



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

<i>List of Tables</i>	4
<i>List of Figures</i>	6
<i>List of Acronyms</i>	11
<i>Waterford Reading Academy</i>	12
<i>Tier 1: Strong Evidence</i>	12
I3 Rural UPSTART Preschool Study	12
I3 WACS Randomized Control Trial.....	14
Randomized Controlled Trial Evaluation of Waterford Early Learning in Indiana.....	16
Randomized Controlled Trial Evaluation of Waterford Upstart in Nevada.....	21
Randomized Controlled Trial Evaluation of the Waterford Summer Learning Path in South Dakota and Wyoming.....	22
Education Innovation and Research Grant Evaluation	23
Multi-Year Randomized Controlled Trial Evaluation of Waterford Upstart in Nevada	26
<i>Tier 2: Moderate Evidence</i>	28
Evaluation of the Waterford Early Reading Program in Ohio	28
Effectiveness of UPSTART.....	30
Evaluation of Indiana UPSTART	47
High-Quality School Readiness Expansion in Utah, 2017-2018	47
High-Quality School Readiness Expansion in Utah, 2018-2019	49
Impact of the Upstart Program on Forestalling Summer Learning Loss	51
<i>Tier 3: Promising Evidence</i>	52
Evaluation of a Tucson Unified School District	52
Evaluation of a School District in Florida, 2014-2015	54
Evaluation of the Waterford Early Reading Program in California	58
Evaluation of the Waterford Early Math and Science Program in Indiana, 2015-2016	59
Effectiveness of UPSTART.....	67
Evaluation of a School District in South Carolina, 2016-2017	78
Evaluation of Waterford Early Learning in Maryland, 2016-2017	79
Evaluation of Waterford Early Learning in Tennessee	84
Evaluation of a School District in South Carolina, 2017-2018	87
Evaluation of Waterford Early Learning in Maryland, 2017-2018	89

Evaluation of the Efficacy of Waterford Summer Learning Path.....	91
Evaluation of Usage Fidelity in the Waterford Summer Learning Path	93
Evaluation of the Waterford Early Math and Science Program in Illinois, 2021-2022.....	95
Evaluation of the Waterford Reading Academy in Illinois	98
Tier 4: Demonstrated Rationale	102
Early Studies	102
Case Studies	103
Later Effectiveness Studies.....	103
Evaluation of the Waterford Early Reading Program in Alabama	105
Evaluation of the Waterford Early Reading Program in Colorado.....	107
Evaluation of the Waterford Early Reading Program in South Carolina	108
Evaluation of the Waterford Early Reading Program in Florida	110
Evaluation of the Waterford Early Reading Program in Indiana, 2015-2016	113
Evaluation across Two School Districts in Texas and Florida, 2015-2016	116
Evaluation across Two School Districts in South Carolina and Indiana, 2015-2016	118
Evaluation across Two School Districts in Texas and Indiana, 2015-2016	119
Evaluation of a School District in Texas, 2015-2016	121
Evaluation of Waterford Early Learning in Idaho, 2017-2018	125
Evaluation of Waterford Early Learning in Three School Districts.....	127
Longitudinal Evaluation of Waterford Early Learning in Florida, 2014-2016	129
Longitudinal Evaluation of Waterford Early Learning in South Carolina, 2015-2017	132
Longitudinal Evaluation of a School District in Maryland, 2015-2018	134
Longitudinal Evaluation of Three School Districts, 2015-2019	137
Longitudinal Evaluation of Waterford Upstart in Mississippi, 2018–2020.....	140
Longitudinal Evaluation of Waterford Early Learning in Maryland, 2015-2018	141
Longitudinal Evaluation of Waterford Early Learning in Texas, 2018-2020	143
Longitudinal Evaluation of a School District in Texas, 2019-2022	147
Effectiveness of UPSTART.....	150
References.....	163

List of Tables

Table 1: Brigance Subscale Posttest Means of Treatment-Control Groups.....	13
Table 2: Brigance Subscale Posttest Effect Size Estimates	13
Table 3: PELI Subscale Posttest Means of Treatment-Control Groups.....	13
Table 4: PELI Subscale Posttest Effect Size Estimates.....	14
Table 5: WACS End of Year Scores Controlling for Beginning of Year Scores	14
Table 6: Overall WACS End of Year Scores by Demographics.....	15
Table 7: Blending WACS End of Year Scores by Demographics	15
Table 8: Letter Sound WACS End of Year Scores by Demographics.....	16
Table 9: Letter Recognition WACS End of Year Scores by Demographics.....	16
Table 10: Baseline Equivalence of Treatment-Control Groups by Construct	27
Table 11: Descriptive Statistics for All Tasks	29
Table 12: Brigance Growth Rate Comparisons	34
Table 13: Bader Growth Rate Comparisons.....	34
Table 14: Effect Size Estimates by Literacy Domain	48
Table 15: Effect Size Estimates by Literacy Domain	50
Table 16: Baseline Equivalence Assessment for UPSTART Summer Program Participants	51
Table 17: Impact Analysis Results for UPSTART Summer Program Cohort 1 Participants.....	52
Table 18: MLL & Non-MLL Students on DIBELS Total Reading Score.....	53
Table 19: ANCOVA & Effect Sizes on All Outcome Measures (ERP Students With 1100 or More Usage Minutes)	53
Table 20: Spring VPK Sub-Strand Scores, Covarying for Fall	54
Table 21: Effects of Demographics on Scores for Each Sub-Strand	55
Table 22: Brigance Posttest Analysis of Treatment-Control Group Differences.....	68
Table 23: Bader Posttest Analysis of Treatment-Control Group Differences	68
Table 24: Treatment-Control Group Differences in Growth Rates on the Brigance	69
Table 25: Treatment-Control Group Differences in Growth Rates on the Bader	69
Table 26: Brigance Posttest Analysis of Treatment-Control Group Differences.....	70
Table 27: Bader Posttest Analysis of Treatment-Control Group Differences	71
Table 28: Treatment-Control Group Differences in Growth Rates on the Brigance	72
Table 29: Treatment-Control Group Differences in Growth Rates on the Bader	73
Table 30: Brigance Posttest Analysis of Treatment-Control Group Differences.....	73
Table 31: Brigance Effect Size Estimates	74
Table 32: Bader Posttest Analysis of Treatment-Control Group Differences	74
Table 33: Bader Effect Size Estimates	75
Table 34: Treatment-Control Group Differences in Growth Rates on the Brigance	75
Table 35: Treatment-Control Group Differences in Growth Rates on the Bader	75
Table 36: Brigance Posttest Analysis of Treatment-Control Group Differences.....	76
Table 37: Brigance Effect Size Estimates	77
Table 38: Bader Posttest Analysis of Treatment-Control Group Differences	77
Table 39: Bader Effect Size Estimates	77
Table 40: Treatment-Control Group Differences in Growth Rates on the Brigance	78
Table 41: Treatment-Control Group Differences in Growth Rates on the Bader	78
Table 42: Kindergarten DRA Gains & End of Year Scores Covarying For Middle of Year Scores.....	79
Table 43: First Grade DRA Gains & End of Year Scores Covarying For Middle of Year Scores.....	79
Table 44: Assessment Scores by Demographics	94

Table 45: MAP Growth Math Beginning of Year Scores by Subskill.....	95
Table 46: MAP Growth Reading Beginning of Year Scores by Subskill	99
Table 47: Performance Measure Means at the Beginning, Middle, & End of the Year	104
Table 48: Beginning of Year First Grade & Second Grade Terra Nova Reading Performance Group Means	105
Table 49: District 1 Overall Kindergarten TPRI Gains by Substrand.....	117
Table 50: District 1 Overall Kindergarten TPRI Gains by MLL Students	117
Table 51: District 1 Overall Second Grade TPRI Gains by Substrand	117
Table 52: District 1 Overall Second Grade TPRI Gains by MLL Students.....	117
Table 53: District 2 Overall VLT End of Year Scores	117
Table 54: District 2 VLT End of Year Scores by MLL Students	118
Table 55: District 1 Kindergarten DRA End of Year Scores.....	118
Table 56: District 1 First Grade DRA Scores	118
Table 57: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores	118
Table 58: District 2 Kindergarten End of Year Scores Covarying for Middle of Year Scores by Low-SES.	119
Table 59: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores	119
Table 60: District 2 Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Low-SES	119
Table 61: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores.....	119
Table 62: District 2 First Grade End of Year Scores Covarying for Beginning of Year Scores by Low-SES	119
Table 63: Kindergarten TPRI Gains by Special Education Services.....	123
Table 64: Kindergarten TPRI Gains by MLL Status	123
Table 65: Kindergarten TPRI Gains by Race/Ethnicity	123
Table 66: First Grade TPRI Gains by Special Education Services.....	124
Table 67: First Grade TPRI Gains by MLL Status	124
Table 68: First Grade TPRI Gains by Race/Ethnicity.....	125
Table 69: District 1 – Kindergarten TPRI Gains by Race/Ethnicity	127
Table 70: District 1 – First Grade TPRI Gains by Race/Ethnicity.....	128
Table 71: District 2 – Second Grade DRA Percent Gains by Race/Ethnicity.....	128
Table 72: District 3 – Kindergarten VLT End of Year Scores by Race/Ethnicity.....	128
Table 73: District 3 – First Grade VLT End of Year Scores by Race/Ethnicity	129
Table 74: End of Kindergarten Scores by Strand & Demographics.....	132
Table 75: End of First Grade DRA Scores	133
Table 76: End of Second Grade DRA Scores	133
Table 77: End of First Grade High Usage DRA Scores	133
Table 78: End of Second Grade High Usage DRA Scores.....	133
Table 79: Participants by District and Usage Group Comparisons	138
Table 80: WACS Question Difficulty Ranges by Grade.....	151
Table 81: WACS Year 5 Posttests	152
Table 82: WACS Year 6 Posttests	154
Table 83: PEEP Exit Raw Scores for Literacy and Numeracy by Support Tier	161

List of Figures

Figure 1: MAP End of Year Scores by Strand	17
Figure 2: MAP RIT Scores by Demographics	18
Figure 3: MAP Reading Foundations Scores by Demographics	18
Figure 4: MAP End of Year Scores - African American/Black Students.....	19
Figure 5: TRC End of Year Lexile Levels.....	19
Figure 6: TRC End of Year Lexile Levels by Demographics	20
Figure 7: TRC End of Year Proficiency Benchmarks	20
Figure 8: Predicted Treatment and Control Math Post-Test Mean Scores Prior to the Start of Kindergarten	21
Figure 9: Predicted Treatment and Control Math Post-Test Mean Scores After the End of Kindergarten	21
Figure 10. Predicted Means of the MCA Growth Scale Value by Treatment Condition and Language Group.....	22
Figure 11: Adjusted Mean Score Comparison at Entrance to Kindergarten.....	23
Figure 12: Math Core Outcomes for Low and High Performers	23
Figure 13: Cohort 1 Upstart Program Effect Sizes	24
Figure 14: Cohort 1 KTEA Literacy and Math Performance	24
Figure 15: Cohort 2 Upstart Program Effect Sizes	25
Figure 16: Cohort 2 KTEA Literacy and Math Performance	25
Figure 17: Cohort 3 Upstart Program Effect Sizes	26
Figure 18: Cohort 3 KTEA Literacy and Math Performance	26
Figure 19: Predicted Treatment and Control Reading Comprehension Mean Scores at Post-Test.....	27
Figure 20: Predicted Treatment and Control Phonological Processing Post-Test Mean Score	27
Figure 21: Predicted Treatment and Control Math Post-Test Mean Score	28
Figure 22: Effect Size Estimates Based on Adjusted Mean Standard Scores by Measure	28
Figure 23: Brigance & Bader Posttest Analysis of Composite Scores	31
Figure 24: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests	31
Figure 25: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores.....	32
Figure 26: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores.....	32
Figure 27: Pre-Literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores	33
Figure 28: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores	33
Figure 29: Brigance & Bader Posttest Analysis of Composite Scores	35
Figure 30: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests	35
Figure 31: Phonological Awareness: Treatment and Control Group Pretest & Posttest Mean Scores	36
Figure 32: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores.....	36
Figure 33: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores	37
Figure 34: Brigance & Bader Posttest Analysis of Composite Scores	38
Figure 35: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests	38
Figure 36: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores.....	39
Figure 37: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores.....	39
Figure 38: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores	40
Figure 39: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores....	40
Figure 40: Brigance & Bader Posttest Analysis of Composite Scores	41
Figure 41: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests	41
Figure 42: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores.....	42
Figure 43: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores.....	42

Figure 44: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores	43
Figure 45: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores	43
Figure 46: Brigance & Bader Posttest Analysis of Composite Scores	44
Figure 47: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests	45
Figure 48: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores.....	45
Figure 49: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores.....	46
Figure 50: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores	46
Figure 51: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores	47
Figure 52: Listening Comprehension Benchmarks at Posttest	48
Figure 53: Post-Program Social-Emotional Development Skills by Program Group	49
Figure 54: Post-Program Social-Emotional Development Skills by Program Group	50
Figure 55: Acadience Reading Composite Scores for UPSTART and Control Groups	50
Figure 56: Spring VPK Sub-Strand Scores, Covarying for Fall.....	55
Figure 57: Oral Language Vocabulary Spring Scores by Demographics.....	56
Figure 58: Phonological Awareness Spring Scores by Demographics.....	56
Figure 59: Print Knowledge Spring Scores by Demographics	57
Figure 60: Math Spring Scores by Demographics	57
Figure 61: Mean Scores by AIMSweb Sub-Strands for Kindergarten Students	58
Figure 62: Mean Scores by AIMSweb Sub-Strand NWF for First Grade Students	58
Figure 63: Kindergarten mCLASS: Math End of Year Scores by Sub-Strand	60
Figure 64: Kindergarten mCLASS: Math Counting End of Year Scores	60
Figure 65: Number Identification End of Year Scores by Demographics	61
Figure 66: Quantity Discrimination End of Year Scores by Demographics	62
Figure 67: First Grade mCLASS: Math End of Year Scores by Sub-Strand.....	63
Figure 68: First Grade mCLASS: Math Counting End of Year Scores.....	63
Figure 69: First Grade Number Facts End of Year Scores by Demographics.....	64
Figure 70: First Grade Quantity Discrimination End of Year Scores by Demographics.....	65
Figure 71: First Grade Missing Number End of Year Scores by Demographics.....	66
Figure 72: First Grade Quantity Discrimination End of Year Scores by Demographics.....	67
Figure 73: UPSTART's Impact on the Brigance in Effect Size Units	71
Figure 74: UPSTART's Impact on the Bader in Effect Size Units	72
Figure 75: Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Strand	80
Figure 76: Kindergarten End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics.....	80
Figure 77: Kindergarten End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics.....	80
Figure 78: Kindergarten End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics.....	81
Figure 79: Kindergarten End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics.....	81
Figure 80: First Grade End of Year Scores Covarying for Beginning of Year Scores by Strand.....	81
Figure 81: First Grade End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics.....	82
Figure 82: First Grade End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics.....	82
Figure 83: First Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics.....	82

