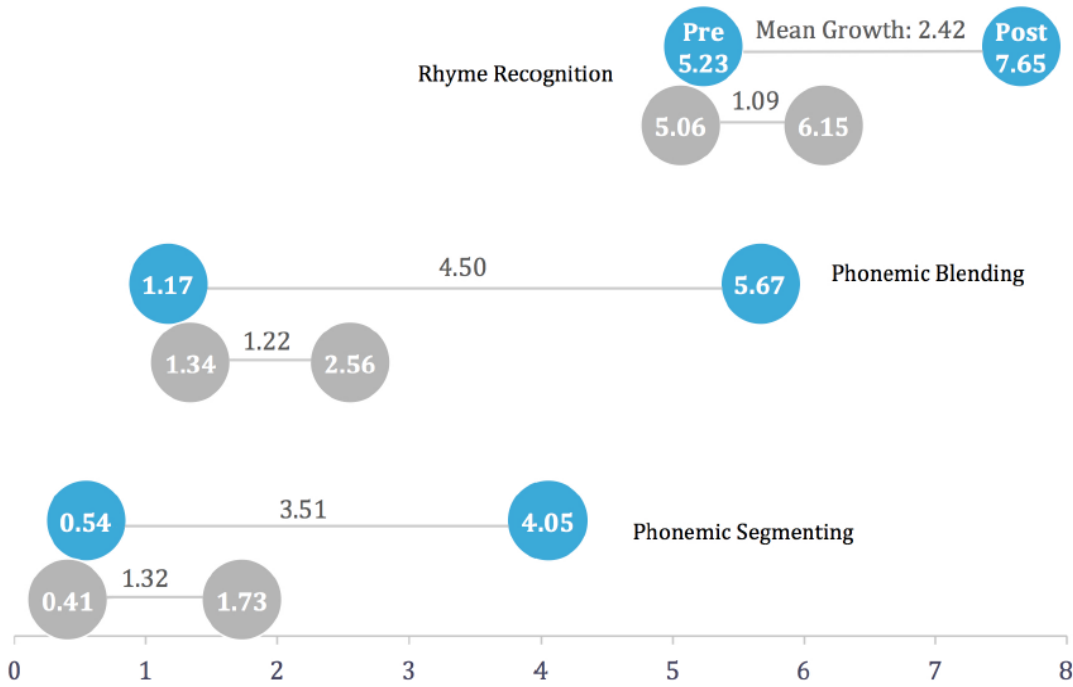


Figure 9: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



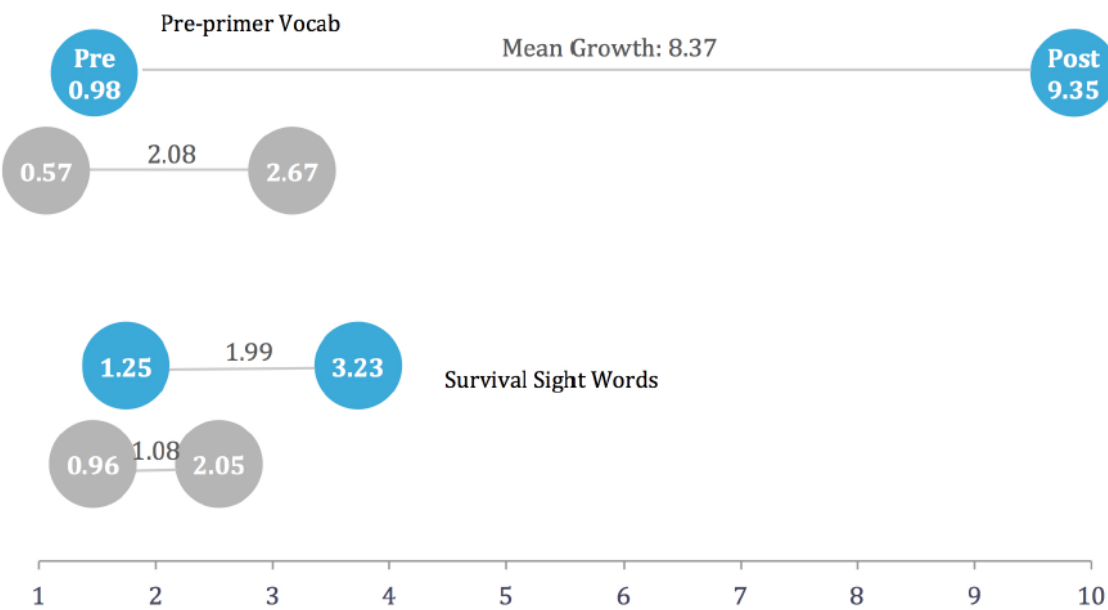
## BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 10: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 11: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 12: Pre-Literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores

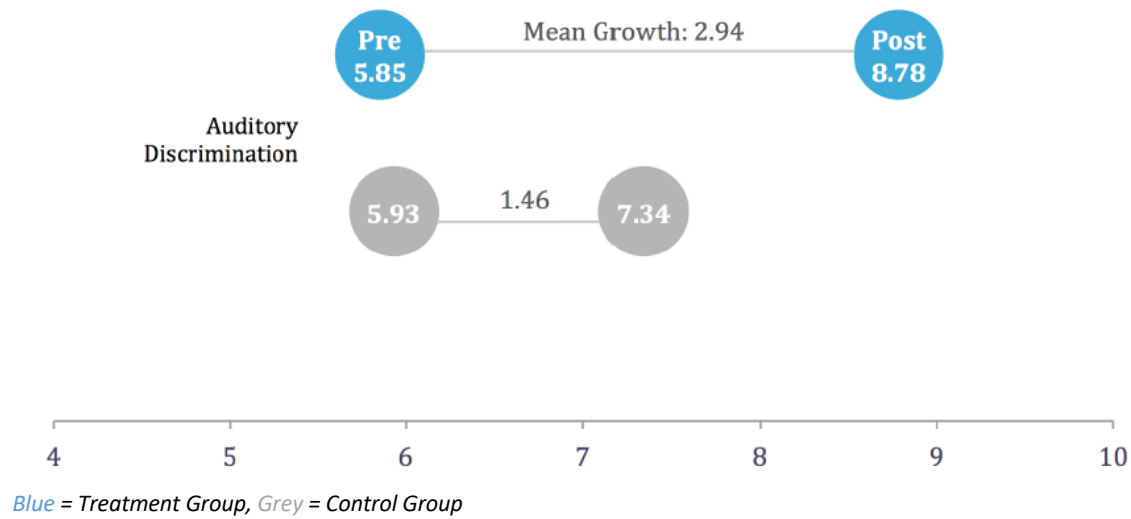


Figure 13: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

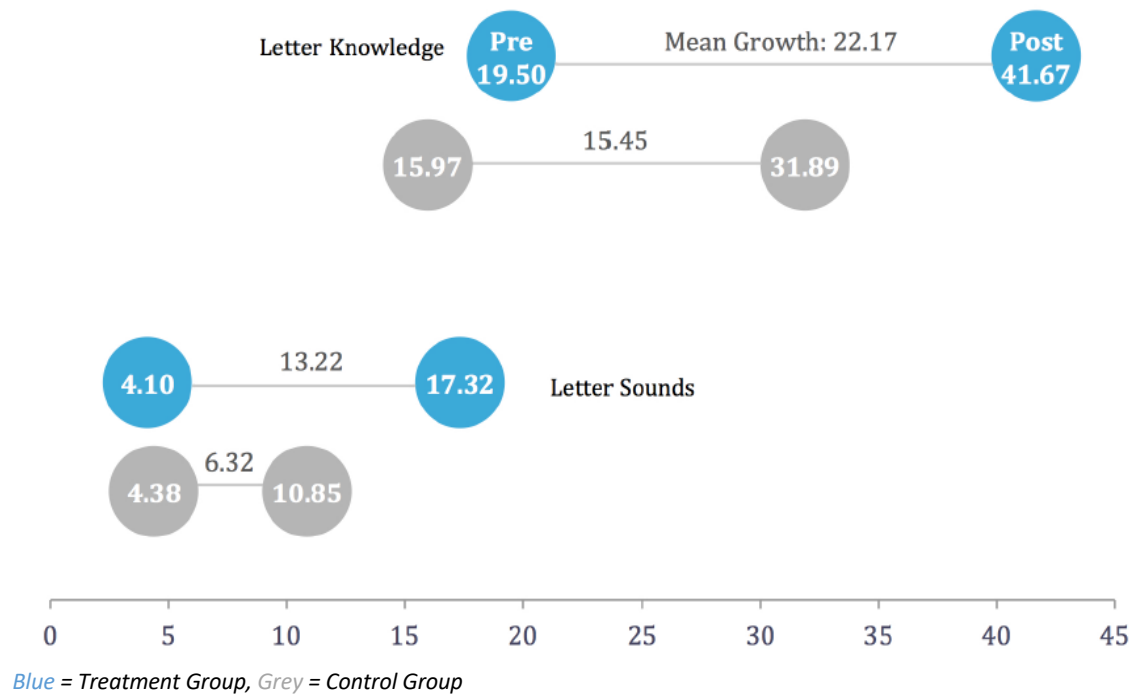


Table 11: Brigance Growth Rate Comparisons

<i>Brigance Test</i>	<u>Control Group</u> ( <i>N = 138</i> ) Mean Growth	<u>Treatment Group</u> ( <i>N = 138</i> ) Mean Growth	<i>T-C Significance</i> $p \leq .05$
Expressive Vocab	0.414	0.906	NS
Receptive Vocab	0.075	0.080	NS
Expressive Grammar	0.474	0.406	NS
Visual Discrimination	2.857	3.565	NS
Recites Alphabet	5.691	7.717	NS
Letter Knowledge	15.451	22.174	*
Letter Sounds	6.316	13.217	*
Auditory Discrimination	1.457	2.935	*
Survival Sight Words	1.075	1.986	*
Basic Vocabulary	2.083	8.370	*
Total Brigance	39.429	76.159	*

\*  $p \leq .05$

Table 12: Bader Growth Rate Comparisons

<i>Bader Test</i>	<u>Control Group</u> ( <i>N = 138</i> ) Mean Growth	<u>Treatment Group</u> ( <i>N = 138</i> ) Mean Growth	<i>T-C Significance</i> $p \leq .05$
Rhyme Recognition	1.0902	2.4203	**
Phoneme Blending	1.2180	4.5000	**
Phoneme Segmenting	1.3233	3.5072	**
Total Bader	3.6316	10.4275	**

\*\*  $p \leq .05$

## YEAR 7

Waterford Institute enrolled 6,639 preschool children in its seventh year of operation during the 2015-16 school year (Evaluation and Training Institute, 2017). Combined posttest results showed that UPSTART participation had a medium impact on students' early literacy skill development: Children enrolled in UPSTART produced significant positive effects ( $ES = 0.52$ ) compared to control children on the Brigance composite. Similarly, UPSTART participants experienced significant positive effects ( $ES = 0.62$ ) on the Bader composite.

Children participating in UPSTART demonstrated significant improvement in word decoding and phonological awareness skills. Medium effect sizes were observed for Pre-Primer Vocabulary (0.74), Phoneme Segmenting (0.64), and Phoneme Blending (0.63). Children participating in UPSTART also demonstrated greater gains on Pre-Primer Vocabulary, Survival Sight Words, and all Phonological Awareness subtests than control students. Participation in UPSTART was associated with improvement in all phonological awareness strands of the Bader, including Rhyme Recognition, Phoneme Blending, and Phoneme Segmenting. Children participating in UPSTART had a 21-point advantage on Brigance posttest scores compared to non-participating children.



## BRIGANCE AND BADER RESULTS

Figure 14: Brigance & Bader Posttest Analysis of Composite Scores

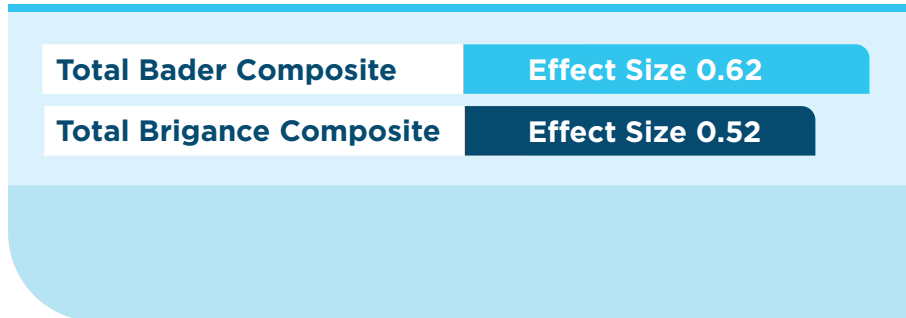
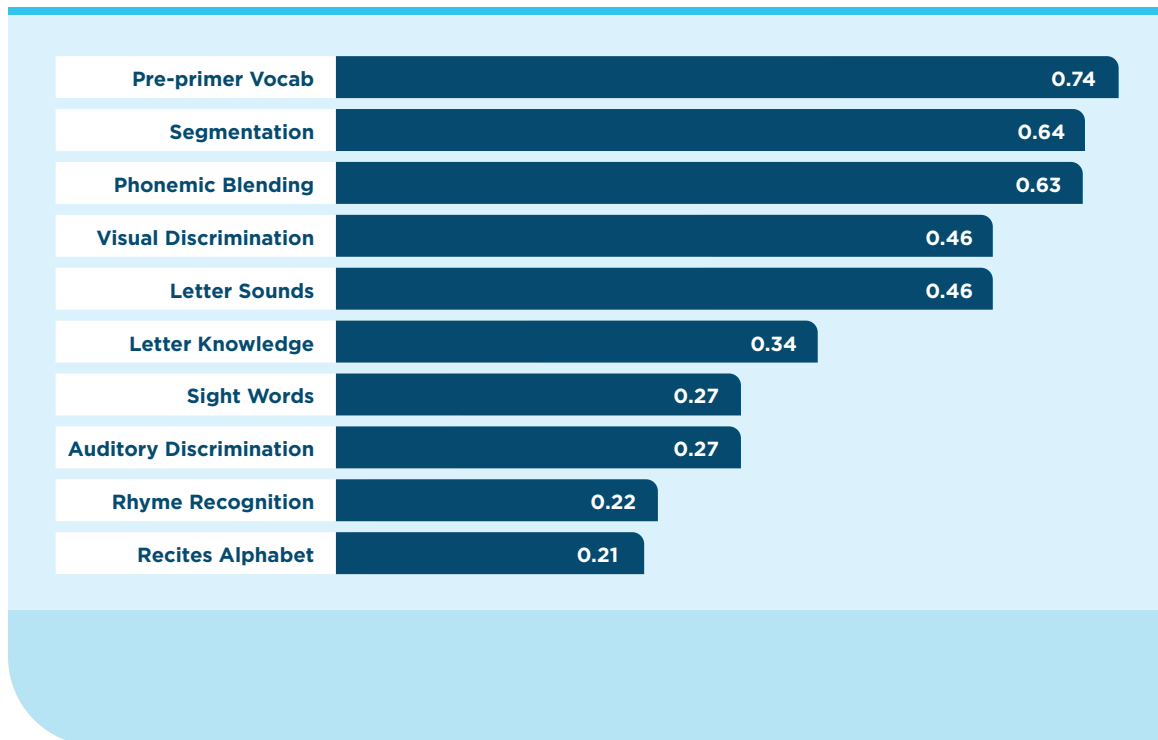
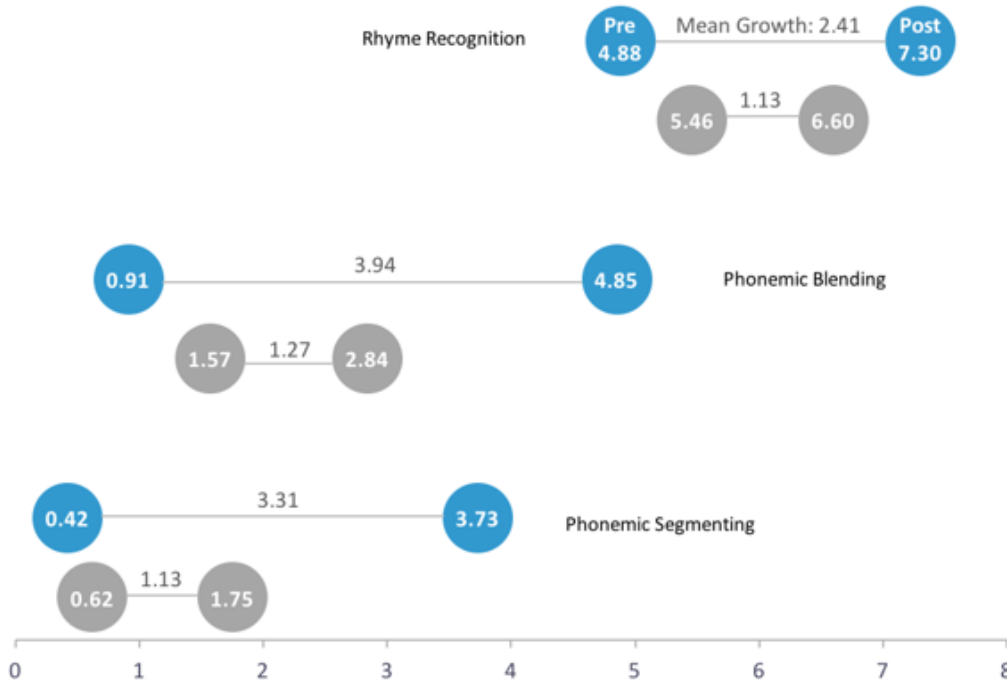


Figure 15: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



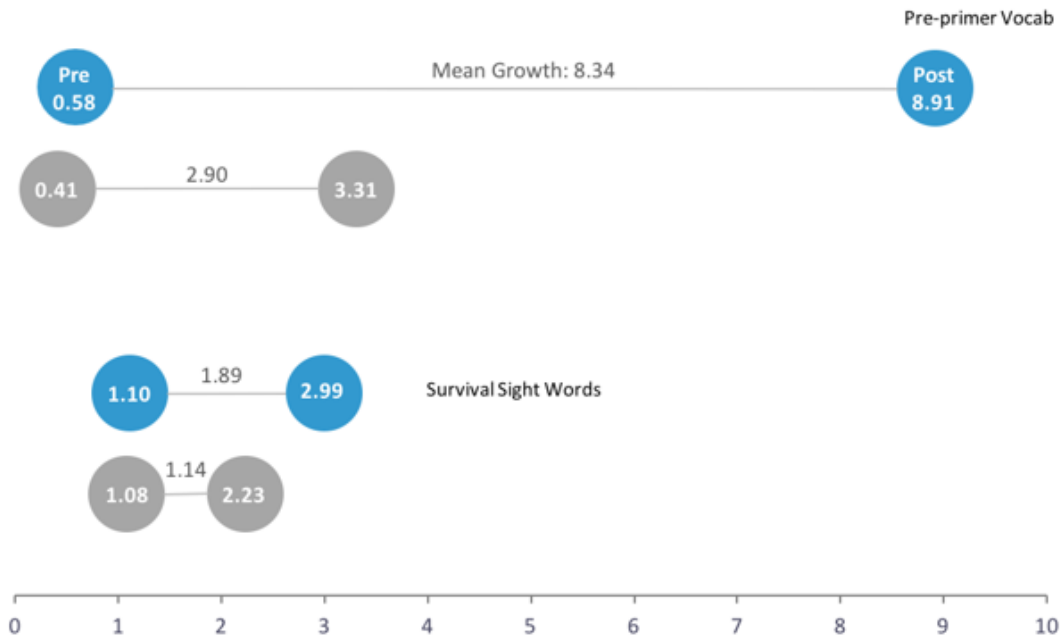
## BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 16: Phonological Awareness: Treatment and Control Group Pretest & Posttest Mean Scores



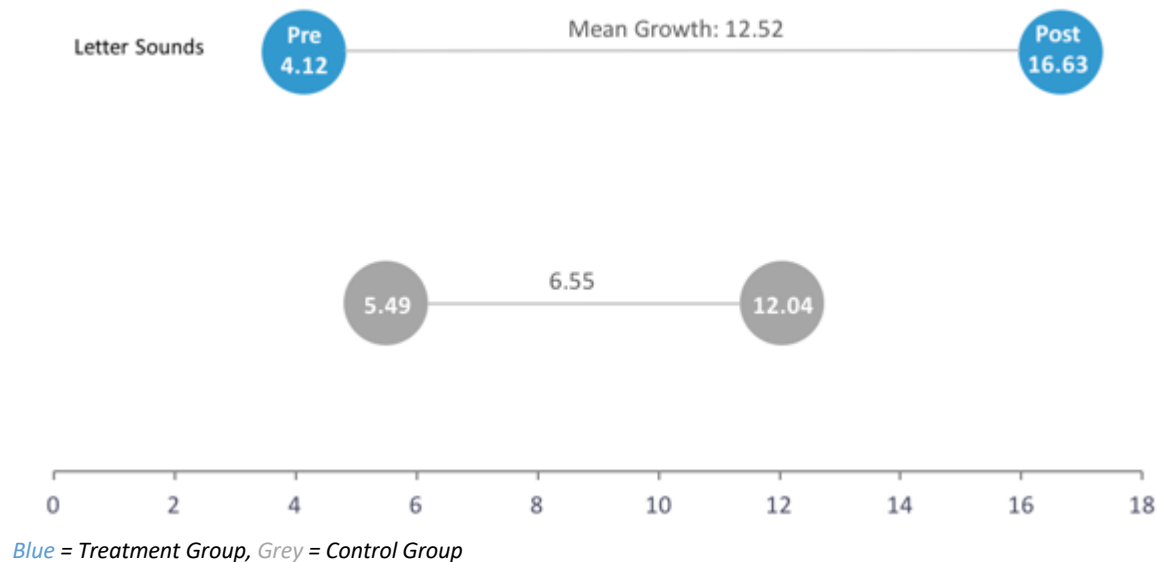
Blue = Treatment Group, Grey = Control Group

Figure 17: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 18: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores



## YEAR 8

Waterford Institute enrolled 10,745 preschool children in its eighth year of operation during the 2016-17 school year (Evaluation and Training Institute, 2018a). Combined posttest results showed that UPSTART participation had a medium impact on students' early literacy skill development. In the matched posttest sample, UPSTART produced strong to medium effects: Children enrolled in UPSTART produced significant positive effects (ES = 0.50) compared to control children on the Brigance composite. Similarly, UPSTART participants experienced significant positive effects (ES = 0.81) on the Bader composite.

Children participating in UPSTART demonstrated significant improvement in word decoding and phonological awareness skills. Medium effect sizes were observed for Phonemic Blending (0.78), Phoneme Segmenting (0.64), and Pre-Primer Vocabulary (0.60). Children participating in UPSTART also demonstrated greater gains on both Phonemic Blending and Phoneme Segmenting subscales than control students.

Participation in UPSTART was associated with significant improvement on both of the phonological awareness strands of the Bader assessed, Phoneme Blending and Phoneme Segmenting. Children participating in UPSTART had a significantly higher growth rate, with a 21-point advantage on Brigance posttest scores compared to non-participating children. Additionally, children who participated in UPSTART during the previous year significantly outperformed their control counterparts on the DIBELS composite at the beginning of first grade. The average beginning of year DIBELS composite score was 7.91 points higher for students who were enrolled in UPSTART compared to students who were not (ES = 0.18).

## BRIGANCE AND BADER RESULTS

Figure 19: Brigance & Bader Posttest Analysis of Composite Scores

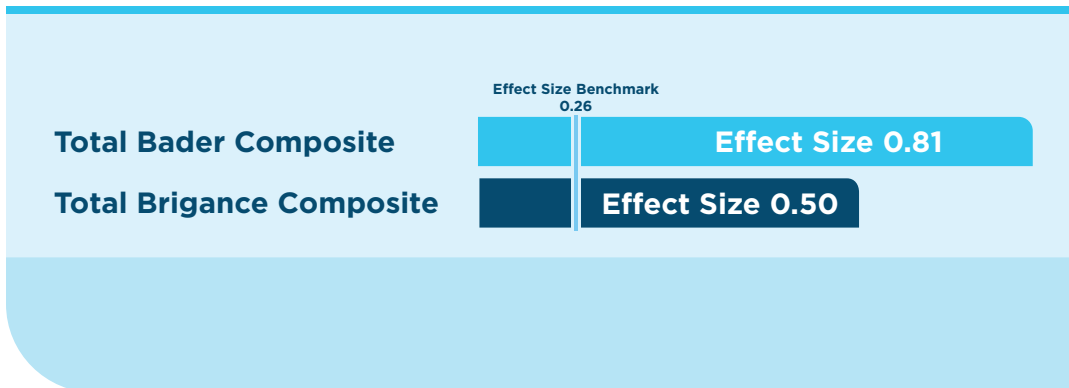
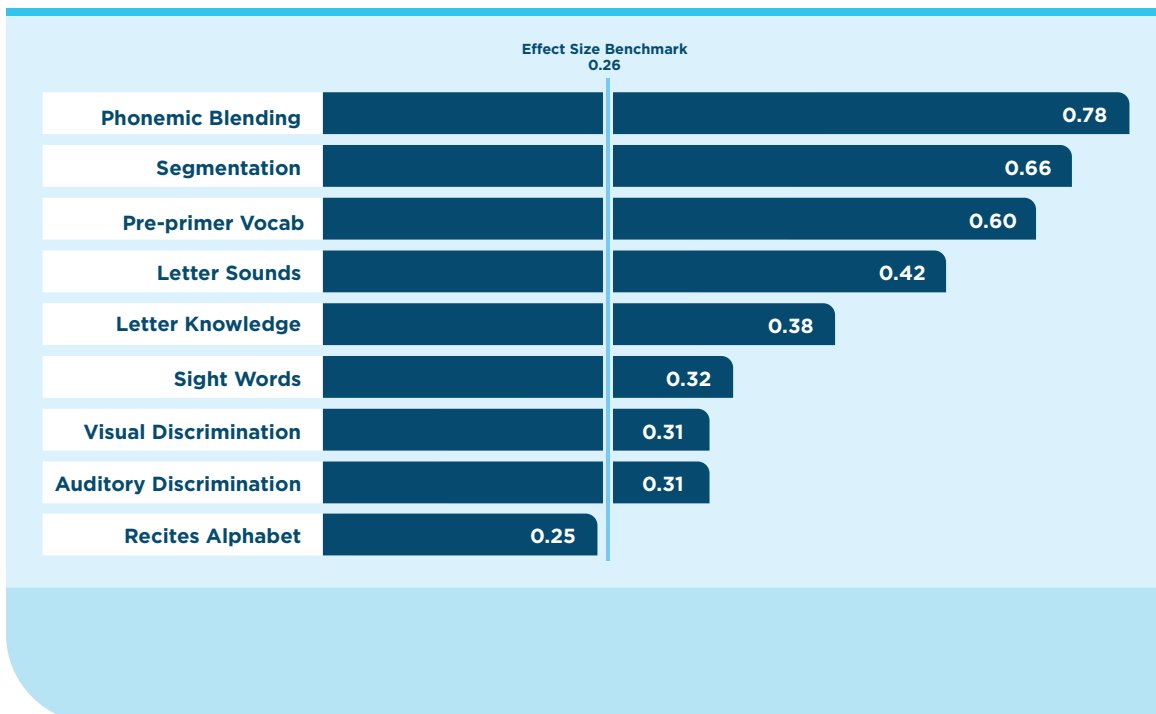
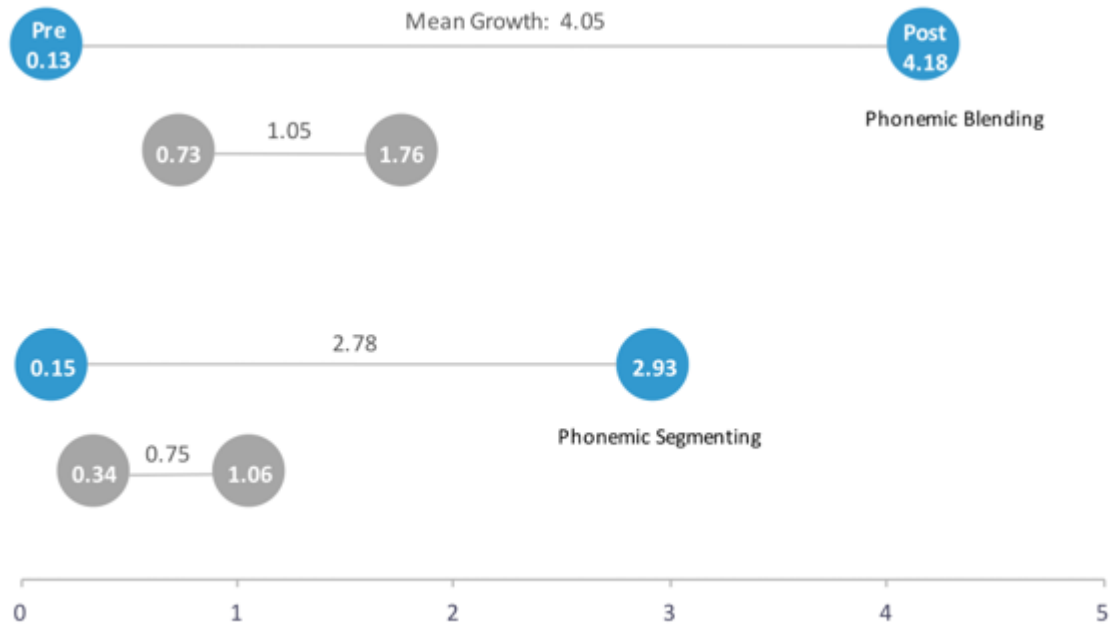


Figure 20: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



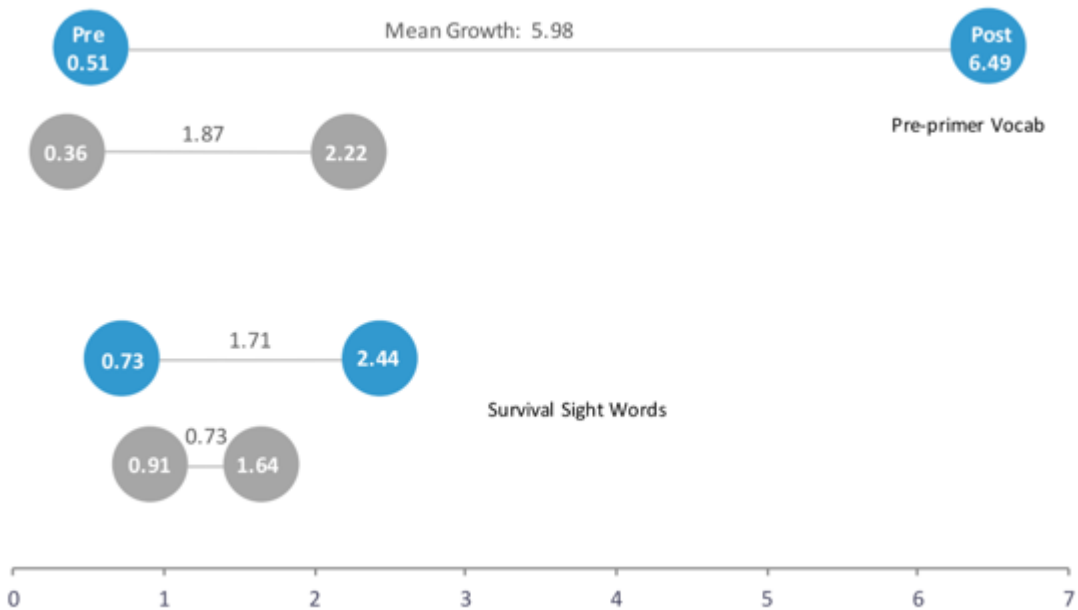
BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 21: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 22: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 23: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

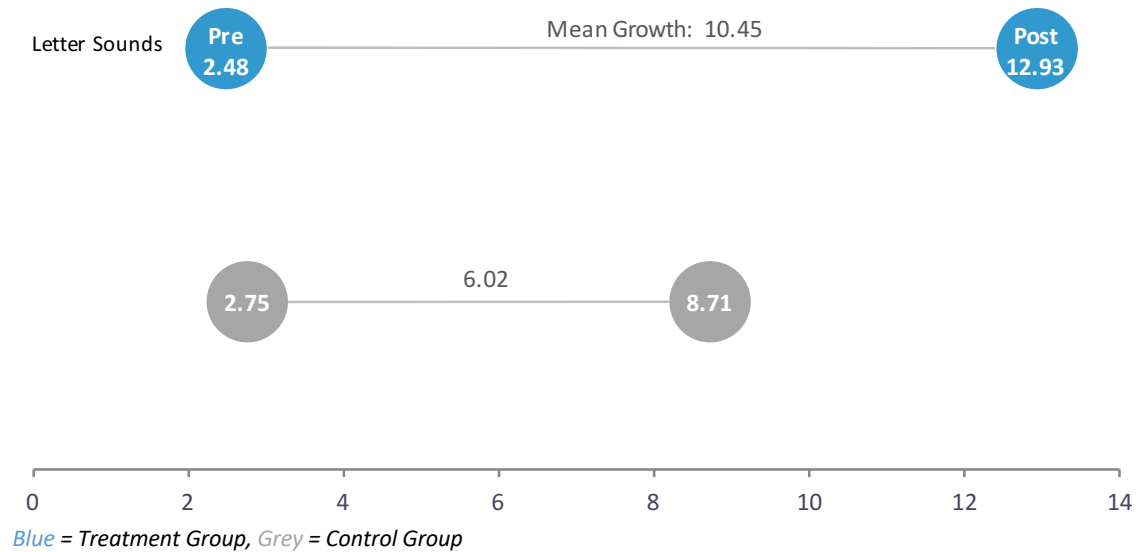
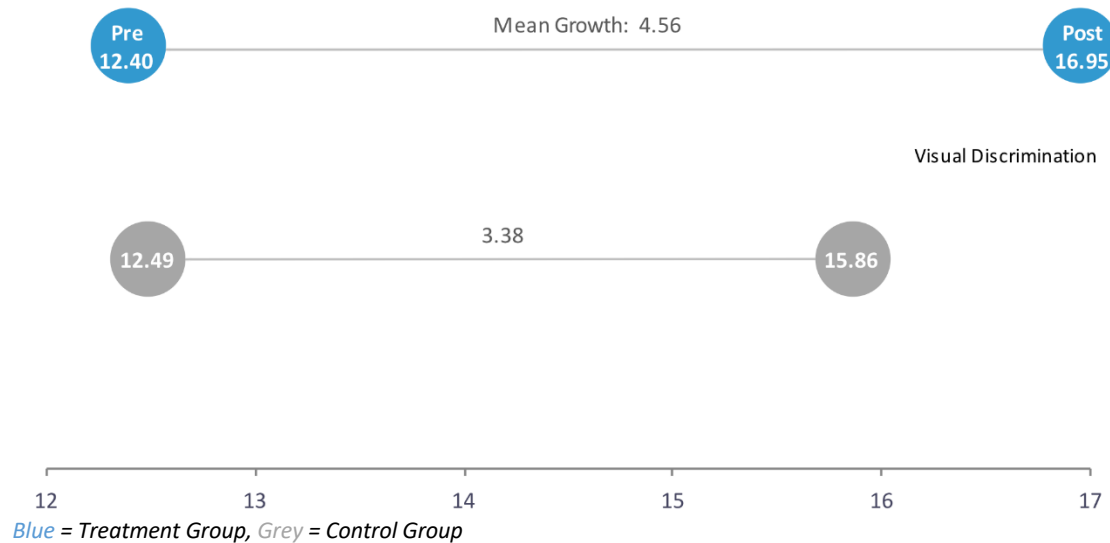


Figure 24: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores



## Evaluation of the UPSTART Elementary Program

This study details the UPSTART summer program provided during the summer following Kindergarten (Hobbs, Overby, & Thomas, 2017). The program intends to prevent the “summer slump” in literacy. The experimental group consisted of students that completed Kindergarten and enrolled in the UPSTART summer program, and the control group consisted of students that completed Kindergarten that chose not to enroll in the UPSTART summer program. Using Coarsened Exact Matching, experimental and control students were matched on gender, ethnicity, poverty status, and baseline literacy skills.

DIBELS Next, a standardized test measuring literacy skills, was used to find the difference between literacy scores at the beginning of first grade with literacy scores at the end of kindergarten. Independent *t*-tests on the DIBELS first grade beginning of year scores indicated significant differences between the experimental group and the control group on the DIBELS Composite score, Letter Naming Fluency, and Phoneme Segmentation. Experimental students scored between one and five points higher than control students on the DIBELS Composite score and all four of the tested strands (Letter Naming Fluency, Phoneme Segmentation, Nonsense Word Fluency: Correct Letter Sounds, and Nonsense Word Fluency: Whole Words Read). Small effect sizes were observed.

Additionally, experimental students lost between one and three points less after participating in the summer program compared to control students. The differences in learning loss were significant for the DIBELS Composite score and Nonsense Word Fluency: Correct Letter Sounds, but not for Letter Naming Fluency or Phoneme Segmentation. Small effect sizes were observed.

## High-Quality School Readiness Expansion in Utah, 2017-2018

This study assessed the efficacy of high-quality school readiness programs (Evaluation and Training Institute, 2018b). Students' literacy scores were analyzed across a control group of students who did not attend preschool identified as high quality by the state and three treatment groups: high-quality private preschool, high-quality public preschool, and UPSTART students.

Seventy-seven percent of UPSTART children had posttest literacy quotients of average or above average, representing a greater level of school readiness than was achieved by either the other intervention groups or children not participating in high-quality school readiness programs. UPSTART children outperformed children not participating in high-quality school readiness programs on Overall Literacy test scores, and subtest scores for UPSTART children were significantly higher in Letter Knowledge, Listening Comprehension, and Phonological Awareness. By the end of the program year, social emotional development (SED) was similar for all three treatment groups.