Figure 84: First Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics.....	83
Figure 85: Second Grade End of Year Scores Covarying for Beginning of Year Scores by Strand.....	83
Figure 86: Second Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics.....	83
Figure 87: Second Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics.....	84
Figure 88: Kindergarten Beginning & End of Year Overall RIT Scores.....	84
Figure 89: Kindergarten Beginning & End of Year RIT Scores by Strand.....	85
Figure 90: Kindergarten End of Year RIT Scores while Covarying for Beginning of Year Scores	85
Figure 91: First Grade Beginning & End of Year RIT Scores	86
Figure 92: First Grade Beginning & End of Year RIT Scores by Strand	86
Figure 93: First Grade End of Year RIT Scores while Covarying for Beginning of Year Scores	87
Figure 94: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Strand.....	88
Figure 95: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Special Education Services	88
Figure 96: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Lunch Status.....	89
Figure 97: Second Grade MAP End of Year Scores by Strand	90
Figure 98: Second Grade MAP End of Year RIT to Reading Scores	90
Figure 99: Second Grade MAP End of Year Scores Controlling for Beginning of Year Scores by Strand ...	91
Figure 100: Second Grade MAP End of Year RIT to Reading Scores Controlling for Beginning of Year Scores	91
Figure 101: Average WACS from the Beginning of the SLP Program and the End of the Program	92
Figure 102: Before and After SLP WACS Scores by Demographic Factors	93
Figure 103: Before and After SLP WACS Scores by State.....	93
Figure 104: Beginning and End of Program Assessment Scores	94
Figure 105: Map Growth Math End of Year Scores by Subskill.....	96
Figure 106: Overall RIT Math End of Year Scores by Special Education Services	96
Figure 107: Operations and Algebraic Thinking End of Year Scores by Special Education Services	97
Figure 108: Number and Operations End of Year Scores by Special Education Services	97
Figure 109: Measurement and Data End of Year Scores by Special Education Services	98
Figure 110: Geometry End of Year Scores by Special Education Services	98
Figure 111: MAP Growth Reading End of Year Scores by Subskill	99
Figure 112: Overall RIT Reading End of Year Scores by Demographics	100
Figure 113: Foundational Skills End of Year Scores by Demographics.....	100
Figure 114: Language and Writing End of Year Scores by Demographics	101
Figure 115: Literature and Informational Text End of Year Scores by Demographics	101
Figure 116: Vocabulary Use and Functions End of Year Scores by Demographics	102
Figure 117: Kindergarten Students' Phonological Awareness Growth.....	104
Figure 118: Reading Gains for Students with Low, Moderate, & High Reading Skills at First Grade.....	105
Figure 119: Mean Gains by STAR Sub-Strands for First Sample Group.....	106
Figure 120: Mean Gains by STAR Sub-Strands for Second Sample Group.....	106
Figure 121: Mean Posttest Scores for DIBELS Sub-Strands for Third Sample Group.....	107
Figure 122: Mean Reading Gains by DRA Overall Kindergarten Scores	107
Figure 123: Kindergarten DRA End of Year Scores.....	108

Figure 124: Kindergarten DRA End of Year Scores by Demographics.....	109
Figure 125: First Grade DRA End of Year Scores.....	109
Figure 126: First Grade DRA End of Year Scores by Demographics	110
Figure 127: End of Year VLT Scores by Grade	111
Figure 128: Kindergarten End of Year VLT Scores by Demographics.....	111
Figure 129: First Grade End of Year VLT Scores by Demographics	112
Figure 130: Second Grade End of Year VLT Scores by Demographics	112
Figure 131: First Grade DRA Percent Gains	113
Figure 132: First Grade DRA Percent Gains by Demographics.....	114
Figure 133: Second Grade DRA Percent Gains.....	115
Figure 134: Second Grade DRA Percent Gains by Demographics.....	116
Figure 135: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Whole Words Read.....	120
Figure 136: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Letter Naming Fluency & Correct Letter Sounds	120
Figure 137: mCLASS DIBELS Next and Math Scores by Strand by Special Education Control—Number Identification & Quantity Discrimination.....	120
Figure 138: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Whole Words Read	120
Figure 139: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Letter Naming Fluency & Correct Letter Sounds.....	120
Figure 140: mCLASS DIBELS Next and Math Scores by Strand by Non-Special Education Control—Number Identification & Quantity Discrimination.....	120
Figure 141: TPRI Gain Scores by Strand by Special Education Control	121
Figure 142: TPRI Gain Scores by Strand by Non-Special Education Control	121
Figure 143: Kindergarten TPRI Gains by Substrand	122
Figure 144: Kindergarten TPRI Gains by Substrand	122
Figure 145: First Grade TPRI Gains by Substrand	124
Figure 146: STAR Early Literacy Scaled Score Gains from Beginning of Year to End of Year	126
Figure 147: STAR Early Literacy Gains from Beginning of Year to End of Year by Strand	126
Figure 148: IRI Gains from Beginning of Year to End of Year by Strand	127
Figure 149: KRT Overall End of Year Scores by Demographics	130
Figure 150: KRT Letter ID & Sounds End of Year Scores by Demographics	130
Figure 151: KRT Concepts of Print End of Year Scores by Demographics.....	131
Figure 152: KRT Phonemic Awareness End of Year Scores by Demographics	131
Figure 153: Three Years of Usage vs. No Usage End of Year Scores by Strand.....	134
Figure 154: Three Years of Usage vs. No Usage RIT to Reading End of Year Scores.....	135
Figure 155: Two Years of Usage vs. No Usage End of Year Scores by Strand	135
Figure 156: Two Years of Usage vs. No Usage RIT to Reading End of Year Scores	136
Figure 157: Three Years of Usage vs. Two Years of Usage End of Year Scores by Strand.....	136
Figure 158: Three Years of Usage vs. Two Years of Usage RIT to Reading End of Year Scores.....	137
Figure 159: Overall Effects - Second Grade Scores	138
Figure 160: Long-Term Effects - End of Second Grade Scores.....	139
Figure 161: Early Effects - End of Second Grade Scores	139
Figure 162: KTEA-3 End of Kindergarten Literacy Scores for Treatment and Comparison Groups	140
Figure 163: KTEA-3 End of Kindergarten Math Scores for Treatment and Comparison Groups	141
Figure 164: Three Years of Usage Vs. No Usage End of Second Grade Scores	142

Figure 165: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores.....	142
Figure 166: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores	143
Figure 167: Three Years of Usage Vs. No Usage End of Second Grade Scores	144
Figure 168: Three Years of Usage Vs. No Usage End of Second Grade Scores by Demographics	144
Figure 169: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores.....	145
Figure 170: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores by Demographics.....	145
Figure 171: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores	146
Figure 172: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores by Demographics	146
Figure 173: 2020-2021 End of First Grade TPRI Word Reading Scores.....	147
Figure 174: 2020-2021 End of First Grade TPRI Word Reading Scores by Demographics	147
Figure 175: 2021-2022 End of First Grade TPRI Word Reading Scores.....	148
Figure 176: 2021-2022 End of First Grade TPRI Word Reading Scores by Demographics	148
Figure 177: Combined End of First Grade TPRI Word Reading Scores	149
Figure 178: Combined End of First Grade TPRI Word Reading Scores by Demographics.....	149
Figure 179: Year 1-6 WACS Overall Scores	151
Figure 180: Year 5 WACS Reading Scores, Grouped by Subtests.....	152
Figure 181: WACS Reading Score Gains Grouped by Race/Ethnicity.....	153
Figure 182: WACS Reading Score Gains Grouped by Other Preschool Attendance	153
Figure 183: WACS Reading Score Gains Grouped by SES	153
Figure 184: Year 6 WACS Reading Scores, Grouped by Subtests.....	154
Figure 185: WACS Reading Score Gains Grouped by Race/Ethnicity.....	155
Figure 186: WACS Reading Score Gains Grouped by Other Preschool Attendance	155
Figure 187: WACS Reading Score Gains Grouped by SES	156
Figure 188: WACS Reading Score Gains Grouped by SES	156
Figure 189: UPSTART Students & State Average DIBELS Scores, Grades 1 through 3	157
Figure 190: UPSTART Students & State Average SAGE Overall Scores, Grades 3 & 4	157
Figure 191: UPSTART Students & State Average SAGE SPED Scores, Grades 3 & 4.....	158
Figure 192: UPSTART Students & State Average SAGE Minority Scores, Grades 3 & 4	158
Figure 193: UPSTART Students & State Average SAGE Low Income Scores, Grades 3 & 4	159
Figure 194: UPSTART Students & State Average SAGE MLL Scores, Grades 3 & 4	159
Figure 195: Upstart Student Proficiency Level at PEEP Entry	160
Figure 196: Upstart Student Performance Level at PEEP Exit.....	160
Figure 197: Kindergarten Readiness Among Those “Needing Support” at Entry for Literacy	161
Figure 198: Kindergarten Readiness Among Those “Needing Support” at Entry for Numeracy	162

List of Acronyms

ACRONYM	MEANING
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DRA	Developmental Reading Assessment
MLL	Multi-lingual Learners
EMS	Waterford Early Math and Science Program
ERP	Waterford Early Reading Program
ES	Effect Size
ILS	Integrated Learning System
KRT	Kindergarten Readiness Test
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
WRA	Waterford Reading Academy
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 1: 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.

I3 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, 2019a). 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	
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

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

I3 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. Additional analysis was conducted to investigate the impact of assigned group and demographics on the end of program WACS scores. The improvement in literacy was seen across gender, socioeconomic status, special education services, and whether students attended another preschool. The interaction between race/ethnicity and treatment was not significant, which indicates that the Waterford reading program had a similar impact on WACS scores for Caucasian/White and Latino/a students. 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

	<u>Experimental</u>			<u>Control</u>			<i>p</i>	<i>g</i>
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>		
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			<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	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
Gender								
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
Other Preschool								
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
Race/Ethnicity								
Caucasian/White	2624.48	358.80	243	2483.10	365.82	221	.00**	0.41
Latino/a	2401.61	424.90	19	2304.63	338.36	19	.36	0.09
Special Education Services								
No Active Special Education Services	2602.98	375.34	256	2468.22	363.34	236	.00**	0.39
Active Special Education Services	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			<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	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
Gender								
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
Other Preschool								
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
Race/Ethnicity								
Caucasian/White	2694.42	702.92	240	2513.20	688.76	220	.01**	0.25
Latino/a	2473.85	944.77	19	2181.40	590.41	19	.19	0.11
Special Education Services								
No Active Special Education Services	2681.43	729.88	256	2504.63	689.90	236	.01**	0.25
Active Special Education Services	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			<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	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
Gender								
Female	2557.28	543.79	141	2273.01	551.91	114	.00**	0.37
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
Race/Ethnicity								
Caucasian/White	2560.42	551.22	242	2275.63	537.37	220	.00**	0.51
Latino/a	2282.17	585.34	19	2219.49	512.13	19	.71	0.03
Special Education Services								
No Active Special Education Services	2527.30	565.56	256	2262.73	530.32	236	.00**	0.49
Active Special Education Services	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
Race/Ethnicity								
Caucasian/White	2021.44	209.73	243	1955.44	224.57	220	.00**	0.30
Latino/a	1964.15	331.11	19	1915.54	252.05	19	.48	0.06
Special Education Services								
No Active Special Education Services	2016.96	223.44	256	1957.05	224.55	236	.00**	0.28
Active Special Education Services	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

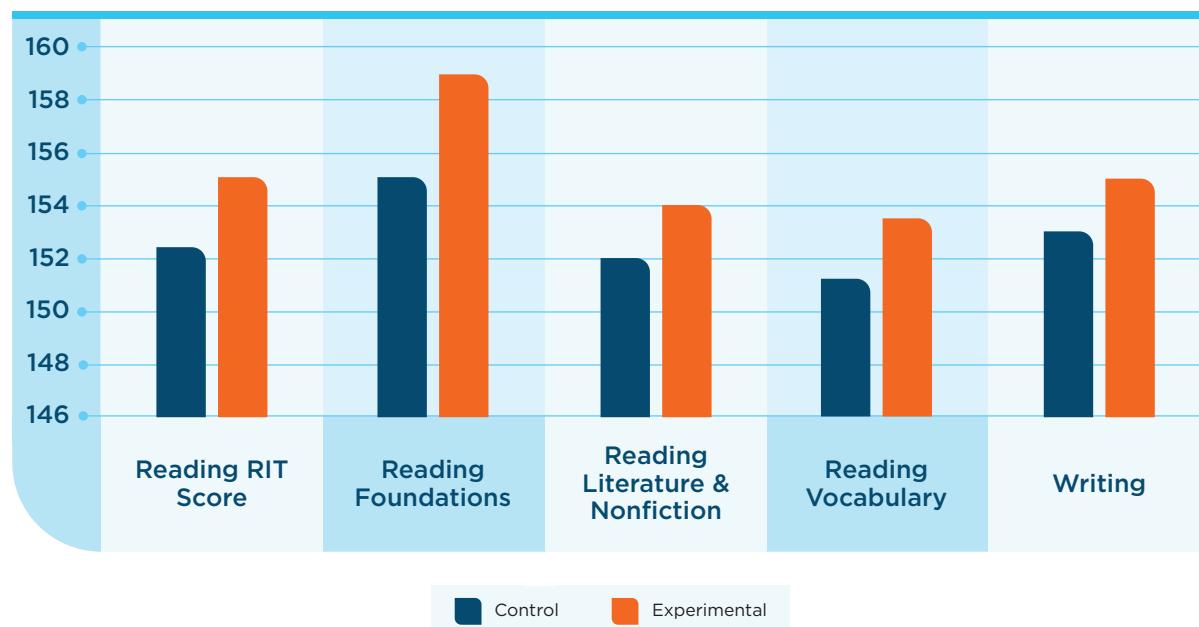
2017-2018 school year for 15 minutes per day, five days per week. Eleven kindergarten classes ($n = 263$) were assigned to the control condition: Students received traditional literacy instruction for the same amount of time that the experimental group received WEL. Thus, overall exposure to literacy 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. Two-way ANOVAs investigating the impact of Waterford curriculum and demographics on MAP and TRC showed that experimental-group students across all races/ethnicities and types of lunch status analyzed outperformed their control group counterparts.

Additional independent *t*-tests were conducted to examine end of year MAP scores for the subsample of African American/Black students. The results showed that within the subsample, students in the experimental group scored significantly higher than the control group on the Overall MAP RIT score and the Reading Foundations, Literature and Nonfiction and Vocabulary strands; the effect sizes were all in the medium range ($d = 0.58$, Literature and Nonfiction to $d = 0.68$, Overall score) (see Figure 4).

MAP

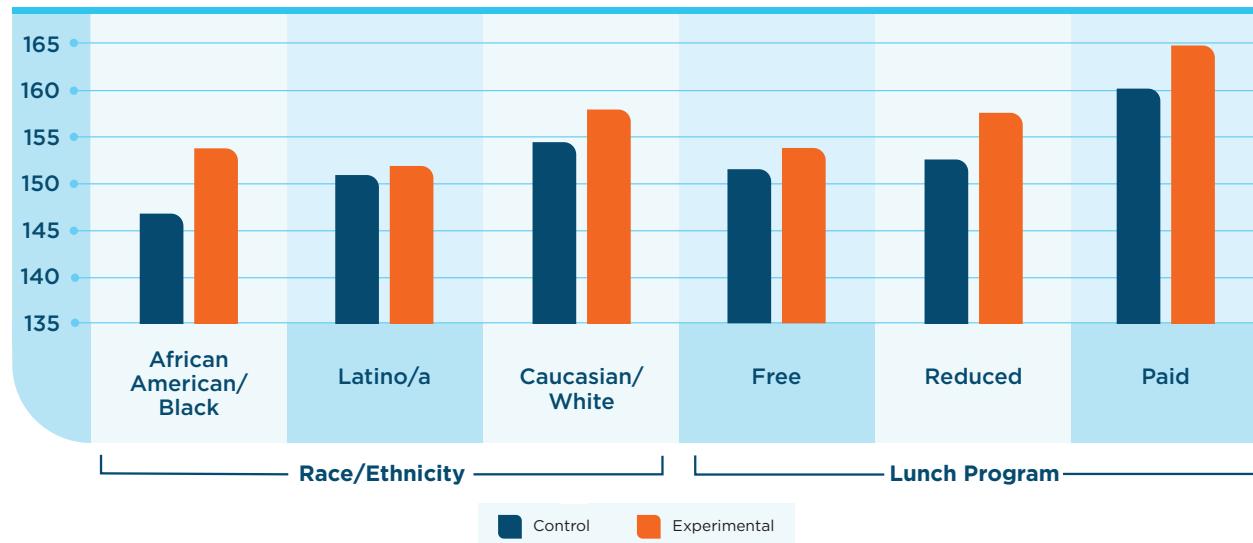
GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS

Figure 1: MAP End of Year Scores by Strand



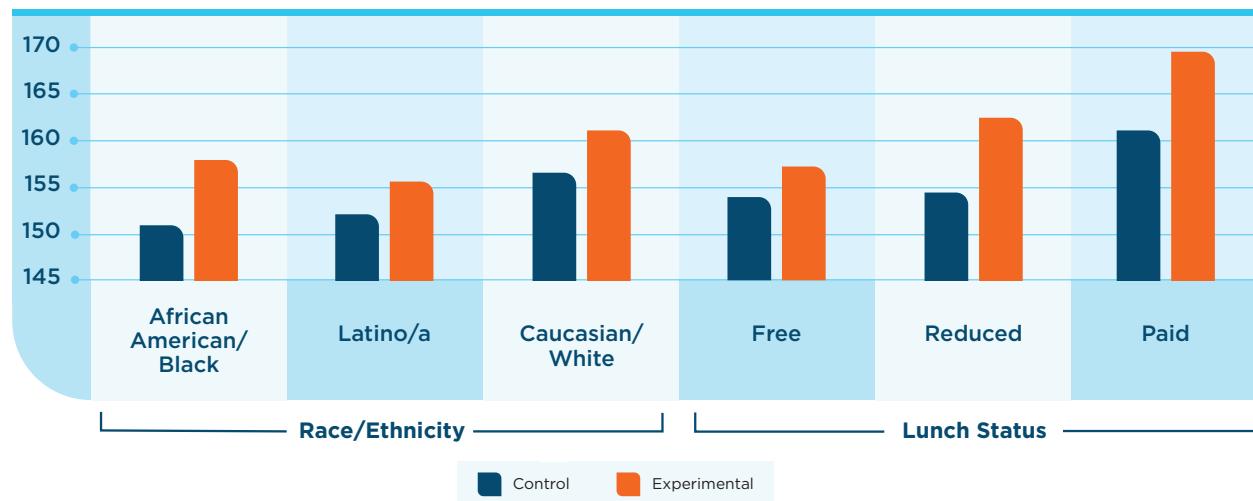
GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS – RIT SCORE

Figure 2: MAP RIT Scores by Demographics



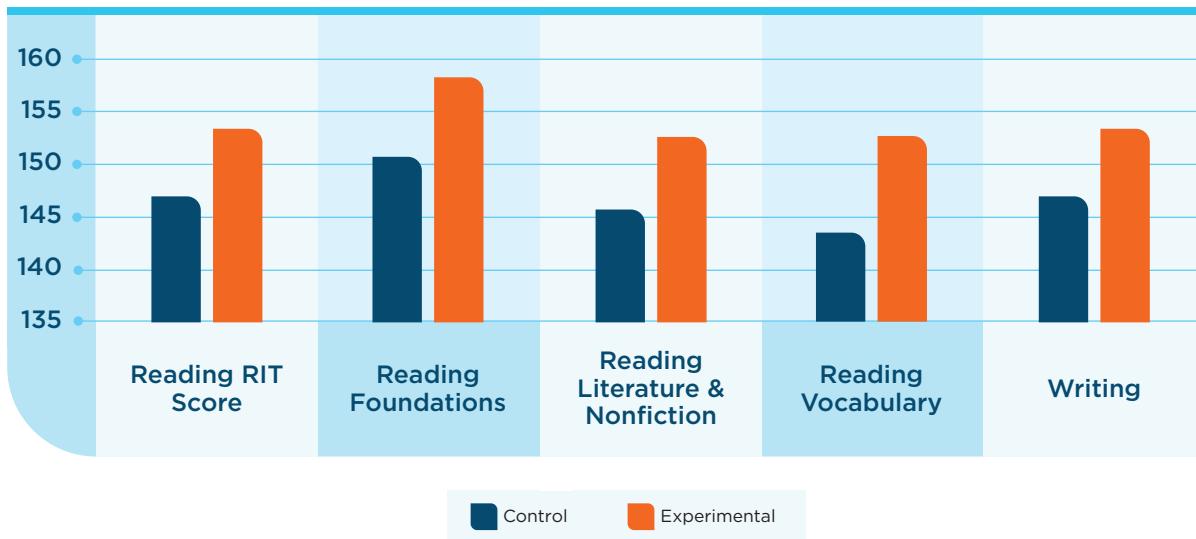
GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS – READING FOUNDATIONS

Figure 3: MAP Reading Foundations Scores by Demographics



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

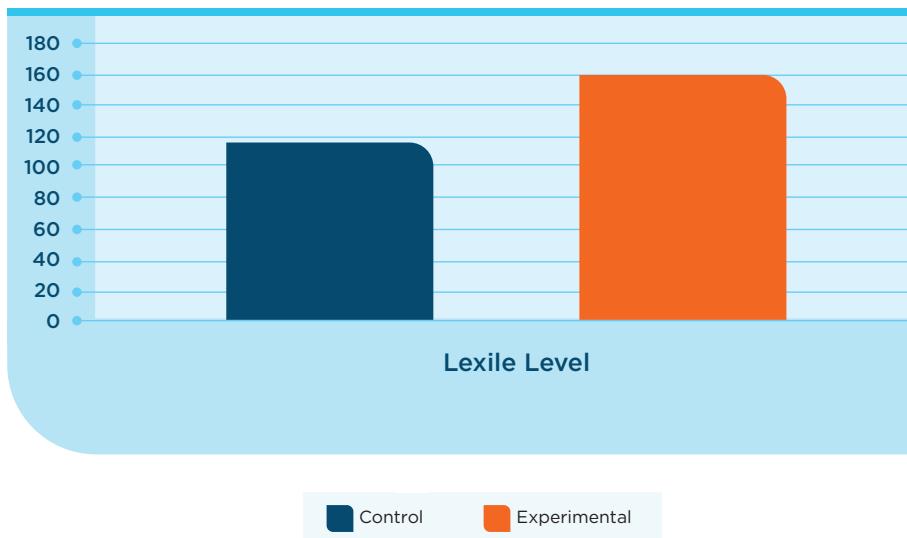
Figure 4: MAP End of Year Scores - African American/Black Students



TRC

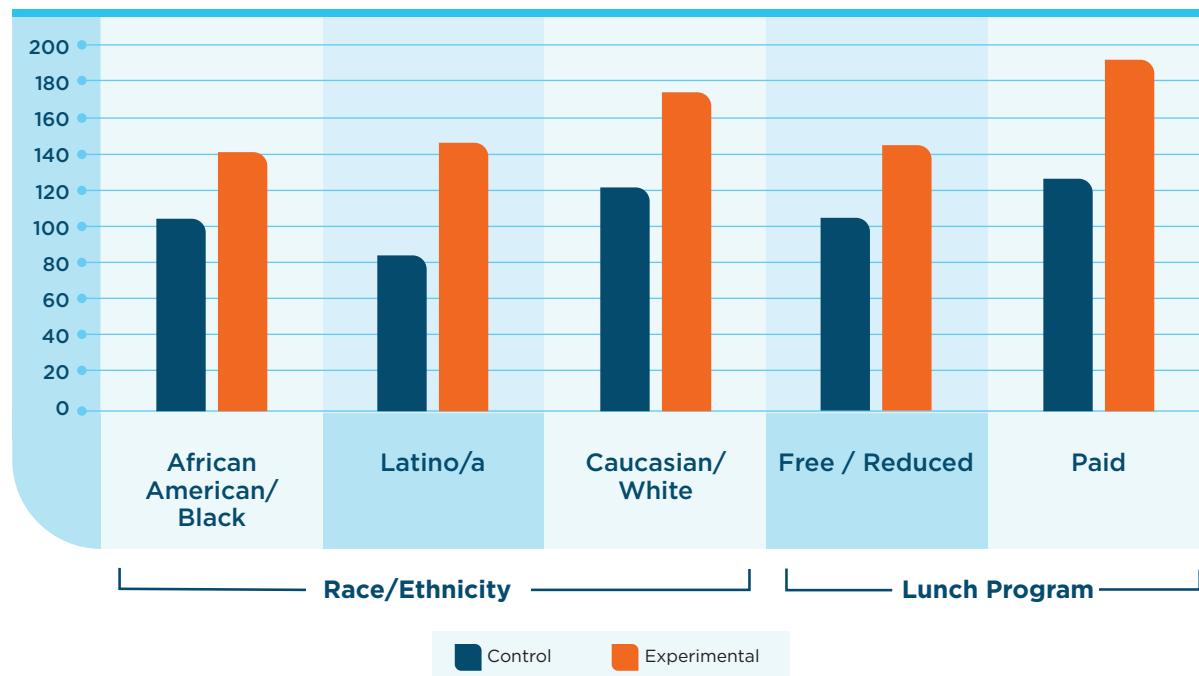
GROUP DIFFERENCES USING AN INDEPENDENT SAMPLES T-TEST

Figure 5: TRC End of Year Lexile Levels



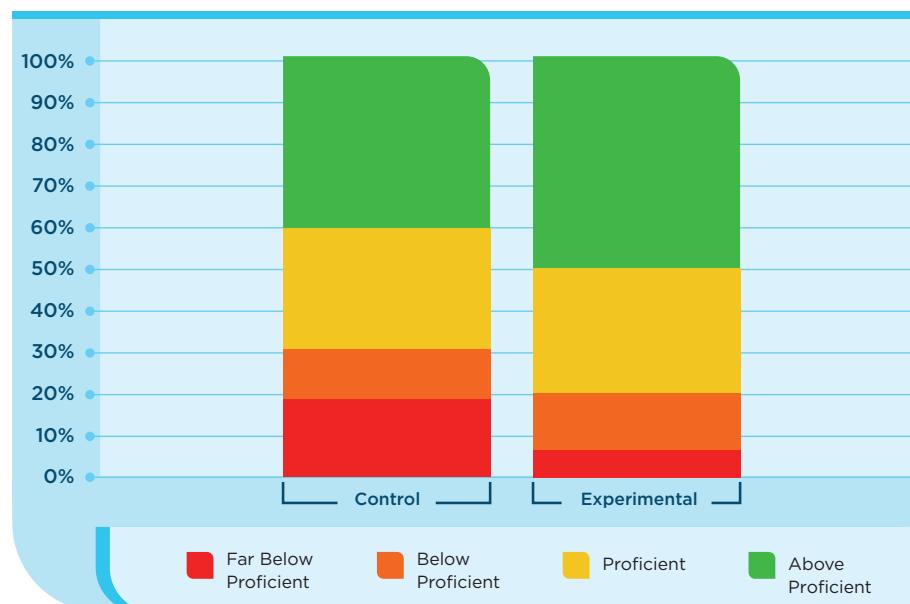
GROUP DIFFERENCES BY DEMOGRAPHICS USING ANOVAS

Figure 6: TRC End of Year Lexile Levels by Demographics



GROUP DIFFERENCES USING CHI-SQUARE

Figure 7: TRC End of Year Proficiency Benchmarks



Randomized Controlled Trial Evaluation of Waterford Upstart in Nevada

The following randomized controlled trial (Evaluation and Training Institute, 2023a) assessed the efficacy of Waterford Upstart in improving math skills. Pre-kindergarten students ($N = 367$) were randomly assigned to either the Upstart Math program (Treatment sample; $n = 176$; English = 84 and MLL = 92) or the Upstart Reading (Control sample; $n = 191$; English = 101 and MLL = 90). Students were assessed using the KTEA-3 prior to the start of the program as well during the summer of 2021 (after completing the program) and the spring of 2022 (after completing kindergarten). Upstart Math had a significant influence on children's early math skill development, with treatment students outperforming the control group on math concepts and applications in the posttest, both at the start of ($g = 0.48$) and at the end of kindergarten ($g = 0.35$). At both posttests, there was no significant interaction effect between the treatment condition and MLL status, indicating that both MLL students and non-MLL students benefited from the use of the program.

Figure 8: Predicted Treatment and Control Math Post-Test Mean Scores Prior to the Start of Kindergarten

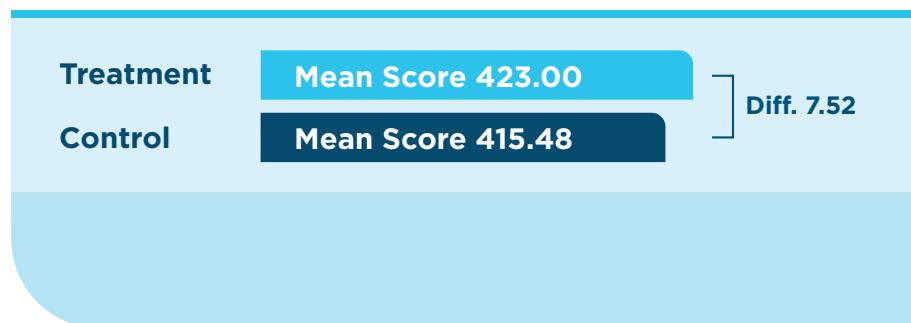


Figure 9: Predicted Treatment and Control Math Post-Test Mean Scores After the End of Kindergarten