Figure 25: Listening Comprehension Benchmarks at Posttest

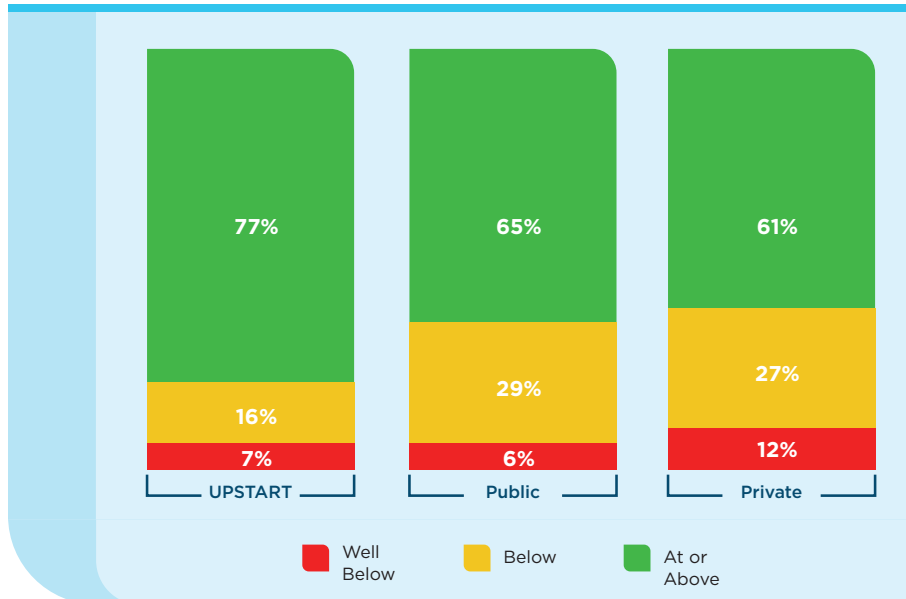
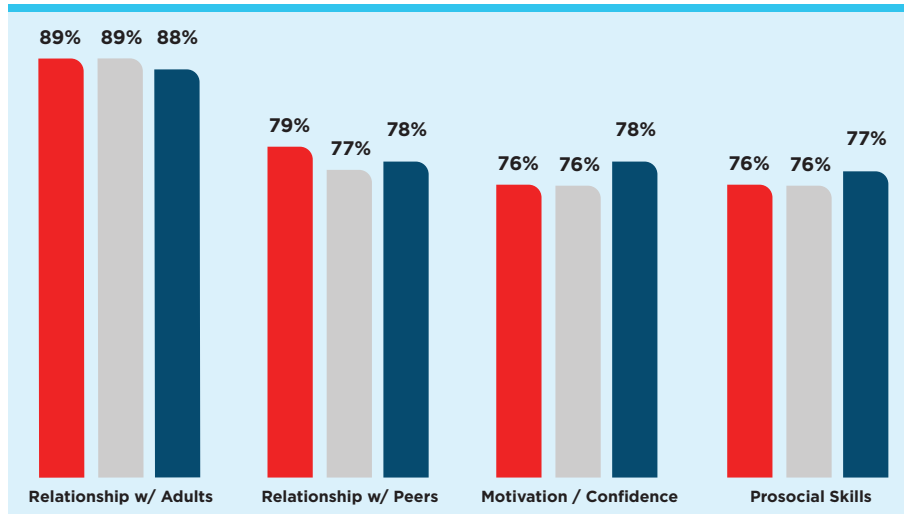


Table 13: Effect Size Estimates by Literacy Domain

Literacy Domain	Skills Tested	UPSTART	Public	Private	Instrument
Letter Knowledge	Letter Sounds	<b>.60**</b>	NS	NS	Brigance
	Recites Alphabet				
	Letter Knowledge				
Listening Comprehension	Inference/Prediction	<b>.29*</b>	NS	NS	PELI
	Recollection				
	Cloze				
Phonological Awareness	Rhyme Recognition	<b>.84**</b>	<b>-.26*</b>	NS	Bader
	Phonemic Blending				
	Phonemic Segmenting				



Figure 26: Post-Program Social-Emotional Development Skills by Program Group



## Tier III: Promising Evidence

For a study to be considered promising evidence, it needs to be a well-designed and well-implemented correlational study, which uses statistical controls for selection bias. This includes sampling and/or analytic methods to account for the differences between the treatment and control group. Additionally, the intervention in the study must demonstrate a statistically significant and positive effect on a student outcome, without having statistically significant and negative outcomes in other studies that meet the WWC Evidence Standards with or without reservations.

### Evaluation of a Tucson Unified School District

Powers and Price-Johnson (2006) completed a large-scale study of the software among 15 kindergarten classes from Tucson, Arizona’s Unified School District. Results showed that the students who used Waterford ( $N = 358$ ) significantly outperformed a large comparison group ( $N = 1480$ ) on both the Dynamic Indicators of Basic Early Literacy Skills (effect size = 0.42) and Core Curriculum Standard Assessment tests (effect size = 0.28). The researchers disaggregated the data by school, gender, ethnicity, primary home language, and other measures. Waterford was found particularly effective for ELL students; these groups demonstrated greater gains than the English-proficient group in the comparison schools ( $F[1, 1045] = 8.62, p = .003$ ).

Table 14: ELL & Non-ELL Students on DIBELS Total Reading Score

	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
ELL students								
ERP	164	9.12	6.65	31.23	12.95	22.11	32.37	.000
Comparison	329	6.38	5.66	22.77	12.04	<u>16.39</u>	32.43	.000
ERP vs. Comparison						5.72***		
Non-ELL students								
ERP	170	12.07	7.00	35.86	11.49	23.79	36.94	.000
Comparison	882	11.66	8.66	31.77	12.97	<u>20.11</u>	65.26	.000
ERP vs. Comparison						3.68***		

Note. ERP students selected with 1100 minutes (6 months) or more usage of ERP Reading Program.  
 \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$  from independent  $t$  tests comparing gains.

Table 15: ANCOVA & Effect Sizes on All Outcome Measures (ERP Students With 1100 or More Usage Minutes)

Measures	N	Covariate		AdjPosttest		ES	F	p
		M	SD	M	SD			
DIBELS: ISF								
ERP	334	4.87	5.71	24.14	13.88	0.56	81.57	.000
Comparison	1218	6.42	6.82	17.35	12.24			
DIBELS: LNF								
ERP	334	4.44	8.17	44.41	16.32	0.25	16.33	.000
Comparison	1155	6.30	10.15	40.61	16.36			
DIBELS: WUF								
ERP	325	3.57	7.15	33.41	20.53	0.06	0.89	.345
Comparison	998	4.94	10.40	32.18	20.87			
DIBELS: PSF								
ERP	355	21.05	15.93	44.58	15.05	0.31	26.22	.000
Comparison	1219	17.10	15.87	39.88	18.69			
DIBELS: NWF								
ERP	355	18.26	14.61	37.06	20.59	0.26	1.16	.282
Comparison	1217	14.66	15.02	31.94	20.21			
DIBELS: Total Reading								
ERP	334	10.62	6.98	33.22	12.42	0.42	46.16	.000
Comparison	1211	10.22	8.29	29.43	13.33			

<i>Measures</i>	<i>N</i>	<u>Covariate</u>		<u>AdjPosttest</u>		<i>ES</i>	<i>F</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
TUSD: CCSA Reading								
ERP	311	1.09	0.49	2.67	0.64	0.28	20.04	.000
Comparison	1263	1.07	0.59	2.41	1.02			

*Note.* ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. The effect size is the adjusted mean posttest difference divided by the square root of the ANCOVA mean squared residual.

## Evaluation of a School District in Florida, 2014-2015

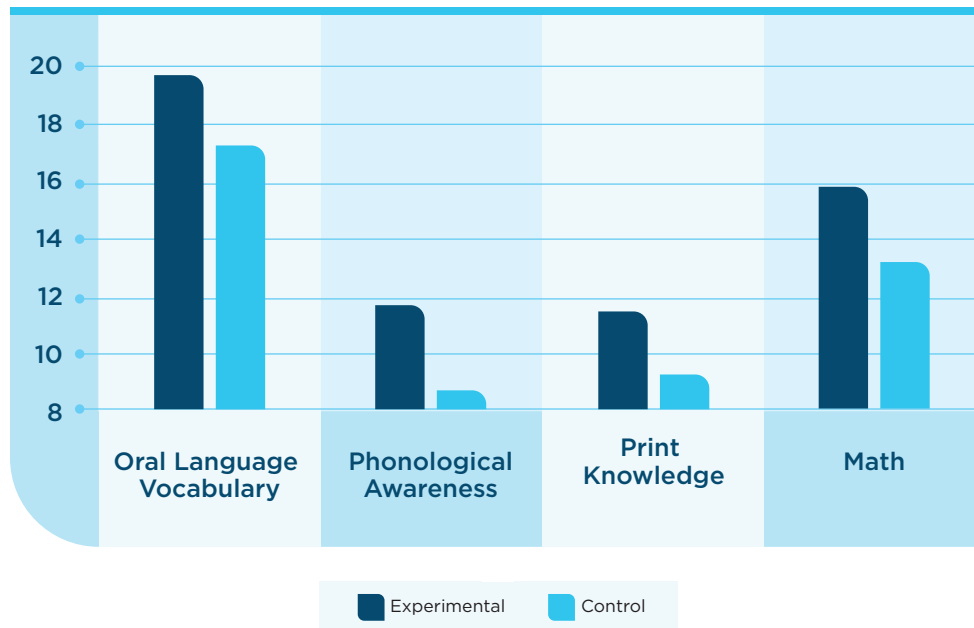
The following study investigates the effect of adding Waterford Early Reading Program (ERP) and Early Math and Science Program (EMS) to an existing Head Start curriculum and its impact on low socioeconomic students’ reading and math scores (Shamir, Feehan, & Yoder, 2016).

Students were administered Florida’s Voluntary Pre-kindergarten Assessment (VPK), a standardized assessment during the fall, winter, and spring of the 2014-2015 school year. The assessment included sub-strands for Oral Language Vocabulary, Phonological Awareness, Print Knowledge, and Math. An ANCOVA examining group differences in spring scores while covarying for fall scores was conducted, revealing significant differences between experimental and control groups (see Table 16 and Figure 27).

*Table 16: Spring VPK Sub-Strand Scores, Covarying for Fall*

Sub Strand	F	p	Experimental (M)	Control (M)	Effect Size (d)
Oral Language Vocabulary	2.62	<0.01	19.64	17.42	0.72
Phonological Awareness	3.81	<0.01	11.92	8.59	1.32
Print Knowledge	2.105	<0.01	11.13	8.78	1.12
Math	0.685	<0.01	15.93	13.11	0.77

Figure 27: Spring VPK Sub-Strand Scores, Covarying for Fall



Further analysis was conducted to examine the effects of gender, ELL status, ethnicity, and special education status on spring scores. Four separate ANCOVA were conducted for each of the four sub strands, which examined the effect of demographics and Waterford curriculum on spring scores while covarying for fall scores (see Table 17 and Figures 28-31).

Table 17: Effects of Demographics on Scores for Each Sub-Strand

Demographic	Oral Language Vocabulary		Phonological Awareness		Print Knowledge		Math	
	F	p	F	p	F	p	F	P
Gender	0.38	0.538	4.798	<.05	16.137	<0.01	2.62	0.106
ELL Status	1.319	0.251	0.05	0.823	8.447	<0.01	3.81	0.051
Ethnicity	2.43	<0.05	1.928	0.104	2.8	<0.05	2.105	0.076
Special Education Status	2.802	<0.06	2.652	0.071	4.326	<0.05	0.685	0.562

Figure 28: Oral Language Vocabulary Spring Scores by Demographics

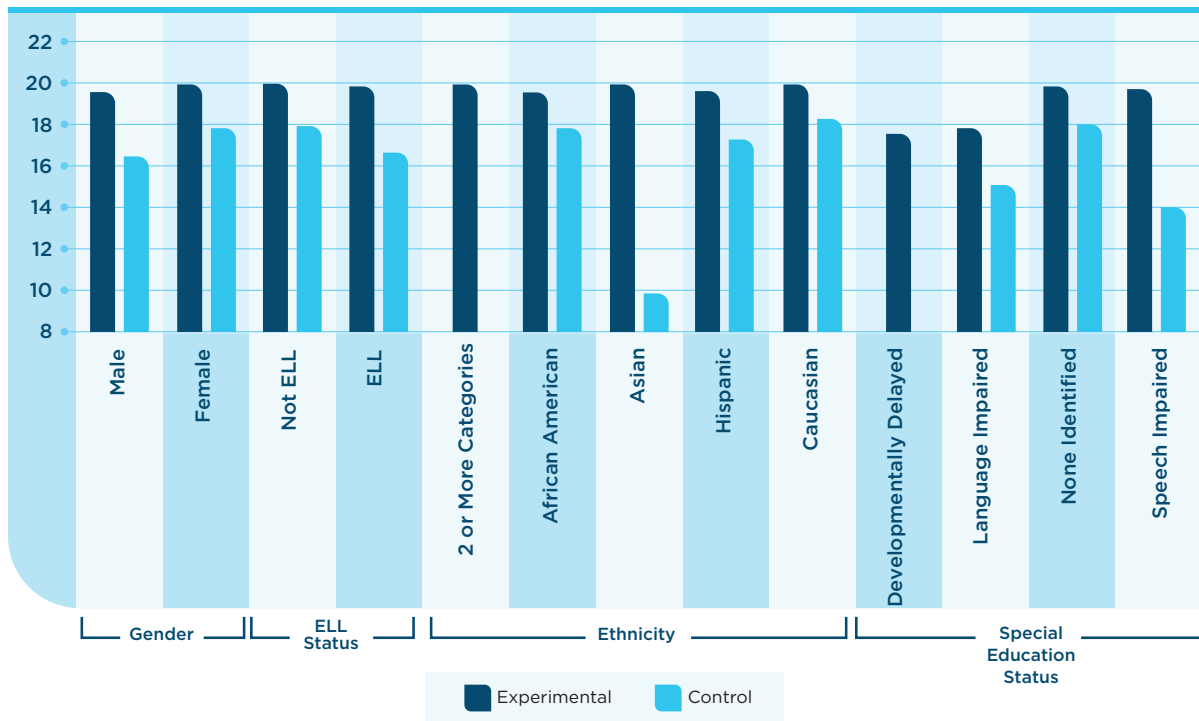


Figure 29: Phonological Awareness Spring Scores by Demographics

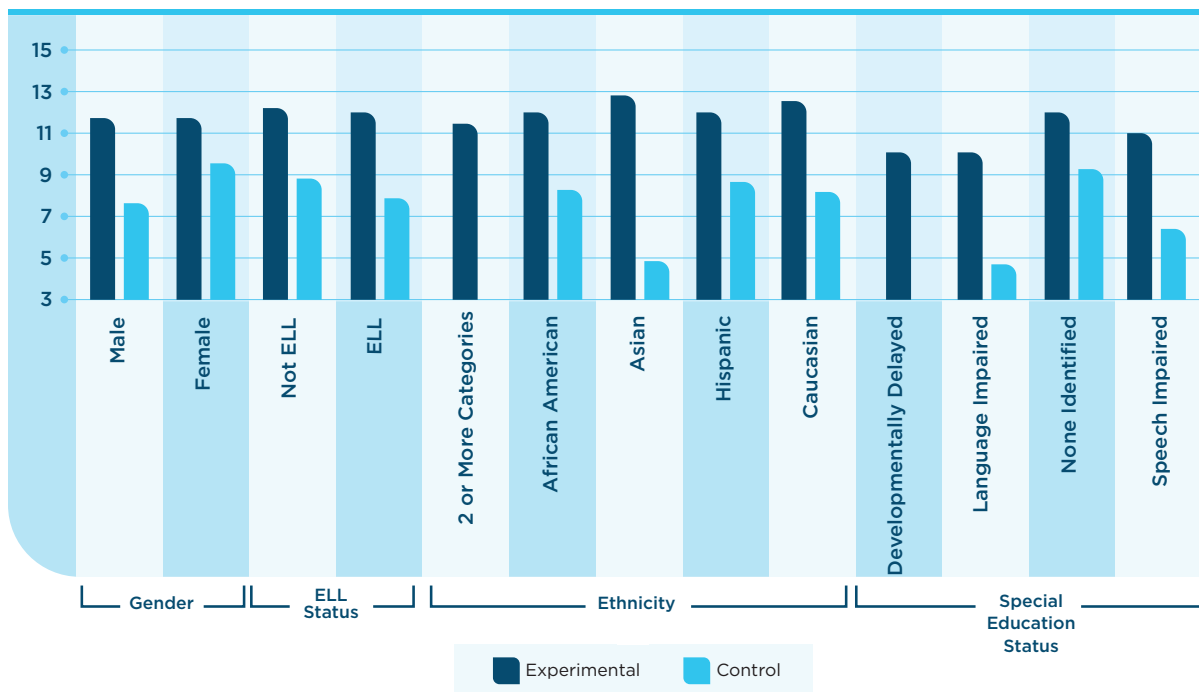


Figure 30: Print Knowledge Spring Scores by Demographics

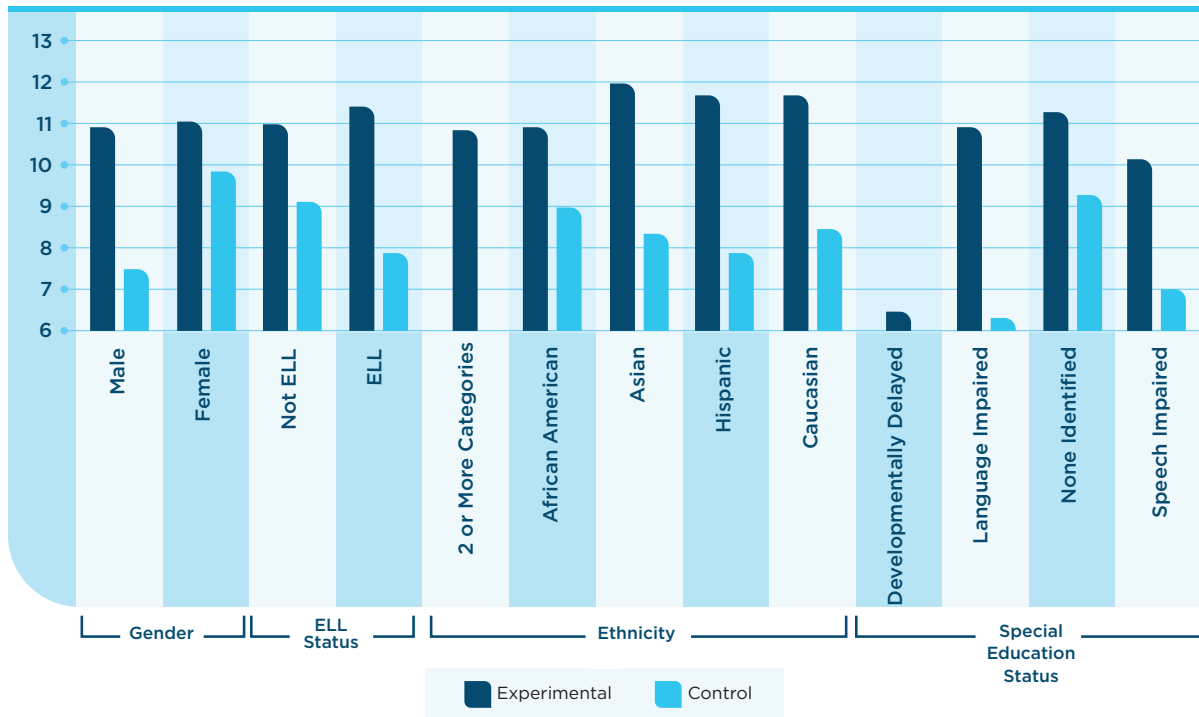
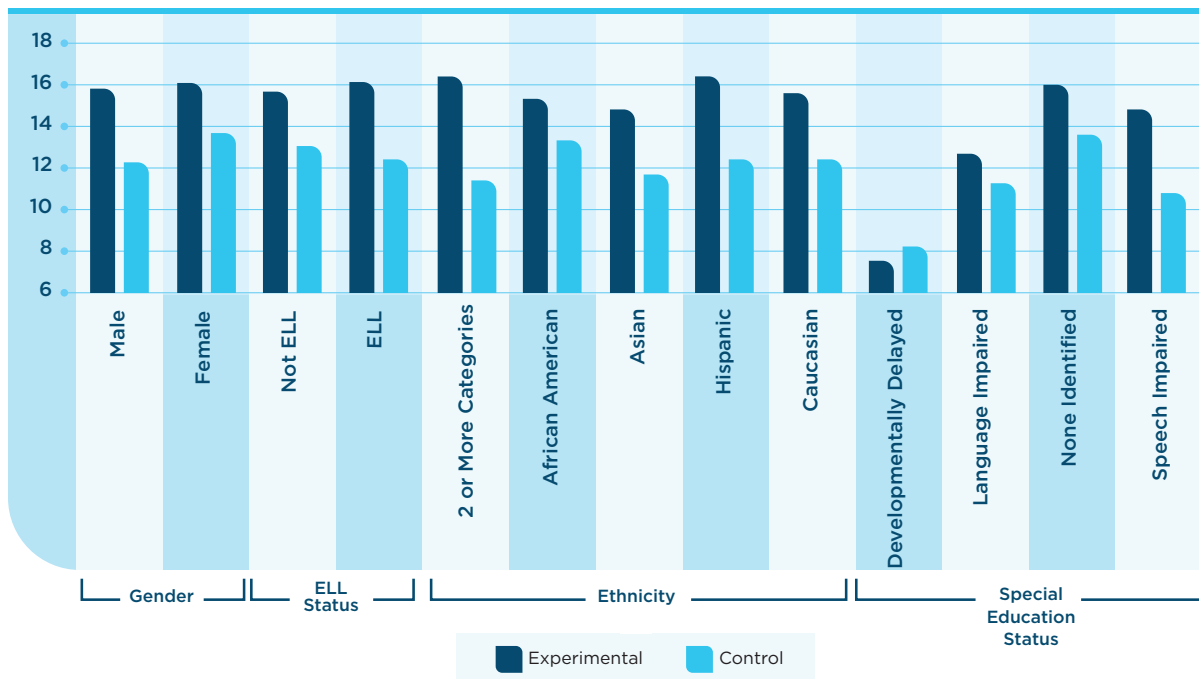


Figure 31: Math Spring Scores by Demographics



## Evaluation of the Waterford Early Reading Program in California

The sample in the current study was composed of 80% Hispanic students and 20% African American students (Shamir & Goethe, 2015). Ninety percent of students qualified for free/reduced lunch. The treatment group consisted of students who used the Waterford Early Reading Program (ERP). The control group consisted of students who did not use ERP. Students were administered the AIMSweb reading assessment.

An ANCOVA on posttest AIMSweb scores with the pretest or the middle of year test (depending on the skill) as a covariate was conducted to compare treatment and control students' scores. The students using ERP significantly outperformed the comparison group on three of the four sub-stands: Letter Sound Fluency (LSF), Letter Name Fluency (LNF), and Nonsense Word Fluency (NWF).

Figure 32: Mean Scores by AIMSweb Sub-Strands for Kindergarten Students

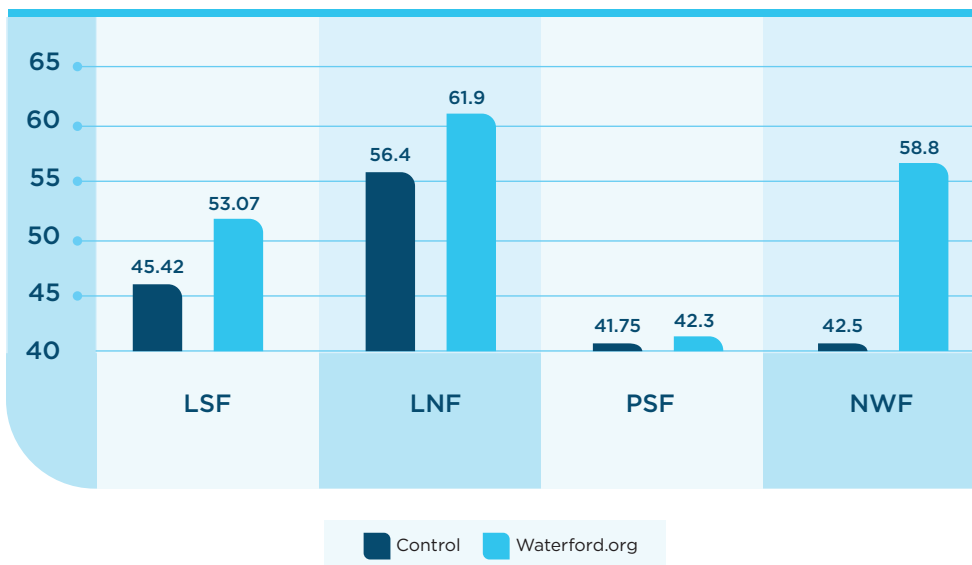
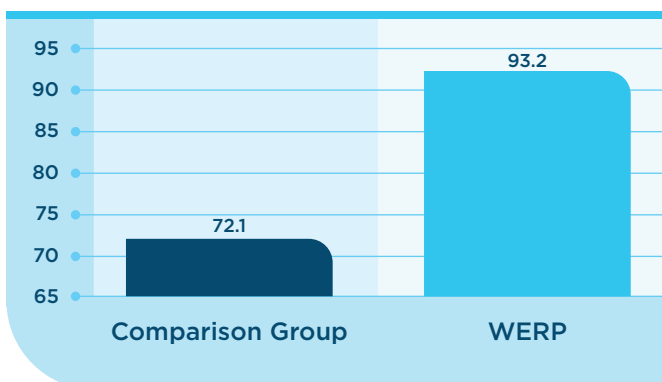


Figure 33: Mean Scores by AIMSweb Sub-Strand NWF for First Grade Students



## Evaluation of the Waterford Early Math and Science Program in Indiana, 2015-2016

The following study (Shamir, Feehan, & Yoder, 2017b) assessed the efficacy of the Waterford Early Math and Science Program (EMS), a computer-adaptive program that was assigned to kindergarten and first grade students in a school district in Indiana during the 2015-2016 school year. Kindergarten students in the experimental group were expected to use EMS for 15 minutes per day, five days per week, and first grade students in the experimental group were expected to use EMS for 30 minutes per day, five days per week. The control groups consisted of students who did not use EMS during the 2015-2016 school year. The experimental group for kindergarten consisted of 114 students, and the control group consisted of 58 students. For first grade, the experimental group consisted of 68 students, and the control group consisted of 255 students. ANCOVAs examining group differences in mCLASS: Math end of year scores while covarying for beginning of year scores were conducted (see Figures 34-35).

### KINDERGARTEN

Analysis of Number Identification end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 168) = 7.34, p < .01$ , due to higher end of year scores made by students who used Waterford ( $M = 32.38$ ) than by control students ( $M = 28.25$ ). Effect size ( $d = 0.33$ ).

Analysis of Quantity Discrimination end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 168) = 4.30, p < .05$ , due to higher end of year scores made by students who used Waterford ( $M = 30.80$ ) than by control students ( $M = 28.12$ ). Effect size ( $d = 0.29$ ).

Analysis of Counting end of year scores, while covarying for beginning of year scores, did not reveal a significant difference between groups,  $F(1, 168) = 3.43, p = .066$ ; however, Waterford students had higher end of year scores ( $M = 88.64$ ) than control students ( $M = 84.33$ ).

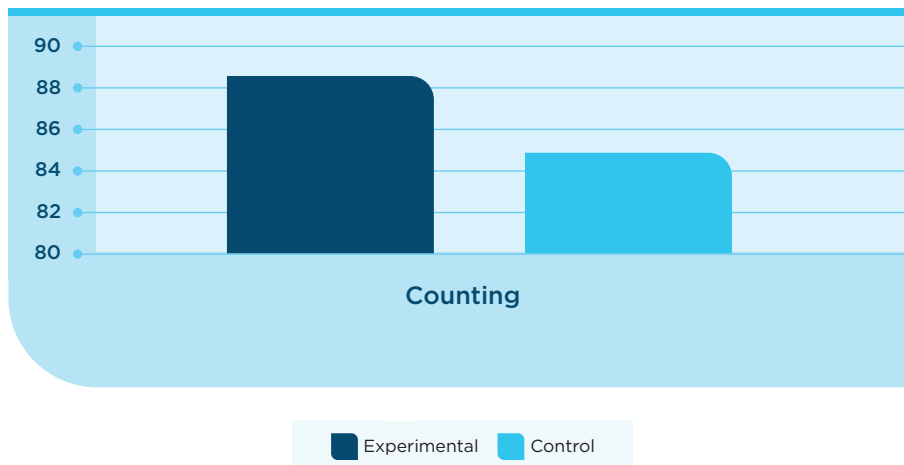
Analysis of Missing Number end of year scores, while covarying for beginning of year scores, did not reveal a significant difference between groups,  $F(1, 168) = 0.04, p = .839$ ; however, Waterford students had higher end of year scores ( $M = 15.70$ ) than control students ( $M = 15.53$ ).



Figure 34: Kindergarten mCLASS: Math End of Year Scores by Sub-Strand



Figure 35: Kindergarten mCLASS: Math Counting End of Year Scores



### GROUP DIFFERENCES BY DEMOGRAPHICS USING ANCOVAS

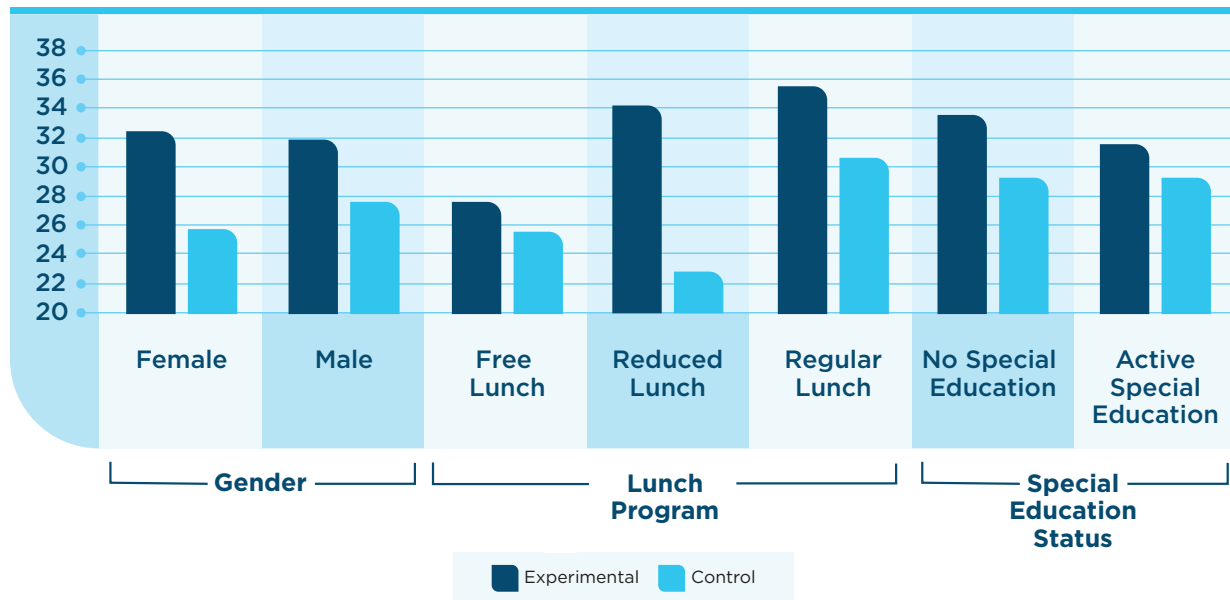
Further analysis was conducted to examine the effects of gender, lunch program, and special education status on Number Identification end of year scores (see Figure 36).

There was no significant interaction between the effects of gender and Waterford curriculum on Number Identification end of year scores, covarying for beginning of year scores,  $F(1, 166) = 2.90, p = .091$ . Simple effects analysis showed that for females, students in the experimental group significantly outperformed students in the control group. Male students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Number Identification end of year scores, covarying for beginning of year scores,  $F(2, 164) = 1.10, p = .334$ . Simple effects analysis showed that for reduced lunch, students in the experimental group significantly outperformed students in the control group. Free lunch and regular lunch students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of special education status and Waterford curriculum on Number Identification end of year scores, covarying for beginning of year scores,  $F(1, 166) = 0.53, p = .468$ . Simple effects analysis showed that for students with no special education status, the experimental group significantly outperformed the control group. For students with active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 36: Number Identification End of Year Scores by Demographics



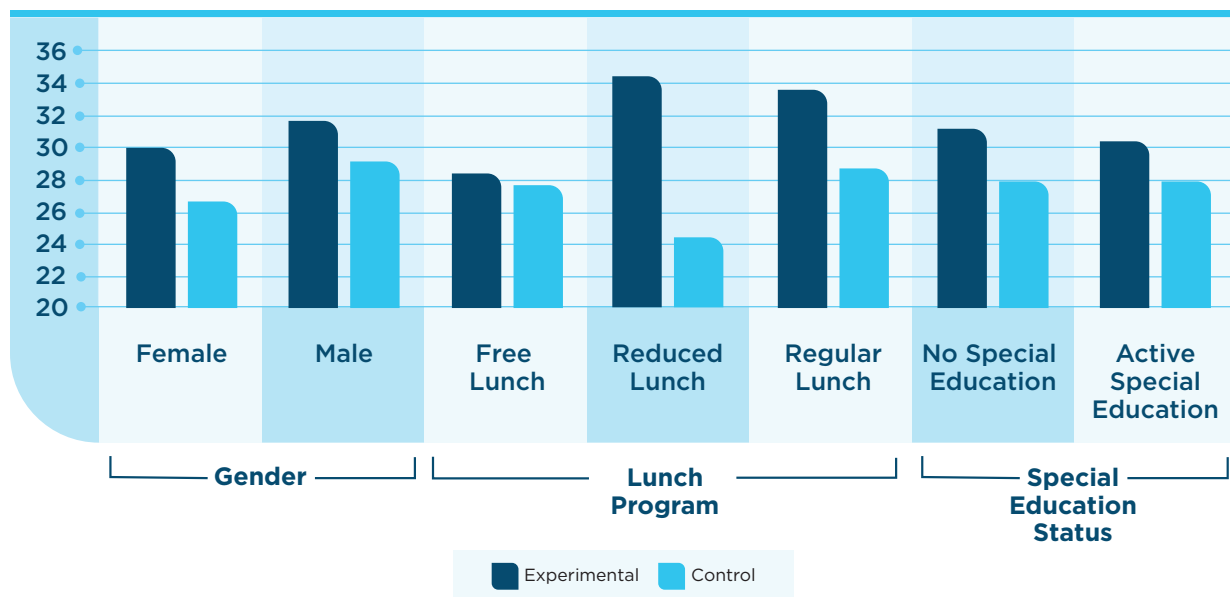
Further analysis was conducted to examine the effects of gender, lunch program, and special education status on Quantity Discrimination end of year scores (see Figure 37).

There was no significant interaction between the effects of gender and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(1, 166) = 0.12, p = .729$ . Simple effects analysis showed that for males and females, students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(2, 164) = 2.41, p = .093$ . Simple effects analysis showed that for reduced lunch and regular lunch, students in the experimental group significantly outperformed students in the control group. Free lunch students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of special education status and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(1, 166) = 0.17, p = .677$ . Simple effects analysis showed that for students with no special education status and active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 37: Quantity Discrimination End of Year Scores by Demographics



## FIRST GRADE

### GROUP DIFFERENCES USING ANCOVAS

ANCOVAs examining group differences in mCLASS: Math end of year scores while covarying for beginning of year scores were conducted.

Analysis of Number Identification end of year scores, while covarying for beginning of year scores, did not reveal a significant difference between groups,  $F(1, 320) = 0.06, p = .813$ ; however, Waterford students ( $M = 52.40$ ) scored slightly higher than control students ( $M = 52.12$ ).

Analysis of Number Facts end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 320) = 9.06, p < .01$ , due to higher end of year scores made by students who used Waterford ( $M = 14.02$ ) than by control students ( $M = 12.69$ ). Effect size ( $d = 0.36$ ). Analysis of Quantity Discrimination end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 320) = 5.88, p < .05$ , due to higher end of year scores made by students who used Waterford ( $M = 42.17$ ) than by control students ( $M = 39.78$ ). Effect size ( $d = 0.26$ ).

Analysis of Counting end of year scores, while covarying for beginning of year scores, did not reveal a significant difference between groups,  $F(1, 320) = 0.66, p = .416$ ; however, Waterford students ( $M = 107.08$ ) scored slightly higher than control students ( $M = 106.03$ ).

Analysis of Missing Number end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 320) = 15.07, p < .01$ , due to higher end of year scores made by students who used Waterford ( $M = 25.90$ ) than by control students ( $M = 23.12$ ). Effect size ( $d = 0.42$ ).

Analysis of Next Number end of year scores, while covarying for beginning of year scores, revealed a significant difference between groups,  $F(1, 320) = 6.18, p < .05$ , due to higher end of year scores made by students who used Waterford ( $M = 23.77$ ) than by control students ( $M = 22.09$ ). Effect size ( $d = 0.28$ ).