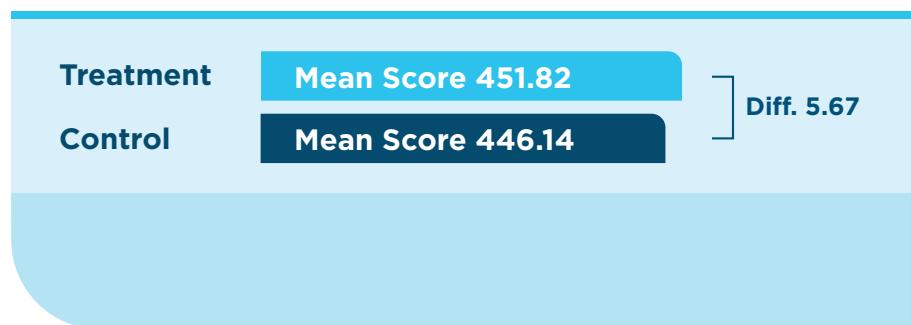
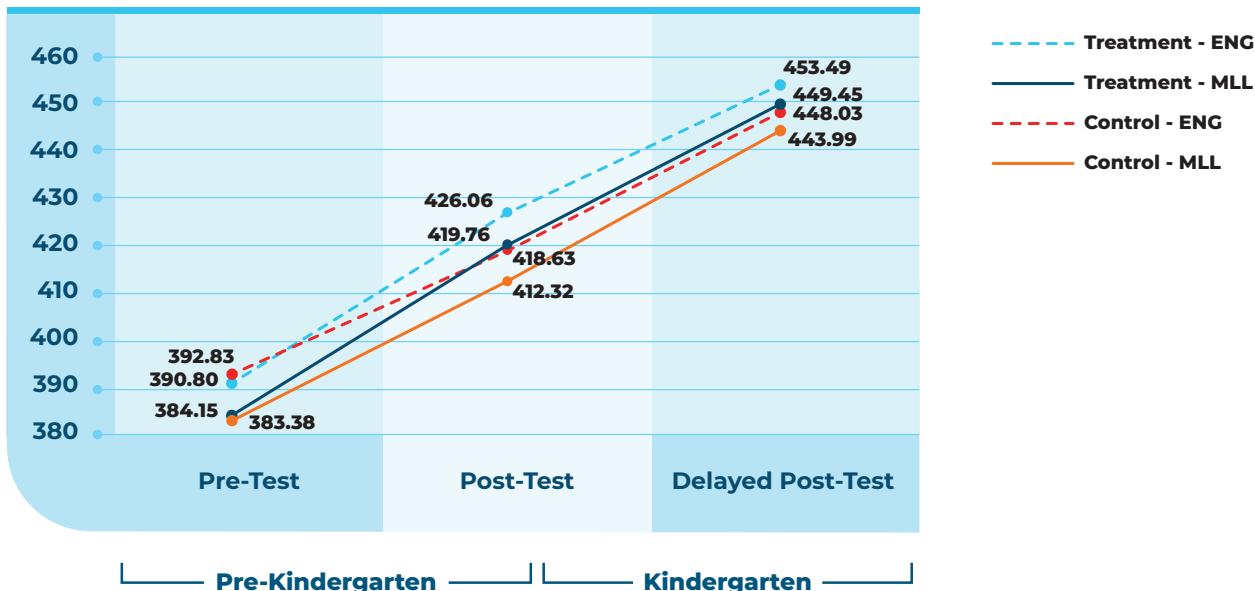


Figure 10. Predicted Means of the MCA Growth Scale Value by Treatment Condition and Language Group



Randomized Controlled Trial Evaluation of the Waterford Summer Learning Path in South Dakota and Wyoming

In a randomized control trial, 329 students were assigned to use either the SLP Reading or SLP Math programs during the summer of 2022 before entering kindergarten (Evaluation and Training Institute, 2023b). Students were assessed at the beginning and end of the program using the Kaufman Test of Educational Achievement Third Edition (KTEA-3). When assessed at the end of the program, students participating in the SLP math program scored significantly higher than their reading counterparts on measures of early math skills, including Concepts and Applications, Computation, and Core Composite. Effect sizes, reported in Hedge's g , ranged from $g = 0.25$ (Concepts and Application) to $g = 0.37$ (Core Composite). There were no statistically significant differential treatment effects when analyzing the influence of the demographic and socio-economic factors; however, math treatment students who began the program with lower math scores (low performance) had a larger average gain in math learning by the end of the program.

Figure 11: Adjusted Mean Score Comparison at Entrance to Kindergarten

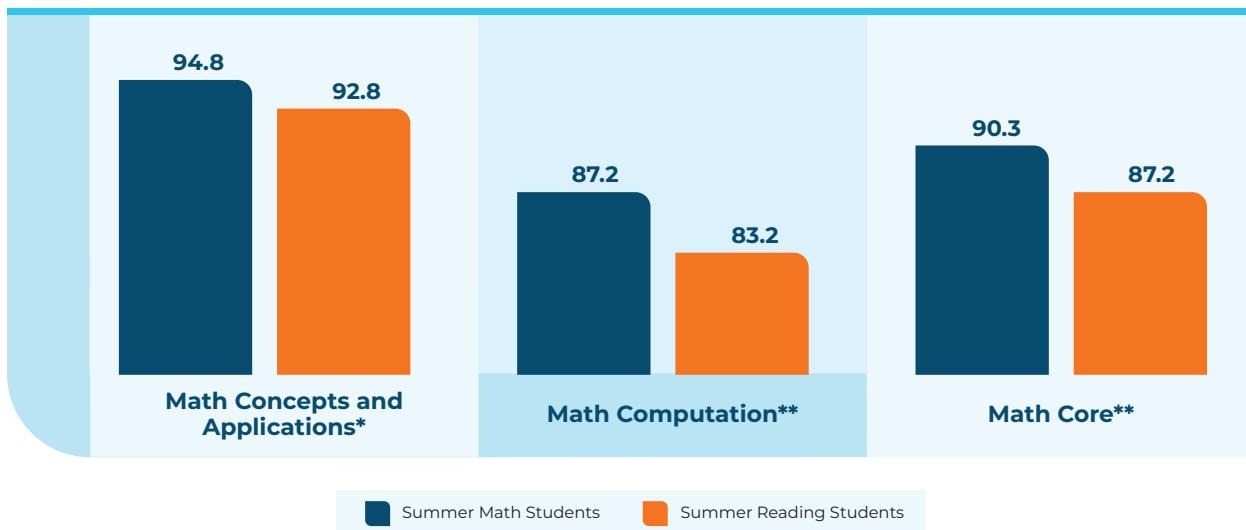
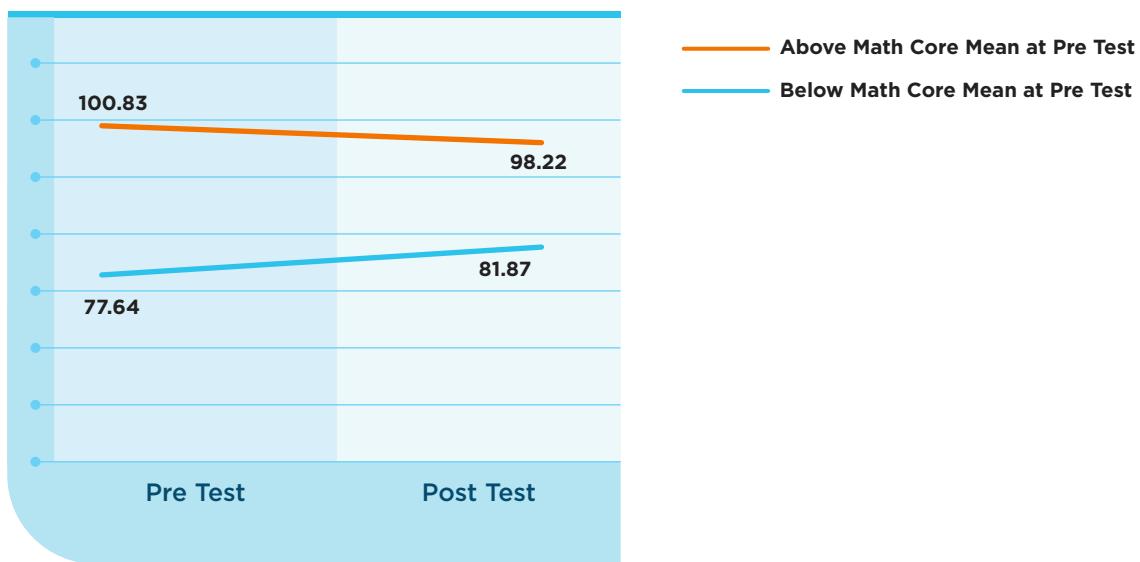


Figure 12: Math Core Outcomes for Low and High Performers



Education Innovation and Research Grant Evaluation

As part of an Education Innovation and Research (EIR) Grant Evaluation grant, a multi cohort randomized controlled trial study was designed to assess the efficacy of the Upstart Reading program. Three cohorts of preschool students in Idaho, Montana, North Dakota, South Dakota, and Wyoming were randomly assigned to the treatment and control conditions. Treatment children receive the Waterford Upstart reading program component as well as Upstart's Collaborative for Academic, Social, and Emotional Learning (CASEL)-aligned lessons (both parents and students) while control children receive the Upstart math/science program component. Participants were assessed with the Kaufman Test of Educational Achievement Third Edition (KTEA-3) as well as the Social Skills Improvement System

Rating Scales (SSIS RS). Each cohort was tested in the summer two years before kindergarten entry, the summer before kindergarten entry, and the summer after kindergarten.

YEAR 1

In the first year of the EIR grant (Evaluation and Training Institute, 2020d), 279 pre-kindergarten students were randomly assigned. Analysis of this cohort indicated that Upstart curricula had a strong impact on children's emerging literacy and math skills. Children who were assigned to use Upstart Reading significantly outperformed children assigned to use Upstart Math on measures for Letter and Word Recognition ($ES = 0.56$), Reading Comprehension ($ES = 0.32$), and Phonological Processing ($ES = 0.43$). Similarly, children assigned to use Upstart Math significantly outperformed comparison children on math concepts and applications ($ES = 0.55$).

Figure 13: Cohort 1 Upstart Program Effect Sizes

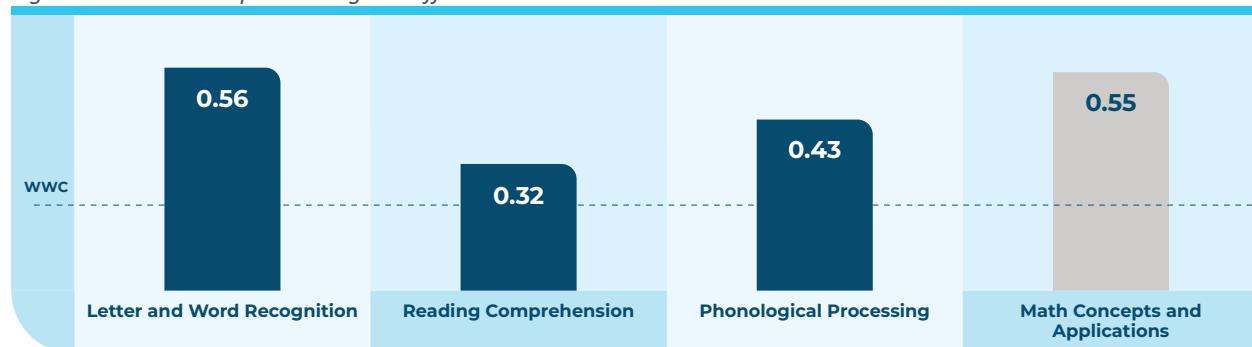
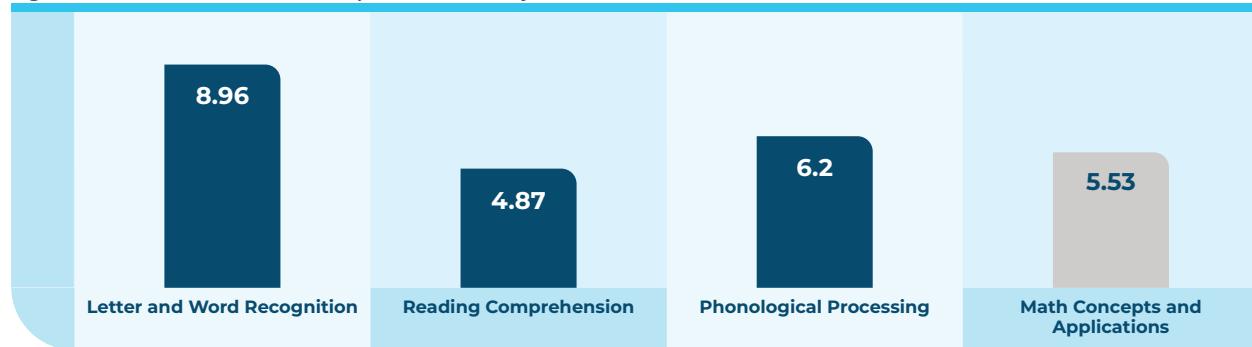


Figure 14: Cohort 1 KTEA Literacy and Math Performance



**Note: Blue bars represent points higher for treatment vs. control children on literacy subtests; gray bar represents points higher for control vs. treatment children on math concepts and applications.*

YEAR 2

In the second year of the EIR grant (Evaluation and Training Institute, 2021), 651 pre-kindergarten students were randomly assigned. Analysis of this cohort indicated that Upstart curricula had a strong impact on children's emerging literacy and math skills. Children who were assigned to use Upstart Reading significantly outperformed children assigned to use Upstart Math on measures for Letter and Word Recognition ($ES = 0.39$), Reading Comprehension ($ES = 0.33$), and Phonological Processing ($ES = 0.38$). Similarly, children assigned to use Upstart Math significantly outperformed comparison children

on math concepts and applications ($ES = 0.33$).

Figure 15: Cohort 2 Upstart Program Effect Sizes

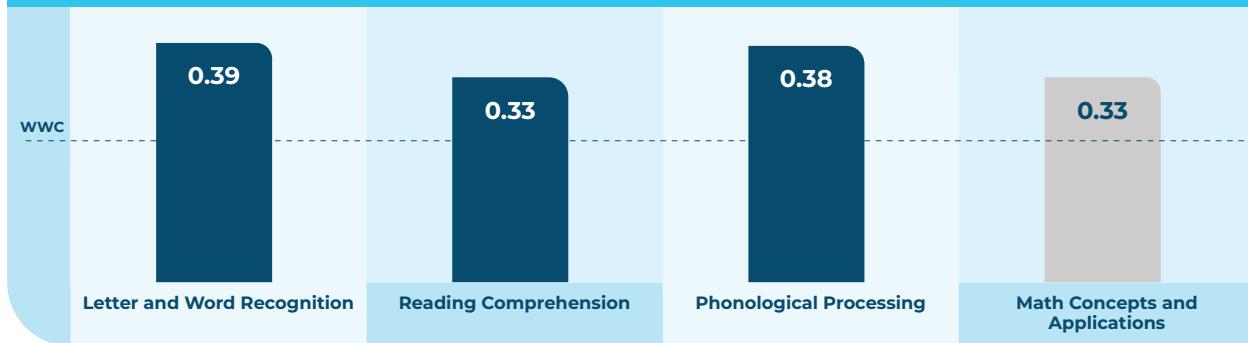
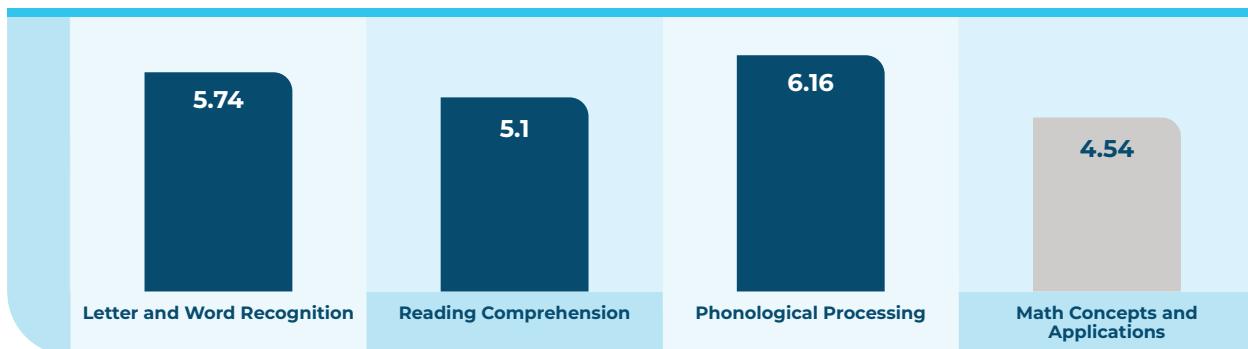


Figure 16: Cohort 2 KTEA Literacy and Math Performance
Treatment Delta



*Note: Blue bars represent points higher for treatment vs. control children on literacy subtests; gray bar represents points higher for control vs. treatment children on math concepts and applications.