Figure 38: First Grade mCLASS: Math End of Year Scores by Sub-Strand

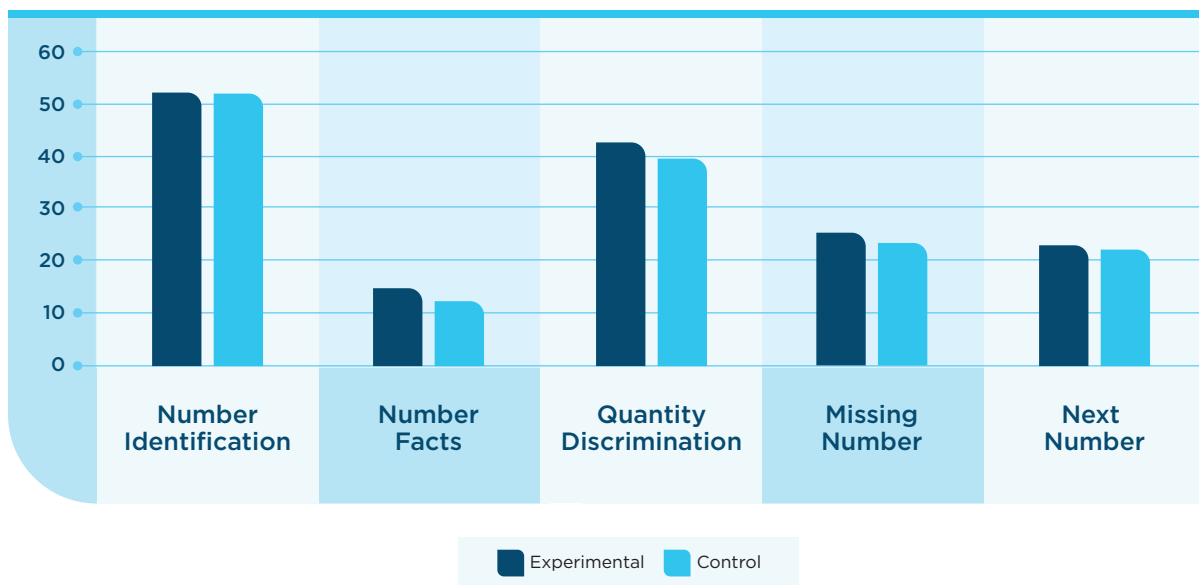
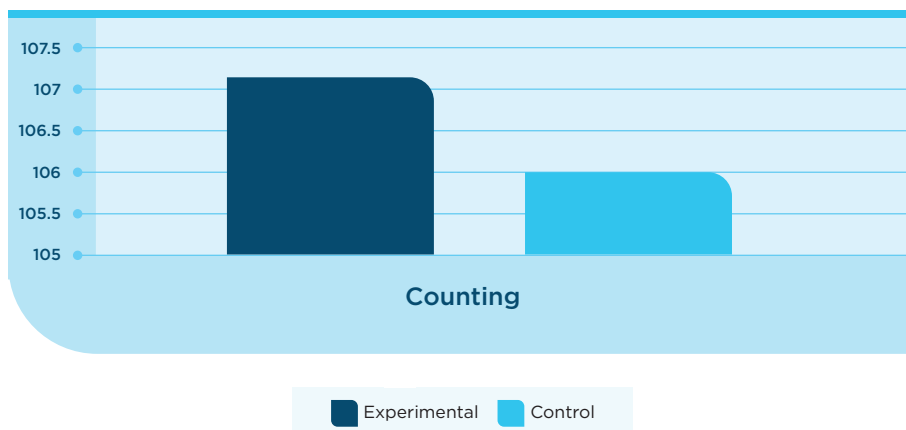


Figure 39: First Grade mCLASS: Math Counting End of Year Scores



## GROUP DIFFERENCES BY DEMOGRAPHICS USING ANCOVAS

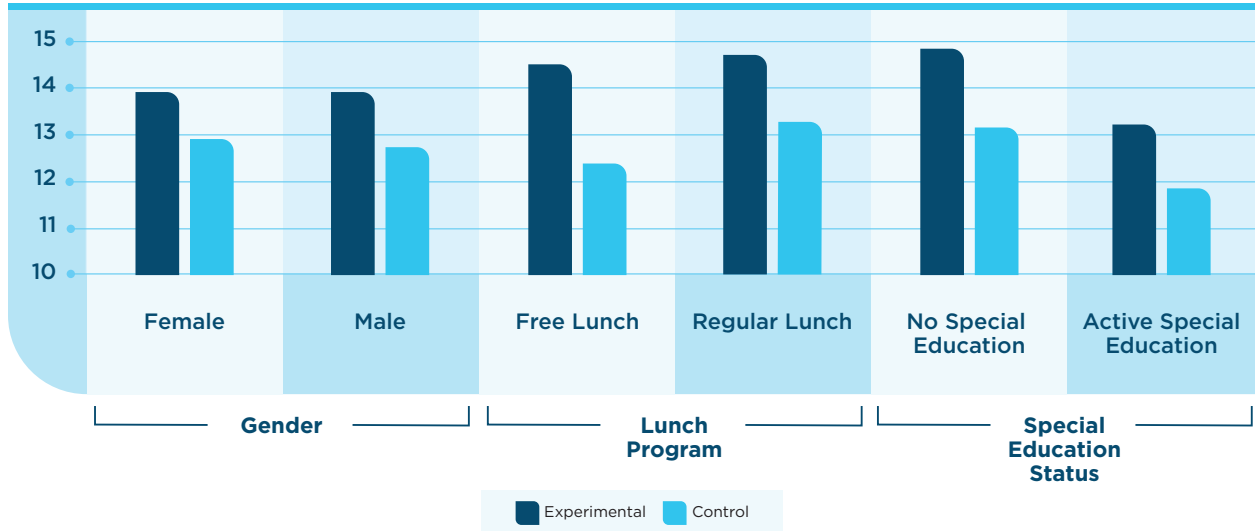
Further analysis was conducted to examine the effects of gender, lunch program, and special education status on Number Facts end of year scores.

There was no significant interaction between the effects of gender and Waterford curriculum on Number Facts end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.05, p = .818$ . Simple effects analysis showed that for males, students in the experimental group significantly outperformed students in the control group. Female students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Number Facts end of year scores, covarying for beginning of year scores,  $F(2, 310) = 2.86, p = .059$ . Simple effects analysis showed that for free lunch and regular lunch, students in the experimental group significantly outperformed students in the control group.

There was no significant interaction between the effects of special education status and Waterford curriculum on Number Facts end of year scores, covarying for beginning of year scores,  $F(1, 317) = .00, p = .982$ . Simple effects analysis showed that for students with no special education status, the experimental group significantly outperformed the control group. For students with active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 40: First Grade Number Facts End of Year Scores by Demographics



Further analysis was conducted to examine the effects of gender, Limited English Proficiency (LEP) status, lunch program, and special education status on end of year Quantity Discrimination scores.

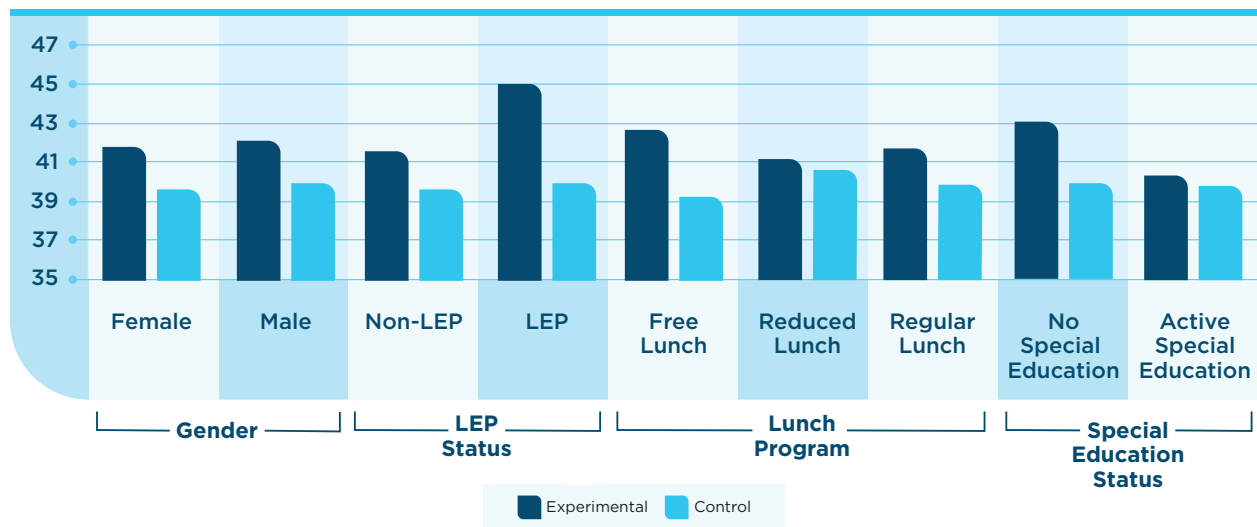
There was no significant interaction between the effects of gender and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.01, p = .918$ . Simple effects analysis showed that for males and females, students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of LEP status and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.56, p = .457$ . Simple effects analysis showed that Non-LEP students' scores in the experimental group were slightly higher than in the control group, approaching significance. LEP students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(2, 310) = 0.37, p = .694$ . Simple effects analysis showed that for free lunch, students in the experimental group significantly outperformed students in the control group. Reduced lunch and regular lunch students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of special education status and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores,  $F(1, 317) = 2.01, p = .158$ . Simple effects analysis showed that for students with no special education status, the experimental group significantly outperformed the control group. For students with active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 41: First Grade Quantity Discrimination End of Year Scores by Demographics



Further analysis was conducted to examine the effects of gender, LEP status, lunch program, and special education status on end of year Missing Number scores.

There was no significant interaction between the effects of gender and Waterford curriculum on Missing Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.17, p = .682$ . Simple effects analysis showed that for males and females, students in the experimental group significantly outperformed students in the control group.

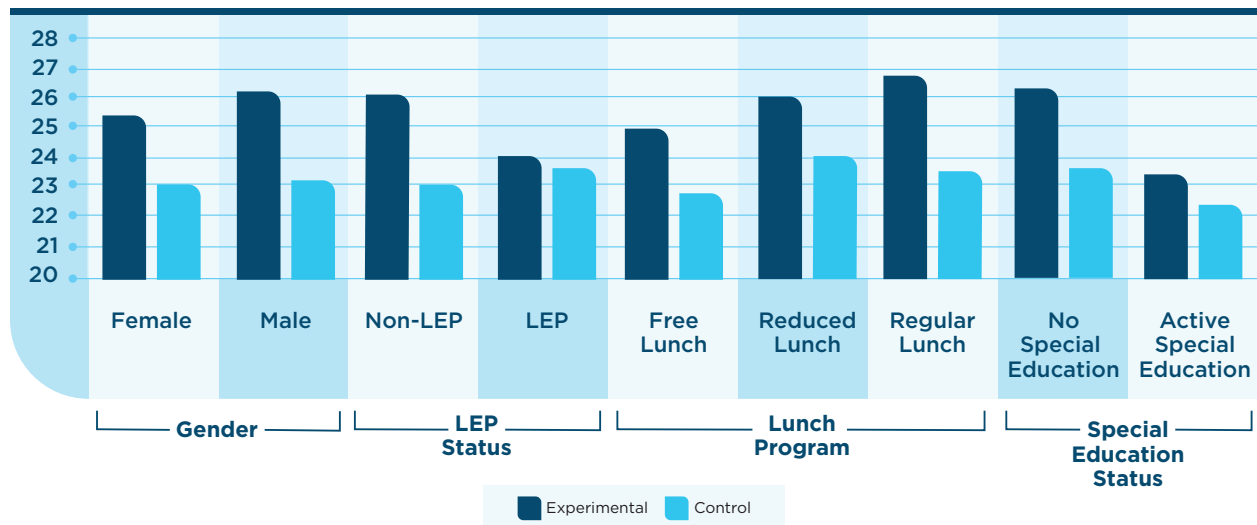
There was no significant interaction between the effects of LEP status and Waterford curriculum on Missing Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 1.47, p = .227$ . Simple effects analysis showed that Non-LEP students in the experimental group significantly

outperformed students in the control group. LEP students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Missing Number end of year scores, covarying for beginning of year scores,  $F(2, 310) = 0.32, p = .730$ . Simple effects analysis showed that for free lunch and regular lunch, students in the experimental group significantly outperformed students in the control group. Reduced lunch students' scores in the experimental group were higher than in the control group, approaching significance.

There was no significant interaction between the effects of special education status and Waterford curriculum on Missing Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.32, p = .574$ . Simple effects analysis showed that for students with no special education status, the experimental group significantly outperformed the control group. For students with active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 42: First Grade Missing Number End of Year Scores by Demographics



Further analysis was conducted to examine the effects of gender, LEP status, lunch program, and special education status on Next Number end of year scores.

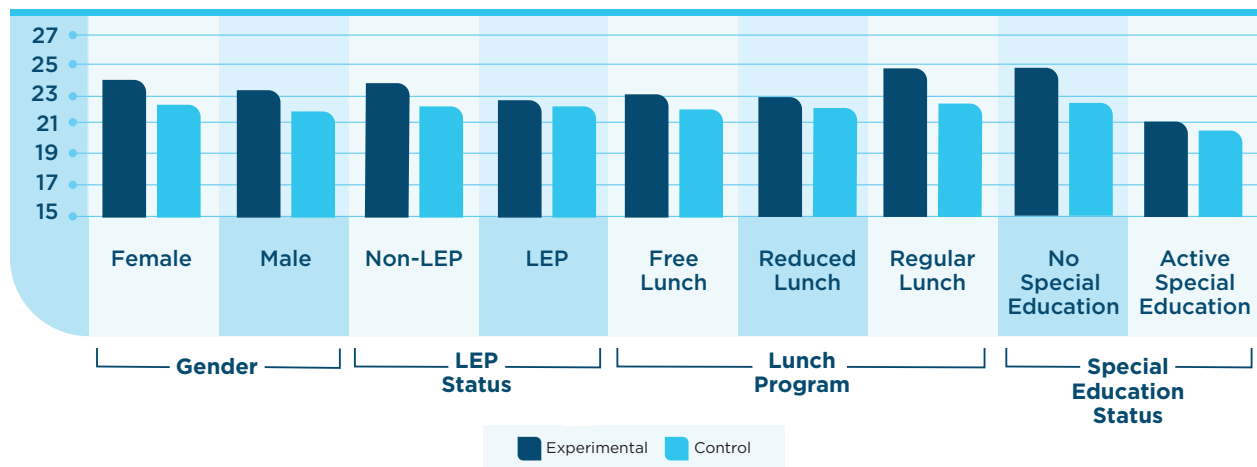
There was no significant interaction between the effects of gender and Waterford curriculum on Next Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.07, p = .787$ . Simple effects analysis showed that for males, students in the experimental group significantly outperformed students in the control group. Female students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of LEP status and Waterford curriculum on Next Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 0.65, p = .421$ . Simple effects analysis showed that Non-LEP students in the experimental group significantly outperformed students in the control group. LEP students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of lunch program and Waterford curriculum on Next Number end of year scores, covarying for beginning of year scores,  $F(2, 310) = 0.26, p = .775$ . Simple effects analysis showed that for regular lunch, students in the experimental group significantly outperformed students in the control group. Free lunch and reduced lunch students' scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

There was no significant interaction between the effects of special education status and Waterford curriculum on Next Number end of year scores, covarying for beginning of year scores,  $F(1, 317) = 1.03, p = .312$ . Simple effects analysis showed that for students with no special education status, the experimental group significantly outperformed students in the control group. For students with active special education status, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 43: First Grade Quantity Discrimination End of Year Scores by Demographics



## Effectiveness of UPSTART

### YEAR 2

Waterford Institute enrolled 1,018 preschool children in its second year of operation during the 2010-11 school year (Evaluation and Training Institute, 2012). The UPSTART treatment group performed significantly better than the control group on the Total Brigance posttest, with an average difference of 7.9 points. The UPSTART treatment group also performed significantly better than the control group on the overall Bader and the Bader Phoneme Blending subtest. Growth rates between the treatment group and the control group were significantly different at the 99% Confidence Interval (CI) for the Total Brigance and the Total Bader.

Children participating in UPSTART demonstrated moderately strong improvements in literacy skills on the Brigance and small improvements in literacy skills on the Bader compared to control children on the Brigance. Children participating in UPSTART scored higher on strands for Lower Case Letters and Lower Case Letter Sounds even when taking initial literacy skills into account. Improvements on the Bader were driven by performance on the Phoneme Blending subtest.



## BRIGANCE POSTTEST RESULTS

Table 18: Brigance Posttest Analysis of Treatment-Control Group Differences

Brigance Posttest	Group	N	Mean	SEM	t	Significance
Expressive Objects	Treatment	77	25.95	0.12	-0.09	NS
	Control	82	25.96	0.12		
Receptive Objects	Treatment	77	26.97	0.08	0.38	NS
	Control	82	26.96	0.02		
Expressive Grammar	Treatment	77	08.90	0.02	0.97	NS
	Control	82	10.10	0.17		
Visual Discrimination	Treatment	77	17.18	0.40	0.77	NS
	Control	82	16.76	0.38		
Recites Alphabet	Treatment	77	18.77	1.06	1.51	NS
	Control	82	16.45	1.10		
Lowercase Letter Knowledge	Treatment	77	41.31	1.67	2.63	**
	Control	82	34.21	2.13		
Sounds of Lowercase Letters	Treatment	77	16.69	1.09	2.56	**
	Control	82	12.61	1.17		
Auditory Discrimination	Treatment	77	7.74	0.31	1.24	NS
	Control	82	7.15	0.36		
Survival Sight Words	Treatment	77	3.74	0.45	1.61	NS
	Control	82	2.85	0.32		
Basic Preprimer Vocabulary	Treatment	77	7.94	1.04	2.28	NS
	Control	82	4.90	0.83		
Total Brigance	Treatment	77	176.39	4.72	2.71	**
	Control	82	157.76	4.96		

\*\* $p < .01$

## BADER RESULTS

Table 19: Bader Posttest Analysis of Treatment-Control Group Differences

Bader Posttest	Group	N	Mean	SEM	t	Significance
Rhyme Recognition	Treatment	76	6.95	0.40	0.59	NS
	Control	82	6.63	0.36		
Phoneme Blending	Treatment	76	4.62	0.38	2.55	**
	Control	82	3.26	0.40		
Phoneme Segmenting	Treatment	76	2.71	0.37	1.58	NS
	Control	82	1.93	0.33		
Total Bader	Treatment	76	14.28	0.93	1.97	NS
	Control	82	11.82	0.84		

\*\* $p < .01$

## BRIGANCE GROWTH SCORE RESULTS

Table 20: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance	<u>Treatment Group</u>		<u>Control Group</u>		Significance <i>p</i> < .01
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Expressive Objects	0.545	0.15 – 0.94	0.915	0.47 – 1.36	NS
Receptive Objects	0.234	0.01 – 0.46	0.305	-0.10 – 0.71	NS
Expressive Grammar	1.208	0.75 – 1.66	1.04	0.51 – 1.57	NS
Visual Discrimination	5.260	3.62 – 6.90	3.67	2.37 – 4.95	NS
Recites Alphabet	9.740	6.46 – 13.03	4.00	1.15 – 6.85	NS
Lowercase Letter Knowledge	19.948	14.46 – 25.44	11.54	7.30 – 15.77	NS
Lowercase Letter Sounds	10.442	7.75 – 13.13	5.43	3.21 – 7.64	**
Auditory Discrimination	3.558	2.27 – 4.85	1.00	-0.31 – 2.31	NS
Survival Sight Words	2.247	1.28 – 3.22	1.31	0.69 – 1.92	NS
Basic Preprimer Vocabulary	6.403	3.94 – 8.87	3.83	2.11 – 5.55	NS
Total Brigance	59.84	49.18 – 69.99	33.01	25.88 – 40.15	**

## BADER GROWTH SCORE RESULTS

Table 21: Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	<u>Treatment Group</u>		<u>Control Group</u>		Significance <i>p</i> < .01
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Rhyme Recognition	2.18	1.02 – 3.35	0.87	-0.28 – 2.01	NS
Phoneme Blending	3.21	2.26 – 4.16	1.06	0.29 – 1.83	**
Phoneme Segmenting	2.18	1.08 – 3.29	0.95	0.24 – 1.67	NS
Total Bader	7.58	5.18 – 9.97	2.88	1.13 – 4.62	**

## YEAR 3

Waterford Institute enrolled 1,168 preschool children in its third year of operation during the 2011-12 school year (Evaluation and Training Institute, 2013). The UPSTART treatment group significantly outperformed the control group on both the Total Brigance and Total Bader. Effect sizes range from 0.33 to 0.85. Favoring the UPSTART treatment group, growth rates between the treatment group and the control group were significantly different at the 99% Confidence Interval (CI) for the Total Brigance and the Total Bader.

Consistent with prior cohorts, children participating in UPSTART demonstrated moderately strong improvements in measures of phonics skills: At the beginning of kindergarten, improvement was observed across all but one strand of the Brigance. Participation in UPSTART

was associated with improvement in all phonological awareness strands of the Bader, including Rhyme Recognition, Phoneme Blending, and Phoneme Segmenting. Children participating in UPSTART had a 28-point advantage on Brigance posttest scores compared to non-participating children.

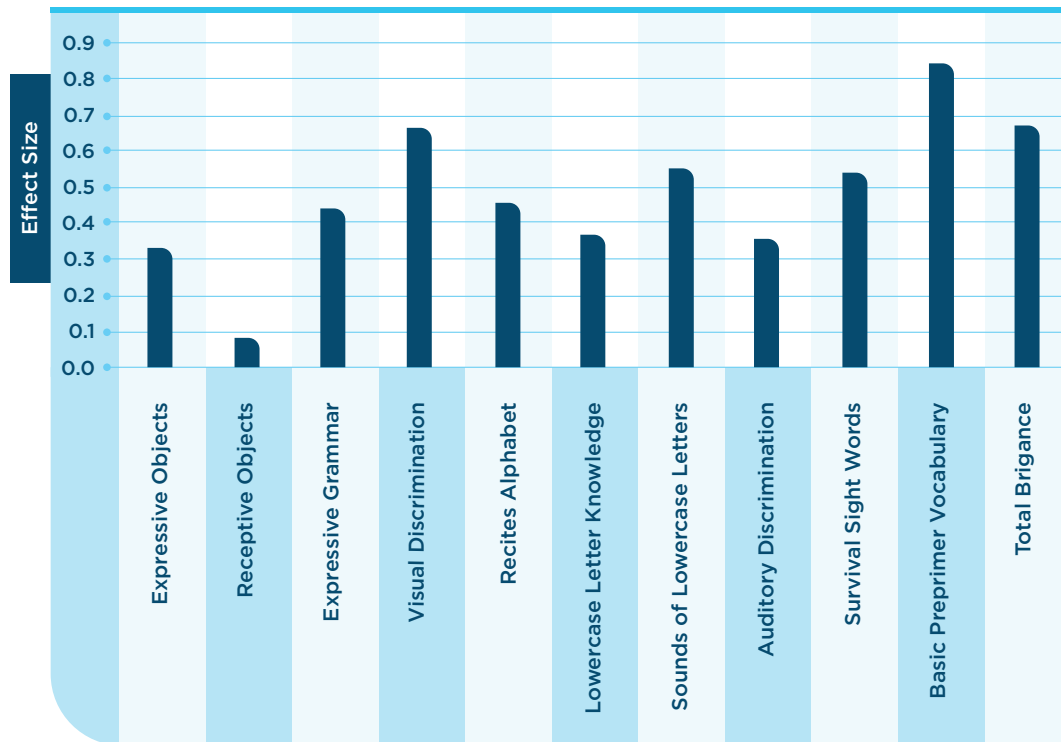
## BRIGANCE RESULTS

Table 22: Brigance Posttest Analysis of Treatment-Control Group Differences

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>Significance</i>
Expressive Objects	Treatment	129	26.35	.79	3.006	**
	Control	130	26.01	1.02		
Receptive Objects	Treatment	129	26.99	.09	1.000	NS
	Control	130	26.98	.15		
Expressive Grammar	Treatment	129	10.20	1.38	3.475	**
	Control	130	9.62	1.33		
Visual Discrimination	Treatment	129	18.44	2.05	5.985	**
	Control	130	16.61	2.82		
Recites Alphabet	Treatment	129	21.19	8.04	3.635	**
	Control	130	17.29	9.16		
Lowercase Letter Knowledge	Treatment	129	44.60	12.33	3.488	**
	Control	130	38.12	17.19		
Sounds of Lowercase Letters	Treatment	129	20.12	7.99	4.886	**
	Control	130	14.71	9.77		
Auditory Discrimination	Treatment	129	8.22	2.58	3.295	**
	Control	130	7.06	3.04		
Survival Sight Words	Treatment	129	4.88	4.39	3.369	**
	Control	130	3.29	3.06		
Basic Preprimer Vocabulary	Treatment	129	13.25	9.19	6.369	**
	Control	130	6.37	8.16		
Total Brigance	Treatment	129	194.23	36.81	6.369	**
	Control	130	166.05	41.32		

\*\* $p < .01$

Figure 44: UPSTART’s Impact on the Brigance in Effect Size Units



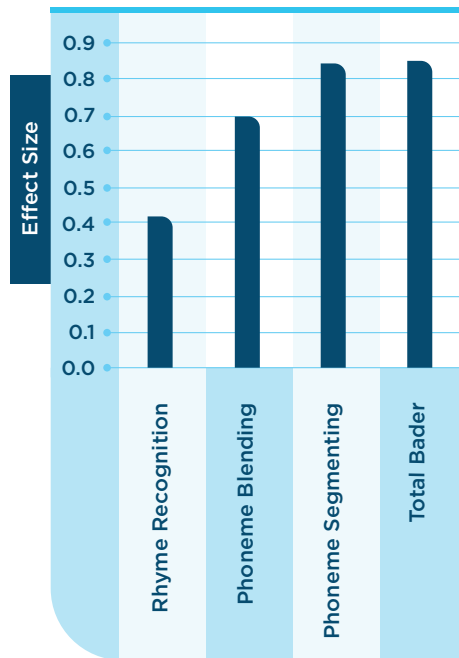
## BADER RESULTS

Table 23: Bader Posttest Analysis of Treatment-Control Group Differences

Bader Posttest	Group	N	Mean	SD	t	Significance
Rhyme Recognition	Treatment	112	8.71	2.14	3.71	**
	Control	120	7.50	2.83		
Phoneme Blending	Treatment	112	6.14	2.61	5.88	**
	Control	120	3.90	3.19		
Phoneme Segmenting	Treatment	112	5.09	3.10	6.29	**
	Control	120	2.59	2.95		
Total Bader	Treatment	112	19.95	6.15	6.89	**
	Control	120	13.99	7.01		

\*\* $p < .01$

Figure 45: UPSTART’s Impact on the Bader in Effect Size Units



### BRIGANCE GROWTH SCORE RESULTS

Table 24: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance Test	Control Group		Treatment Group		Significance <i>p</i> ≤ .01
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Expressive Objects	.476	.196 – .756	.868	.571 – 1.165	NS
Receptive Objects	.115	.018 – .212	.093	.014 – .171	NS
Expressive Grammar	.538	.157 - .919	1.046	.713 – 1.379	NS
Visual Discrimination	2.507	1.555 – 3.459	4.341	3.407 – 5.274	NS
Recites Alphabet	3.307	1.129 -5.486	10.572	8.287 – 12.776	**
Letter Knowledge	12.315	8.720 -15.910	17.333	13.441 – 21.228	NS
Letter Sounds	5.815	4.021 – 7.609	11.100	9.137 – 13.063	**
Auditory Discrimination	.984	.008 – 1.960	3.550	2.664 – 4.435	**
Survival Sight Words	1.530	.951 – 2.110	3.155	2.355 – 3.954	**
Basic Vocabulary	4.923	3.369 – 6.476	11.395	9.406 – 13.384	**
Total Brigance	32.515	26.661 – 38.369	63.410	56.680 – 70.141	**

## BADER GROWTH SCORE RESULTS

Table 25: Treatment-Control Group Differences in Growth Rates on the Bader

<i>Bader Test</i>	<i>Control Group</i>		<i>Treatment Group</i>		Significance <i>p</i> ≤ .01
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Rhyme Recognition	.883	.081 – 1.684	2.883	2.048 – 3.719	**
Phoneme Blending	1.108	.452 – 1.763	4.169	3.406 – 4.932	**
Phoneme Segmenting	1.591	.945 – 2.238	3.973	3.168 – 4.777	**
Total Bader	3.583	2.239 – 4.927	11.026	9.472 – 12.581	**

## YEAR 4

Waterford Institute enrolled 1,250 preschool children in its fourth year of operation during the 2012-13 school year (Evaluation and Training Institute, 2014). The UPSTART treatment group performed significantly better than the control group on both the Total Brigance and Total Bader. Effect sizes range from 0.34 to 0.59. Favoring the UPSTART treatment group, growth rates between the treatment group and the control group were significantly different at the 99% Confidence Interval (CI) for the Total Brigance and the Total Bader.

Preschool children participating in UPSTART improved phonics skills with small to medium effect sizes, as demonstrated on the Overall Brigance Composite as well as on Visual Discrimination, Letter Sounds, and Basic Pre-Primer Vocabulary subtests. Children participating in UPSTART had a 29-point advantage on Brigance posttest scores compared to non-participating children.

## BRIGANCE RESULTS

Table 26: Brigance Posttest Analysis of Treatment-Control Group Differences

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>Significance</i>
Expressive Objects	Treatment	101	25.980	.979	-0.138	NS
	Control	102	26.000	1.062		
Receptive Objects	Treatment	101	26.990	.099	1.000	NS
	Control	102	26.971	.169		
Expressive Grammar	Treatment	101	10.188	1.230	1.522	NS
	Control	102	9.922	1.264		
Visual Discrimination	Treatment	101	18.139	2.069	3.294	**
	Control	102	16.912	3.134		
Recites Alphabet	Treatment	101	18.891	8.956	1.647	NS
	Control	102	16.706	9.931		
Lowercase Letter Knowledge	Treatment	101	39.337	16.638	1.277	NS
	Control	102	36.176	18.566		
Sounds of Lowercase Letters	Treatment	101	17.198	8.918	2.643	**
	Control	102	13.588	10.486		
Auditory Discrimination	Treatment	101	7.594	3.076	1.327	NS
	Control	102	7.000	3.297		

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>Significance</i>
Survival Sight Words	Treatment	101	3.703	3.968	1.330	NS
	Control	102	3.020	3.318		
Basic Pre-Primer Vocabulary	Treatment	101	10.000	9.583	3.815	**
	Control	102	5.284	7.942		
Total Brigance	Treatment	101	178.019	42.974	2.669	**
	Control	102	161.578	44.778		

\* $p < .05$ , \*\* $p < .01$

Table 27: *Brigance Effect Size Estimates*

Test	Effect Size	Significance	Magnitude of Effect
Expressive Objects	-0.018	NS	-
Receptive Objects	0.111	NS	-
Expressive Grammar	0.210	NS	-
Visual Discrimination	0.391	**	Small
Recites Alphabet	0.220	NS	-
Letter Knowledge	0.170	NS	-
Sounds of Lowercase Letters	0.344	**	Small
Auditory Discrimination	0.180	NS	-
Survival Sight Words	0.205	NS	-
Basic Pre-Primer Vocabulary	0.593	**	Medium
Total Brigance	0.367	**	Small

\* $p < .05$ , \*\* $p < .01$

## BADER RESULTS

Table 28: *Bader Posttest Analysis of Treatment-Control Group Differences*

Bader Posttest	Group	N	Mean	SD	t	Significance
Rhyme Recognition	Treatment	79	7.86	2.53	1.27	NS
	Control	93	7.31	3.16		
Phoneme Blending	Treatment	79	5.09	2.96	3.35	**
	Control	93	3.48	3.32		
Phoneme Segmenting	Treatment	79	3.98	3.42	3.45	**
	Control	93	2.24	3.13		
Total Bader	Treatment	79	16.92	6.90	3.52	**
	Control	93	13.03	7.70		

\* $p < .05$ , \*\* $p < .01$

Table 29: Bader Effect Size Estimates

Test	Effect Size	Significance	Magnitude of Effect
Rhyme Recognition	0.174	NS	-
Phonemic Blending	0.484	**	Small
Phoneme Segmentation	0.555	**	Medium
Total Bader	0.519	**	Medium

\* $p < .05$ , \*\* $p < .01$

## BRIGANCE GROWTH SCORE RESULTS

Table 30: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance Test	<u>Control Group</u>		<u>Treatment Group</u>		Significance
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Expressive Objects	.706	.301 – 1.111	1.485	.522 – 2.448	NS
Receptive Objects	.235	-.092 – .563	.168	.030 – .307	NS
Expressive Grammar	.892	.490 – 1.294	1.287	.848 – 1.726	NS
Visual Discrimination	3.431	2.206 – 4.657	4.624	3.447 – 5.800	NS
Recites Alphabet	4.216	1.605 – 6.827	10.267	7.848 – 12.687	**
Letter Knowledge	13.775	9.558 – 17.992	23.332	13.997 – 28.666	NS
Letter Sounds	6.029	3.964 – 8.105	12.693	10.495 – 14.892	**
Auditory Discrimination	1.324	.112 – 2.535	3.079	2.021 – 4.138	NS
Survival Sight Words	1.471	.850 – 2.091	2.475	1.667 – 3.284	**
Basic Vocabulary	3.971	2.351 – 5.591	8.921	6.608 – 11.234	**
Total Brigance	36.049	28.737 – 43.361	68.832	60.294 – 77.369	**

\* $p < .05$ , \*\* $p < .01$

## BADER GROWTH SCORE RESULTS

Table 31: Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	<u>Control Group</u>		<u>Treatment Group</u>		Significance $p \leq .01$
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Rhyme Recognition	1.086	.063 – 2.110	2.620	1.632 – 3.608	NS
Phoneme Blending	1.172	.449 – 1.900	3.987	3.138 – 4.836	**
Phoneme Segmenting	1.301	.567 – 2.046	3.570	2.585 – 4.555	**
Total Bader	3.559	1.927 – 5.191	10.177	8.332 – 12.022	**

\* $p < .05$ , \*\* $p < .01$



## YEAR 5

Waterford Institute enrolled 1,577 preschool children in its fifth year of operation during the 2013-14 school year (Evaluation and Training Institute, 2015). The UPSTART treatment group performed significantly better than the control group on the Total Brigance and Total Bader. Effect sizes range from 0.27 to 0.85. Favoring the UPSTART treatment group, growth rates between the UPSTART treatment group and the control group were significantly different at the 99% Confidence Interval (CI) for three of the Brigance subtests, and for the Total Bader and for two of the Bader subtests: Phoneme Blending and Phoneme Segmenting.

Children participating in UPSTART demonstrated small overall improvements in phonics skills when assessed on the Brigance, indicated by a 12-point advantage on the Overall Brigance Composite compared to control children. Additionally, large effects were found for Pre-Primer Vocabulary. Participation in UPSTART was associated with significant improvement in two of the three phonological awareness strands of the Bader, including Phoneme Blending and Phoneme Segmenting.

### BRIGANCE RESULTS

*Table 32: Brigance Posttest Analysis of Treatment-Control Group Differences*

Brigance Posttest	Group	N	Mean	SD	t	Significance
Expressive Objects	Treatment	94	25.989	1.406	0.568	NS
	Control	100	25.900	1.049		
Receptive Objects	Treatment	94	26.968	0.176	-0.077	NS
	Control	100	26.970	0.171		
Expressive Grammar	Treatment	94	9.787	1.860	-0.384	NS
	Control	100	9.880	1.492		
Visual Discrimination	Treatment	94	18.617	1.814	4.478	**
	Control	100	17.090	2.850		
Recites Alphabet	Treatment	94	19.809	8.425	2.252	*
	Control	100	16.850	9.853		
Lowercase Letter Knowledge	Treatment	94	42.809	13.876	2.387	*
	Control	100	37.430	17.403		
Sounds of Lowercase Letters	Treatment	94	17.585	9.367	3.194	**
	Control	100	13.040	10.445		
Auditory Discrimination	Treatment	94	8.309	2.813	3.788	**
	Control	100	6.640	3.313		
Survival Sight Words	Treatment	94	3.670	3.900	1.544	NS
	Control	100	2.890	3.058		
Basic Pre-Primer Vocabulary	Treatment	94	11.000	9.727	5.037	**
	Control	100	4.730	7.368		
Total Brigance	Treatment	94	169.190	40.066	1.482	NS
	Control	100	160.490	41.592		

\* $p < .05$ , \*\* $p < .01$

Table 33: Brigance Effect Size Estimates

Brigance Posttest	Effect Size	Significance	Magnitude of Effect
Expressive Objects	0.085	NS	-
Receptive Objects	-0.111	NS	-
Expressive Grammar	-0.062	NS	-
Visual Discrimination	0.535	**	Medium
Recites Alphabet	0.300	*	Small
Letter Knowledge	0.309	*	Small
Letter Sounds	0.435	**	Small
Auditory Discrimination	0.504	**	Medium
Survival Sight Words	0.255	NS	(Small)
Basic Pre-Primer Vocabulary	0.850	**	Large
Total Brigance	0.209	NS	(Small)

\* $p < .05$ , \*\* $p < .01$

## BADER RESULTS

Table 34: Bader Posttest Analysis of Treatment-Control Group Differences

Bader Posttest	Group	N	Mean	SD	t	Significance
Rhyme Recognition	Treatment	89	7.96	2.66	2.07	*
	Control	100	7.07	3.22		
Phoneme Blending	Treatment	89	5.21	2.93	4.58	**
	Control	100	3.15	3.27		
Phoneme Segmenting	Treatment	89	4.51	3.24	5.24	**
	Control	100	2.12	3.02		
Total Bader	Treatment	89	17.67	6.67	5.21	**
	Control	100	12.34	7.33		

\* $p < .05$ , \*\* $p < .01$

Table 35: Bader Effect Size Estimates

Bader Posttest	Effect Size	Significance	Magnitude of Effect
Rhyme Recognition	0.27	*	Small
Phonemic Blending	0.63	**	Medium
Phoneme Segmenting	0.79	**	Medium
Total Bader	0.73	**	Medium

\* $p < .05$ , \*\* $p < .01$

## BRIGANCE GROWTH SCORE RESULTS

Table 36: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance Test	Control Group		Treatment Group		Sig
	Mean	99% CI	Mean	99% CI	
	Growth	Growth Rate	Growth	Growth Rate	
Expressive Objects	.940	.461 – 1.419	.723	.404 – 1.043	NS
Receptive Objects	.350	-.105 – .805	.340	.103 – .578	NS
Expressive Grammar	1.040	.572 – 1.508	.851	.340 – 1.362	NS
Visual Discrimination	3.860	2.647 – 5.073	4.872	3.674 – 6.070	NS
Recites Alphabet	5.480	2.863 – 8.097	9.830	6.871 – 12.789	NS
Letter Knowledge	15.720	11.282 – 20.158	23.053	18.313 – 27.793	NS
Letter Sounds	6.780	4.655 – 8.905	11.936	9.526 – 14.346	**
Auditory Discrimination	.820	-.407 – 2.407	2.670	1.645 – 3.695	**
Survival Sight Words	1.230	.685 – 1.775	2.362	1.429 – 3.294	NS
Basic Vocabulary	3.630	2.075 – 5.185	10.394	7.850 – 12.937	**
Total Brigance	38.920	31.753 – 46.087	51.681	43.213 – 60.149	NS

\* $p < .05$ , \*\* $p < .01$

## BADER GROWTH SCORE RESULTS

Table 37: Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	Control Group		Treatment Group		Significance
	Mean	99% CI	Mean	99% CI	
	Growth	Growth Rate	Growth	Growth Rate	
Rhyme Recognition	1.240	.276 – 2.204	1.966	.901 – 3.031	NS
Phoneme Blending	1.290	.597 – 1.983	3.933	3.093 – 4.771	**
Phoneme Segmenting	1.160	.427 – 1.893	3.989	3.064 – 4.914	**
Total Bader	3.690	2.156 – 5.224	9.889	8.075 – 1.670	**

\* $p < .05$ , \*\* $p < .01$

## Evaluation of a School District in South Carolina, 2016-2017

In the following study (Shamir, Pocklington, Feehan, & Yoder, 2018), Waterford Early Learning (WEL), a computer-adaptive curriculum, was administered to kindergarten and first grade students enrolled in a South Carolina public school district during the 2016-2017 school year. Kindergarten students were expected to use WEL for 15 minutes per day, five days per week, and first grade students were expected to use WEL for 30 minutes per day, five days per week. The Developmental Reading Assessment (DRA) was administered at the middle and at the end of the school year. Analysis of gains made in literacy skills from the middle of the school year to the end of the school year indicated a significant positive effect for students in kindergarten and first grade. Analysis of end of year scores while controlling for middle of year scores showed similar evidence of the efficacy of WEL for students in kindergarten and first grade. Examination of available demographics indicated that students of all genders and ethnicities benefitted from WEL.

Table 38: Kindergarten DRA Gains & End of Year Scores Covarying For Middle of Year Scores

	Gains					ANCOVA				
	Experimental		Control		<i>p</i>	Experimental		Control		<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Kindergarten Overall	2.41	1.25	1.73	1.38	.00**	4.84	2.01	4.32	2.00	.01*
<b>Gender</b>										
Male	2.37	1.24	1.75	1.51	.01*	4.82	2.04	4.37	2.16	.07
Female	2.46	1.27	1.67	1.07	.03*	4.86	1.97	4.21	1.62	.07
<b>Ethnicity</b>										
African American	2.40	1.20	1.20	1.32	.00**	4.82	2.11	3.86	1.60	.01**
Hispanic	2.04	1.31	2.00	0.85	.93	4.55	2.32	4.52	0.99	.94

\**p* < .05, \*\**p* < .01

Table 39: First Grade DRA Gains & End of Year Scores Covarying For Middle of Year Scores

	Gains					ANCOVA				
	Experimental		Control		<i>p</i>	Experimental		Control		<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
First Grade Overall	3.53	1.33	2.84	1.63	.00**	11.15	1.96	10.02	3.65	.00**
<b>Gender</b>										
Male	3.59	1.37	2.82	1.72	.00**	11.15	2.08	9.95	3.68	.00**
Female	3.46	1.29	2.86	1.56	.03*	11.14	1.81	10.11	3.65	.00**
<b>Ethnicity</b>										
African American	3.44	1.45	2.60	1.98	.01**	11.04	2.21	9.47	4.18	.00**
Caucasian	3.56	1.30	2.44	1.42	.00**	11.19	1.91	9.89	3.59	.00**

\**p* < .05, \*\**p* < .01

## Evaluation of Waterford Early Learning in Maryland, 2016-2017

The following study (Shamir, Pocklington, Feehan, & Yoder, 2019c) assessed the efficacy of Waterford Early Learning (WEL). Kindergarten, first grade, and second grade students used WEL during the 2016-2017 school year. ANCOVAs examining group differences in end of year scores, covarying for beginning of year scores, between the experimental and control groups were conducted: Students with high usage of WEL significantly outperformed students with low usage on all end of year literacy scores, and students who used WEL significantly outperformed students who did not use WEL on all end of year literacy scores as well. Students who used WEL also outperformed their control counterparts across demographics. These results indicate WEL potentially positively impacts early literacy skills.

## KINDERGARTEN

Figure 46: Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Strand

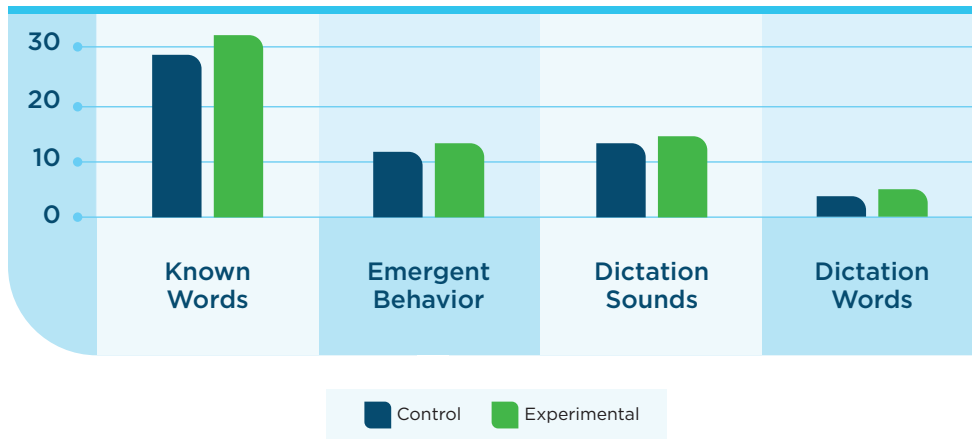


Figure 47: Kindergarten End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics

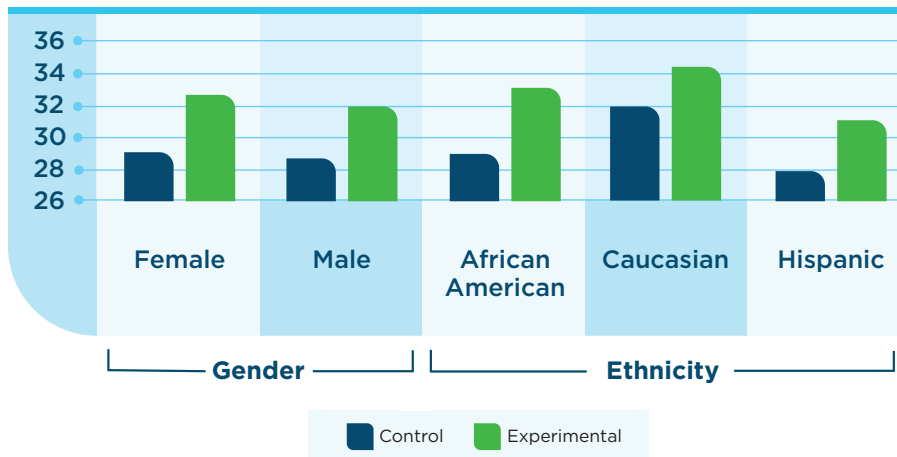


Figure 48: Kindergarten End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics

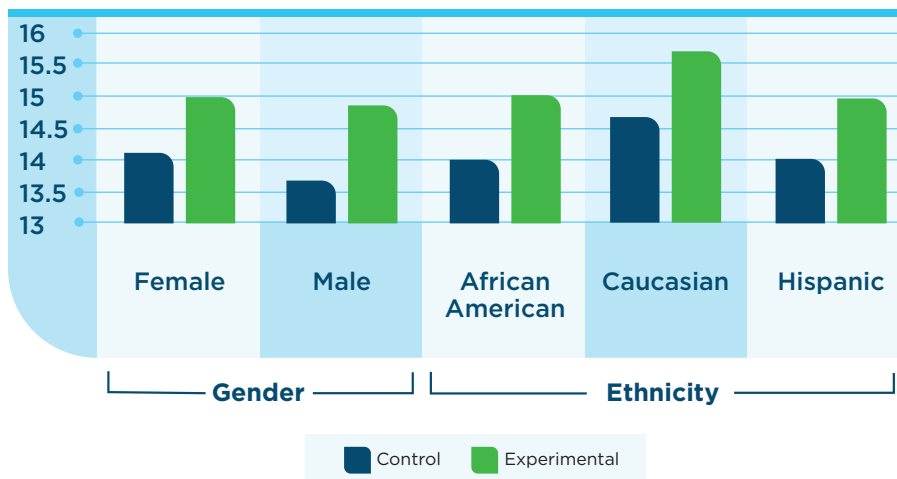


Figure 49: Kindergarten End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

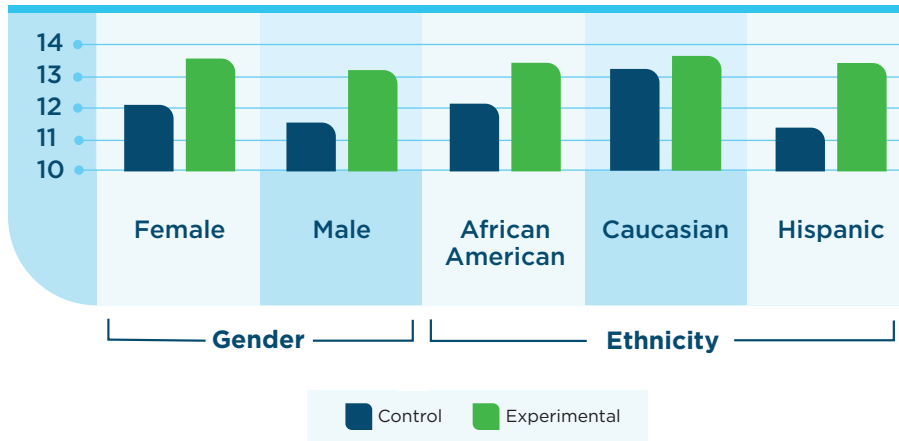
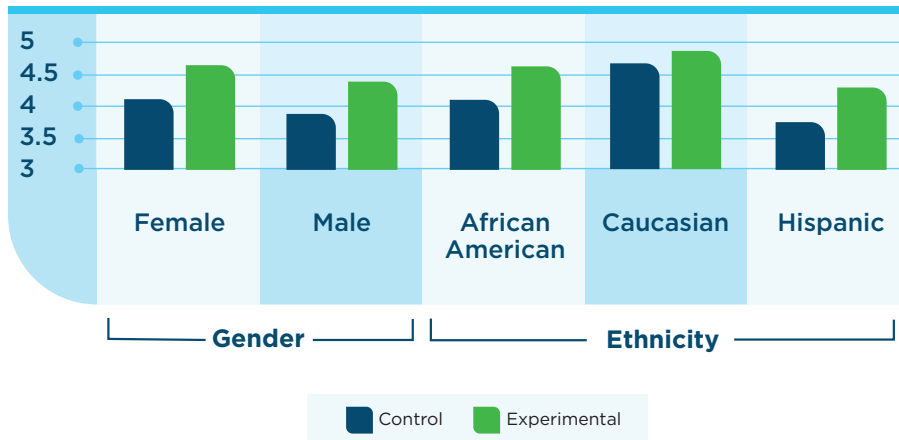


Figure 50: Kindergarten End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



## FIRST GRADE

Figure 51: First Grade End of Year Scores Covarying for Beginning of Year Scores by Strand

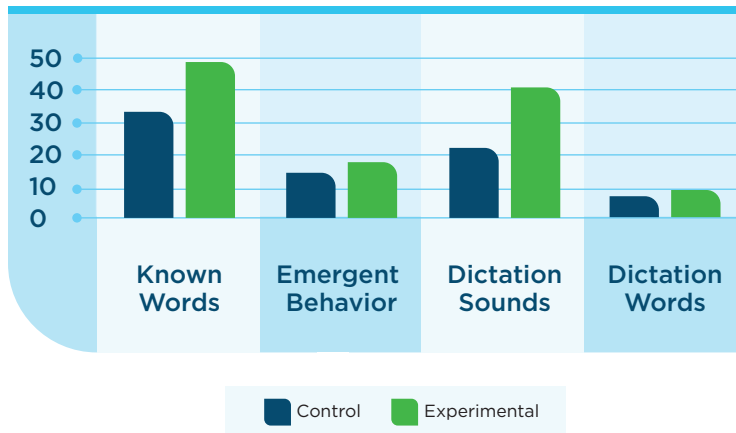


Figure 52: First Grade End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics

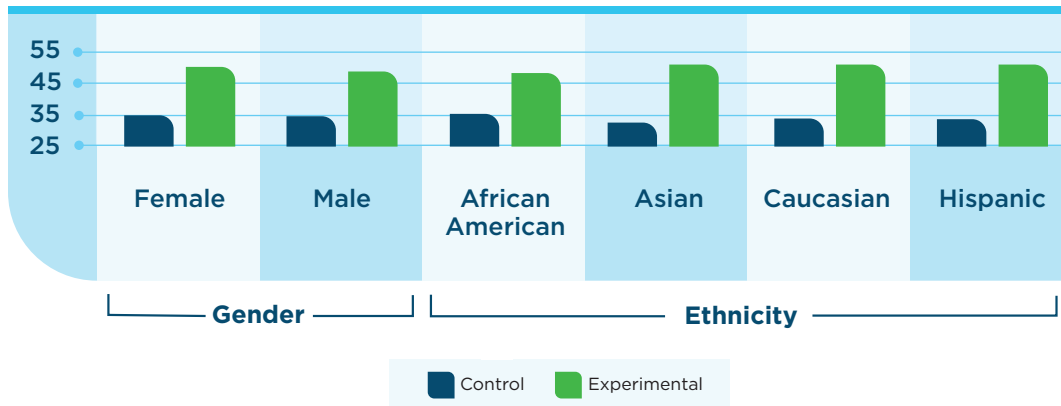


Figure 53: First Grade End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics

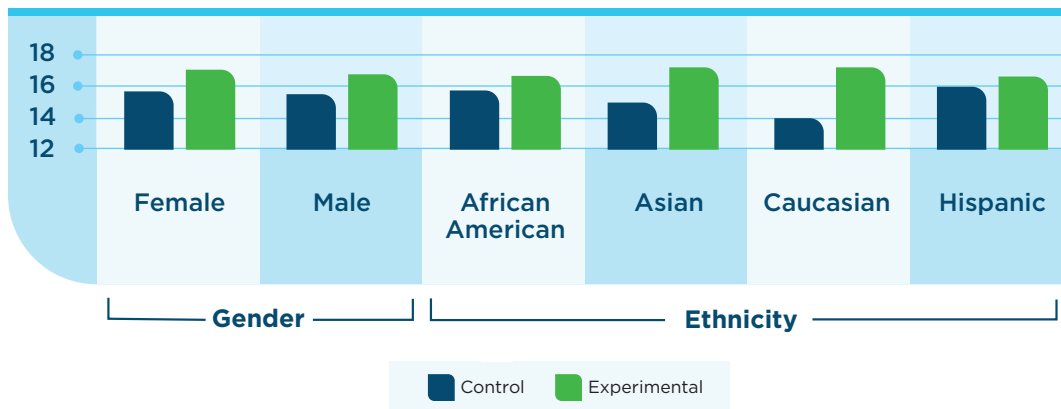


Figure 54: First Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

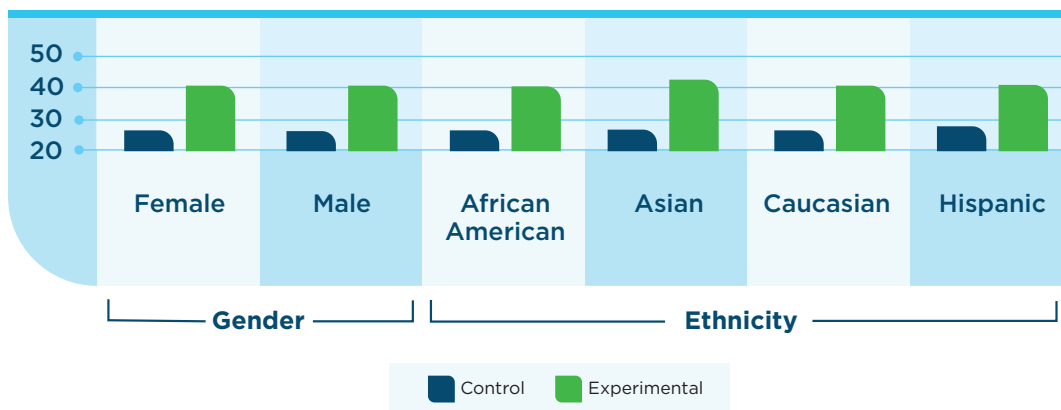
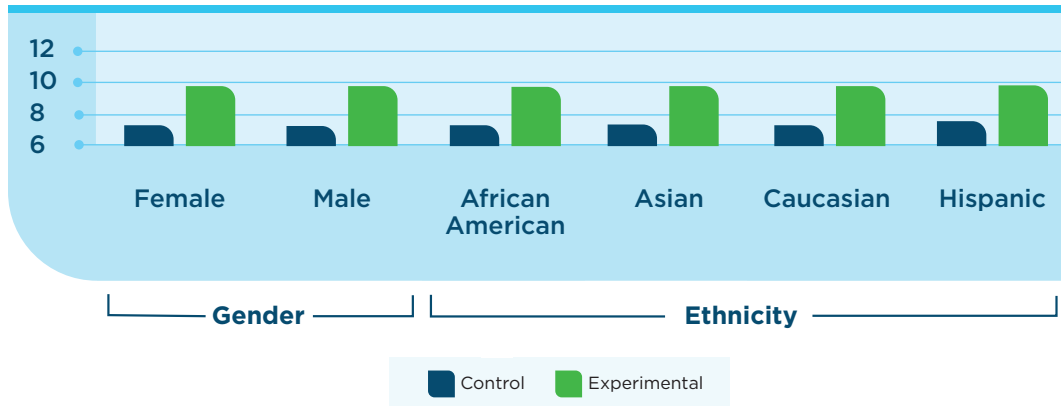


Figure 55: First Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



## SECOND GRADE

Figure 56: Second Grade End of Year Scores Covarying for Beginning of Year Scores by Strand

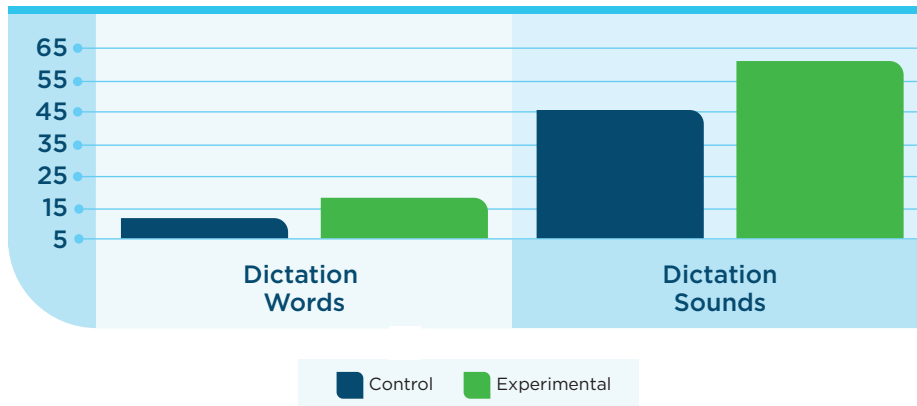


Figure 57: Second Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

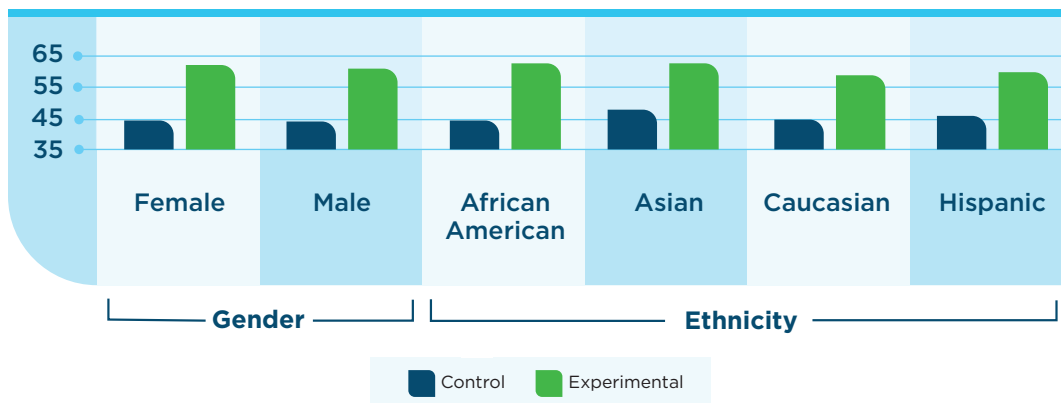
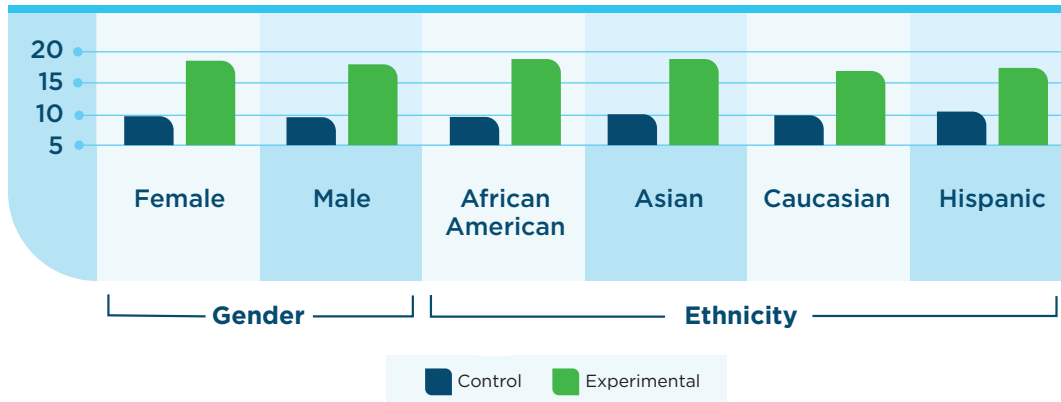




Figure 58: Second Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



## Evaluation of Waterford Early Learning in Tennessee

The following study (Shamir, Pocklington, Feehan, & Yoder, 2019a) assessed the efficacy of Waterford Early Learning (WEL). Kindergarten and first grade students were administered WEL and assessed at the beginning and end of the 2016-2017 school year. Analysis of gains found that students in both grades who used WEL benefited from significantly greater growth for literacy skills than students who did not use WEL. ANCOVAs examining group differences in end of year scores, covarying for beginning of year scores, between the experimental and control groups were conducted: Students who used WEL outperformed their control counterparts on end of year scores despite on most strands having lower beginning of year scores than their control group counterparts.

### KINDERGARTEN

#### GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS

Figure 59: Kindergarten Beginning & End of Year Overall RIT Scores

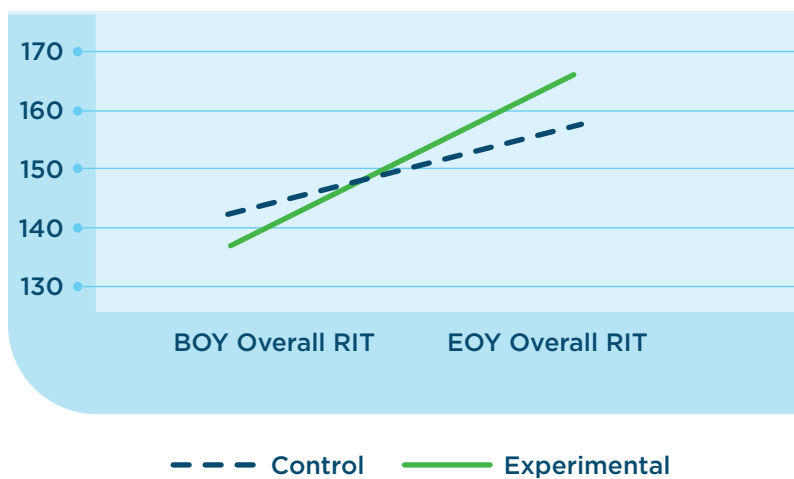
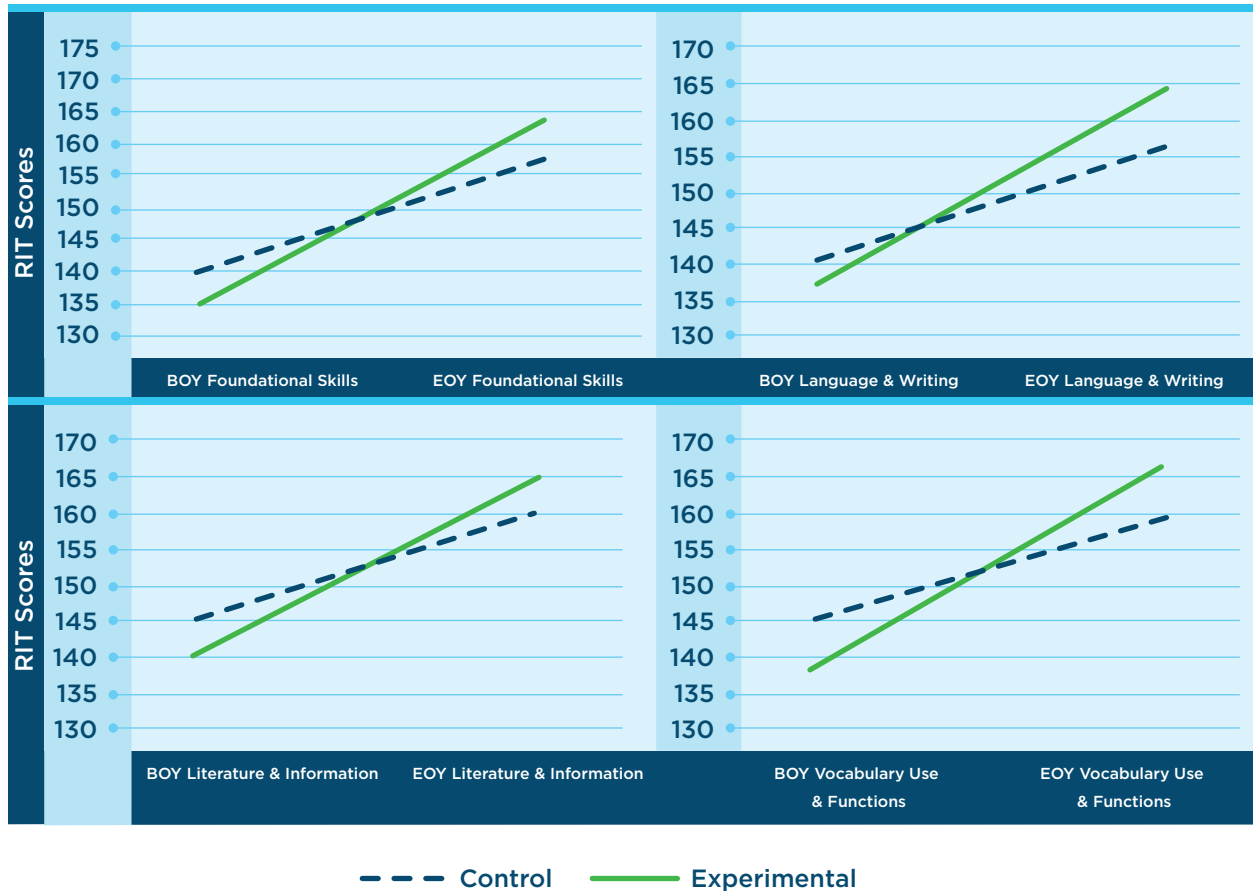
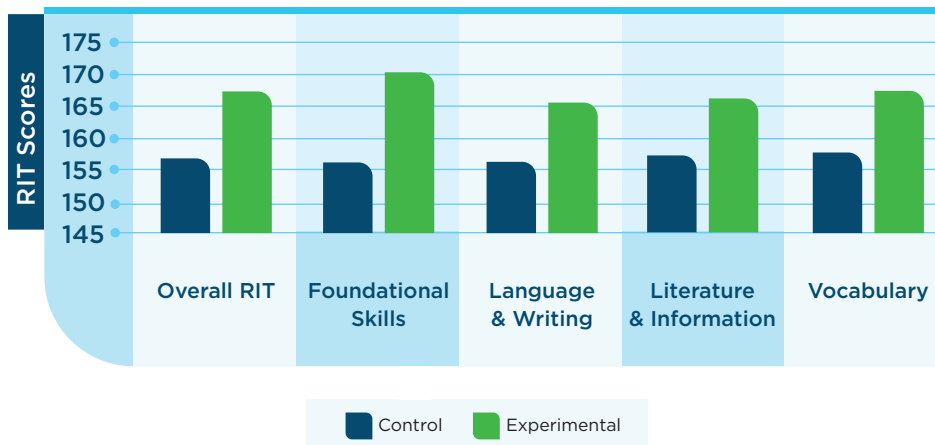


Figure 60: Kindergarten Beginning & End of Year RIT Scores by Strand



GROUP DIFFERENCES USING ANALYSIS OF COVARIANCE (ANCOVA)

Figure 61: Kindergarten End of Year RIT Scores while Covarying for Beginning of Year Scores



# FIRST GRADE

## GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS

Figure 62: First Grade Beginning & End of Year RIT Scores

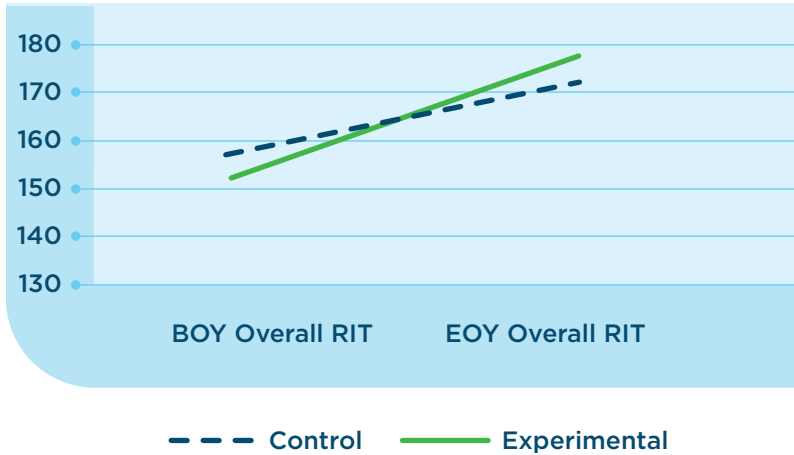
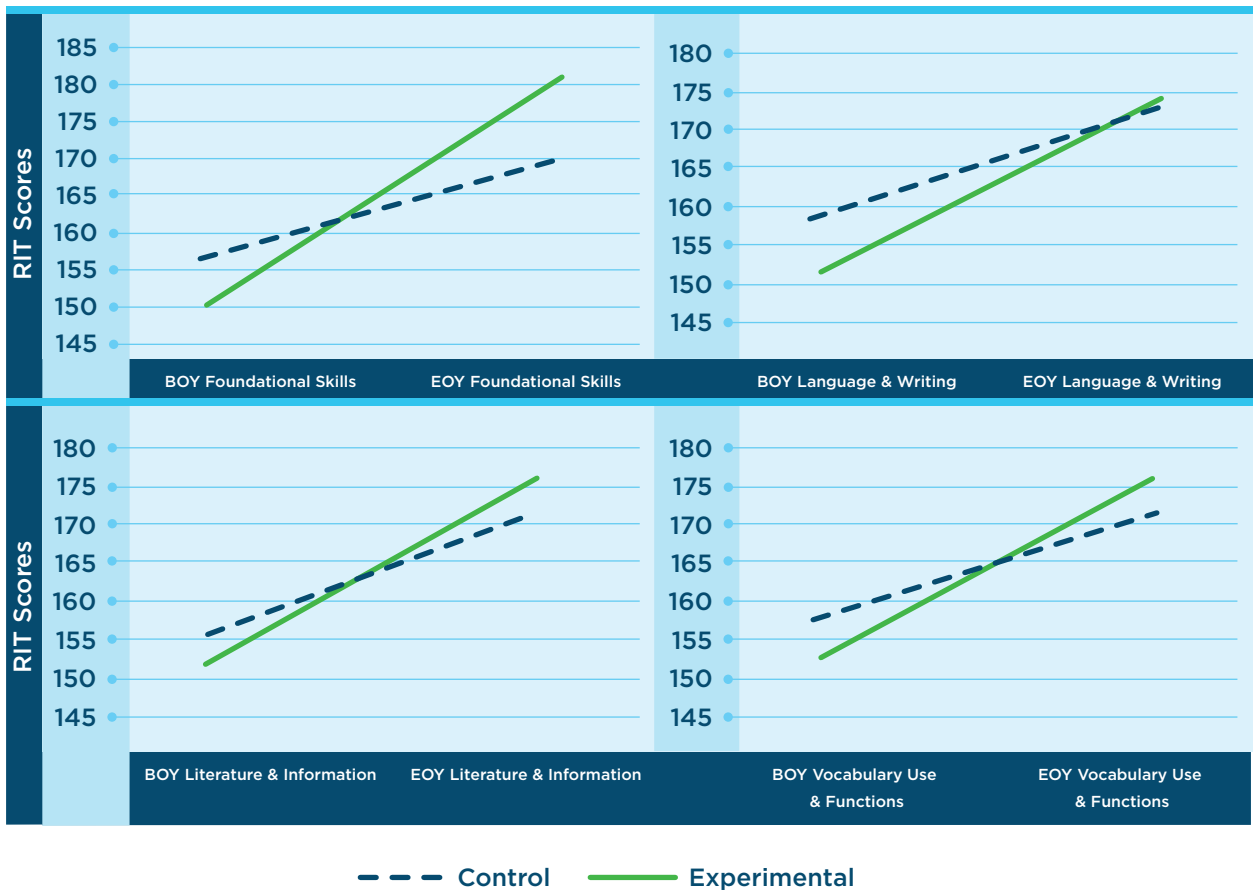
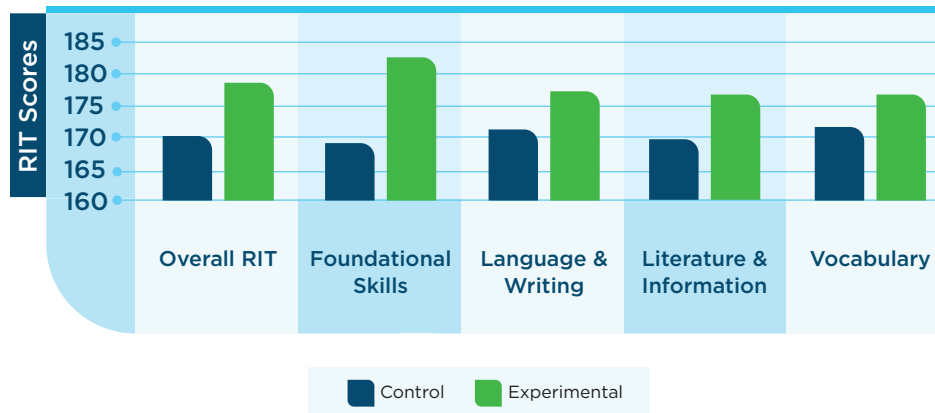


Figure 63: First Grade Beginning & End of Year RIT Scores by Strand



## GROUP DIFFERENCES USING ANALYSIS OF COVARIANCE (ANCOVA)

Figure 64: First Grade End of Year RIT Scores while Covarying for Beginning of Year Scores



## Evaluation of Waterford Early Learning in Maryland, 2017-2018

The following study (Shamir, Pocklington, Yoder, & Feehan, 2019) assessed the efficacy of Waterford Early Learning (WEL) assigned to second grade students in a Maryland school district during the 2017-2018 school year. Students were expected to use WEL for 30 minutes per day, five days per week. The Northwest Evaluation Association’s (NWEA) MAP assessment was administered at the beginning and end of the school year. Independent samples *t*-tests were conducted and revealed the experimental group consistently outperformed the control group on all strands of the MAP assessed, and ANCOVAs to examine group differences in end of year MAP scores between the experimental and control groups while covarying for beginning of year scores were conducted and revealed that students who used WEL significantly outperformed students who did not use WEL on all strands of the MAP assessed.