YEAR 3

In the third year of the EIR grant (Evaluation and Training Institute, 2022a), 352 pre-kindergarten students were randomly assigned. Analysis of this cohort indicated that Upstart curricula had a strong impact on children's emerging literacy and math skills. Children who were assigned to use Upstart Reading significantly outperformed children assigned to use Upstart Math on measures for Letter and Word Recognition ($ES = 0.31$), Reading Comprehension ($ES = 0.27$), and Phonological Processing ($ES = 0.42$). Similarly, children assigned to use Upstart Math significantly outperformed comparison children on math concepts and applications ($ES = 0.41$).

Figure 17: Cohort 3 Upstart Program Effect Sizes

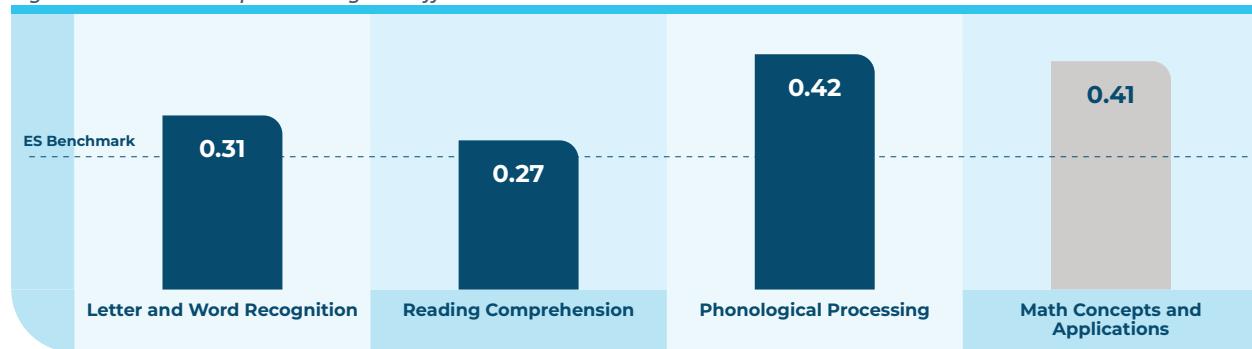


Figure 18: Cohort 3 KTEA Literacy and Math Performance
Treatment Delta



*Note: Blue bars represent points higher for treatment vs. control children on literacy subtests; gray bar represents points higher for control vs. treatment children on math concepts and applications.

Multi-Year Randomized Controlled Trial Evaluation of Waterford Upstart in Nevada

In this randomized controlled trial, two cohorts of children in 2019-2020 and 2020-2021 ($N = 550$) were randomly assigned to use only the Waterford Early Reading program or the Waterford Early Math and Science program (Hobbs & Coordt, 2021). Participants were administered early literacy and math subscales from the Kaufman Test of Educational Achievement, Third Edition (KTEA-3). Parents provided feedback on measures of social skills from the Social Skills Improvement System Rating Scales (SSIS Rating Scales), which were based on their child's behavior. Participation in the treatment condition had a significant impact on posttest scores in measures of Reading Comprehension (ES = 0.44), Phonological Processing (ES = 0.26), and Letter and Word Recognition (ES = 0.42) compared to the control group. Additionally, the control group who received the math software scored significantly higher in measures of math concepts when compared to the control group (ES = 0.44). There were no differences between the treatment and control groups on assessment measures of social-emotional development. Multi-Lingual Learners (MLL) scored lower than non-MLL children on measures of reading comprehension and phonological processing, however, there were no significant interactions between MLL and non-MLL treatment children, indicating that all students that used the program benefitted from it in comparable terms.

Additionally, Upstart had a positive impact on the lowest-performing students. Students performing at or below the 8th percentile were identified as “at-risk for SPED.” Out of the 550 children who completed the pre and posttest (treatment $n = 272$ control $n = 278$), a total of 103 (treatment $n = 51$ control $n = 52$) were categorized as “at-risk for SPED” at pretest. After the posttest, a total of 55% of the treatment group moved out of the risk category compared to 38% of the control group.

Table 10: Baseline Equivalence of Treatment-Control Groups by Construct

Subscale	N	Mean	SD	Pre-test t-value	Difference	St. Difference
Reading Comprehension						
Reading	272	101.53	15.84	-0.142	0.19	0.887
Math	278	101.34	15.12			
Letter and Word Recognition						
Reading	272	97.09	16.85	-0.5524	0.78	0.5809
Math	278	96.32	16.07			
Phonological Processing						
Reading	272	92.03	12.12	0.0553	-0.06	0.9559
Math	278	92.09	13.50			
Math Concepts and Applications						
Reading	156	97.08	16.89	-0.7184	1.35	0.4731
Math	149	95.73	15.76			
Social-Emotional (RS)						
Reading	270	104.89	12.20	-0.3269	0.35	0.7439
Math	277	104.54	12.63			

Figure 19: Predicted Treatment and Control Reading Comprehension Mean Scores at Post-Test

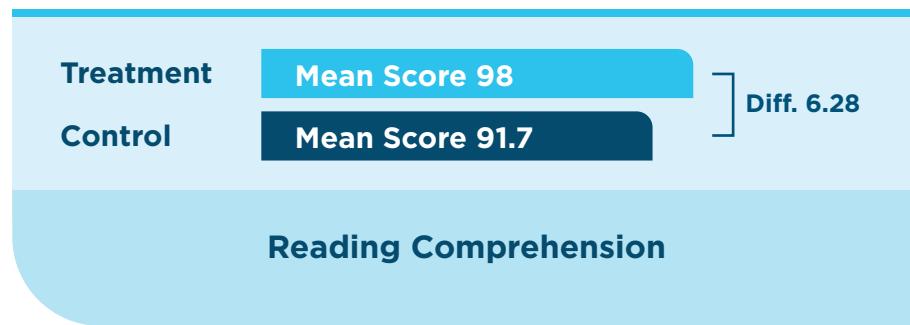


Figure 20: Predicted Treatment and Control Phonological Processing Post-Test Mean Score

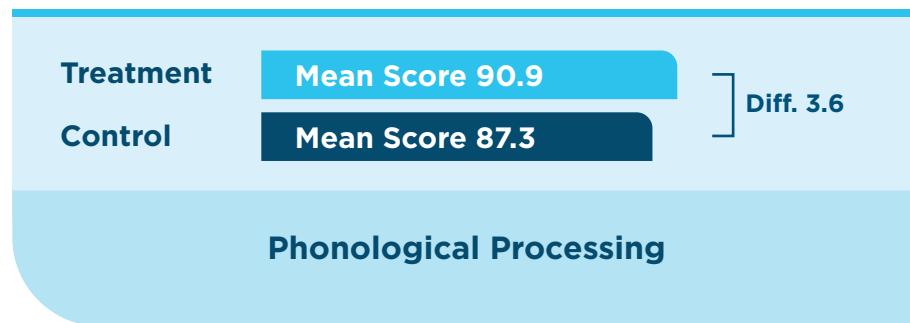


Figure 21: Predicted Treatment and Control Math Post-Test Mean Score

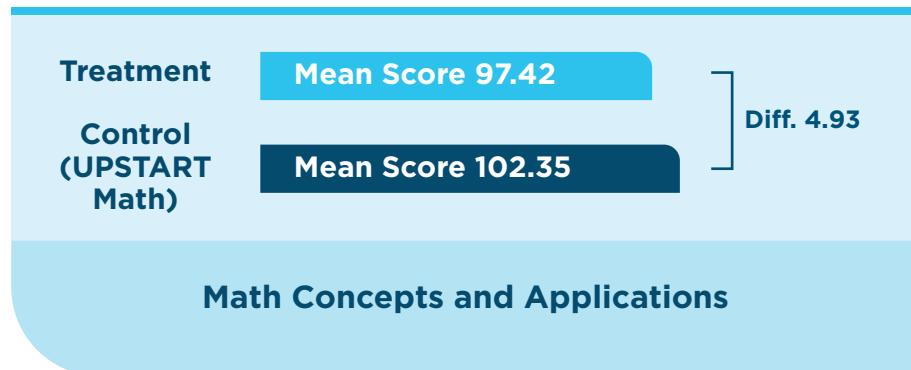
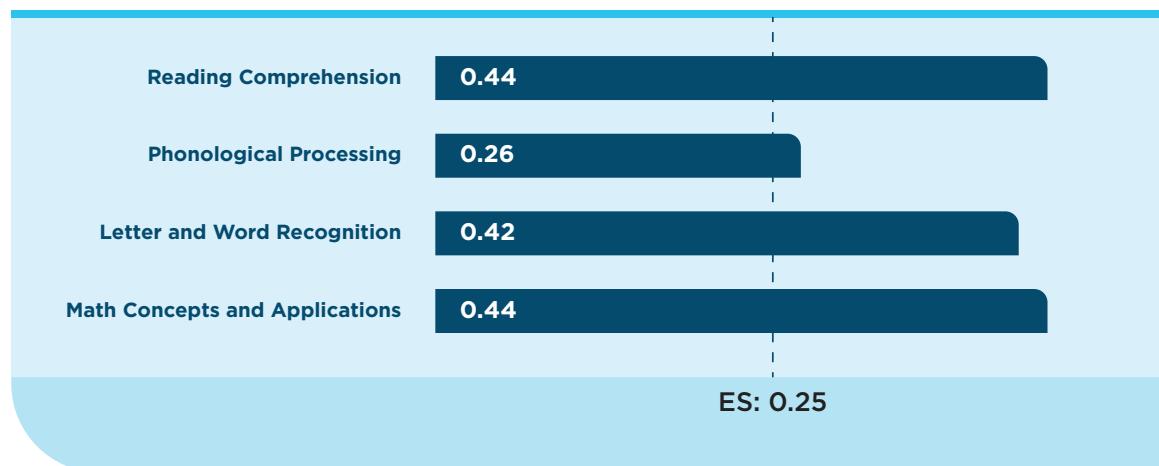


Figure 22: Effect Size Estimates Based on Adjusted Mean Standard Scores by Measure



Tier 2: 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 11: Descriptive Statistics for All Tasks

Pretest Predictors	Training (n = 42)			Control (n = 34)		
	M	SD	Reliability	M	SD	Reliability
Segmenting						
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	—
Blending						
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	—
Letter Name Knowledge						
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	—
Letter Sound Knowledge						
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	—
Letter Writing Knowledge						
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	—
Word Reading						
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	—
Invented Spelling						
Pre	3.55	4.28	.96 ^a	2.41	3.84	.98 ^a
Post	26.71	19.67	.91 ^a	8.09	7.79	.94 ^a
Gain	23.16***	17.75	—	5.68***	5.77	—
Vocabulary Knowledge						
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	—
Print Concepts						
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	—
Total time spent using the WERP-1 (hh:mm:ss)						
		21:25:49			7:17:38	

^aProportion 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 Alphabetics and Comprehension. In evaluating its effectiveness, the Waterford Early Reading Program was found to have potentially positive effects on Alphabetics—+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 students experiencing poverty and minority students, and in the first year of the program 61% of UPSTART participants came from homes experiencing poverty, and 20% classified their races/ethnicities as Latino/a or non-Caucasian/White. 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, when disaggregated by race/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

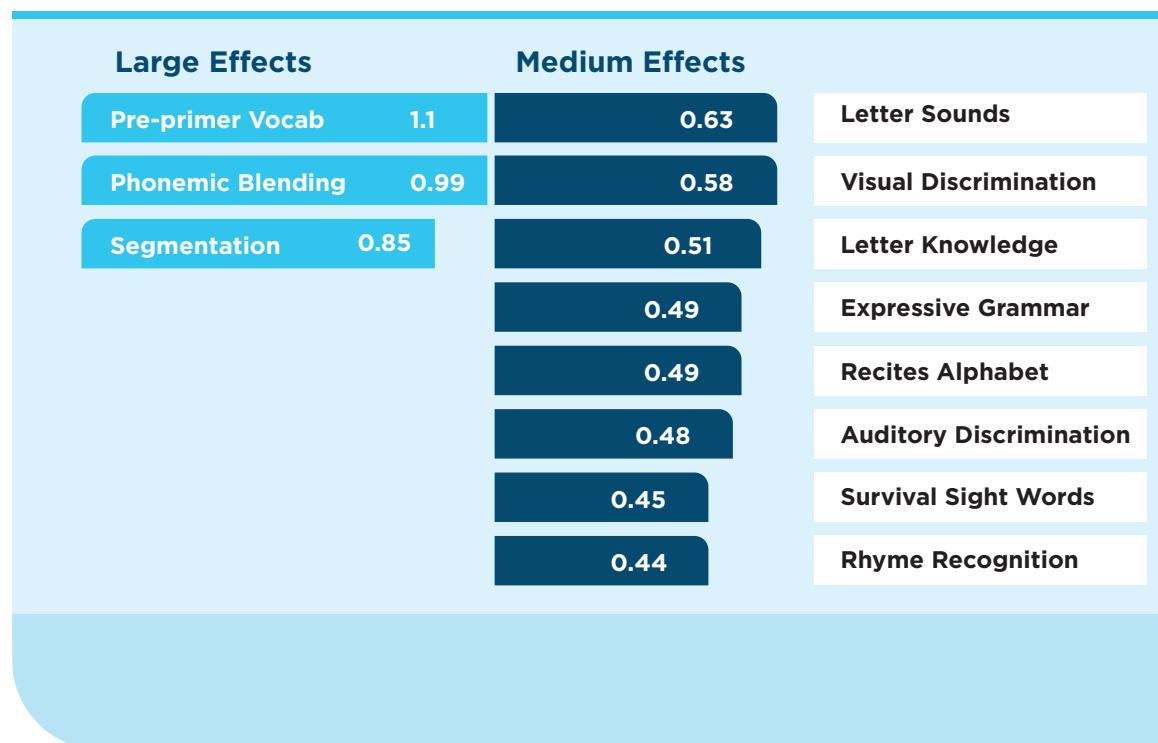
(0.99), and Phoneme Segmenting (0.85). Children participating in UPSTART also demonstrated greater gains on Pre-Primer Vocabulary, Survival Sight Words, and all Phonological Awareness subscales than control students. Participation in UPSTART was associated with improvement in all phonological awareness 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 23: Brigance & Bader Posttest Analysis of Composite Scores