Figure 65: Second Grade MAP End of Year Scores by Strand

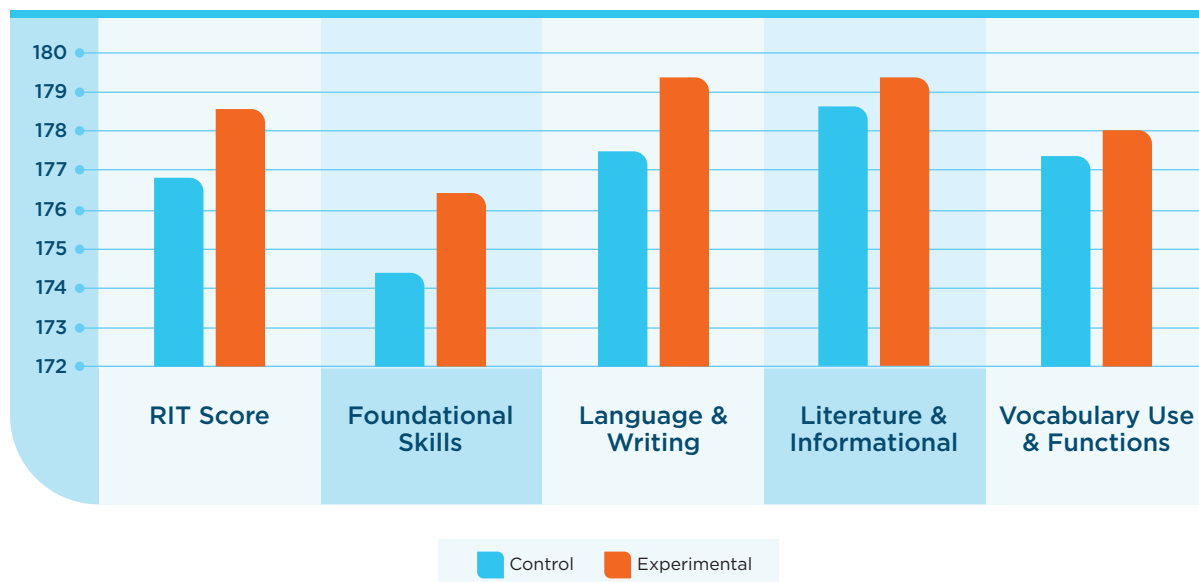


Figure 66: Second Grade MAP End of Year RIT to Reading Scores

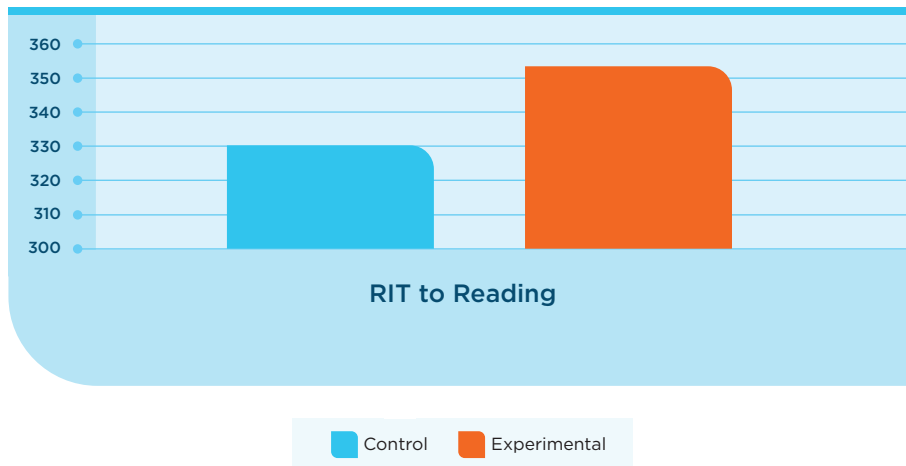


Figure 67: Second Grade MAP End of Year Scores Controlling for Beginning of Year Scores by Strand

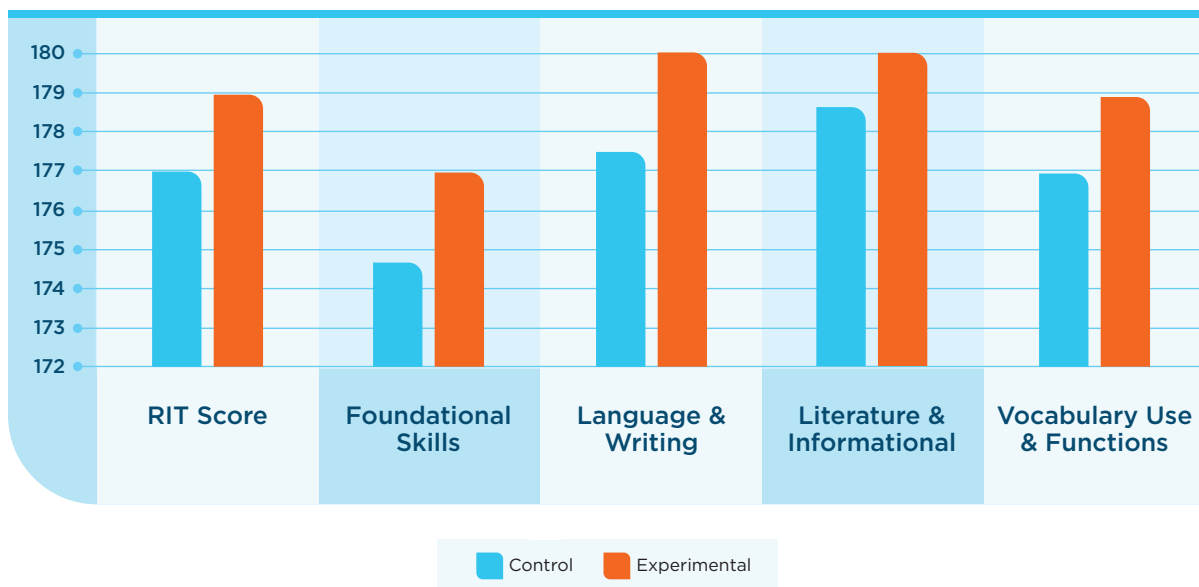
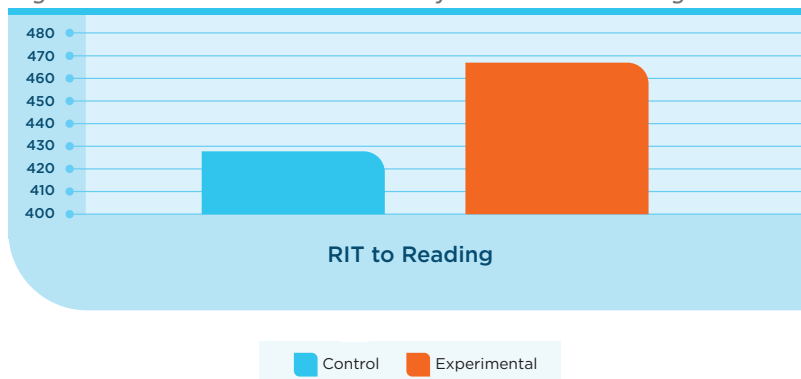


Figure 68: Second Grade MAP End of Year RIT to Reading Scores Controlling for Beginning of Year Scores



## Tier 4: Demonstrated Rationale

For an intervention to demonstrate a rationale, it should have a well-specified logic model that is informed by research or an evaluation suggesting how the intervention is likely to have a positive effect on relevant outcomes. Additionally, there must be an effort to study the effects of the intervention to inform stakeholders about the success of that intervention.

### Early Studies

After developing Level 1, researchers at Waterford Research Institute designed a series of early effectiveness studies using a number of public and private schools in Utah—including Waterford’s own school, then located in Provo—and the New York City Public School System (Waterford Institute, 1996). Testing was performed at the beginning and end of the school year using Waterford’s internal measure, the Waterford Early Reading Instrument (WERI). These initial tests yielded positive results reported in percentages: In every case, classes that used the software made greater gains in pre-literacy skills than comparison classrooms. Waterford students at Timpanogos Elementary, for instance, improved reading test average scores from 50% to 91.8% over the course of the year—compared with score averages of 55% (pretest) to 73% (posttest) among the control group. Each of the New York schools tested included large English Language Learner (ELL) populations. These early results were the first to demonstrate the particular effectiveness of the software for ELL students: In both of the New York schools (PS 43 in the Bronx and PS 1 in Manhattan), where largely ELL classes used the software, these classes performed better on posttests than the English-proficient classes in the control group (60% vs. 47% on the WERI in PS 43 and 85% vs. 68% in PS 1).

### Case Studies

After these initial tests, a number of school districts throughout the country were invited to implement the Waterford software in their kindergarten, first, and second grade classrooms. One of the first of these case studies was conducted during the 1996-1997 school year in the Dallas Independent School District (Waterford Institute, 1997). After a one year trial with 668 kindergarteners, researchers noted highly significant differences ( $p < 0.01$ ) between students using the Waterford program and control classrooms. Results from a study conducted in the Whittier City and Hacienda la Puente Districts in Los Angeles were similar: after a large number of kindergarten students ( $N = 558$ ) used the program for approximately six months, analyses indicated that the average growth scores for Waterford students were significantly higher than those of comparison classes ( $p < .001$ ) (Research, Assessment, & Measurement, Inc., 1999). A third study, commissioned by the State of Ohio in three Columbus-area schools, found significant gains among students using the program in comparison with the control group ( $p < .05$ ) for skills including Letter-Word Identification, Spelling, and Phonological Awareness (Hecht, 2000). In 1997-1998, a study in Newark public schools showed statistically significant differences in favor of the Waterford software student group over the control group ( $p < .02$ ) on the TERA-2 standardized assessment ( $N = 265$ ) (Young & Tracey, 1999). In 1999, administrators in the Decatur Illinois School District implemented the reading software among kindergarten and first grade students during two successive school years ( $N = 700$ ) (Reynolds, 2000). First grade students who used the software significantly outperformed control group students on the Iowa Basic Test of Skills for reading ( $p = .003$ ). After a statewide implementation of the reading software ( $N = 2414$ ) in Idaho kindergartens, evaluators, working in connection with the Albertson Foundation, reported strong evidence from a representative

sample of eight school districts for its effectiveness among academically disadvantaged students (Walberg, 2001). After one year with the program, the effect size for students who had originally tested in the lowest third on standardized reading measures was 1.14, and the overall effect size for students who completed the program was 0.52.

The Los Angeles study also provided another early indication that the software could be particularly effective among ELL populations: the average growth rate for limited English-proficient students was twice that of the English-proficient group in letter recognition and phonological awareness (Electronic Education, 2002). Another study in Maryland (at Glenridge Elementary School) produced similar results: ELL students benefited even more than native speakers from their use of the software, increasing their scores more than 600% (as compared to 283% for the control group) (Research, Assessment, & Measurement, Inc., 1998). Researchers noted that using the software appeared to have significantly reduced the literacy gap between ELL students and other students.

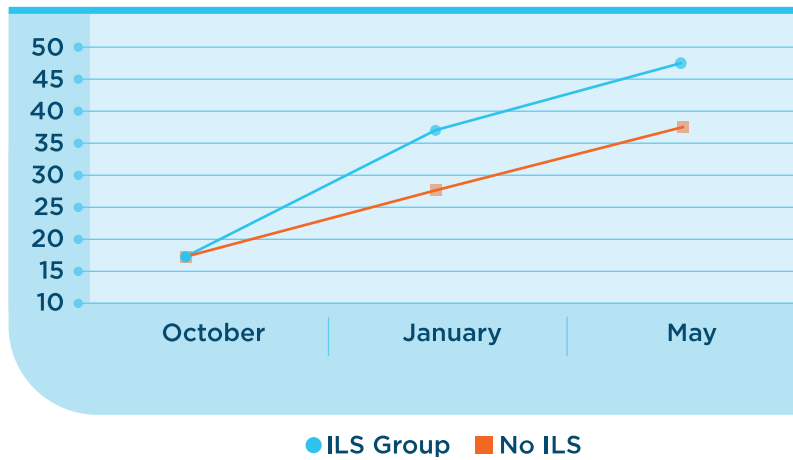
## Later Effectiveness Studies

Cassady and Smith (2003) published the first of their two Waterford-related studies in 2003. An Indiana school implemented Waterford software in its kindergarten classes to work in conjunction with existing literacy instruction (ILS group); the evaluation used the Phonological Abilities Tests (PAT) at the beginning, middle, and end of the trial year to assess student gains for basic literacy skills. Another school in the area, which had not implemented the program at all, served as the control group (No-ILS group). Teachers in both schools, as participants in the Intentional Reading Project (IRP), were engaged in ongoing professional development activities, and both schools received various other resources throughout the year; the researchers were careful to ensure that Waterford software was the principal curricular difference. Despite no significant differences in pretest scores, students using Waterford software experienced a faster acquisition of phonological awareness skills than students who had not used the program,  $F(2, 85) = 3.05, p < .05, \eta^2 = .07$ .

Table 40: Performance Measure Means at the Beginning, Middle, & End of the Year

	October <i>M</i> (SD)	January <i>M</i> (SD)	May <i>M</i> (SD)
	Concepts About Print		
ILS group ( <i>n</i> = 26)	11.69 (3.62)	15.12 (2.81)	17.46 (2.76)
No-ILS ( <i>n</i> = 62)	8.52 (4.74)	13.05 (3.95)	16.24 (3.50)
	Phonological Awareness Test		
ILS group ( <i>n</i> = 26)	17.04 (10.08)	35.85 (12.51)	46.38 (11.24)
No-ILS ( <i>n</i> = 62)	17.14 (11.53)	27.71 (15.57)	37.18 (15.58)

Figure 69: Kindergarten Students' Phonological Awareness Growth



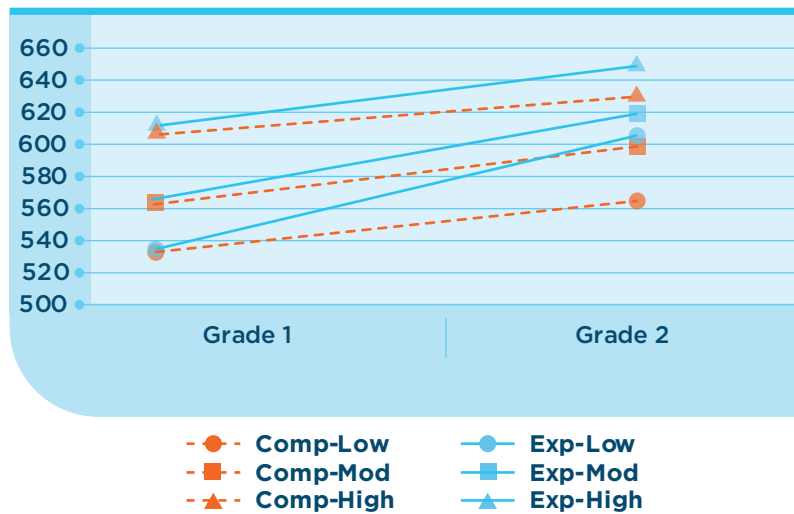
Cassady and Smith (2005) examined the effect on reading achievement gains during the first grade year. Again, students who used the software experienced significantly greater reading skill gains on a standardized test (the CTBS Terra Nova) than the comparison group,  $F(1, 91) = 10.61, p < .01, \eta^2 = .10$ . Researchers also noted that it was the lowest-performing students who benefited most from the program; these students dramatically outperformed the low-performing comparison group ( $F[1, 21] = 15.67, p < .01, \eta^2 = .43$ ). By the end of the first grade year, test scores among this at-risk group were equivalent to those of the moderate-performing students in the comparison group.

Table 41: Beginning of Year First Grade & Second Grade Terra Nova Reading Performance Group Means

	Low Performance		Moderate Performance		High Performance	
	Grade 1	Grade 2	Grade 1	Grade 2	Grade 1	Grade 2
Comparison	535.82 (10.25) <i>n</i> = 11	561.82 (27.77)	559.62 (8.14) <i>n</i> = 26	599.35 (32.14)	612.00 (26.68) <i>n</i> = 10	633.40 (33.46)
Experimental	535.75 (16.09) <i>n</i> = 12	610.83 (26.29)	561.35 (8.93) <i>n</i> = 20	617.95 (36.67)	608.50 (22.06) <i>n</i> = 14	649.71 (38.33)



Figure 70: Reading Gains for Students with Low, Moderate, & High Reading Skills at First Grade



## Evaluation of the Waterford Early Reading Program in Alabama

The following study investigates the impact of adding the Waterford Early Reading Program (ERP) to an existing kindergarten and first grade curriculum and its impact on reading scores (Shamir, Feehan, & Yoder, 2017a). Participants were enrolled in kindergarten and first grade classes from a school district in Alabama in which over 90% of the students were African Americans from economically disadvantaged homes.

Three sample groups were used in the study. The treatment groups for all samples used ERP software; the control groups for all samples did not use ERP software. Students in the first and second sample groups were administered the STAR (Standardized Test for the Assessment of Reading) Early Literacy assessments. Students in the third sample group were administered the DIBELS (Dynamic Indicators of Early Literacy Skills) reading assessment.

On all measured skills, students who used ERP outperformed the control groups. In the first sample group, the treatment group significantly outperformed the control group on six of the ten sub-strands: Phonemic Awareness (PA), Phonics (PH), Structural Analysis (SA), Vocabulary (VO), Sentence-Level Comprehension (SC), and Paragraph-Level Comprehension (PC). In the second sample group, the treatment group significantly outperformed the control group on seven of the ten sub-strands: Phonemic Awareness (PA), Phonics (PH), Structural Analysis (SA), Vocabulary (VO), Sentence-Level Comprehension (SC), Paragraph-Level Comprehension (PC), and Early Numeracy (EN). In the third sample group, the treatment group significantly outperformed the control group on two of the three sub-strands: Letter Naming Fluency (LNF) and Nonsense Words Fluency-Correct Letter Sounds (NWF-CLS).

Figure 71: Mean Gains by STAR Sub-Strands for First Sample Group

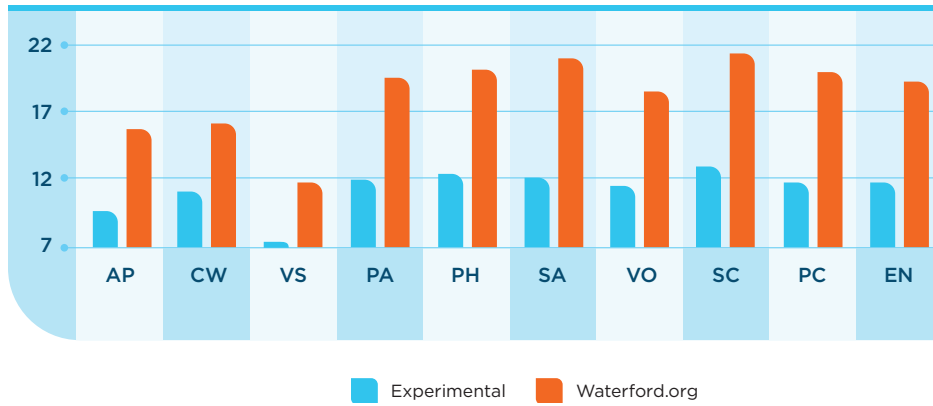


Figure 72: Mean Gains by STAR Sub-Strands for Second Sample Group

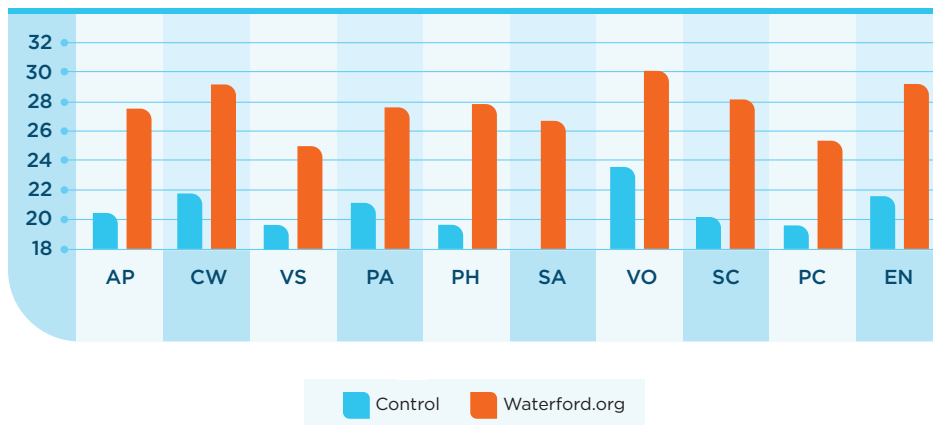
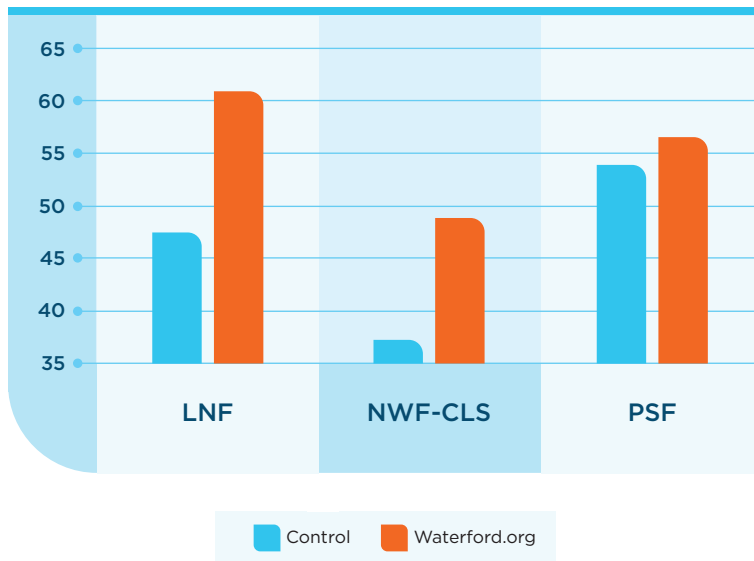


Figure 73: Mean Posttest Scores for DIBELS Sub-Strands for Third Sample Group

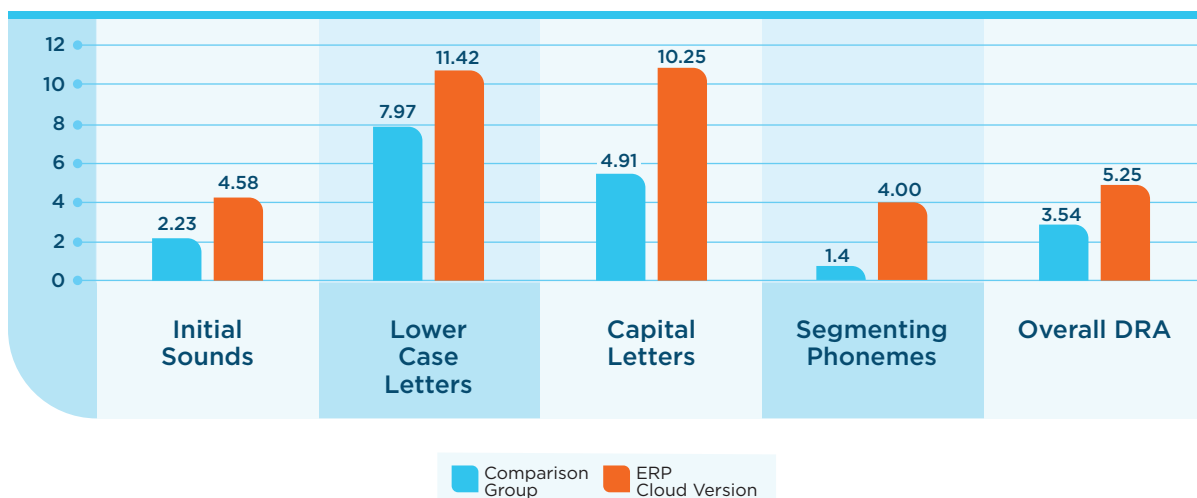


## Evaluation of the Waterford Early Reading Program in Colorado

The following study investigates the impact of adding Waterford Early Reading Program (ERP) to an existing kindergarten and first grade curriculum and its impact on reading scores. The students were administered the DRA (Developmental Reading Assessment) (Shamir & Goethe, 2015).

T-tests were conducted to examine gain differences between the groups. The students using ERP Cloud Version significantly outperformed the comparison group on three of the four sub-strands: Initial Sounds, Capital Letters, and Segmenting Phonemes.

Figure 74: Mean Reading Gains by DRA Overall Kindergarten Scores



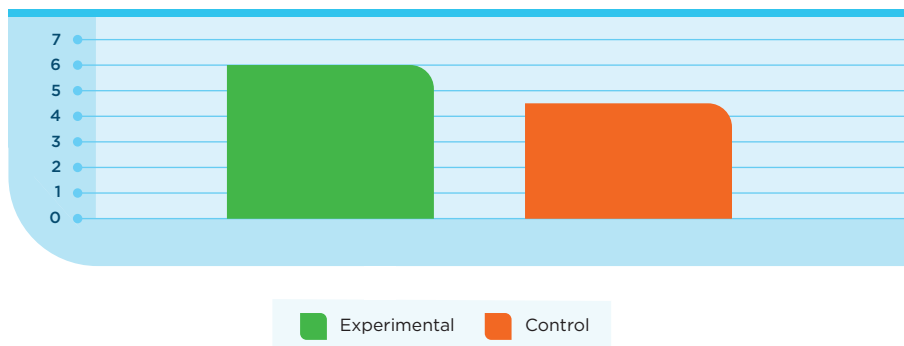
## Evaluation of the Waterford Early Reading Program in South Carolina

The following study (Shamir, Feehan, & Yoder, 2017c) details a recent investigation of the impact of adding the Waterford Early Reading Program (ERP) to an existing school curriculum and its impact on students' literacy scores.

Kindergarten and first grade students were administered the Developmental Reading Assessment (DRA), a standardized assessment during the fall, winter, and spring of the 2015-2016 school year. The experimental group for kindergarten ( $N = 1,004$ ) included students that used Waterford curriculum for more than 1,000 minutes throughout the 2015-2016 school year; the control group ( $N = 28$ ) included students that used Waterford curriculum for less than 500 minutes throughout the 2015-2016 school year. The experimental group for first grade ( $N = 1,064$ ) included students that used Waterford curriculum for more than 1,000 minutes throughout the 2015-2016 school year; the control group ( $N = 52$ ) included students that used Waterford curriculum for less than 500 minutes throughout the 2015-2016 school year.

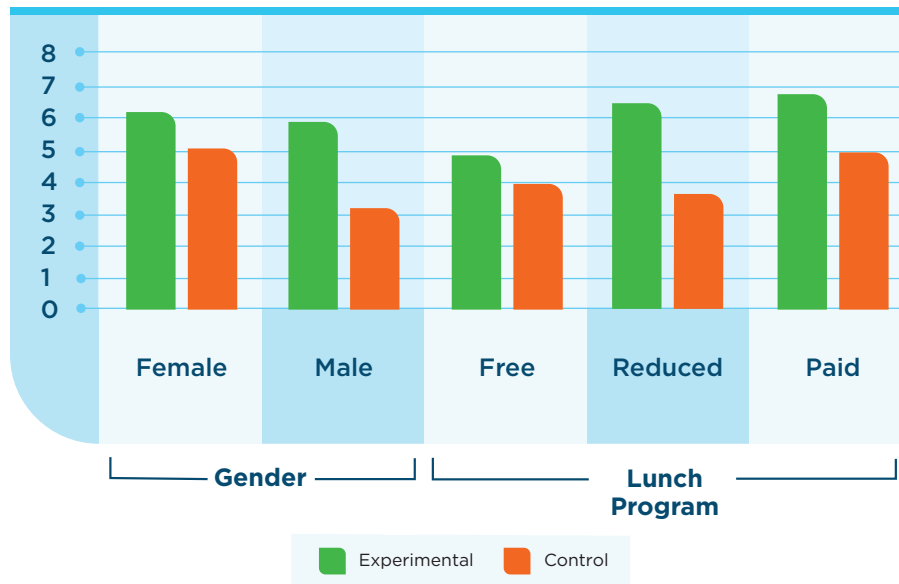
An independent samples  $t$ -test examining group differences in DRA kindergarten end of year scores between the experimental group and the control group was conducted. Analysis of end of year scores revealed a significant difference between groups due to higher end of year scores made by experimental students than by control students.

Figure 75: Kindergarten DRA End of Year Scores



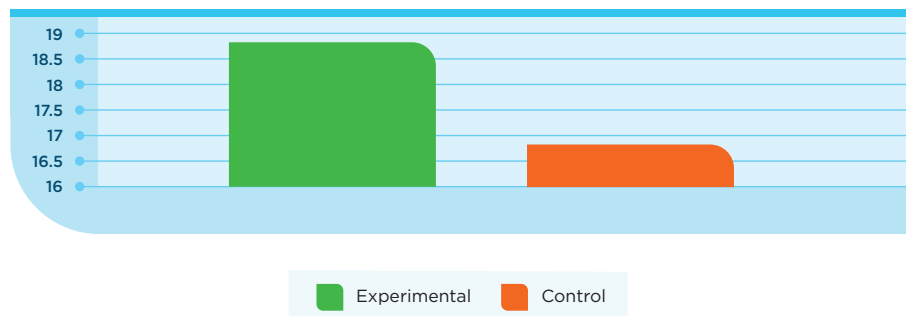
Further analysis was conducted to examine the effects of gender and subsidized lunch on DRA end of year scores, covarying for DRA beginning of year scores. Two separate two-way ANCOVAs were conducted to examine the effect of Waterford curriculum and demographics on DRA end of year scores, covarying for beginning of year DRA scores.

Figure 76: Kindergarten DRA End of Year Scores by Demographics



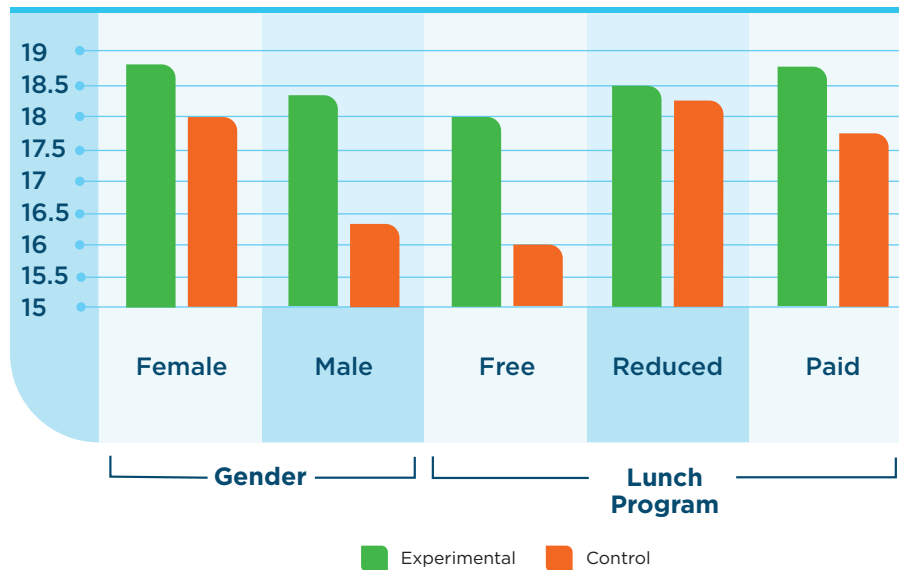
For first grade, analysis of DRA end of year scores, while covarying for DRA beginning of year scores, revealed a significant difference between groups due to higher end of year scores made by experimental students than by control students.

Figure 77: First Grade DRA End of Year Scores



Further analysis was conducted to examine the effects of gender and subsidized lunch on DRA end of year scores, covarying for DRA beginning of year scores. Two separate two-way ANCOVAs were conducted to examine the effect of Waterford curriculum and demographics on DRA end of year scores, covarying for beginning of year DRA scores.

Figure 78: First Grade DRA End of Year Scores by Demographics

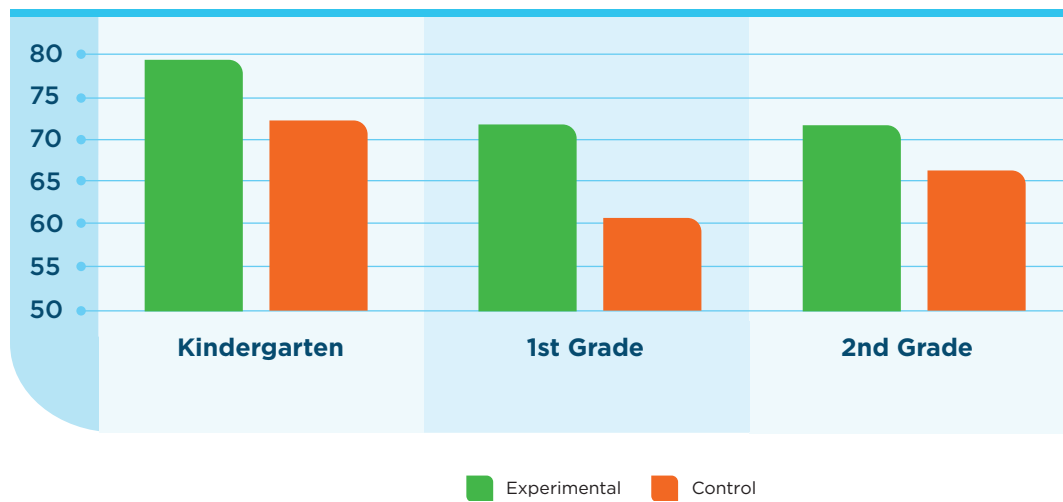


## Evaluation of the Waterford Early Reading Program in Florida

The current study (Shamir, Feehan, & Yoder, 2017d) assessed the impact of augmenting an existing curriculum with the *Waterford Early Reading Program* (ERP) on the reading ability of students in Kindergarten through second grade. One thousand three hundred thirty Kindergarten students, 1,926 first grade students, and 2,259 second grade students used Waterford curriculum throughout the 2015-2016 school year. The VLT was administered to kindergarten students at the end of the third and fourth quarters of the 2015-2016 school year and to the first and second grade students four times throughout the 2015-2016 school year, at the end of each quarter. Students were assessed using high and low usage groups. The experimental group for kindergarten included students with more than 1,000 minutes of usage throughout the school year, and the control group included students with less than 500 minutes of usage. For first grade, the experimental group included students with more than 1,000 minutes of usage, and the control group included students with less than 300 minutes. For second grade, the experimental group included students with more than 1,200 minutes, and the control group included students with less than 500 minutes.

Independent sample *t*-tests showed statistically significant positive effects for high use of Waterford on the end of year VLT scores of Kindergarten students,  $t(1, 1328) = -1.97, p < .05$ , first grade,  $t(1, 1924) = -3.14, p < .01$ , and second grade,  $t(1, 2257) = -2.57, p < .05$ .

Figure 79: End of Year VLT Scores by Grade



## Evaluation of the Waterford Early Reading Program in Indiana, 2015-2016

The following study (Shamir, Feehan, & Yoder, 2017e) assessed the efficacy of the Waterford Early Reading Program (ERP), a computer-adaptive program that was assigned to first and second grade students in a school district in Indiana during the 2015-2016 school year. Students in the experimental group were expected to use ERP for 30 minutes per day, five days per week, throughout the 2015-2016 school year. The control group consisted of students who did not use ERP during the 2015-2016 school year. The Developmental Reading Assessment (DRA) was administered at the beginning and end of the year.

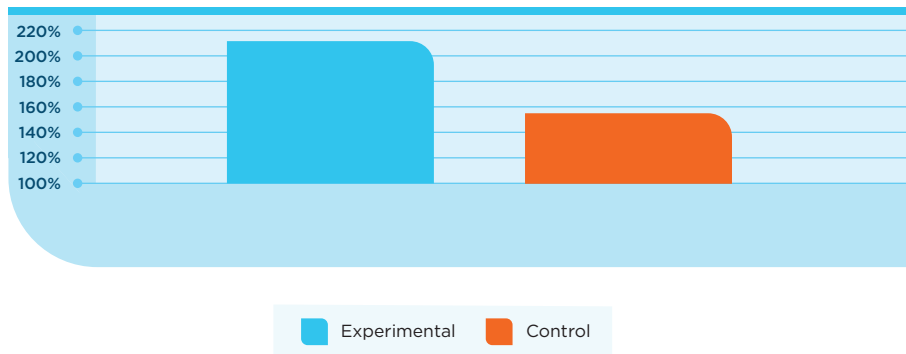
### FIRST GRADE

The experimental group for first grade ( $N = 103$ ) consisted of students that used the Waterford Early Reading Program during the 2015-2016 school year. The control group ( $N = 534$ ) consisted of students that did not use the Waterford Early Reading Program during the 2015-2016 school year.

#### GROUP DIFFERENCES USING AN INDEPENDENT SAMPLES T-TEST

An independent samples  $t$ -test examining percent gains between beginning of year and end of year scores was conducted. Analysis of percent gains revealed a significant difference between groups  $t(1, 635) = -3.44, p < .01$ , due to higher percent gains made by students who used Waterford ( $M = 205.78$ ) than by control students ( $M = 156.06$ ). Effect size ( $d = 0.37$ ).

Figure 80: First Grade DRA Percent Gains



### GROUP DIFFERENCES BY DEMOGRAPHICS USING TWO-WAY ANOVAS

Further analysis was conducted to examine the effects of gender, lunch program, and special education status on DRA percent gains. Three separate ANOVAs were conducted that examined the effects of Waterford curriculum and demographics on DRA percent gains.

#### **Gender**

There was no significant interaction between the effects of gender and Waterford curriculum on DRA percent gains,  $F(1, 632) = 1.27, p = .261$ . Simple effects analysis showed that for females, students' percent gains in the experimental group were significantly higher than in the control group. Male students' percent gains in the experimental group were higher than in the control group, approaching significance.

#### **Lunch Program**

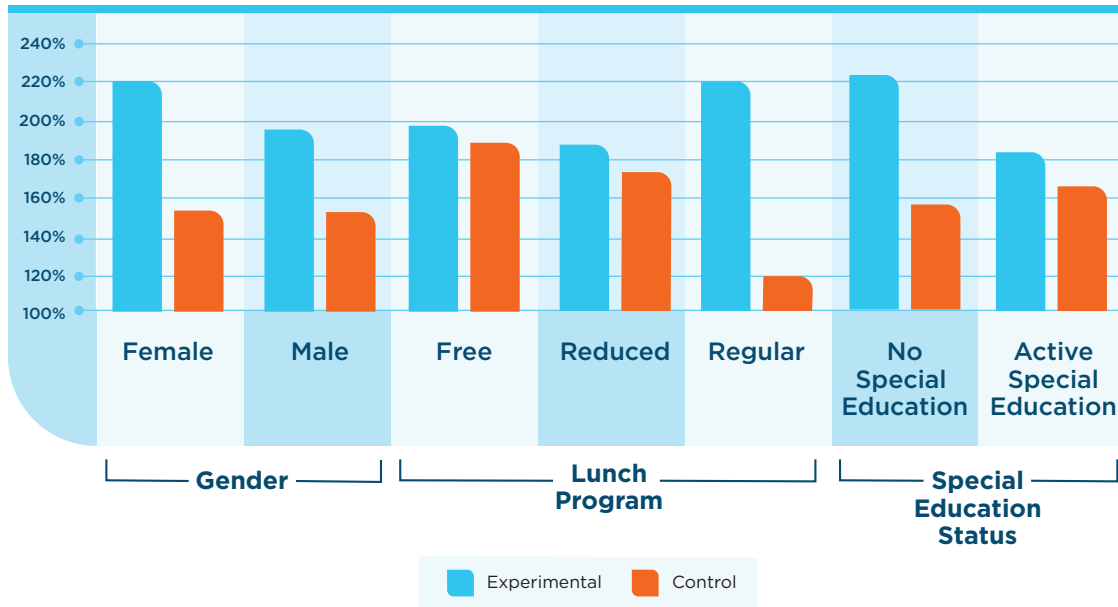
There was a significant interaction between the effects of lunch program and Waterford curriculum on DRA percent gains,  $F(2, 630) = 4.71, p < .01$ . Simple effects analysis showed that regular lunch students' percent gains in the experimental group were significantly higher than in the control group. Free lunch and reduced lunch students' percent gains in the experimental group were slightly higher than in the control group, but the difference was not significant.

#### **Special Education Status**

There was no significant interaction between the effects of special education status and Waterford curriculum on DRA percent gains,  $F(1, 631) = 2.49, p = .115$ . Simple effects analysis showed that students with no special education status in the experimental group made significantly higher percent gains than the control group. Students with an active special education status in the experimental group made slightly higher percent gains than the control group, but the difference was not significant.



Figure 81: First Grade DRA Percent Gains by Demographics



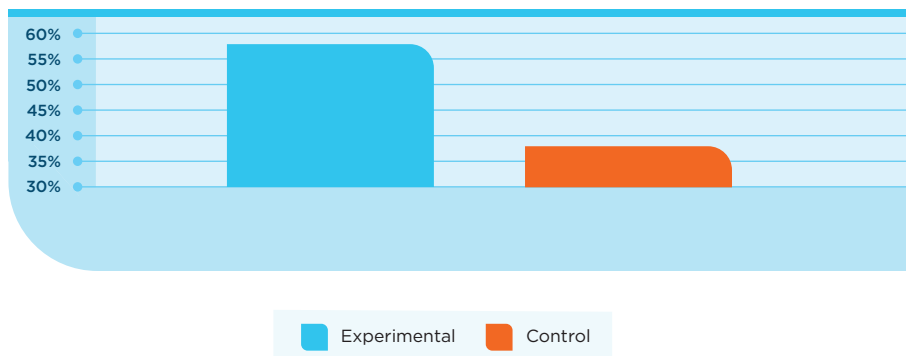
## SECOND GRADE

The experimental group for second grade ( $N = 70$ ) consisted of students that used the Waterford Early Reading Program during the 2015-2016 school year. The control group ( $N = 407$ ) consisted of students that did not use the Waterford Early Reading Program during the 2015-2016 school year.

### GROUP DIFFERENCES USING AN INDEPENDENT SAMPLES T-TEST

An independent samples  $t$ -test examining percent gains between beginning of year and end of year scores was conducted. Analysis of percent gains revealed a significant difference between groups  $t(1, 475) = -5.44, p < .01$ , due to higher percent gains made by students who used Waterford ( $M = 56.32$ ) than by control students ( $M = 37.70$ ). Effect size ( $d = 0.70$ ).

Figure 82: Second Grade DRA Percent Gains



## GROUP DIFFERENCES BY DEMOGRAPHICS USING TWO-WAY ANOVAS

Further analysis was conducted to examine the effects of gender, ethnicity, lunch program, LEP status, and special education status on DRA percent gains. Five separate ANOVAs were conducted that examined the effects of Waterford curriculum and demographics on DRA percent gains.

### *Gender*

There was no significant interaction between the effects of gender and Waterford curriculum on DRA percent gains,  $F(1, 472) = 0.25, p = .619$ . Simple effects analysis showed that for females and males, students' percent gains in the experimental group were significantly higher than in the control group.

### *Ethnicity*

There was a significant interaction between the effects of ethnicity and Waterford curriculum on DRA percent gains,  $F(4, 464) = 2.74, p < .05$ . Simple effects analysis showed that Multiracial and Caucasian students' percent gains in the experimental group were significantly higher than in the control group. Asian and African American students' percent gains in the experimental group were slightly higher than in the control group, but the difference was not significant.

### *Lunch Program*

There was a significant interaction between the effects of lunch program and Waterford curriculum on DRA percent gains,  $F(2, 470) = 3.77, p < .05$ . Simple effects analysis showed that free, reduced, and regular lunch students' percent gains in the experimental group were significantly higher than in the control group.

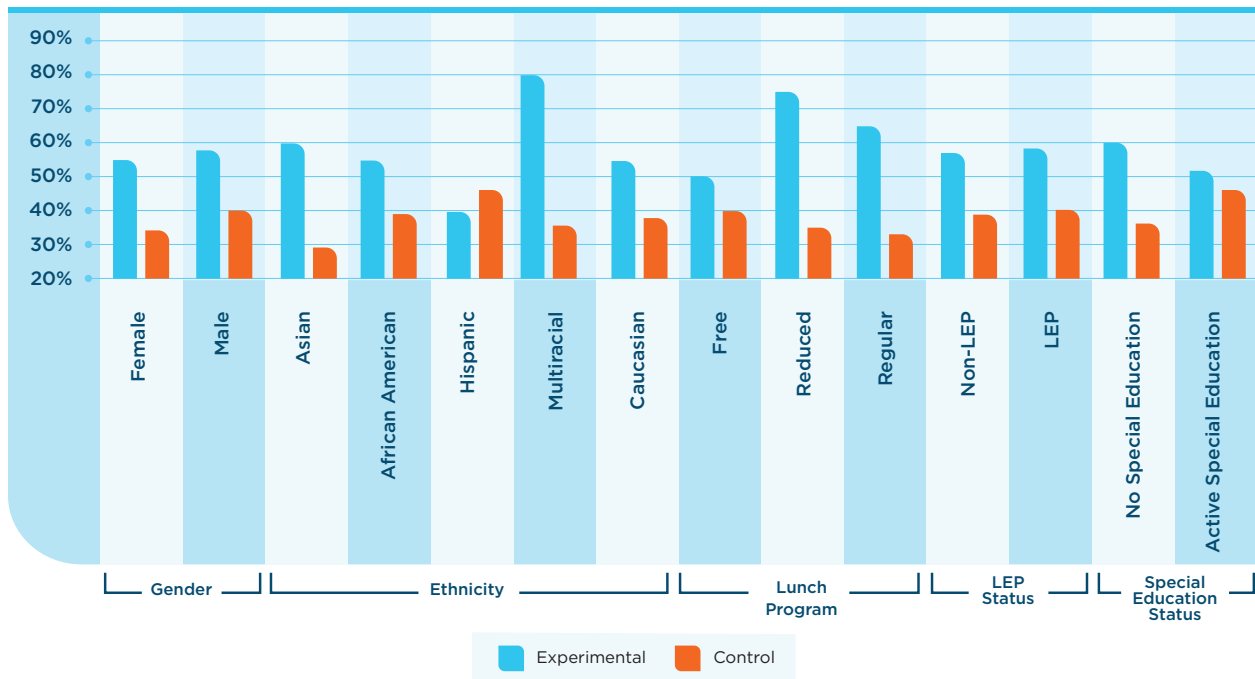
### *LEP Status*

There was no significant interaction between the effects of LEP status and Waterford curriculum on DRA percent gains,  $F(1, 472) = 0.01, p = .930$ . Simple effects analysis showed that Non-LEP students' percent gains in the experimental group were significantly higher than in the control group. LEP students' percent gains in the experimental group were slightly higher than in the control group, but the difference was not significant.

### *Special Education Status*

There was a significant interaction between the effects of special education status and Waterford curriculum on DRA percent gains,  $F(1, 472) = 6.12, p < .05$ . Simple effects analysis showed that students with no special education status in the experimental group made significantly higher percent gains than the control group. Students with an active special education status in the experimental group made slightly higher percent gains than the control group, but the difference was not significant.

Figure 83: Second Grade DRA Percent Gains by Demographics



## Evaluation across Two School Districts in Texas and Florida, 2015-2016

The following study investigates the effect of adding Waterford Early Learning (WEL) to school districts' curricula and its impact on early literacy skills of English Language Learners (ELL) (Shamir, Feehan, Yoder, & Pocklington, 2018b). Students attended two different districts in Texas and Florida during the 2015-2016 school year. The experimental groups used WEL, and the control groups either had low usage of WEL or did not use WEL depending on the district. In the first district, the Texas Primary Reading Inventory (TPRI) was administered at the beginning and end of the year. In the second district, the VLT was administered at the end of the year. Analyses revealed that ELL students in the experimental group outperformed students in the control group on standardized literacy assessments (Tables 42-47).

### DISTRICT 1

Table 42: District 1 Overall Kindergarten TPRI Gains by Substrand

	Experimental			Control			P
	M	SD	N	M	SD	N	
Letter Name Identification	12.03	8.51	71	5.63	8.61	369	.00**
Letter to Sound Linking	4.63	3.41	27	2.80	3.63	276	.01*
Inferring Word Meaning	0.23	0.65	207	0.05	0.51	1458	.00**
Linking Details	0.11	1.03	211	-0.35	0.92	1463	.00**
Recalling Details	1.58	0.82	210	1.36	0.74	1462	.00**
Listening Comprehension Total	1.96	1.56	201	1.07	1.51	1371	.00**

Table 43: District 1 Overall Kindergarten TPRI Gains by ELL Students

	Experimental			Control			P
	M	SD	N	M	SD	N	
Letter Name Identification	6.86	7.98	14	3.42	7.45	59	.18
Inferring Word Meaning	0.33	0.61	43	0.08	0.56	175	.01**
Linking Details	0.43	1.17	44	0.09	1.06	176	.03*
Recalling Details	1.82	0.92	44	1.53	0.81	177	.02*
Listening Comprehension Total	2.57	1.71	42	1.70	1.70	167	.00**

Table 44: District 1 Overall Second Grade TPRI Gains by Substrand

	Experimental			Control			P
	M	SD	N	M	SD	N	
Graphophonemic Knowledge -Set 1	1.26	1.32	136	0.88	1.21	1470	.00**
Word Reading - Set 1	1.23	1.26	126	0.55	1.04	1313	.00**
Word Reading - Set 2	1.67	1.50	131	0.98	1.32	1319	.00**
Word Reading - Set 3	1.55	1.35	131	1.09	1.36	1324	.00**
Word Reading - Set 4	1.65	1.46	131	1.27	1.50	1324	.01**

Table 45: District 1 Overall Second Grade TPRI Gains by ELL Students

	Experimental			Control			P
	M	SD	N	M	SD	N	
Graphophonemic Knowledge - Set 1	1.41	1.23	34	0.73	1.15	175	.00**
Word Reading - Set 1	0.93	1.41	27	0.69	1.30	160	.28
Word Reading - Set 2	1.55	1.43	31	0.91	1.38	160	.02*
Word Reading - Set 3	1.13	1.45	31	1.03	1.29	160	.71

## DISTRICT 2

Table 46: District 2 Overall VLT End of Year Scores

	Experimental			Control			P
	M	SD	N	M	SD	N	
Kindergarten	79.35	22.06	1287	72.56	26.91	43	.05*
Second Grade	71.24	20.02	2150	66.19	20.48	109	.01*

Table 47: District 2 VLT End of Year Scores by ELL Students

	Experimental			Control			P
	M	SD	N	M	SD	N	
Kindergarten	69.65	24.42	172	50.00	32.07	8	.01*
Second Grade	63.78	21.08	365	58.59	20.81	22	.23

## Evaluation across Two School Districts in South Carolina and Indiana, 2015-2016

The following study investigates the effect of adding Waterford Early Learning (WEL) to school districts' curricula and its impact on early literacy and mathematics skills of low socioeconomic (SES) students (Shamir, Feehan, Yoder, & Pocklington, 2018a). Students attended two different districts in South Carolina and Indiana during the 2015-2016 school year; the experimental groups used WEL, and the control groups either had low usage of WEL or did not use WEL. In the first district, the DRA was administered at the beginning and end of the year. In the second district, the mCLASS: DIBELS Next assessment and the mCLASS: Math assessment were administered at the beginning, middle, and end of the year. Analyses revealed that low socioeconomic status (SES) students in the experimental group outperformed students in the control group on multiple reading and math assessments (Tables 48-54).

### DISTRICT 1

Table 48: District 1 Kindergarten DRA End of Year Scores

	Experimental			Control			P
	M	SD	N	M	SD	N	
Overall	5.99	3.52	1004	4.39	3.21	28	.02*
Low-SES	4.85	2.82	381	4.00	4.08	10	.44

Table 49: District 1 First Grade DRA Scores

	Experimental			Control			P
	M	SD	N	M	SD	N	
Overall	18.54	5.05	1064	16.87	5.26	52	.00**
Low-SES	18.02	5.48	377	16.01	5.09	29	.01**

### DISTRICT 2

Table 50: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores

	Experimental			Control			Statistical Measure	P
	M	SD	N	M	SD	N		
Letter Naming Fluency	58.36	12.14	108	53.26	21.33	27	5.77	.02*
NWF - Correct Letter Sounds	52.74	30.20	108	44.72	17.92	27	4.06	.05*
NWF - Whole Words Read	9.21	12.40	108	4.11	6.32	27	7.39	.01**

Table 51: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores by Low-SES

	Experimental			Control			P
	M	SD	N	M	SD	N	
Letter Naming Fluency	58.33	12.64	48	51.81	16.17	14	.03*
NWF - Correct Letter Sounds	47.90	18.47	48	47.78	19.68	14	.98
NWF - Whole Words Read	6.95	6.21	48	3.68	6.92	14	.22

Table 52: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores

	Experimental			Control			Statistical Measure	P
	M	SD	N	M	SD	N		
Number Identification	32.38	12.81	114	28.25	11.08	57	7.34	.01**
Quantity Discrimination	30.80	9.58	114	28.12	7.28	57	4.30	.04*

Table 53: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Low-SES

	Experimental			Control			P
	M	SD	N	M	SD	N	
Number Identification	29.89	11.63	59	27.40	10.95	35	.21
Quantity Discrimination	28.57	7.83	59	27.74	7.45	35	.60

Table 54: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores

	Experimental			Control			Statistical Measure	P
	M	SD	N	M	SD	N		
Number Facts	14.02	4.05	68	12.69	3.61	255	9.06	.00**
Quantity Discrimination	42.17	9.58	68	39.78	9.09	255	5.88	.02*
Missing Number	25.90	6.89	68	23.12	6.49	255	15.07	.00**
Next Number	23.77	5.99	68	22.09	6.04	255	6.18	.01*

Table 55: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores by Low-SES

	Experimental			Control			P
	M	SD	N	M	SD	N	
Number Facts	14.12	4.06	32	12.24	3.55	123	.00**
Quantity Discrimination	42.54	8.12	32	39.33	9.56	123	.03*
Missing Number	25.07	5.37	32	22.80	6.58	123	.03*
Next Number	23.08	4.47	32	21.76	5.84	123	.19

## Evaluation across Two School Districts in Texas and Indiana, 2015-2016

The following study investigates the effect of adding Waterford Early Learning (WEL) to school districts' curricula and its impact on early literacy skills of special education students (Shamir, Yoder, Pocklington, & Feehan, 2018b). Students attended two different districts in Texas and Indiana during the 2015-2016 school year. Kindergarten students in the experimental group were expected to use WEL for 15 minutes per day, five days per week, and first grade students in the experimental group were expected to use WEL for 30 minutes per day, five days per week. Students in the control groups received traditional literacy instruction for the same amount of time that the experimental group received WEL. Thus, overall exposure to literacy and math instruction was the same for both groups. In the first district, the mCLASS: DIBELS Next assessment and the mCLASS: Math assessment were administered three times throughout the school year, at the beginning, middle, and end of the year. In the second district, the TPRI was administered at the beginning and end of the year. In all cases, the experimental group outperformed not only the control group of students with active special education status, but also the control group of students without active special education status.

### DISTRICT 1

ANCOVAs examining group differences in mCLASS: DIBELS Next end of year scores while covarying for middle of year scores were conducted, and ANCOVAs examining group differences in mCLASS: Math end of year scores while covarying for beginning of year scores were conducted.

Figure 84: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Whole Words Read

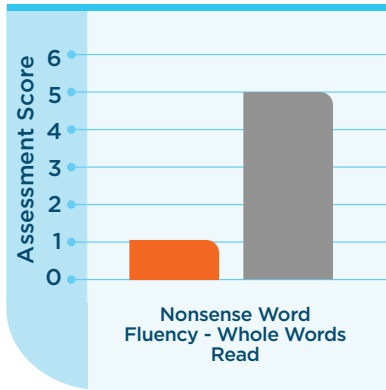


Figure 85: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Letter Naming Fluency & Correct Letter Sounds

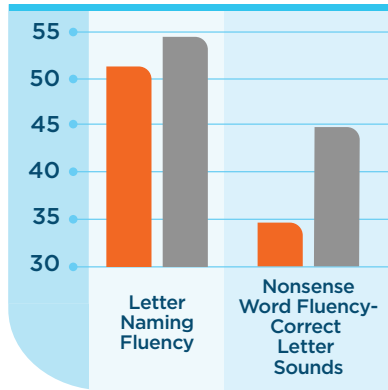
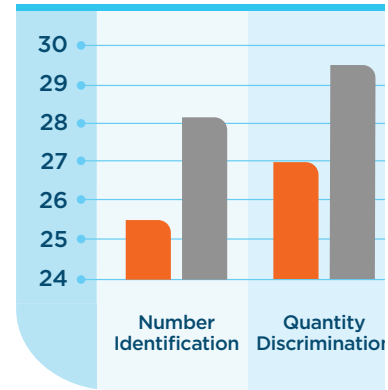


Figure 86: mCLASS DIBELS Next and Math Scores by Strand by Special Education Control—Number Identification & Quantity Discrimination



Special Education Control      Special Education Experimental

Figure 87: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Whole Words Read

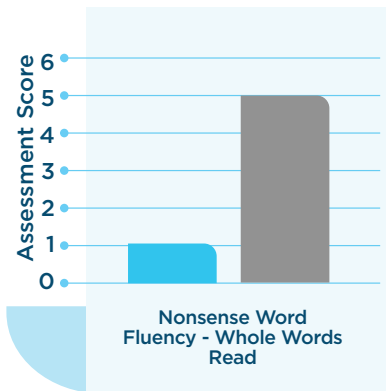


Figure 88: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Letter Naming Fluency & Correct Letter Sounds

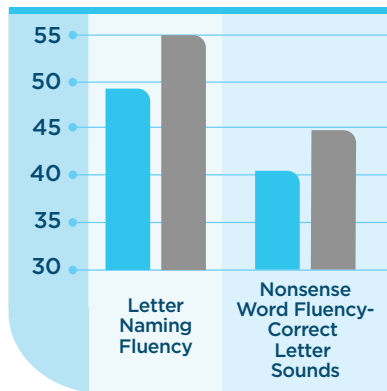
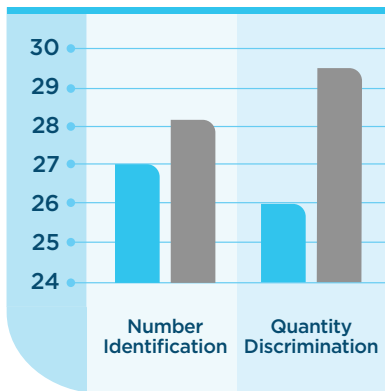


Figure 89: mCLASS DIBELS Next and Math Scores by Strand by Non-Special Education Control—Number Identification & Quantity Discrimination



Non-Special Education Control      Special Education Experimental

## DISTRICT 2

One-way ANOVAs were conducted to examine the effects of WEL and special education status on gains for each strand.

Figure 90: TPRI Gain Scores by Strand by Special Education Control

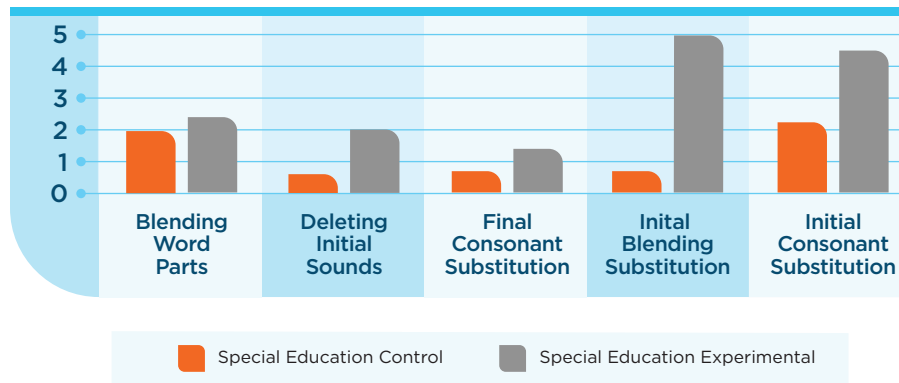
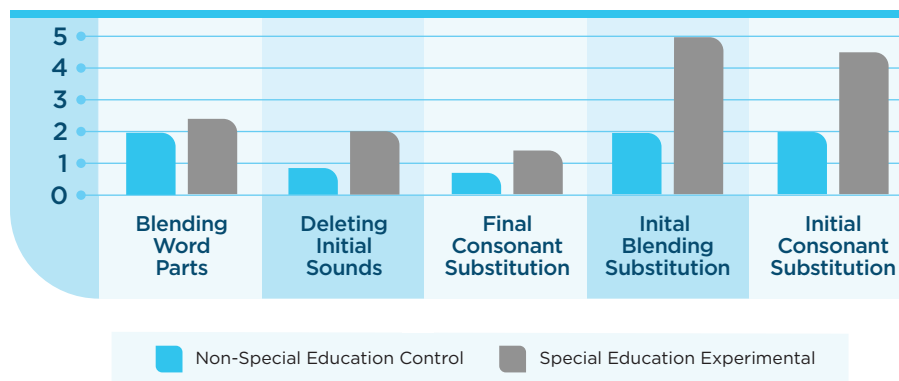


Figure 91: TPRI Gain Scores by Strand by Non-Special Education Control



## Evaluation of a School District in Texas, 2015-2016

This study investigates the impact of using a computer-adaptive program in early childhood education (Shamir, Yoder, Pocklington, & Feehan, 2018a). Waterford Early Learning (WEL) is a computer-adaptive program that was assigned to kindergarten and first grade students in a school district in Texas for the 2015-2016 school year. The Texas Primary Reading Inventory (TPRI) assessment was administered to students at the beginning and end of the school year to assess reading skills across multiple strands. Analysis revealed consistently higher gains made by students who used WEL, across demographics.



## KINDERGARTEN

Figure 92: Kindergarten TPRI Gains by Substrand

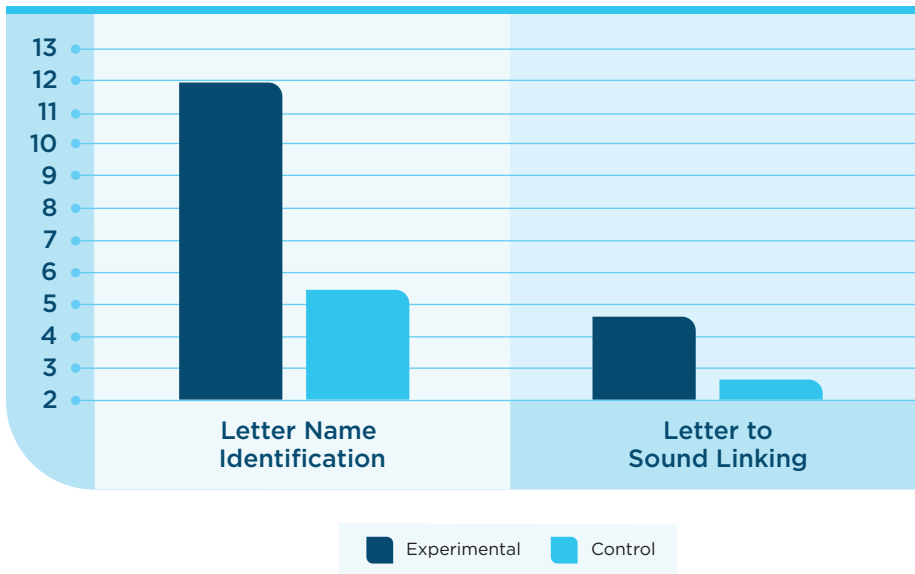


Figure 93: Kindergarten TPRI Gains by Substrand

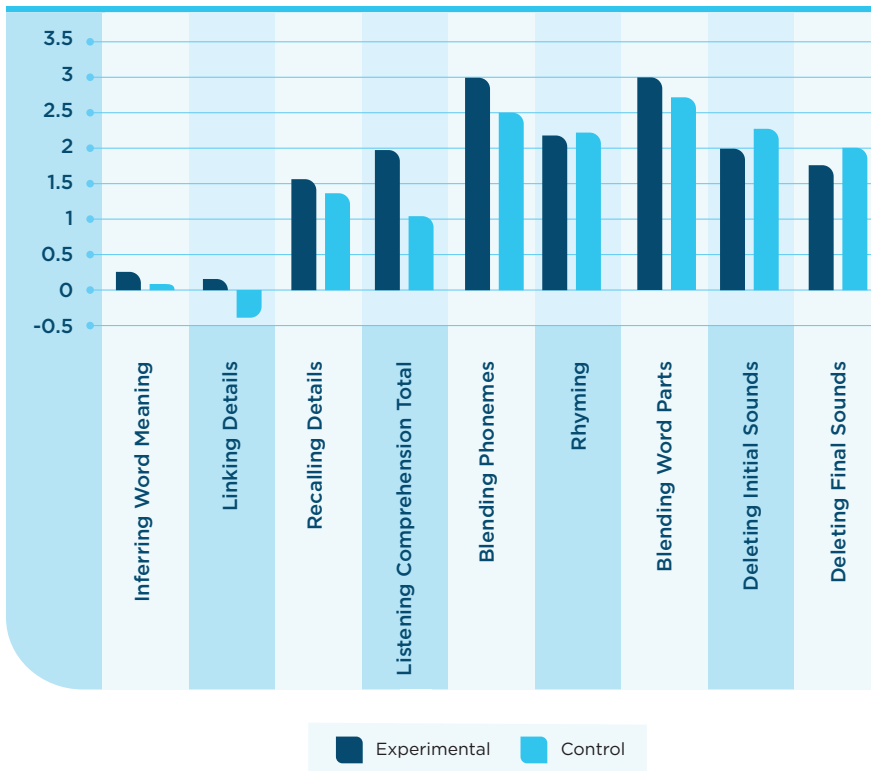


Table 56: Kindergarten TPRI Gains by Special Education Status

Kindergarten	Active Special Education				No Special Education			
	Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD
Letter Name Identification	13.0	1.41	6.21	9.85	12.0	8.63	5.60	8.56
Letter to Sound Linking	10.0	.	4.50	3.71	4.42	3.30	2.73	3.62
Inferring Word Meaning	0.13	0.64	-0.15	0.61	0.24	0.65	0.06	0.51
Linking Details	0.25	1.04	0.10	1.05	0.11	1.03	-0.37	0.91
Recalling Details	1.50	0.53	1.45	0.91	1.58	0.83	1.36	0.73
Listening Comprehension Total Score	1.75	1.58	1.40	1.70	1.96	1.56	1.06	1.50

Table 57: Kindergarten TPRI Gains by EL Status

Kindergarten	EL				Non-EL			
	Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD
Letter Name Identification	6.86	7.98	3.42	7.45	13.30	8.22	6.05	8.77
Letter to Sound Linking	4.67	3.78	2.43	3.54	4.62	3.40	2.89	3.65
Inferring Word Meaning	0.33	0.61	0.08	0.56	0.21	0.66	0.05	0.51
Linking Details	0.43	1.17	0.09	1.06	0.03	0.98	-0.41	0.89
Recalling Details	1.82	0.92	1.53	0.81	1.52	0.78	1.34	0.72
Listening Comprehension Total Score	2.57	1.71	1.70	1.70	1.79	1.48	0.98	1.46

Table 58: Kindergarten TPRI Gains by Ethnicity

Kindergarten	African American				Hispanic				Caucasian			
	Treatment		Control		Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Letter Name Identification	9.73	8.42	4.12	7.38	11.47	8.82	6.73	9.32	14.5	7.92	5.20	7.67
Letter to Sound Linking	4.14	4.34	2.33	3.36	5.00	3.25	3.11	3.72	4.50	3.42	2.67	3.73
Inferring Word Meaning	0.27	0.64	0.03	0.56	0.23	0.65	0.01	0.55	0.21	0.69	0.11	0.45
Linking Details	0.20	0.92	-0.27	0.97	0.14	1.10	-0.23	1.03	-0.04	0.96	-0.53	0.79
Recalling Details	1.77	0.68	1.42	0.72	1.63	0.81	1.42	0.80	1.37	0.88	1.24	0.67
Listening Comprehension Total Score	2.31	1.49	1.17	1.54	2.04	1.65	1.21	1.70	1.55	1.44	0.85	1.30

## FIRST GRADE

Figure 94: First Grade TPRI Gains by Substrand

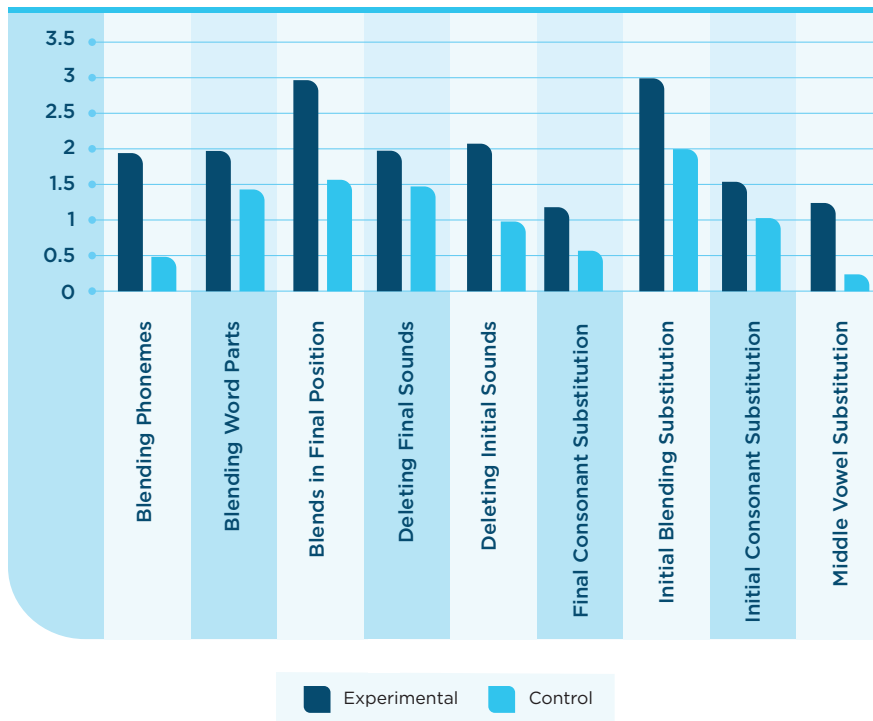


Table 59: First Grade TPRI Gains by Special Education Status

First Grade	Active Special Education				No Special Education			
	Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD
Blending Word Parts	2.40	1.52	1.92	1.72	2.26	1.60	1.89	1.52
Blends in Final Position	3.50	2.12	0.00	.	2.81	1.87	1.67	1.73
Deleting Initial Sounds	2.00	2.83	0.50	1.00	2.18	1.74	1.00	1.76
Final Consonant Substitution	1.50	1.91	0.60	1.34	1.20	1.74	0.61	1.27
Initial Blending Substitution	5.00	0.00	0.60	0.89	2.89	1.73	2.10	1.97
Initial Consonant Substitution	4.50	0.71	2.29	1.82	2.46	1.69	1.93	1.61

Table 60: First Grade TPRI Gains by EL Status

First Grade	EL				Non-EL			
	Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD
Blending Phonemes	1.50	2.66	0.50	1.40	2.06	1.80	0.73	1.28
Blends in Final Position	1.40	2.30	0.83	1.64	3.46	1.33	1.80	1.71
Deleting Initial Sounds	1.14	1.21	1.14	1.96	2.59	1.80	0.96	1.71
Final Consonant Substitution	1.25	0.96	0.67	1.59	1.25	1.86	0.60	1.23
Initial Blending Substitution	1.86	1.77	0.94	1.91	3.43	1.60	2.20	1.93
Initial Consonant Substitution	1.33	1.87	1.09	1.62	1.75	1.88	1.04	1.55
Middle Vowel Substitution	0.50	1.00	0.46	1.13	1.50	1.83	0.26	0.87

Table 61: First Grade TPRI Gains by Ethnicity

First Grade	African American				Hispanic				Caucasian			
	Treatment		Control		Treatment		Control		Treatment		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Blending Phonemes	3.00	1.58	0.69	1.23	1.87	2.13	0.88	1.53	1.50	2.12	0.44	1.00
Blending Word Parts	2.67	1.56	1.33	1.48	1.79	1.77	1.58	1.66	2.17	1.33	1.29	1.40
Blends in Final Position	4.00	1.41	2.00	1.93	2.30	2.00	2.04	1.65	3.67	1.53	1.11	1.37
Deleting Initial Sounds	2.75	1.50	1.07	1.76	1.80	1.82	1.13	1.99	3.50	1.00	0.76	1.45
Final Consonant Substitution	1.83	2.04	0.14	0.55	1.31	1.80	1.04	1.67	0.00	0.00	0.50	0.89
Initial Blending Substitution	4.40	0.89	2.40	1.82	2.33	1.68	2.28	2.03	4.00	1.41	1.71	2.02
Initial Consonant Substitution	2.00	2.31	0.62	1.18	1.77	1.85	1.47	1.71	1.00	1.41	0.62	1.26
Middle Vowel Substitution	2.25	2.22	0.35	0.77	1.17	1.75	0.41	1.21	1.00	1.00	0.07	0.36

## Evaluation of Waterford Early Learning in Idaho, 2017-2018

The following study (Shamir, Feehan, Pocklington, & Yoder, 2019a) assessed the efficacy of Waterford Early Learning (WEL). Kindergarten students enrolled in a public school district in Idaho were randomly assigned on the class level to an experimental or control condition during the 2017-2018 school year: Students in the experimental condition used WEL for at least 1,500 minutes, and students in the control condition received traditional literacy instruction for the same amount of instructional time. Experimental students significantly outperformed their control counterparts on all strands of STAR Early Literacy.

### STAR EARLY LITERACY

Independent samples *t*-tests examining gains made from beginning of year to end of year were conducted and revealed that gains were significantly higher on most strands for experimental group students that met the recommended usage of WEL than for control group students.