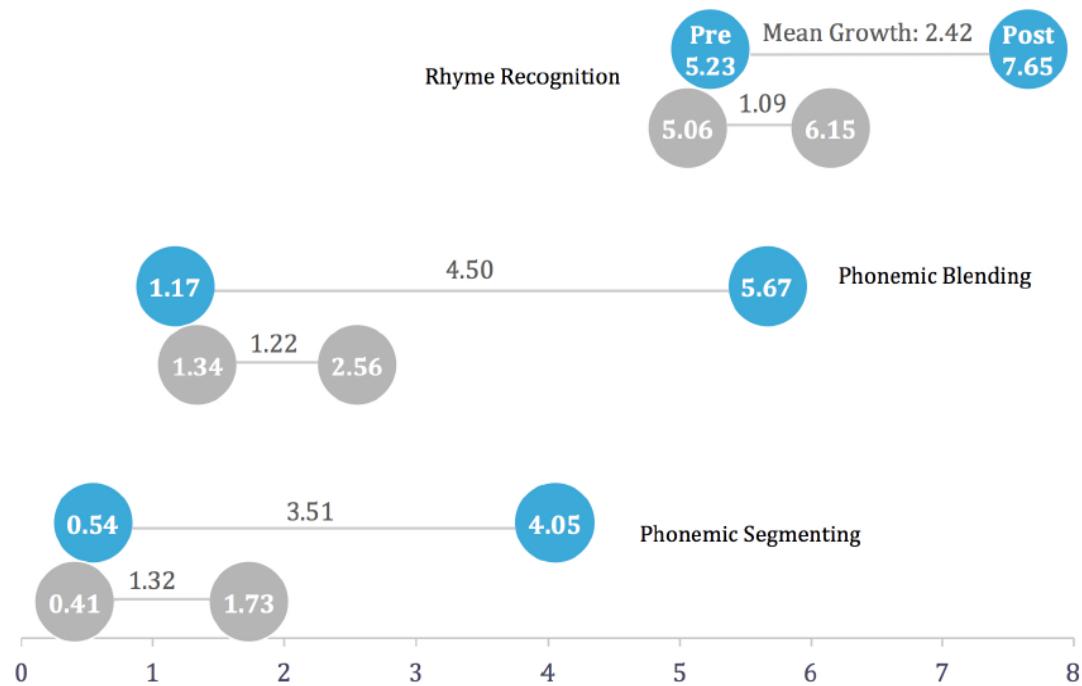


Figure 24: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



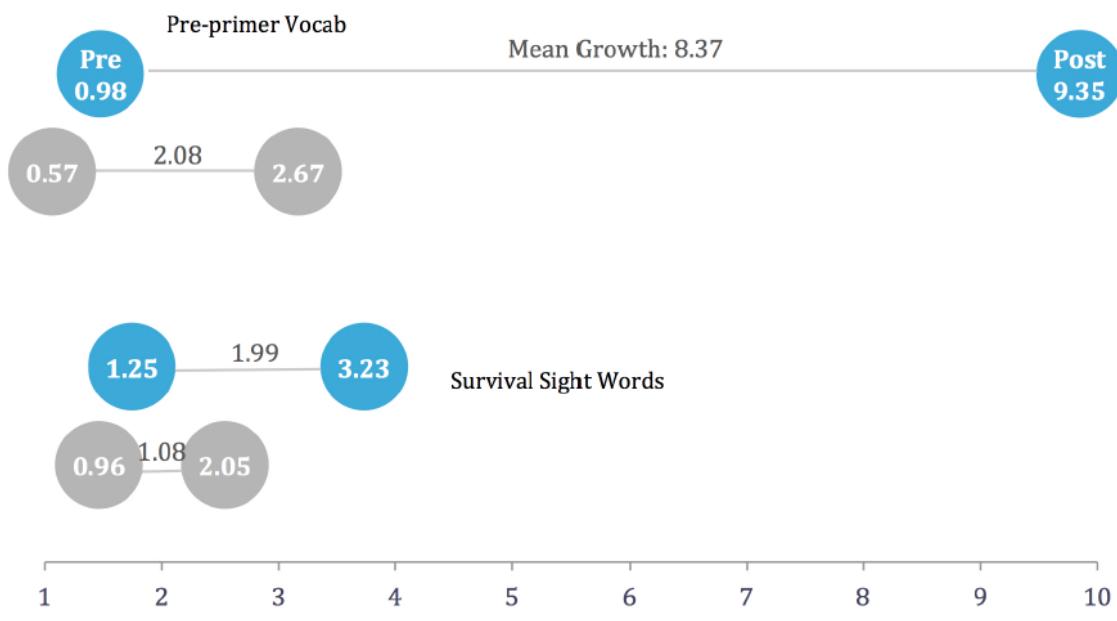
BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 25: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 26: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 27: Pre-Literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores

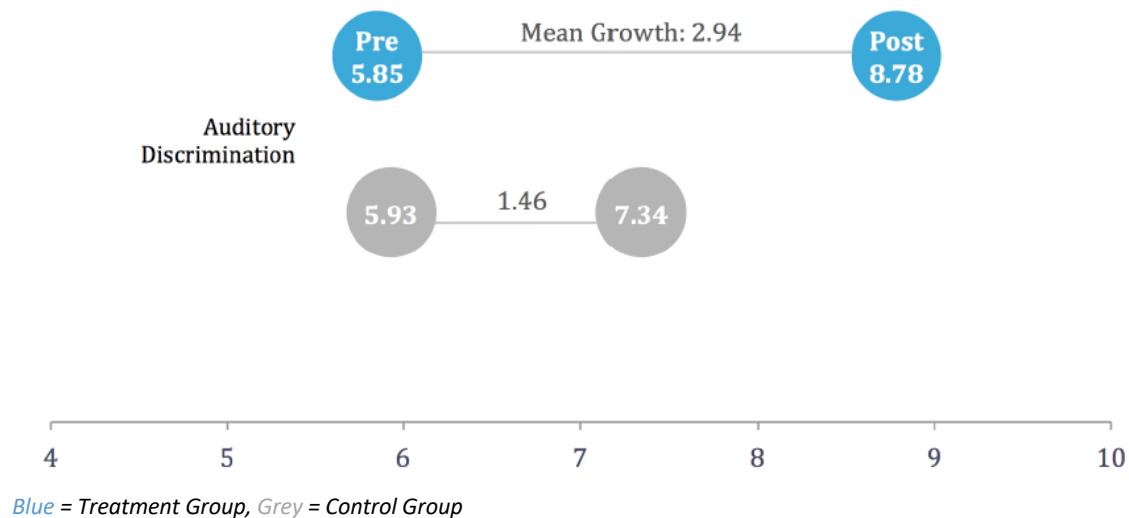


Figure 28: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

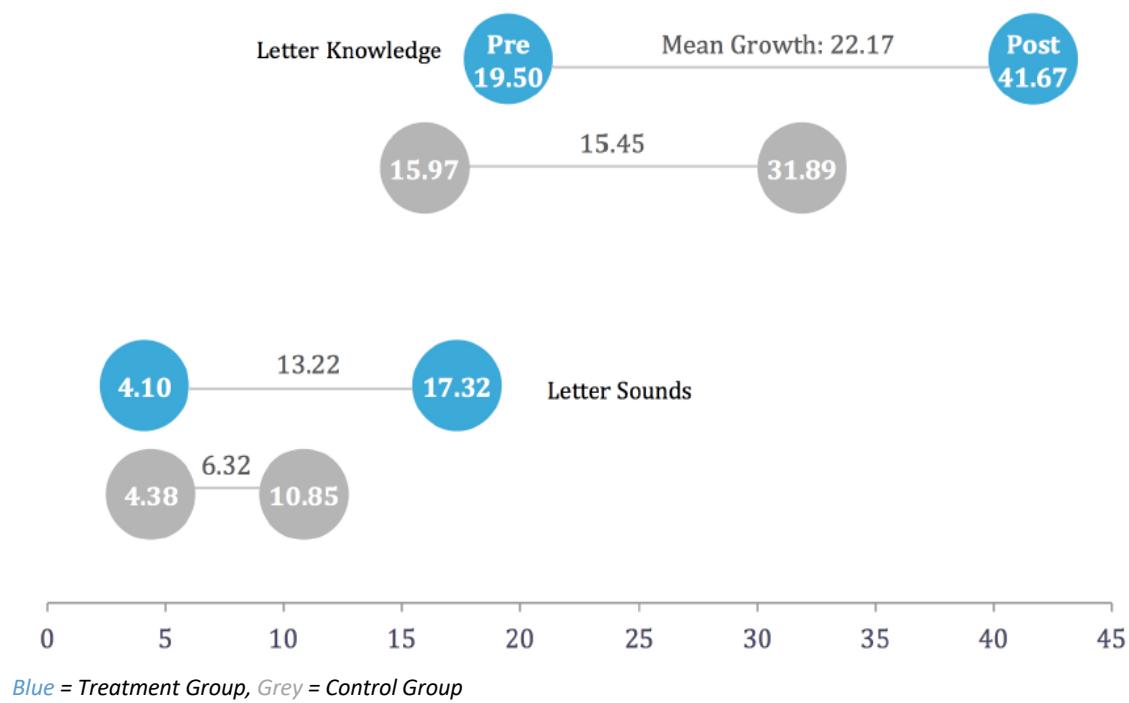


Table 12: Brigance Growth Rate Comparisons

<i>Brigance Test</i>	<i>Control Group (N = 138)</i>	<i>Treatment Group (N = 138)</i>	<i>T-C Significance p ≤ .05</i>
	Mean Growth	Mean Growth	
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 13: Bader Growth Rate Comparisons

<i>Bader Test</i>	<i>Control Group (N = 138)</i>	<i>Treatment Group (N = 138)</i>	<i>T-C Significance p ≤ .05</i>
	Mean Growth	Mean Growth	
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 29: Brigance & Bader Posttest Analysis of Composite Scores

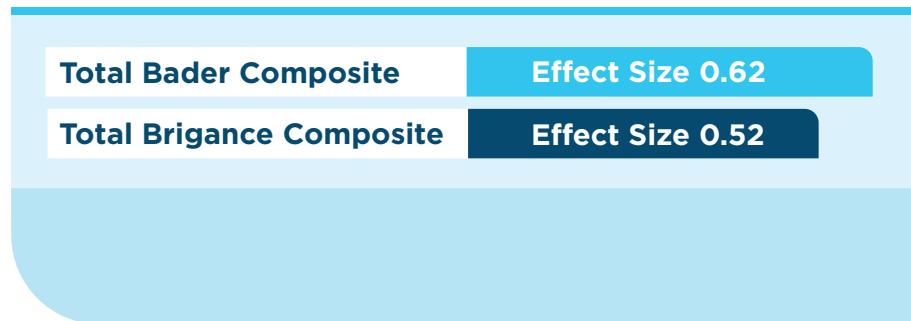
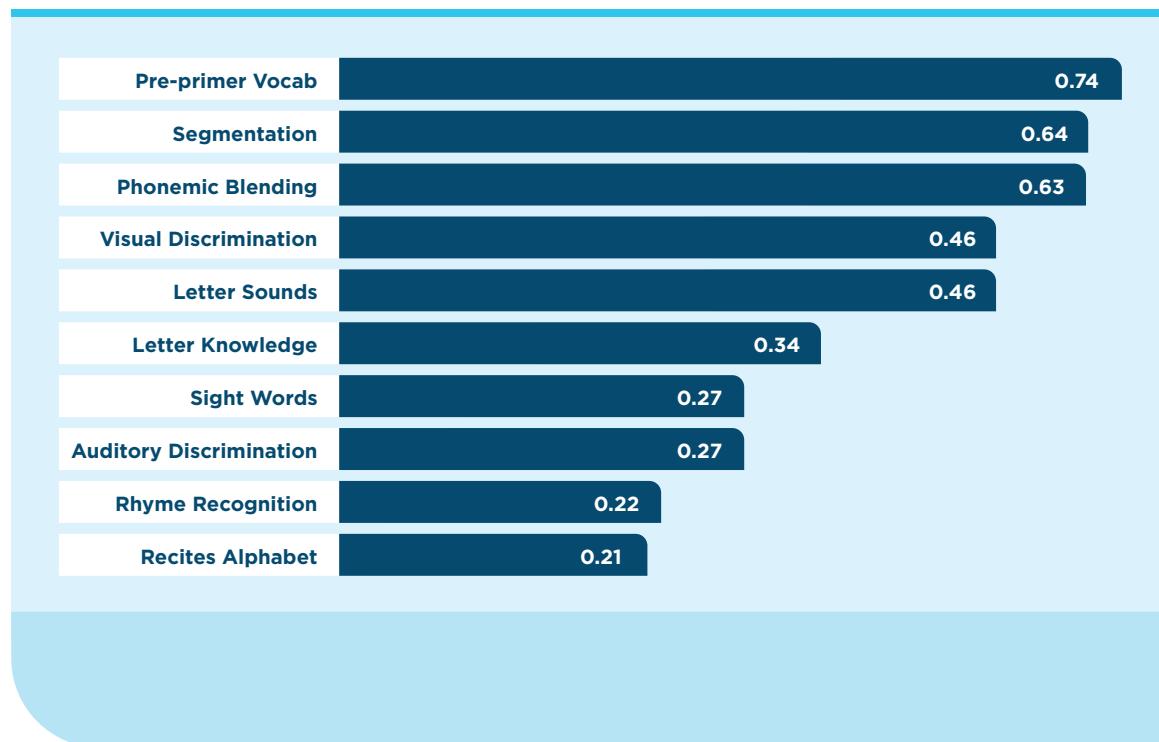
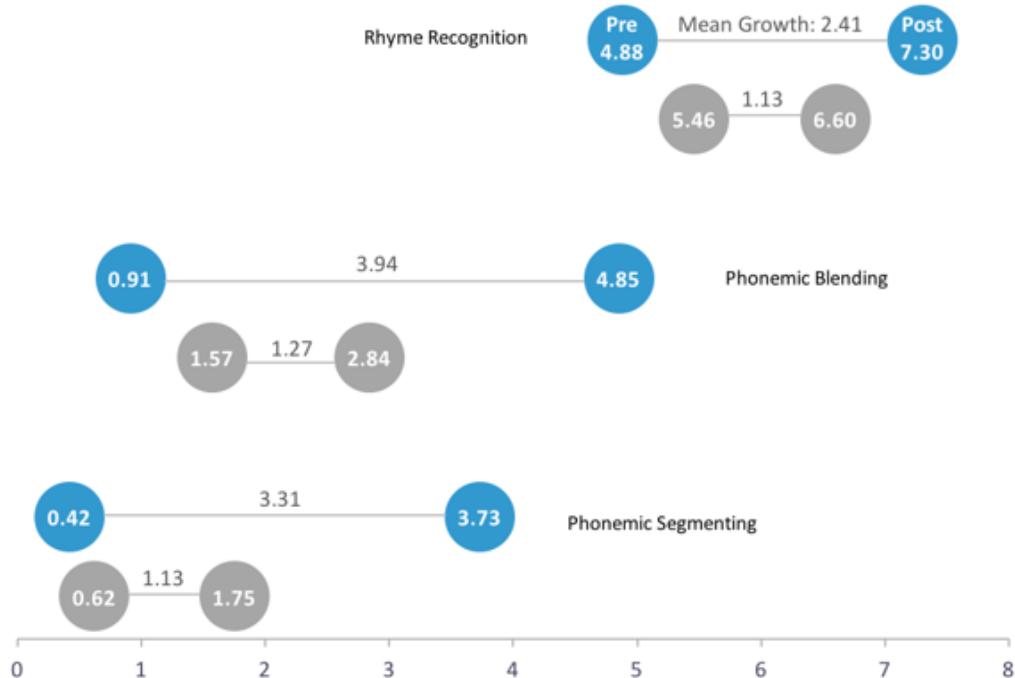