Figure 95: STAR Early Literacy Scaled Score Gains from Beginning of Year to End of Year

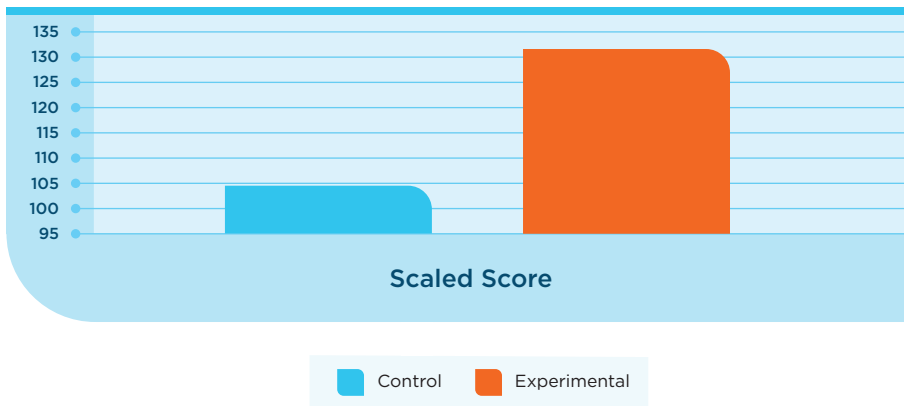
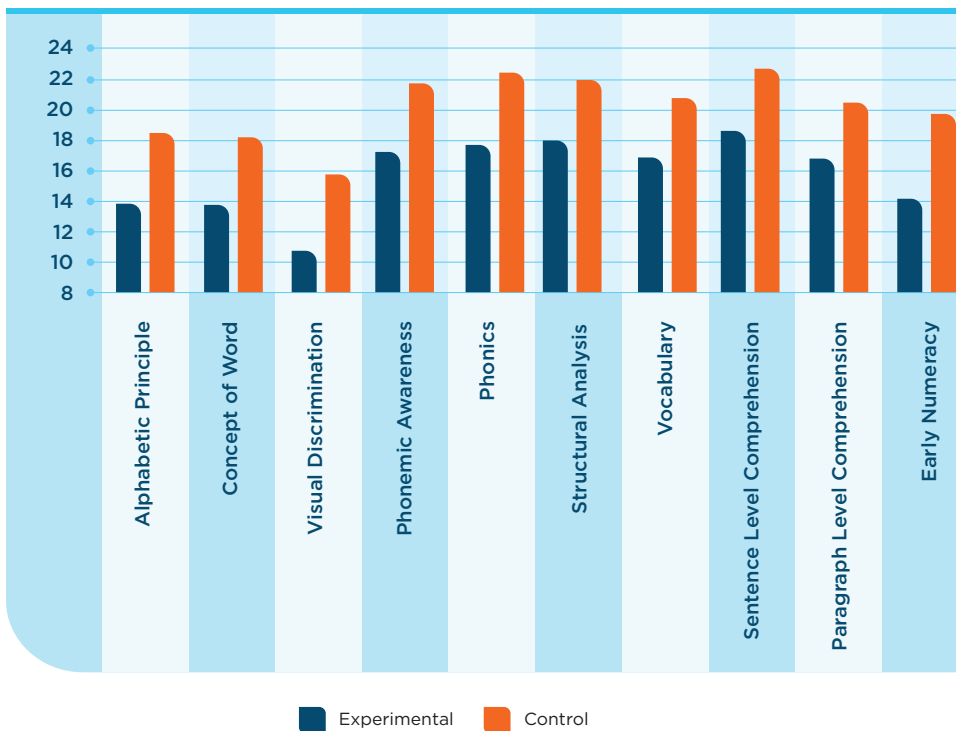
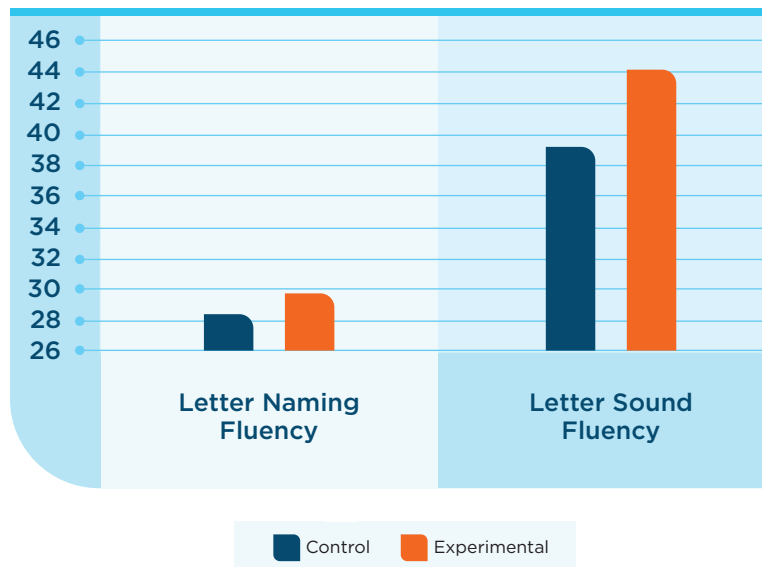


Figure 96: STAR Early Literacy Gains from Beginning of Year to End of Year by Strand



## IRI

Figure 97: IRI Gains from Beginning of Year to End of Year by Strand



## Evaluation of Waterford Early Learning in Three School Districts

The following study (Shamir, Yoder, Pocklington, & Feehan, 2019b) assessed the efficacy of Waterford Early Learning (WEL). The study reports findings for young learners in kindergarten through second grade using WEL in three geographically diverse school districts. In all districts, students who used WEL benefited from significantly higher gains, percent gains, and end of year scores than students in the control group. For reference, Caucasian students’ scores are included in the tables.

### DISTRICT 1

#### KINDERGARTEN

Table 62: District 1 – Kindergarten TPRI Gains by Ethnicity

Kindergarten	African American				Hispanic				Caucasian			
	Treatment		Control		Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Letter Name Identification.	9.73	8.42	4.12	7.38	11.47	8.82	6.73	9.32	14.50	7.92	5.20	7.67
Letter to Sound Linking.	4.14	4.34	2.33	3.36	5.00	3.25	3.11	3.72	4.50	3.42	2.67	3.73
Inferring Word Meaning.	0.27	0.64	0.03	0.56	0.23	0.65	0.01	0.55	0.21	0.69	0.11	0.45
Linking Details.	0.20	0.92	-0.27	0.97	0.14	1.10	-0.23	1.03	-0.04	0.96	-0.53	0.79
Recalling Details.	1.77	0.68	1.42	0.72	1.63	0.81	1.42	0.80	1.37	0.88	1.24	0.67
Listening Comprehension - Total Score.	2.31	1.49	1.17	1.54	2.04	1.65	1.21	1.70	1.55	1.44	0.85	1.30

## FIRST GRADE

Table 63: District 1 – First Grade TPRI Gains by Ethnicity

First Grade	<u>African American</u>				<u>Hispanic</u>				<u>Caucasian</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Blending Phonemes.	3.00	1.58	0.69	1.23	1.87	2.13	0.88	1.53	1.50	2.12	0.44	1.00
Blending Word Parts.	2.67	1.56	1.33	1.48	1.79	1.77	1.58	1.66	2.17	1.33	1.29	1.40
Blends in Final Position.	4.00	1.41	2.00	1.93	2.30	2.00	2.04	1.65	3.67	1.53	1.11	1.37
Deleting Initial Sounds.	2.75	1.50	1.07	1.76	1.80	1.82	1.12	1.99	3.50	1.00	0.76	1.45
Final Consonant Substitution.	1.83	2.04	0.14	0.55	1.31	1.80	1.04	1.67	0.00	0.00	0.50	0.89
Initial Blending Substitution.	4.40	0.89	2.40	1.82	2.33	1.68	2.27	2.03	4.00	1.41	1.71	2.02
Initial Consonant Substitution.	2.00	2.31	0.62	1.18	1.77	1.85	1.47	1.71	1.00	1.41	0.62	1.26
Middle Vowel Substitution.	2.25	2.22	0.35	0.77	1.17	1.75	0.40	1.21	1.00	1.00	0.06	0.36

## DISTRICT 2

### SECOND GRADE

Table 64: District 2 – Second Grade DRA Percent Gains by Ethnicity

Second Grade	<u>African American</u>				<u>Multiracial</u>				<u>Caucasian</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
DRA.	53.57	37.46	38.95	36.32	80.57	56.21	33.30	24.68	52.84	28.90	37.83	23.33

## DISTRICT 3

### KINDERGARTEN

Table 65: District 3 – Kindergarten VLT End of Year Scores by Ethnicity

Kindergarten	<u>Hispanic</u>				<u>Caucasian</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VLT.	73.49	23.19	55.00	29.76	80.93	21.71	74.62	26.27

### FIRST GRADE

Table 66: District 3 – First Grade VLT End of Year Scores by Ethnicity

First Grade	<u>African American</u>				<u>Hispanic</u>				<u>Caucasian</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VLT.	66.78	20.31	63.00	19.05	70.53	19.54	52.29	26.29	73.30	19.47	63.87	23.78

## Longitudinal Evaluation of Waterford Early Learning in Florida, 2014-2016

The following study (Shamir, Yoder, Pocklington, & Feehan, 2019a) investigated the longitudinal impact of computer-adaptive technology on early literacy skills. Pre-kindergarten students from low socioeconomic status homes in Florida used Waterford Early Learning (WEL) for five days per week for 15 minutes per day during the 2014-2015 school year. Students did not use WEL during the 2015-2016 school year while in kindergarten. At the end of the 2015-2016 school year, these students (experimental) were assessed on a literacy assessment, and their scores were compared to the scores of kindergarten students who did not have access to WEL (control). Differences between experimental and control groups were analyzed and parsed in terms of demographic factors, including English learner (EL) status, ethnicity, and socioeconomic status (SES). Students who used WEL in pre-kindergarten outperformed students who did not use WEL in pre-kindergarten at the end of their kindergarten year.

Independent samples *t*-tests were conducted to examine group differences on end of kindergarten year scores for each strand of Kindergarten Readiness Test (KRT) scores for English learner (EL) students, students of minority ethnicities, and students with free lunch status.

Figure 98: KRT Overall End of Year Scores by Demographics

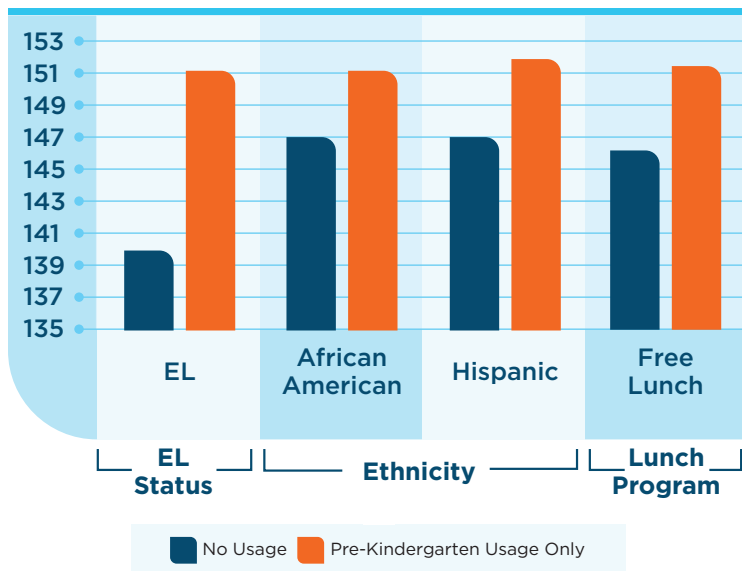




Figure 99: KRT Letter ID & Sounds End of Year Scores by Demographics

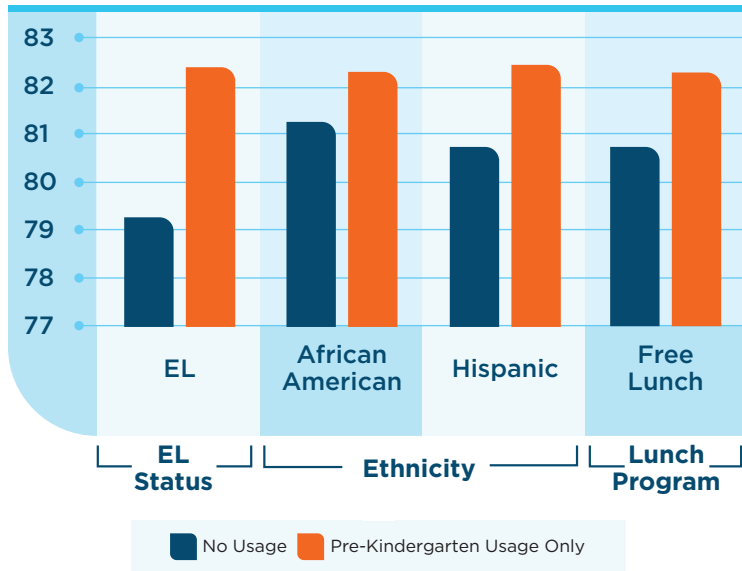


Figure 100: KRT Concepts of Print End of Year Scores by Demographics

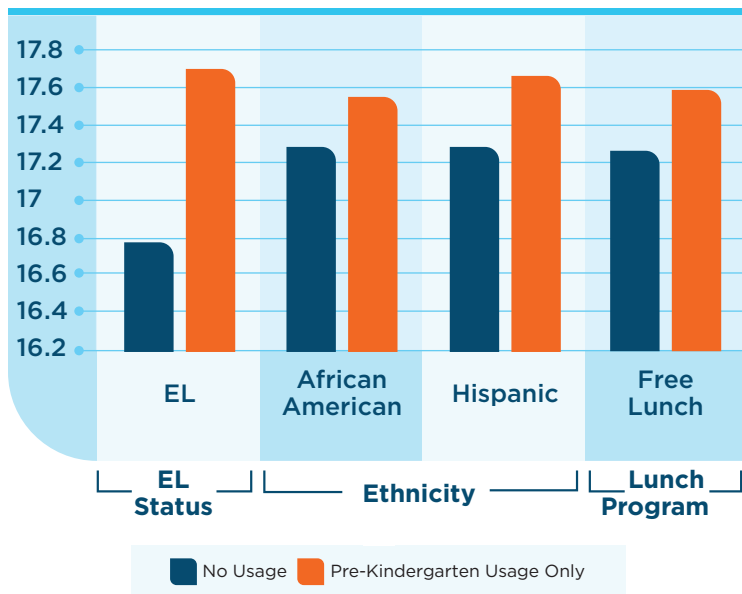


Figure 101: KRT Phonemic Awareness End of Year Scores by Demographics

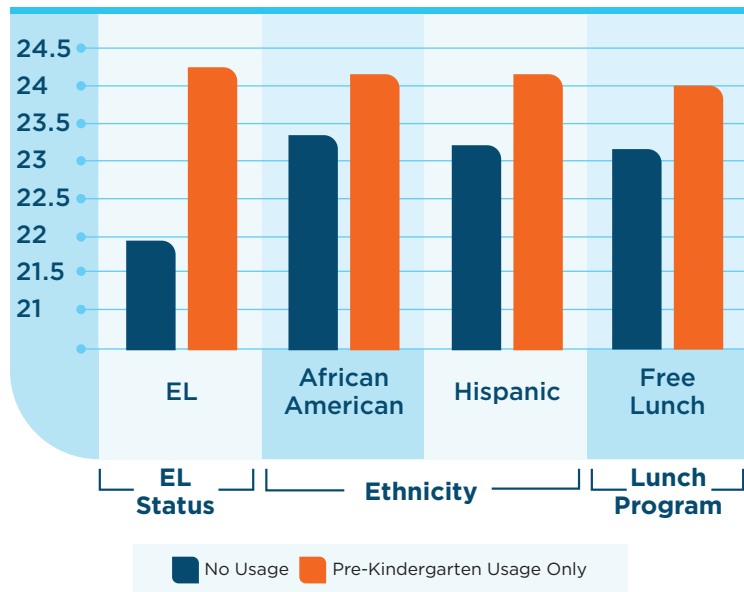


Table 67: End of Kindergarten Scores by Strand & Demographics

	Experimental			Control			t	p
	M	SD	N	M	SD	N		
<b>Overall</b>								
EL Status	151.43	15.98	111	140.34	32.27	2627	-6.75	.000**
African American	151.67	18.16	100	147.25	26.37	2528	-2.34	.021*
Hispanic	152.06	16.28	141	147.04	27.52	4693	-3.52	.001**
Free Lunch	151.44	17.67	241	146.21	27.09	7240	-4.42	.000**
<b>Letter ID and Sounds</b>								
EL Status	82.50	2.30	111	79.23	12.14	2613	-10.13	.000**
African American	82.14	3.24	100	81.16	7.27	2509	-2.76	.006**
Hispanic	82.45	2.38	141	80.72	9.35	4668	-7.13	.000**
Free Lunch	82.25	2.99	241	80.81	8.62	7192	-6.60	.000**
<b>Concepts of Print</b>								
EL Status	17.69	1.32	111	16.78	3.17	2612	-6.55	.000**
African American	17.53	2.04	100	17.35	2.18	2506	-0.81	.421
Hispanic	17.69	1.31	141	17.29	2.41	4667	-3.45	.001**
Free Lunch	17.58	1.79	241	17.32	2.25	7185	-2.16	.032*
<b>Phonemic Awareness</b>								
EL Status	24.14	3.45	111	22.06	6.32	2623	-5.95	.000**
African American	24.10	3.17	100	23.39	4.72	2521	-2.15	.033*
Hispanic	24.15	3.39	141	23.21	5.04	4687	-3.18	.002**
Free Lunch	24.04	3.54	241	23.25	4.88	7224	-3.38	.001**

\* $p < .05$ , \*\* $p < .01$

## Longitudinal Evaluation of Waterford Early Learning in South Carolina, 2015-2017

The following longitudinal study (Shamir, Pocklington, Feehan, & Yoder, 2019b) explored the long-term impact of a computer-adaptive program on young students’ literacy skills. Two cohorts of elementary school students used Waterford Early Learning (WEL) in kindergarten, first grade, or both kindergarten and first grade during the 2015-2016 and the 2016-2017 school years. The Developmental Reading Assessment (DRA) was administered to students at the end of the 2016-2017 and 2017-2018 school years when students were in first or second grade. Scores of students in both cohorts who used WEL only during kindergarten or first grade (for one year only) or in kindergarten and first grade (for two years) were compared to scores of students who received traditional, teacher-directed classroom instruction. Analysis indicated a salient and persistent effect of WEL: One to two years after students stopped using the program, students who used WEL for one or two years outperformed students who did not use WEL. Additionally, evidence was found for a dosage effect: While all students who used WEL had higher end of year scores than students who did not use WEL, the largest effects were found for students with high WEL usage (Tables 68-71).

### GROUP DIFFERENCES IN END OF FIRST GRADE DRA SCORES USING INDEPENDENT SAMPLES T-TESTS

Table 68: End of First Grade DRA Scores

		Treatment			Control			P
		M	SD	N	M	SD	N	
K and 1 <sup>st</sup> Usage	Cohort 1	18.47	5.12	1416	14.94	5.41	31	.00**
	Cohort 2	18.11	5.10	1381	12.63	7.94	32	.00**
K Only Usage	Cohort 1	16.85	7.92	26	14.94	5.41	31	.29
	Cohort 2	20.15	4.29	39	12.63	7.94	32	.00**

\* $p < .05$ , \*\* $p < .001$

Table 69: End of Second Grade DRA Scores

		Treatment			Control			p
		M	SD	N	M	SD	N	
K and 1 <sup>st</sup> Usage	Cohort 2	28.11	5.74	1235	24.88	7.61	281	.00**
K Only Usage	Cohort 2	25.47	9.19	49	24.88	7.61	281	.63
1 <sup>st</sup> Only Usage	Cohort 1	27.15	5.69	1529	24.58	7.34	323	.00**
	Cohort 2	27.11	7.20	237	24.88	7.61	281	.00**

\* $p < .05$ , \*\* $p < .001$

GROUP DIFFERENCES IN END OF FIRST GRADE DRA SCORES USING INDEPENDENT SAMPLES T-TESTS FOR HIGH USAGE VS NO USAGE GROUPS

Table 70: End of First Grade High Usage DRA Scores

		Treatment			Control			p
		M	SD	N	M	SD	N	
K and 1 <sup>st</sup> Usage	Cohort 1	19.00	4.82	1094	14.94	5.41	31	.00**
	Cohort 2	18.33	5.10	1158	12.63	7.94	32	.00**
K Only Usage	Cohort 1	19.06	7.61	18	14.94	5.41	31	.03*
	Cohort 2	21.36	3.93	28	12.63	7.94	32	.00**

\*p < .05, \*\*p < .001

GROUP DIFFERENCES IN END OF SECOND GRADE DRA SCORES USING INDEPENDENT SAMPLES T-TESTS FOR HIGH USAGE VS NO USAGE GROUPS

Table 71: End of Second Grade High Usage DRA Scores

		Treatment			Control			p
		M	SD	N	M	SD	N	
K and 1 <sup>st</sup> Usage	Cohort 2	28.56	5.29	967	24.88	7.61	281	.00**
K Only Usage	Cohort 2	27.70	8.00	33	24.88	7.61	281	.05*
1 <sup>st</sup> Only Usage	Cohort 1	27.54	5.40	1230	24.58	7.34	323	.00**
	Cohort 2	26.99	7.32	176	24.88	7.61	281	.00**

\*p < .05, \*\*p < .001

## Longitudinal Evaluation of a School District in Maryland, 2015-2018

The following longitudinal study (Shamir, Feehan, Pocklington, & Yoder, 2019b) assessed the efficacy of Waterford Early Learning (WEL), a computer-adaptive program that was assigned to students in a Maryland school district during the three school years between 2015, when students were in kindergarten, and 2018, when students were in second grade. While in kindergarten students were expected to use WEL for 15 minutes per day five days per week, and while in first and second grade students were expected to use WEL for 30 minutes per day, five days per week. The NWEA MAP was administered at the end of the 2017-2018 school year while students were in the second grade. This study tested three hypotheses: (1) If use of WEL has an overall effect on literacy scores, then students who used WEL over multiple school years will have higher literacy assessment scores than students who did not use WEL. (2) If use of WEL has a long-term effect on literacy scores, then students who used WEL will have higher literacy assessment scores than students who did not use WEL, when assessed a year after use of WEL had stopped. (3) If early use of WEL has an effect on literacy scores, then students who used WEL in kindergarten through second grade will have higher literacy assessment scores than students who used WEL in first and second grade only.

### THREE YEARS OF USAGE VS. NO USAGE: OVERALL WEL EFFECTS

For three years of usage compared to no usage, the experimental group included students who used WEL for more than 100 minutes during kindergarten (2015-2016), first grade (2016-2017), and second grade (2017-2018). The control group included students who used WEL for less than 100 minutes during all three school years. Independent samples *t*-tests were conducted and revealed that the experimental group consistently outperformed the control group on all strands of the MAP.

Figure 102: Three Years of Usage vs. No Usage End of Year Scores by Strand

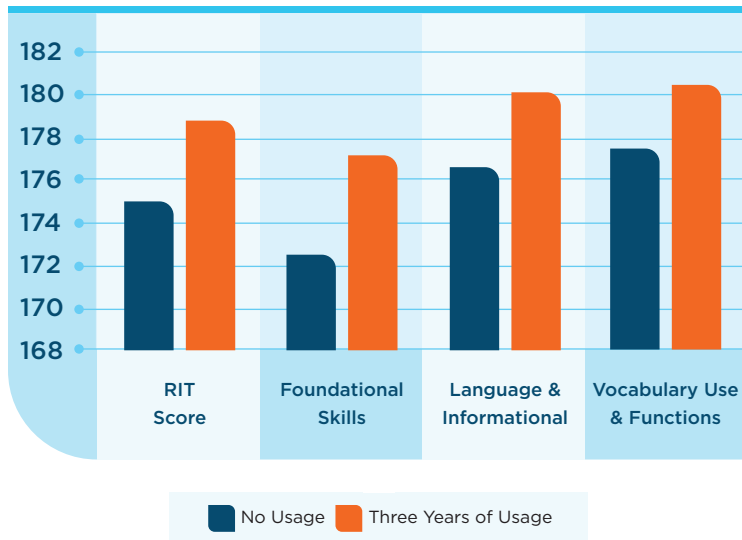
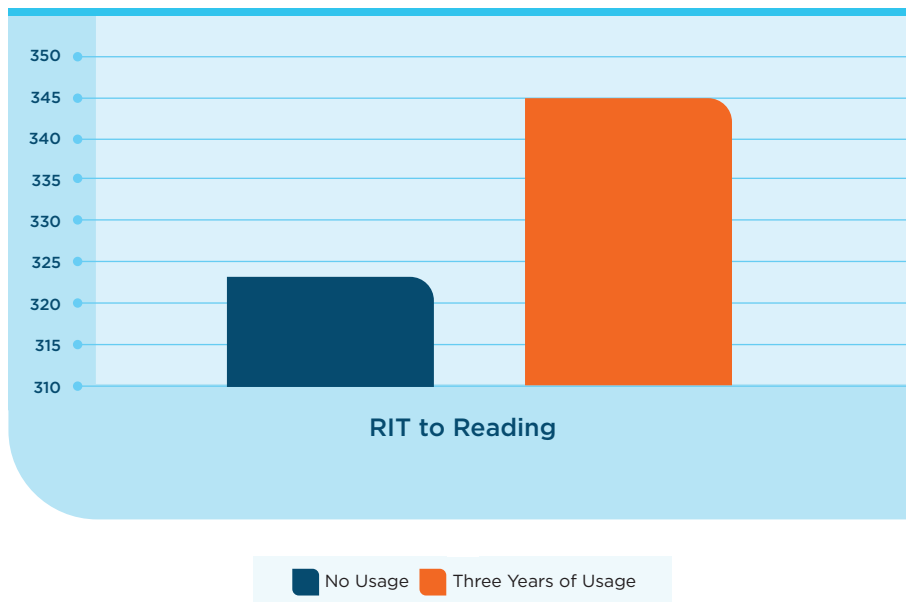


Figure 103: Three Years of Usage vs. No Usage RIT to Reading End of Year Scores



### TWO YEARS OF USAGE VS. NO USAGE: LONG-TERM EFFECTS

For two years of usage compared to no usage, the experimental group included students who used WEL for more than 100 minutes during kindergarten (2015-2016) and first grade (2016-2017) only. The control group included students who used WEL for less than 100 minutes during all three school years. Independent samples *t*-tests were conducted and revealed that the experimental group consistently outperformed the control group on all strands of the MAP, including significant differences between experimental and control groups on the Literature and Informational strand.

Figure 104: Two Years of Usage vs. No Usage End of Year Scores by Strand

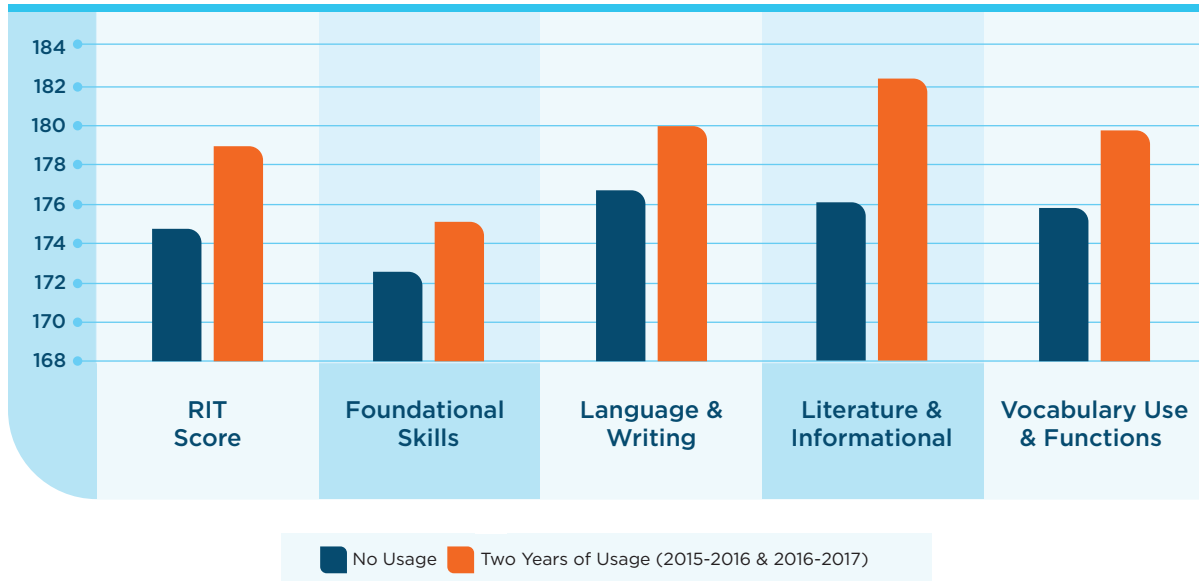
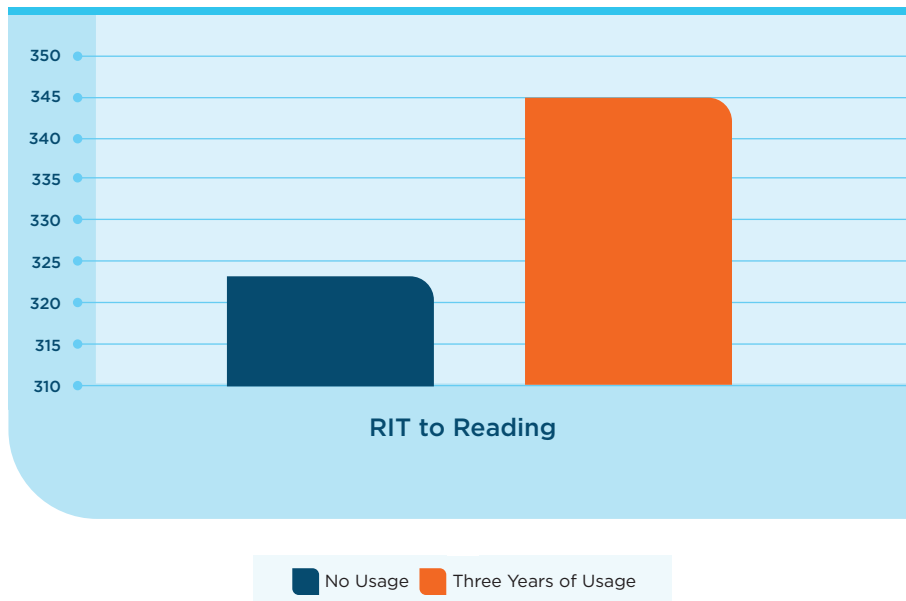


Figure 105: Two Years of Usage vs. No Usage RIT to Reading End of Year Scores



### THREE YEARS OF USAGE VS. TWO YEARS OF USAGE: EARLY EFFECTS

For three years of usage compared to two years of usage, the experimental group included students who used WEL for more than 100 minutes during kindergarten (2015-2016), first grade (2016-2017), and second grade (2017-2018). The control group included students who used WEL for more than 100 minutes during first grade (2016-2017) and second grade (2017-2018) only. Independent samples *t*-tests were conducted and revealed significant differences between experimental and control groups on MAP end of second grade scores.

Figure 106: Three Years of Usage vs. Two Years of Usage End of Year Scores by Strand

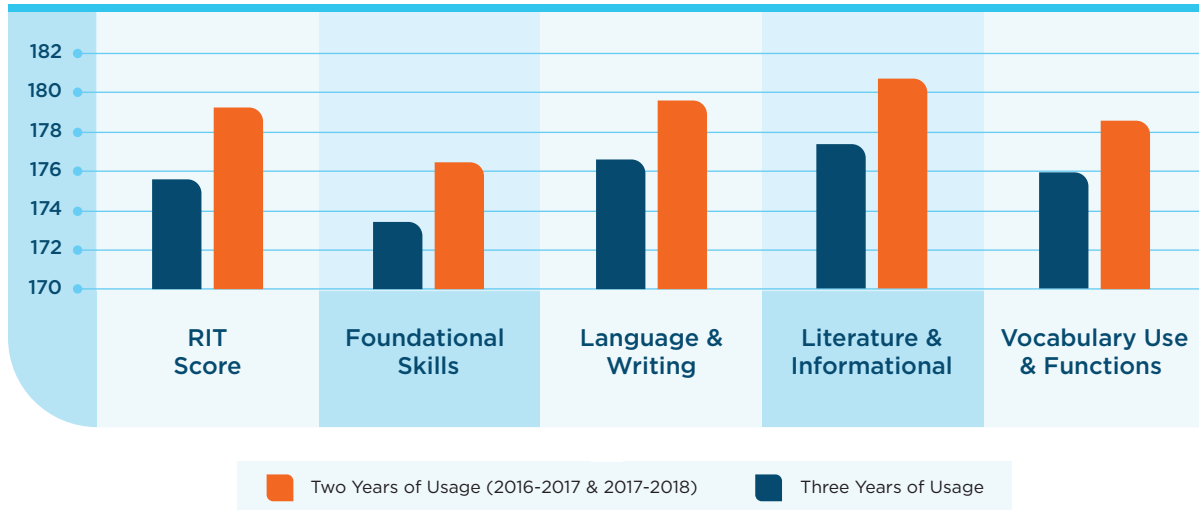
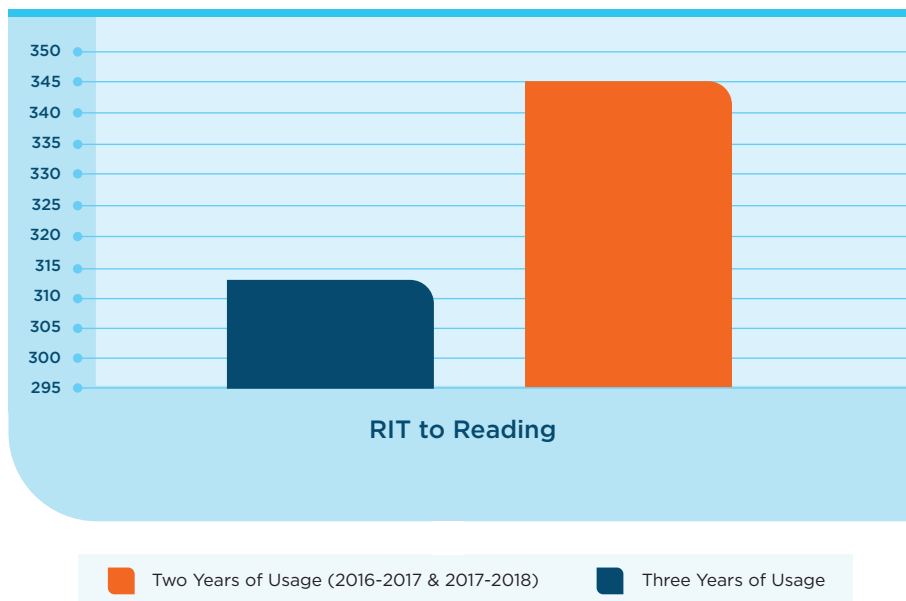


Figure 107: Three Years of Usage vs. Two Years of Usage RIT to Reading End of Year Scores



## Effectiveness of UPSTART

### YEAR 1

One thousand two hundred forty-eight students participated in the UPSTART Year 1 program. One hundred thirty-seven children from seven Utah public school districts participated in the UPSTART research study during its first year of operation during the 2009-2010 school year (Evaluation and Training Institute, 2011). Findings revealed that children who had participated in UPSTART during preschool scored almost 18 points higher in reading proficiency on the DIBELS Next (DN) Composite compared to beginning kindergarten children who did not participate in UPSTART prior to enrolling in public school. Additionally, middle kindergarten children who had participated in the UPSTART preschool program scored approximately 19 points higher in reading proficiency on the DN Composite compared to middle kindergarten children who did not participate in UPSTART prior to enrolling in public school.

When assessed at the beginning of kindergarten by the DIBELS Next Beginning Kindergarten Composite, children participating in UPSTART demonstrated moderately strong improvements in reading proficiency compared to children who did not participate in UPSTART and demonstrated higher gains when assessed on the DIBELS Next Middle Kindergarten Composite.

### WACS UPSTART

WACS results are provided for years five and six (Shamir, Miner, Izzo, Pocklington, Feehan, & Yoder, 2018). Results for years one through four are very similar and can be provided upon request.

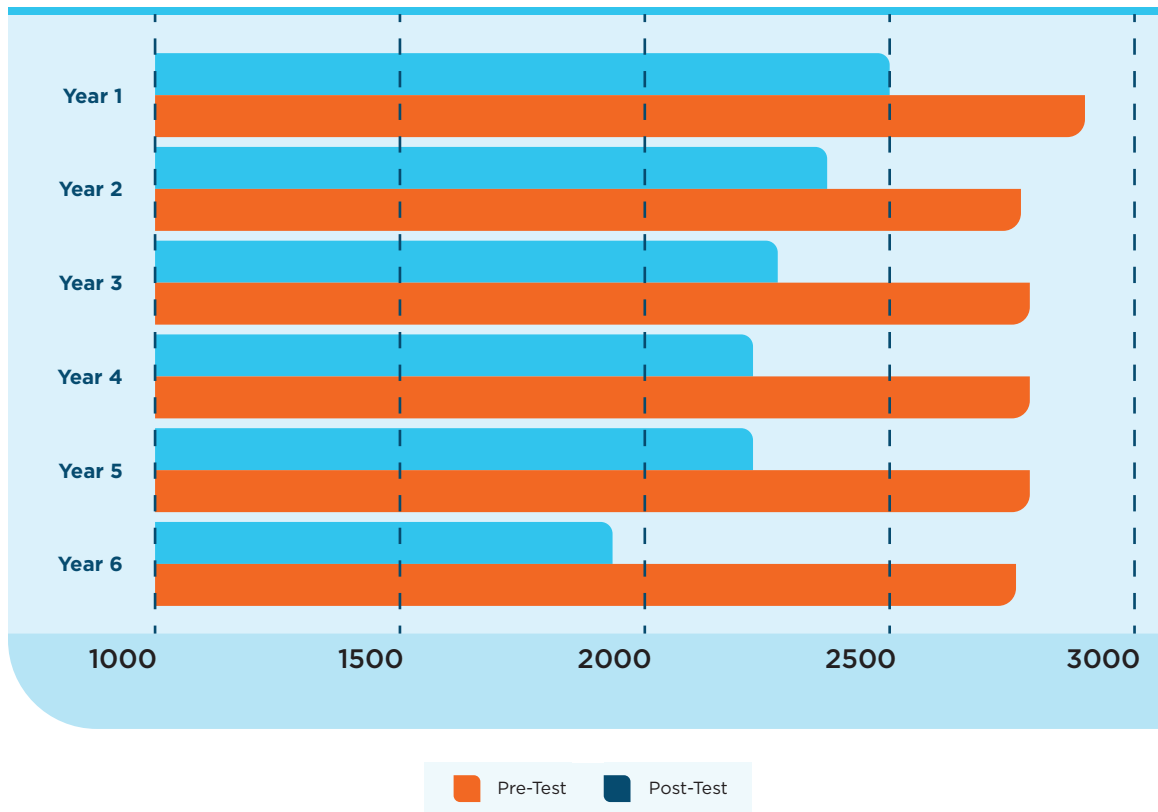
Waterford Assessment of Core Skills™ (WACS) measures learning gains across eleven key pre-literacy and reading skills (Miner, 2014). Administered at the beginning and end of the UPSTART preschool year, WACS is internally consistent and has strong test-retest reliability ( $r = .90$ ). Initial content validity for WACS was established against state and national standards for the 11 subtests. All items were then calibrated for Item Response Theory to determine item difficulty. Student performance on WACS was compared to performance on five commonly used standardized tests also measuring early reading skills in order to establish concurrent validity and predictive validity, and all correlations between the tests were found to be significant, ranging from  $r = .41$  to  $r = .78$  (median  $r = .63$ ). Figure 108 shows the improvement in academic achievement as measured by WACS pretest and posttest scores.