Figure 30: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



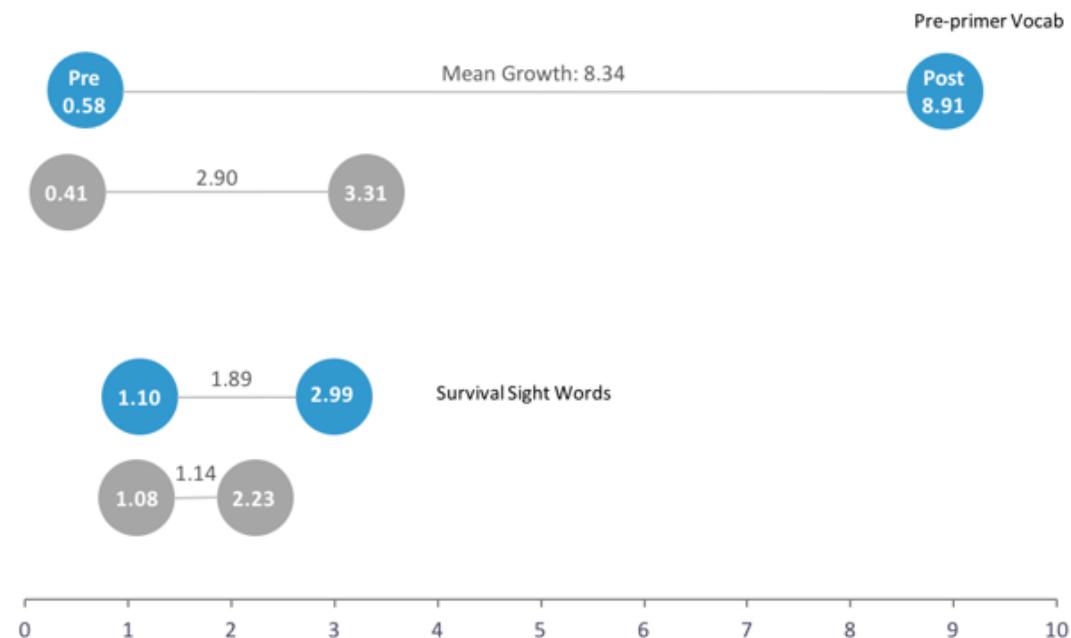
BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 31: Phonological Awareness: Treatment and Control Group Pretest & Posttest Mean Scores



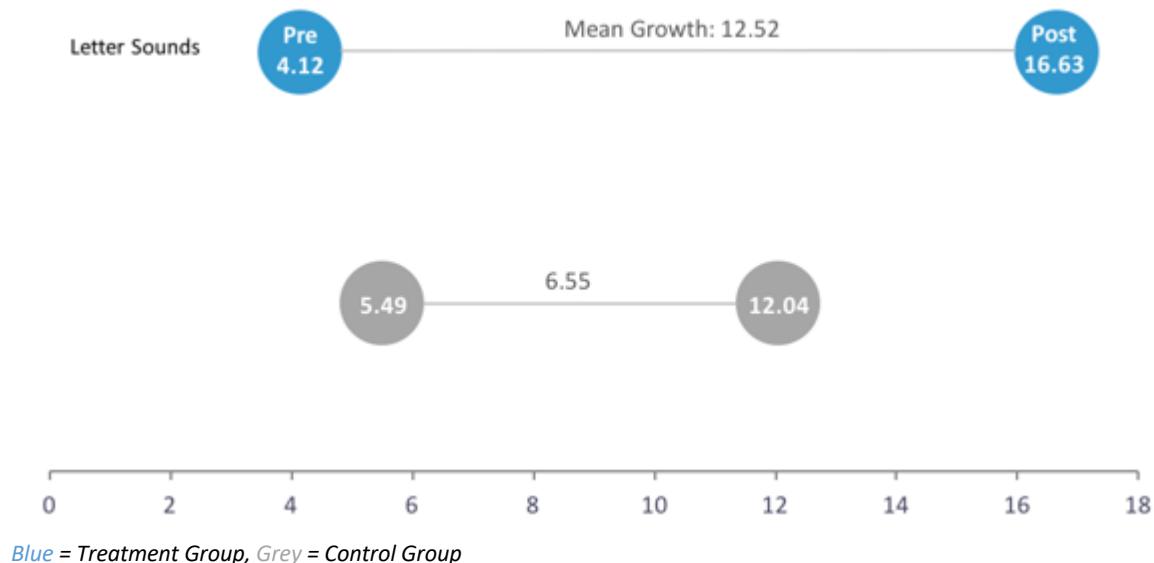
Blue = Treatment Group, Grey = Control Group

Figure 32: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 33: 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 pre-kindergarten 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 34: Brigance & Bader Posttest Analysis of Composite Scores

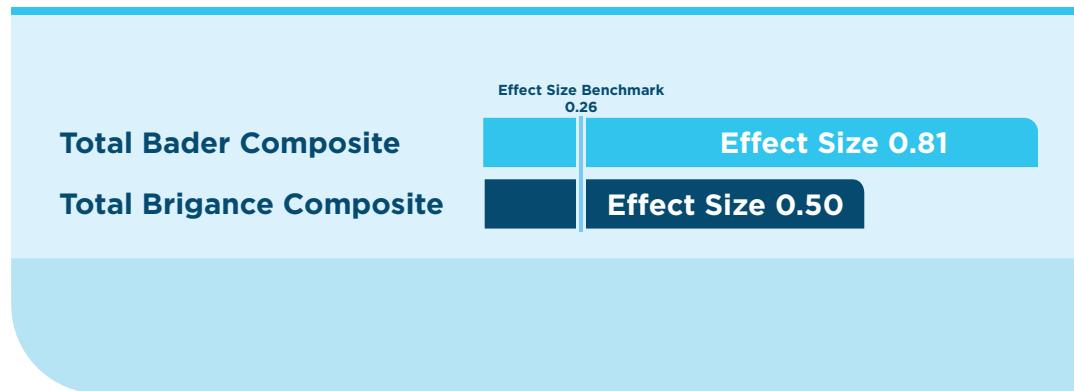
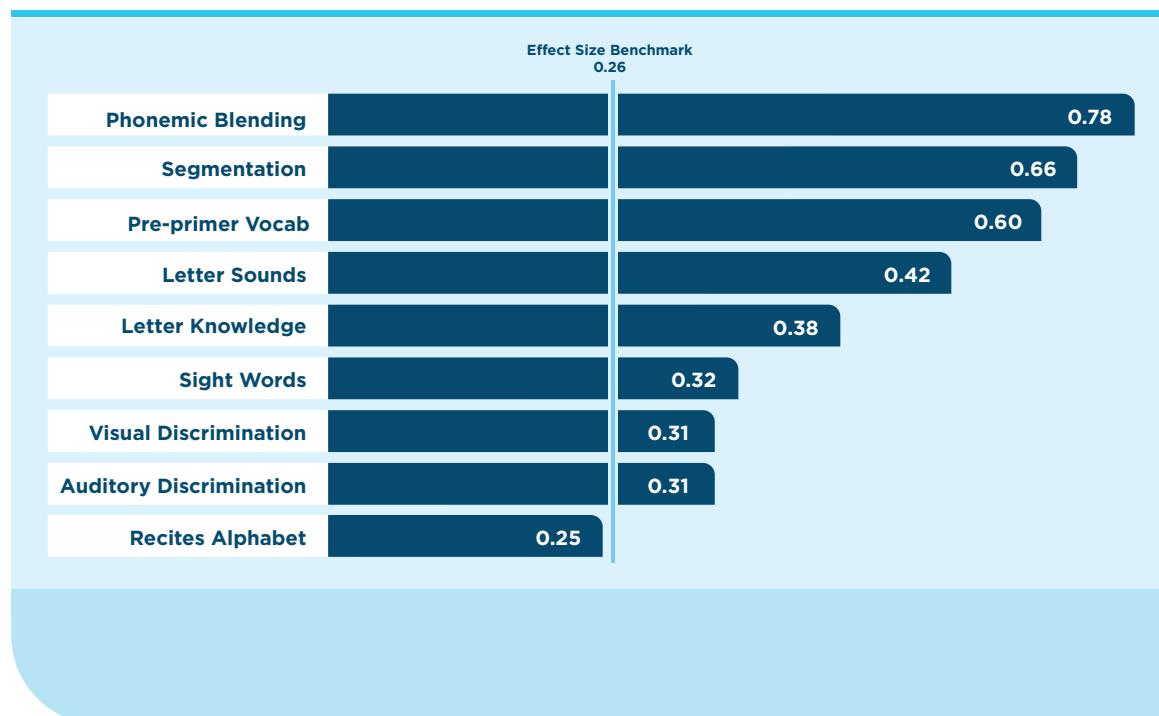


Figure 35: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 36: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores

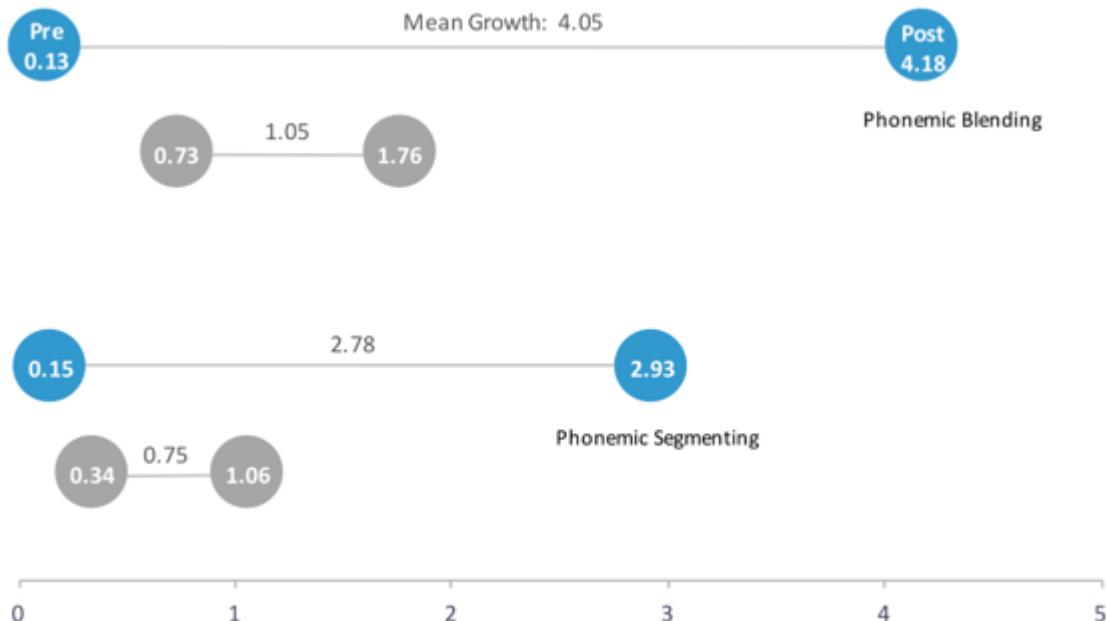
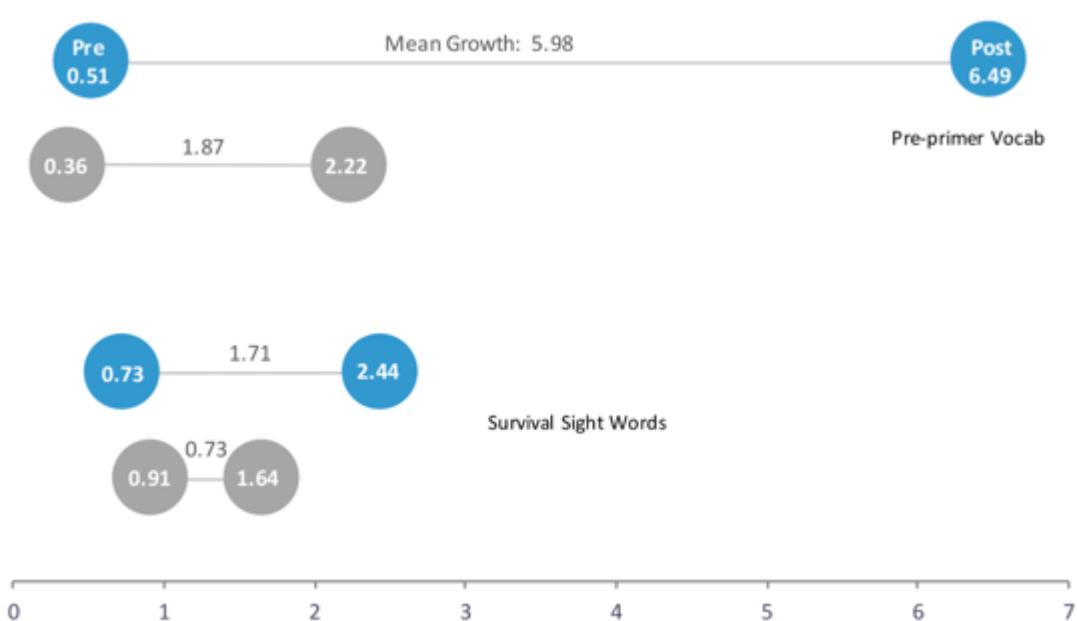


Figure 37: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 38: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

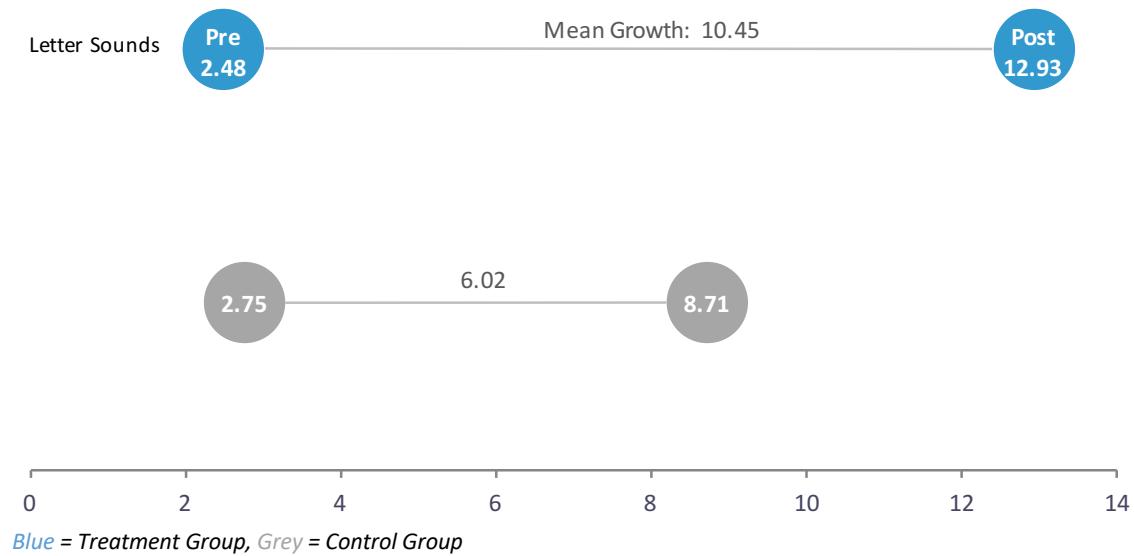
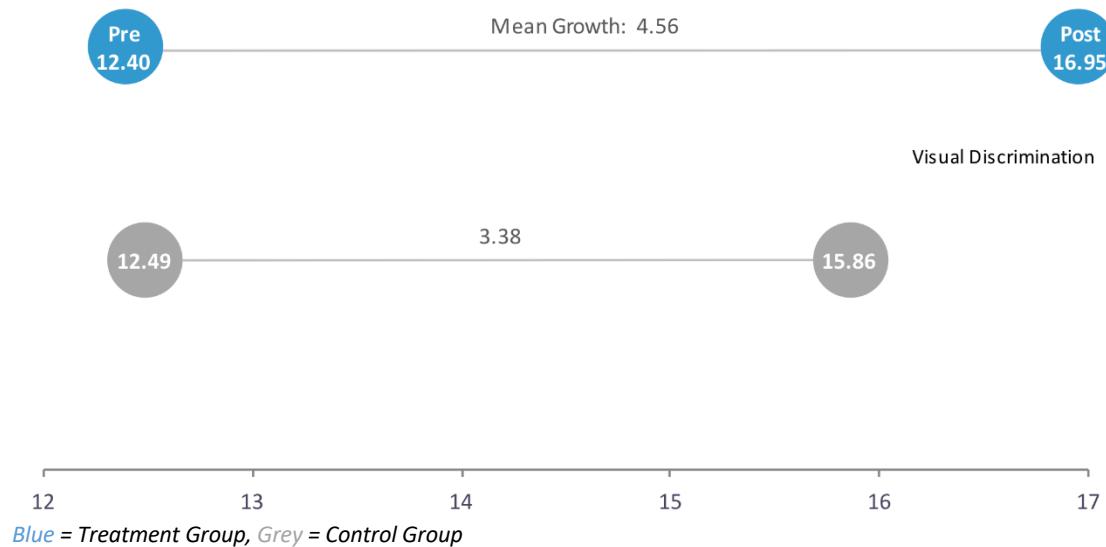


Figure 39: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores



YEAR 9

Waterford Institute enrolled 14,278 preschool children in its ninth year of operation during the 2017-18 school year (Evaluation and Training Institute, 2019a). Combined post-test results showed that UPSTART participation had a medium impact on students' early literacy skill development. In the matched post-test sample, UPSTART produced strong to medium effects: Children enrolled in UPSTART produced significant positive effects ($ES = 0.53$) compared to control children on the Brigance composite. Similarly, UPSTART participants experienced significant positive effects ($ES = 0.56$) 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 (0.71), Pre-Primer Vocabulary (0.67), and Phoneme Segmenting (0.48).