Table 72: WACS Question Difficulty Ranges by Grade

<i>Grade</i>	<i>Beginning</i>	<i>Intermediate</i>	<i>Advanced</i>
Preschool	1001 - 1333	1334 - 1666	1667 - 2000
K	2001 - 2333	2334 - 2666	2667 - 3000
1	3001 - 3333	3334 - 3666	3667 - 4000
2	4001 - 4333	4334 - 4666	4667 - 5000
3	5001 - 5333	5334 - 5666	5667 - 6000
4	6001 - 6333	6334 - 6666	6667 - 7000



Figure 108: Year 1-6 WACS Overall Scores



UPSTART participants completed the program at the WACS Kindergarten Advanced level on average, indicating that the ability levels of the UPSTART students are similar to the ability levels of the top third of kindergarteners nationwide. This average includes students from rural and urban settings, of all ethnicities, and from upper and lower SES levels, indicating that the UPSTART program closes the achievement gap between groups of students (Miner, 2014).

Figure 109: Year 5 WACS Reading Scores, Grouped by Subtests

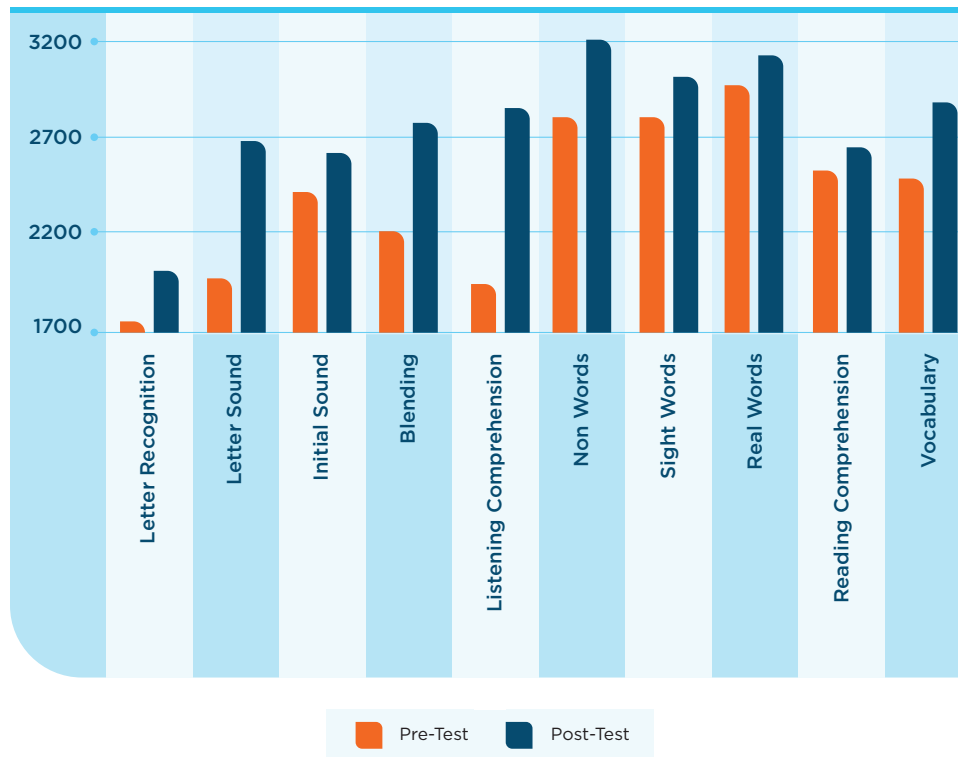


Table 73: WACS Year 5 Posttests

Measured Individual Skill	Final WACS Score	Grade Equivalent
Letter Recognition	2070	K Beginning
Letter Sound	2693	K Advanced
Initial Sound	2609	K Intermediate
Blending	2786	K Advanced
Listening Comprehension	2896	K Advanced
Non Words	3193	1st Beginning
Sight Words	3045	1st Beginning
Real Words	3129	1st Beginning
Reading Comprehension	2671	K Advanced
Vocabulary	2871	K Advanced

The following figures display the gains in WACS reading scores for UPSTART students by demographics.

Figure 110: WACS Reading Score Gains Grouped by Ethnicity

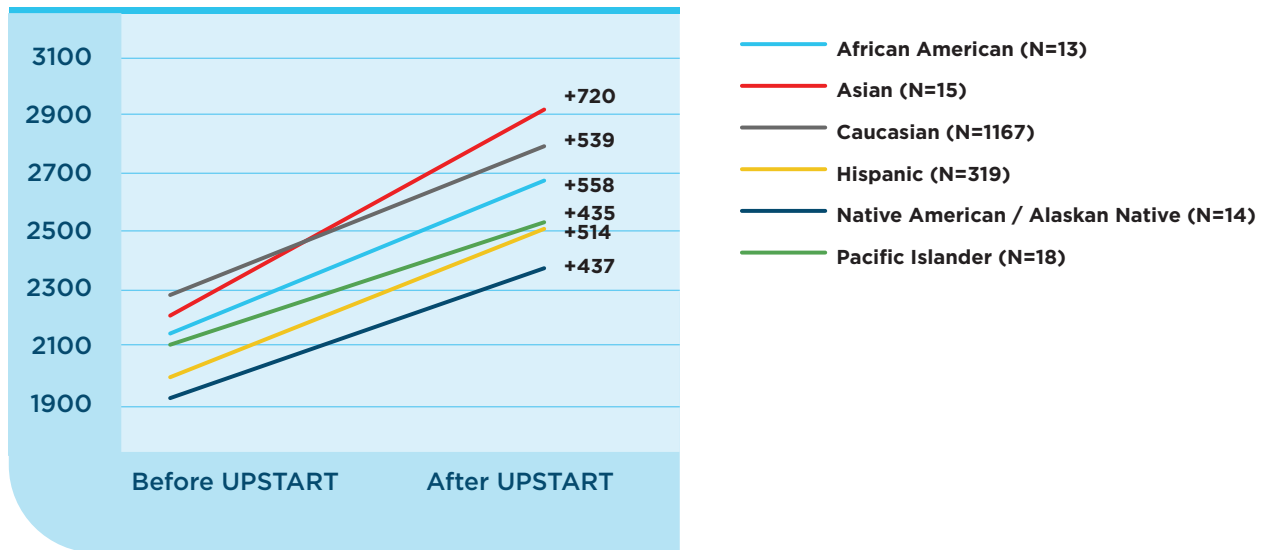


Figure 111: WACS Reading Score Gains Grouped by Other Preschool Attendance

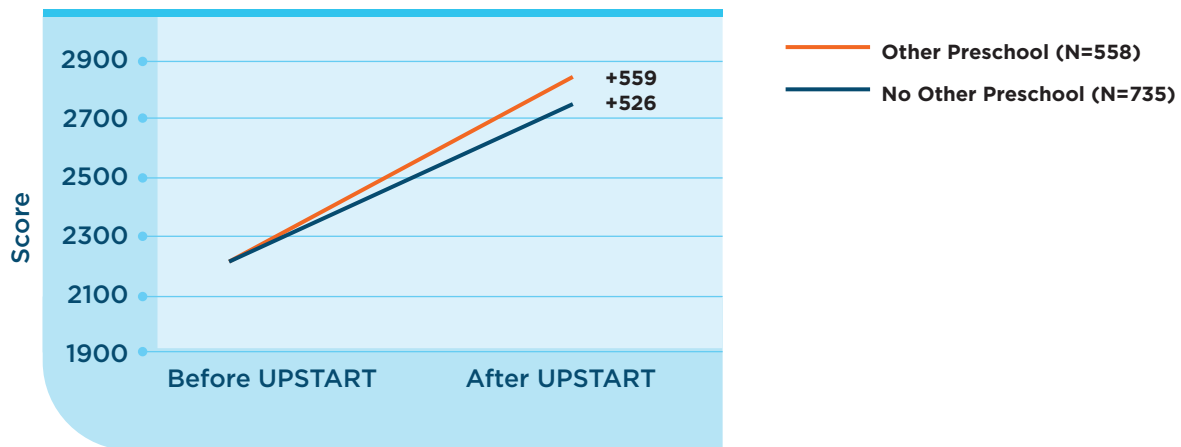


Figure 112: WACS Reading Score Gains Grouped by SES

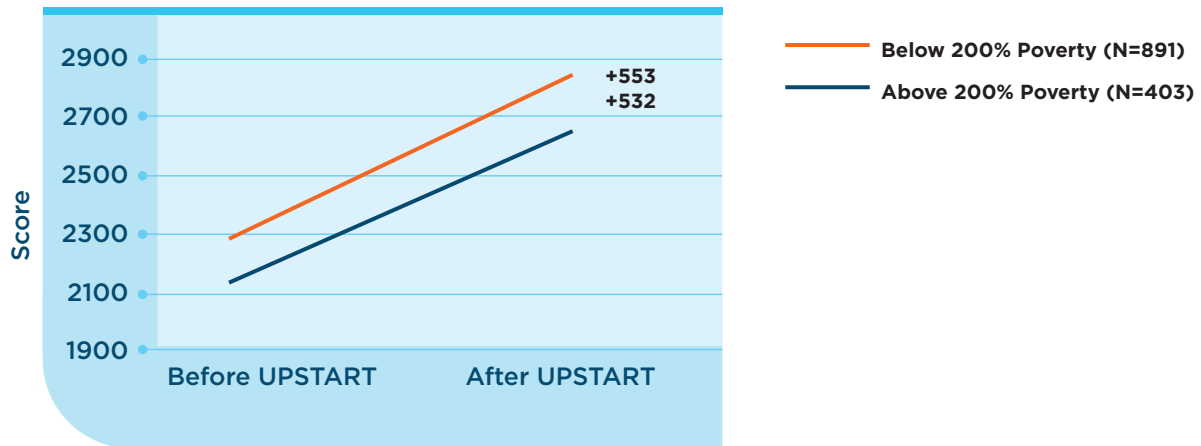


Figure 113: Year 6 WACS Reading Scores, Grouped by Subtests

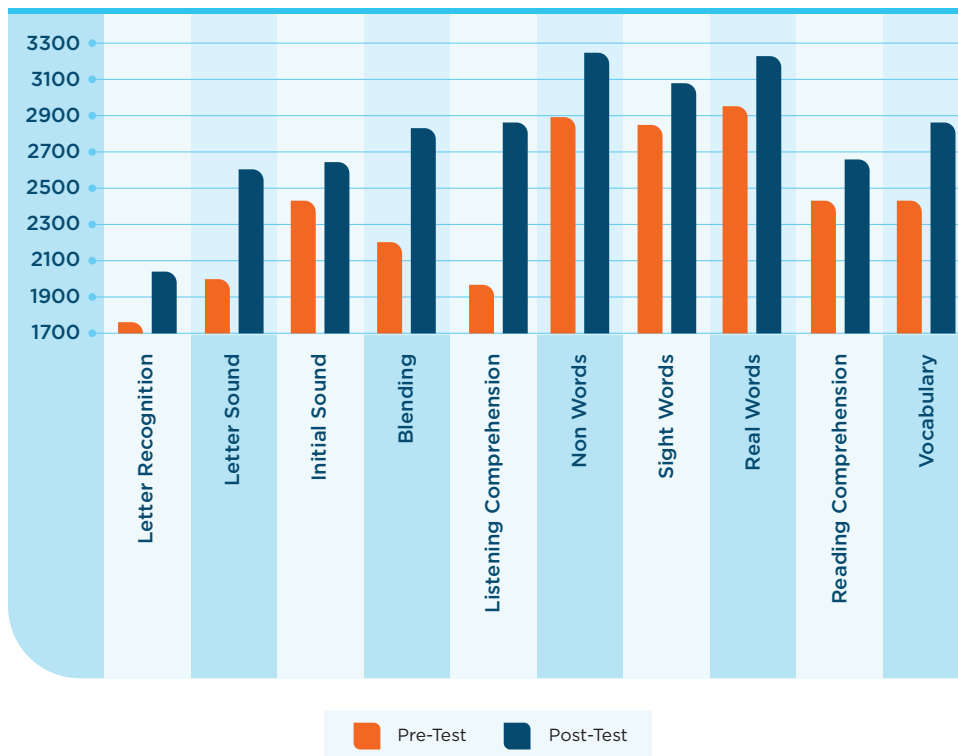


Table 74: WACS Year 6 Posttests

Measured Individual Skill	Final WACS Score	Grade Equivalent
Letter Recognition	2063	K Beginning
Letter Sound	2641	K Intermediate
Initial Sound	2616	K Intermediate
Blending	2815	K Advanced
Listening Comprehension	2838	K Advanced
Non Words	3204	1st Beginning
Sight Words	3036	1st Beginning
Real Words	3186	1st Beginning
Reading Comprehension	2649	K Intermediate
Vocabulary	2862	K Advanced

The following figures display the gains in WACS reading scores for UPSTART students by demographics.

Figure 114: WACS Reading Score Gains Grouped by Ethnicity

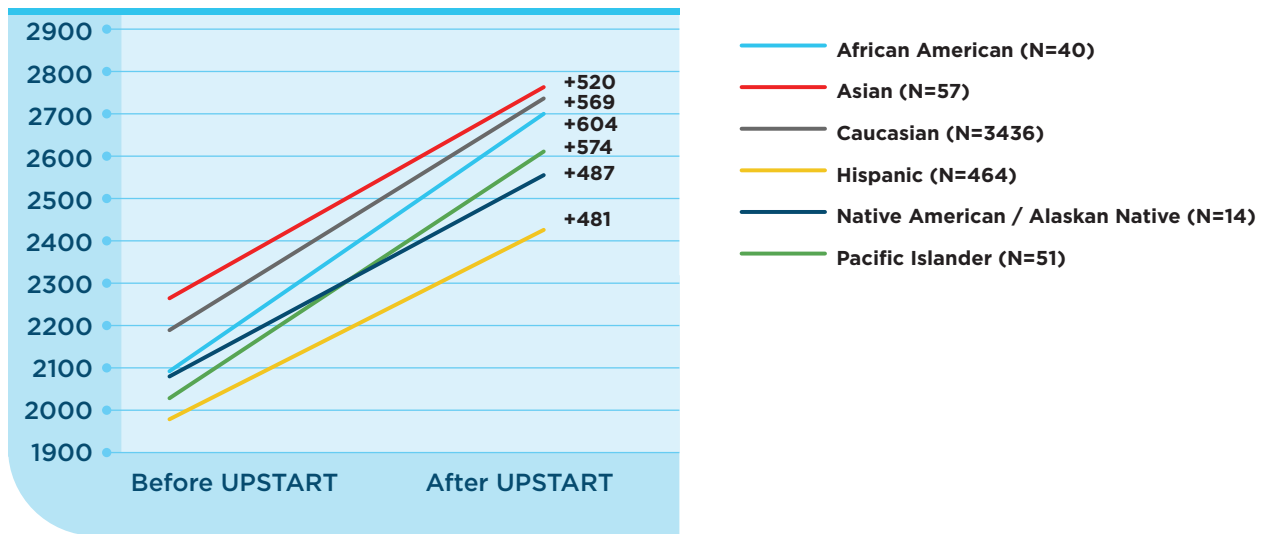


Figure 115: WACS Reading Score Gains Grouped by Other Preschool Attendance

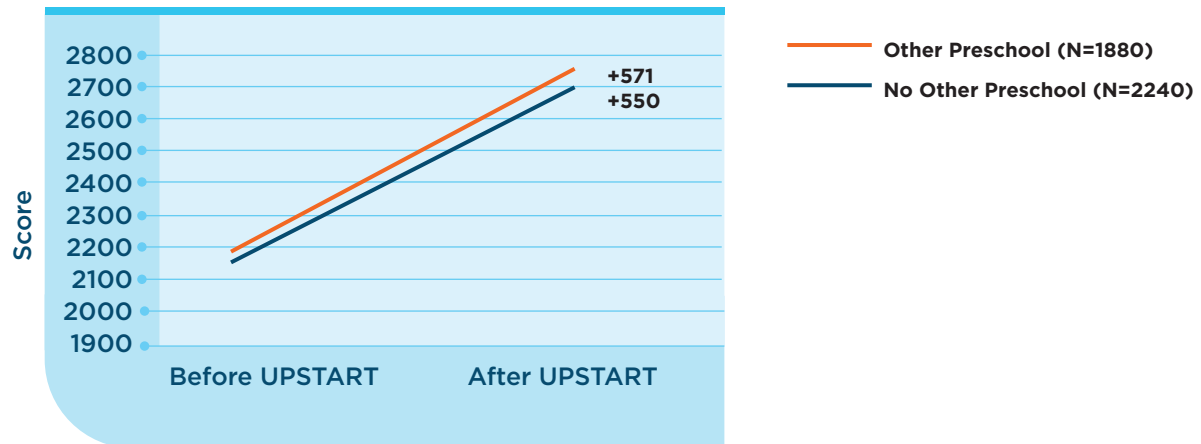


Figure 116: WACS Reading Score Gains Grouped by SES

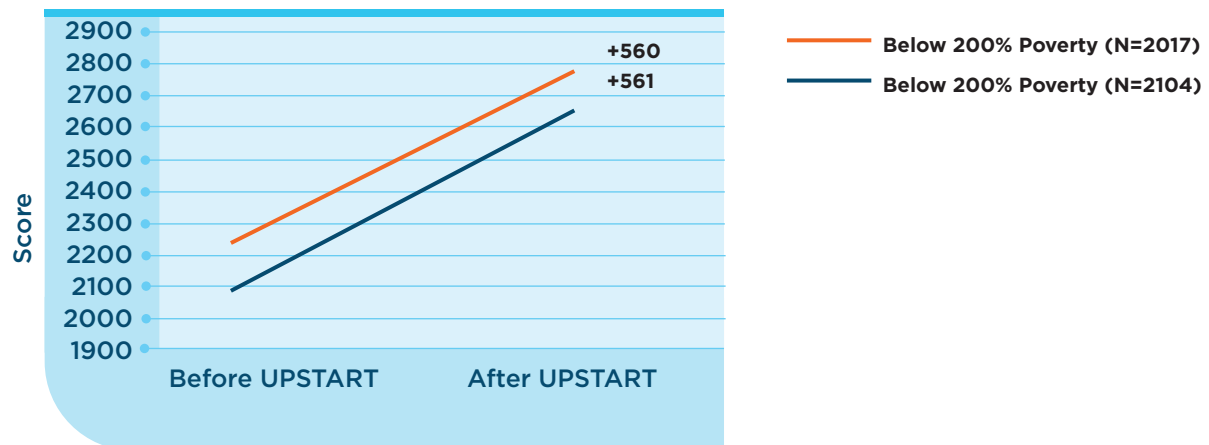
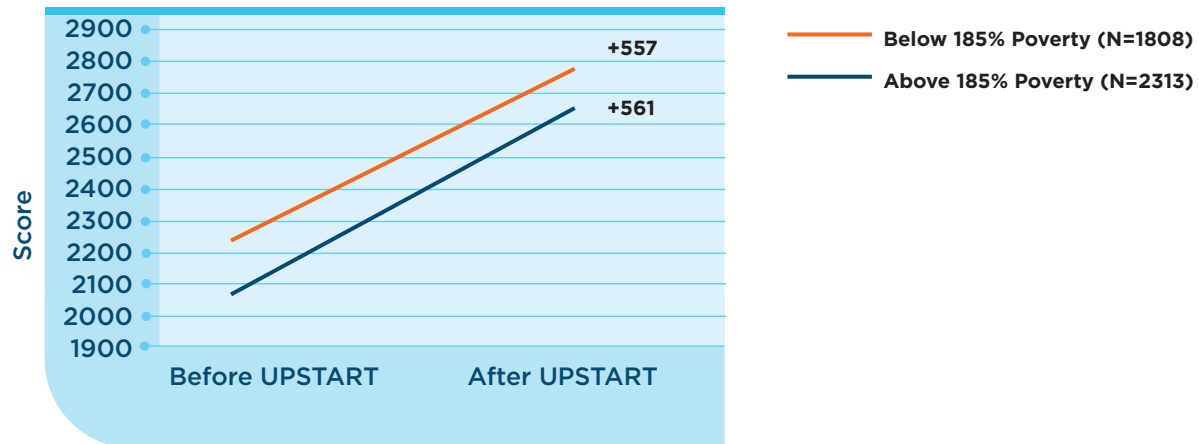


Figure 117: WACS Reading Score Gains Grouped by SES



## LONGITUDINAL EFFECTS

The following figure depicts longitudinal data from UPSTART compared to state averages on The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test scores and on Student Assessment of Growth and Excellence (SAGE) test scores (Suddreth, Throndsen, & Wiebke, 2016). This study shows that students who participated in the UPSTART program as preschoolers maintained their gains longitudinally on state testing, outscoring non-UPSTART students on state testing in Grades 1-4.

Figure 118: UPSTART Students & State Average DIBELS Scores, Grades 1 through 3

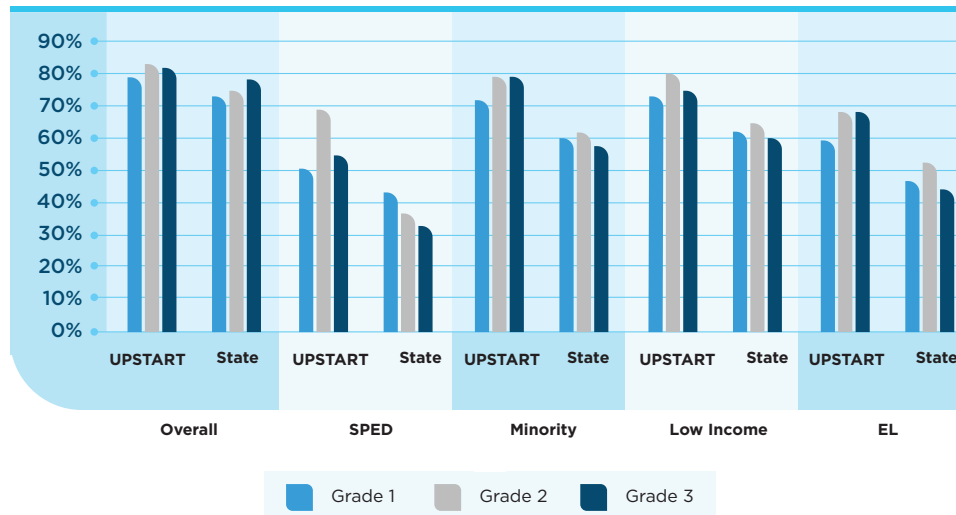


Figure 119: UPSTART Students & State Average SAGE Overall Scores, Grades 3 & 4

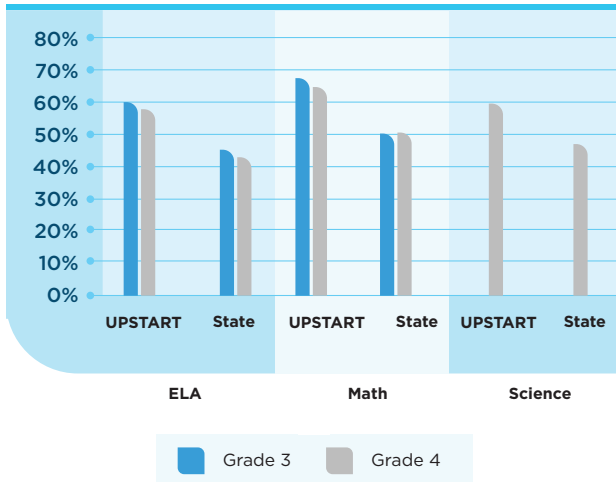


Figure 120: UPSTART Students & State Average SAGE SPED Scores, Grades 3 & 4

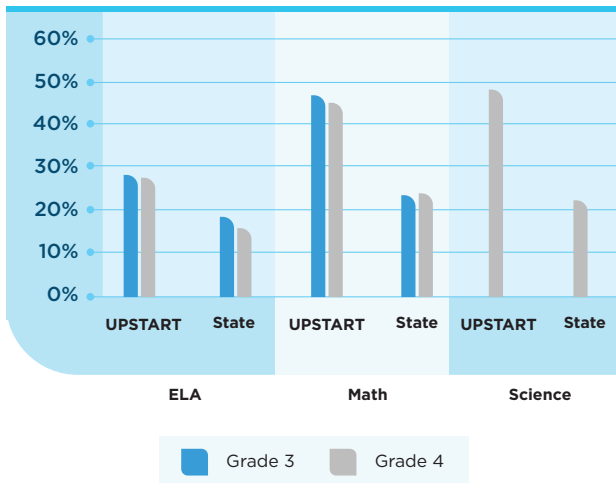




Figure 121: UPSTART Students & State Average SAGE Minority Scores, Grades 3 & 4

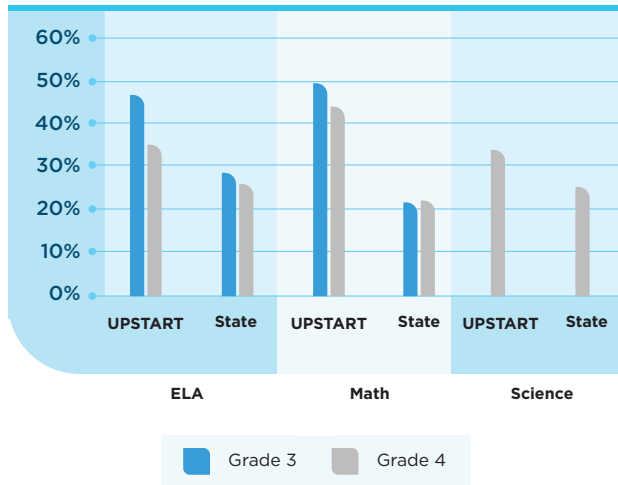


Figure 122: UPSTART Students & State Average SAGE Low Income Scores, Grades 3 & 4

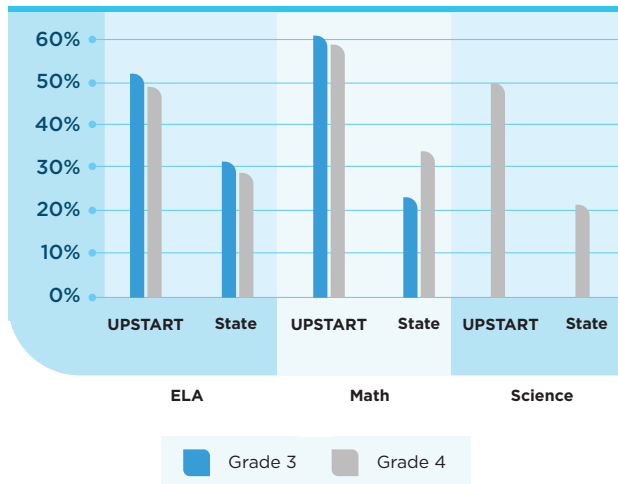
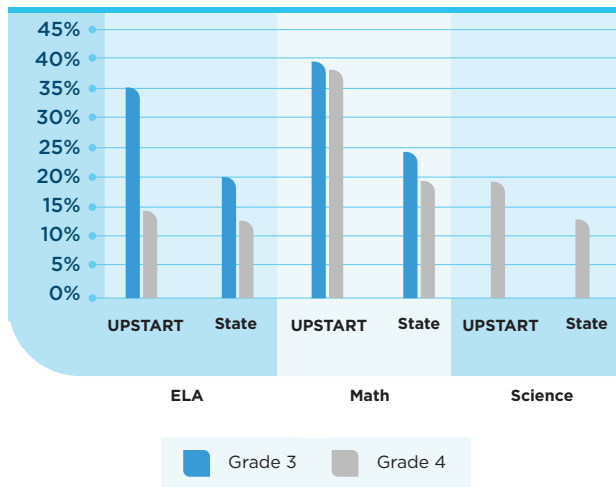


Figure 123: UPSTART Students & State Average SAGE EL Scores, Grades 3 & 4



## References

- Cassady, J. C., & Smith, L. L. (2003). The impact of a reading-focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research, 35*(4), 947-964.
- Cassady, J. C., & Smith, L. L. (2005). The impact of a structured integrated learning system on first grade students' reading gains. *Reading and Writing Quarterly, 21*, 361-376.
- Electronic Education. (2002). Los Angeles Unified school district - Waterford Early Reading Program initial implementation findings. Research Compendium, 2002.
- Evaluation and Training Institute. (2011). Kindergarten Outcomes: Program Impacts on Reading Proficiency. *First Year Results: Utah UPSTART Education Program Evaluation Technical Report*.
- Evaluation and Training Institute. (2012). Utah UPSTART Education Program Evaluation: Program Impacts on Early Literacy. *Cohort 2 Results: Technical Report*.
- Evaluation and Training Institute. (2013). Utah UPSTART Program Evaluation: Program Impacts on Early Literacy. *Third Year Results: Cohort 3 Technical Report*.
- Evaluation and Training Institute. (2014). Utah UPSTART Program Evaluation: Program Impacts on Early Literacy. *Fourth Year Results: Cohort 4 Technical Report*.
- Evaluation and Training Institute. (2015). Utah UPSTART Program Evaluation: Program Impacts on Early Literacy. *Year 5 Results: Cohort 5 Technical Report*.
- Evaluation and Training Institute. (2016). UPSTART Program Evaluation: Year 6 Program Results.
- Evaluation and Training Institute. (2017). UPSTART Program Evaluation: Year 7 Program Results.
- Evaluation and Training Institute. (2018a). UPSTART Program Evaluation: Year 8 Program Results.
- Evaluation and Training Institute. (2018b). Utah High-Quality School Readiness Expansion (HQSR-E) Program Evaluation.
- Hecht, S. A. (2000). Waterford Early Reading Program in Ohio: An Evaluation. Research Compendium, 2000.
- Hecht, S. A., & Close, L. (2002). Emergent literacy skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. *Journal of Experimental Child Psychology, 82*, 93-115.
- Hobbs, L. J., & Overby, M. (2019). Impact of UPSTART reading participation for rural students while controlling for prior reading achievement and school district. Retrieved from [https://docs.wixstatic.com/ugd/796969\\_313351e185694f4c84ab8cc1eae59dfa.pdf](https://docs.wixstatic.com/ugd/796969_313351e185694f4c84ab8cc1eae59dfa.pdf)

Hobbs, L. J., Overby, M., & Thomas, A. (2017). Evaluating the impact of the i3 UPSTART summer program on rural elementary school students' literacy skills. Manuscript in preparation (draft available at <https://www.eticonsulting.org/i3>).

Miner, C. (2014). 2014 Annual Report to the Utah Legislature.

Powers, S., & Price-Johnson, C. (2006). Evaluation of the Waterford Early Reading Program in Kindergarten, 2005-06. *Online Submission*.

Research, Assessment, & Measurement, Inc. (1998). A study of the effectiveness of the Waterford program at Glenridge elementary school. Research Compendium, 1998.

Research, Assessment, & Measurement, Inc. (1999). Evaluation of Waterford Early Reading Program: Hacienda la Puente and Whittier School Districts. Los Angeles, California. Research Compendium, 1999.

Reynolds, C. (2000). An evaluation of the Waterford Early Reading Program in the Decatur, Ill., school district: Analysis of impact on vocabulary development. Research Compendium, 2000.

Shamir, H., Feehan, K., Pocklington, D., & Yoder, E. (2019a, June). Dosage effects of CAI on literacy skills. Paper to be presented at the International KES Conference on Smart Education and E-Learning, St. Julians, Malta.

Shamir, H., Feehan, K., Pocklington, D., & Yoder, E. (2019b). Effects of long-term and early use of CAI on students' literacy skills. Paper to be presented at the EdMedia 2019—World Conference on Educational Media and Technology, Amsterdam, The Netherlands.

Shamir, H., Feehan, K., & Yoder, E. (2016). Using technology to improve reading and math scores for the digital native. In *Proceedings of EdMedia 2016--World Conference on Educational Media and Technology* (pp. 1405-1412). Vancouver, BC, Canada: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Feehan, K., & Yoder, E. (2017a). Computer-assisted instruction in early literacy for African American, economically disadvantaged children. *Journal of Educational Multimedia and Hypermedia*, 26(2), 179-192.

Shamir, H., Feehan, K., & Yoder, E. (2017b). Does CAI improve early math skills? In *Proceedings of the 9<sup>th</sup> International Conference on Computer Supported Education (CSEDU 2017) - Volume 2* (pp. 285-292). Porto, Portugal: Institute for Systems and Technologies of Information, Control and Communication.

Shamir, H., Feehan, K., & Yoder, E. (2017c). Effects of personalized learning on kindergarten and first grade students' early literacy skills. In *Proceedings of the 9<sup>th</sup> International Conference on Computer Supported Education (CSEDU 2017) - Volume 2* (pp. 273-279). Porto, Portugal: Institute for Systems and Technologies of Information, Control and Communication.

Shamir, H., Feehan, K., & Yoder, E. (2017d). Literacy improvement in early learners using technology. In *Proceedings of the 11<sup>th</sup> International Technology, Education and Development Conference* (pp. 7889-7896). Valencia, Spain: International Academy of Technology, Education and Development.

Shamir, H., Feehan, K., & Yoder, E. (2017e). Using computer-assisted instruction to improve early literacy skills for struggling students. In *Proceedings of the 11<sup>th</sup> International Technology, Education and Development Conference* (pp. 7897-7903). Valencia, Spain: International Academy of Technology, Education and Development.

Shamir, H., Feehan, K., Yoder, E., & Pocklington, D. (2018a). Can CAI improve literacy and mathematics achievement in young low socioeconomic status students? In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 993-998). Washington, D.C.: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Feehan, K., Yoder, E., & Pocklington, D. (2018b, May). *Can CAI improve reading achievement in young ELL students?* Paper presented at the Conference on Education, Teaching, and E-Learning, Prague, Czech Republic.

Shamir, H., & Goethe, R. (2015). Common Core State Standards: Is computer assisted instruction up for the challenge? In *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2015* (pp. 220-227). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Miner, C., Izzo, A., Feehan, K., Yoder, E., & Pocklington, D. (2019). Improving early literacy skills using technology at home. *International Journal of Learning and Teaching* (In Press).

Shamir, H., Miner, C., Izzo, A., Pocklington, D., Feehan, K., & Yoder, E. (2018). Preparing students for kindergarten using UPSTART at home. *Journal of Educational Multimedia and Hypermedia*, 27(2), 209-230.

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2018). Educational equity using computer-assisted instruction. In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 717-722). Washington, D.C.: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2019a). Bridging the achievement gap for low-performing students using computer-adaptive instruction. *International Journal of Information and Education Technology*, 9(3), 196-200.

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2019b). Evidence for dosage and long-term effects of computer-assisted instruction. *International Journal of Learning and Teaching* (In Press).

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2019c). Game-based learning for young learners. *International Journal of Learning and Teaching* (In Press).

Shamir, H., Pocklington, D., Yoder, E., & Feehan, K. (2019). Reinforcing second grade literacy skills using a computer-adaptive reading program (In Review).

- Shamir, H., Yoder, E., Feehan, K., & Pocklington, D. (2019, June). Randomized controlled trial of kindergarten students using literacy technology. Paper to be presented at International KES Conference on Smart Education and E-Learning, St. Julians, Malta.
- Shamir, H., Yoder, E., Pocklington, D., & Feehan, K. (2018a). Using adaptive CAI to supplement literacy development in early learners. *Journal of Educational Multimedia and Hypermedia*, 27(3), 367-389.
- Shamir, H., Yoder, E., Pocklington, D., & Feehan, K. (2018b). Using CAI for improving academic skills of students with special needs. In M. Carmo (Ed.), *Proceedings of the International Conference on Education and New Developments (END 2018)* (pp. 557-561). Budapest, Hungary: World Institute for Advanced Research and Science.
- Shamir, H., Yoder, E., Pocklington, D., & Feehan, K. (2019a). Computer-assisted instruction: Long-term effects on early literacy skills of low socioeconomic status students. *International Journal for Information and Education Technology*, 9(4), 263-267.
- Shamir, H., Yoder, E., Pocklington, D., & Feehan, K. (2019b). Technology improving literacy skills for all students: Findings from three districts. *International Journal of Information and Education Technology*, 9(4), 280-285.
- Suddreth, D., Throndsen, J., & Wiebke, S. (2016). UPSTART Program: Report of FY 2016. *Utah State Office of Education*.
- Walberg, H. J. (2001). Final evaluation of the reading initiative. Research Compendium, 2002.
- Waterford Institute. (1996). Preliminary research: Waterford Institute's Early Research in Utah and New York schools. Research Compendium, 1998.
- Waterford Institute. (1997). A preliminary report of the 1996-97 test results from elementary schools in the Dallas ISD on the effectiveness of the Waterford Early Reading Program. Research Compendium, 1998.
- What Works Clearinghouse. (2007). What Works Clearinghouse Intervention Report: Waterford Early Reading Program. Retrieved from What Works Clearinghouse: [https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/WWC\\_Waterford\\_071607.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/WWC_Waterford_071607.pdf).
- Young, J. W., & Tracey, D. H. (1999). An evaluation of the Waterford Early Reading Program: Newark, New Jersey. Research Compendium, 1998.