Participation in UPSTART was associated with significant improvement on both phonological awareness strands of the Bader assessment, Phoneme Blending and Phoneme Segmenting. Children participating in UPSTART had a significantly higher growth rate, with a 23-point advantage on Brigance post-test scores compared to non-participating children. Additionally, first grade students who previously participated in UPSTART during pre-kindergarten significantly outperformed their control counterparts on the DIBELS composite at the beginning of first grade. The average beginning of year DIBELS composite score was 7.97 points higher for students who were enrolled in UPSTART compared to students who were not (ES = 0.17).

BRIGANCE AND BADER RESULTS

Figure 40: Brigance & Bader Posttest Analysis of Composite Scores

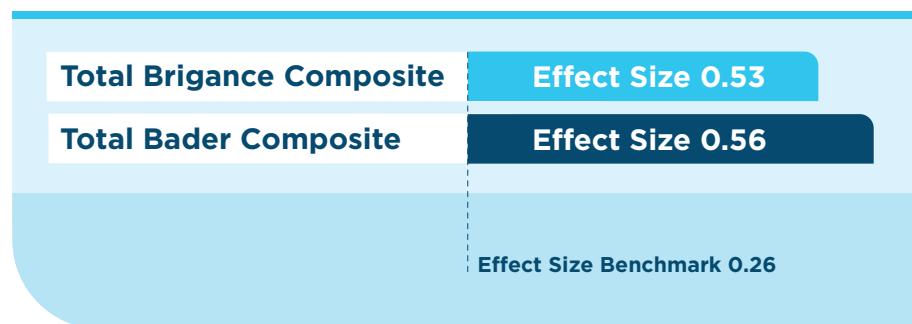
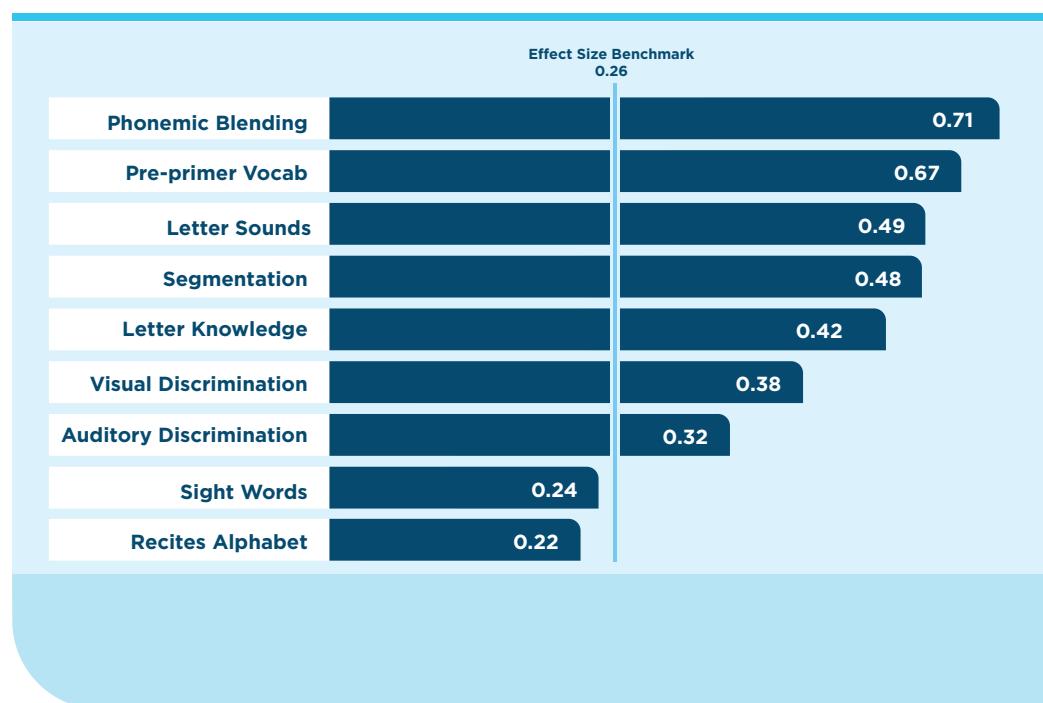
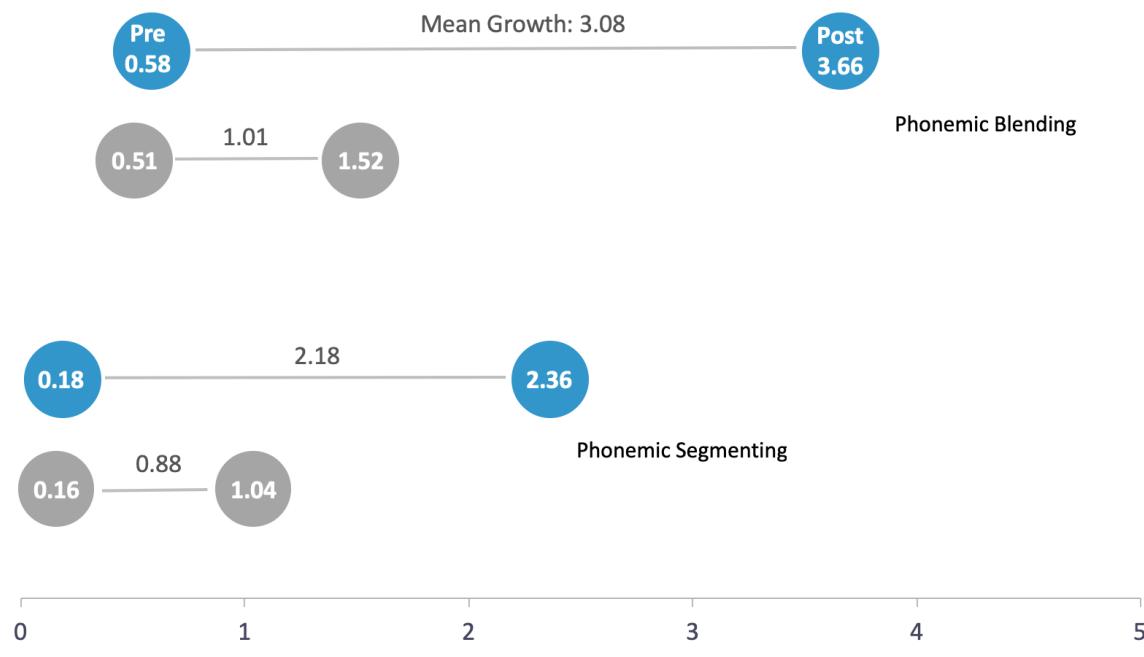


Figure 41: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



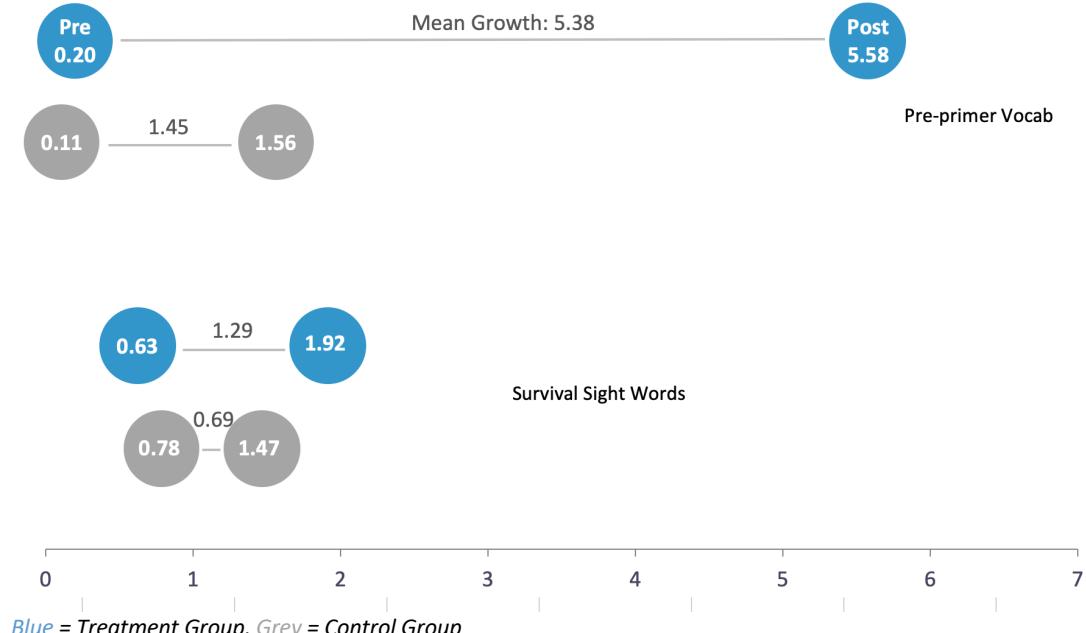
BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 42: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 43: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores



Blue = Treatment Group, Grey = Control Group

Figure 44: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

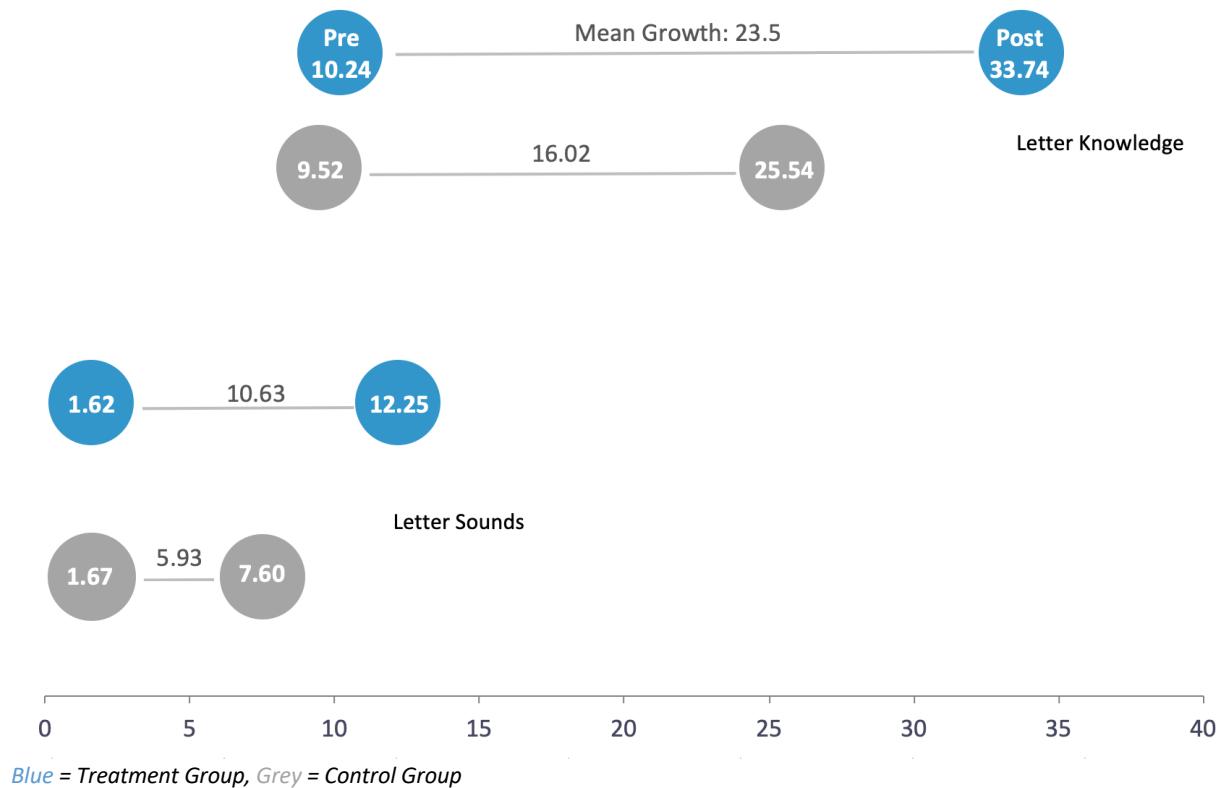
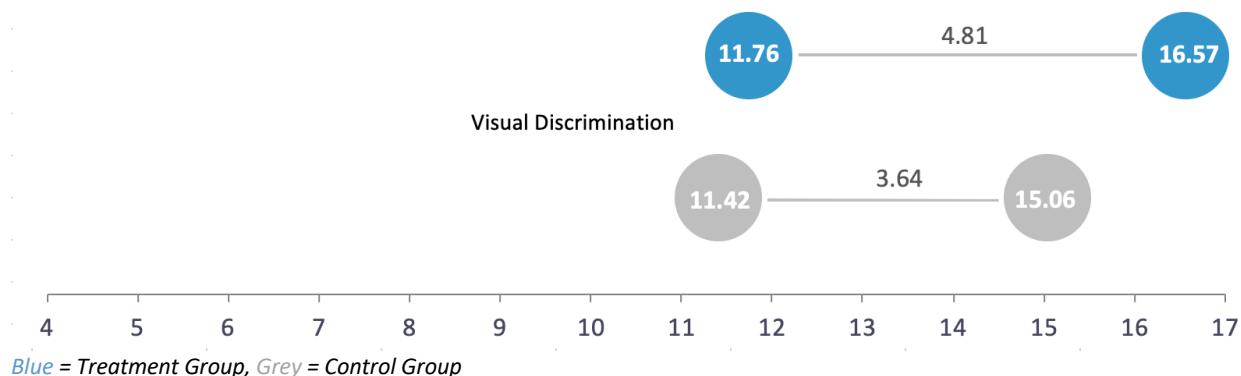


Figure 45: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores



YEAR 10

Waterford Institute enrolled 14,125 preschool children in its tenth year of operation during the 2018-19 school year (Evaluation and Training Institute, 2020a). Combined post-test results showed that UPSTART participation had a medium impact on students' early literacy skill development. In the matched post-test sample, UPSTART produced strong to medium effects: Children enrolled in UPSTART produced

significant positive effects ($ES = 0.53$) compared to control children on the Brigance composite. Similarly, UPSTART participants experienced significant positive effects ($ES = 0.35$) on the Bader composite.

Children participating in UPSTART demonstrated significant improvement in word decoding and phonological awareness skills. Medium effect sizes were observed for Phoneme Segmenting (0.71), Phonemic Blending (0.68), and Pre-Primer Vocabulary (0.59).

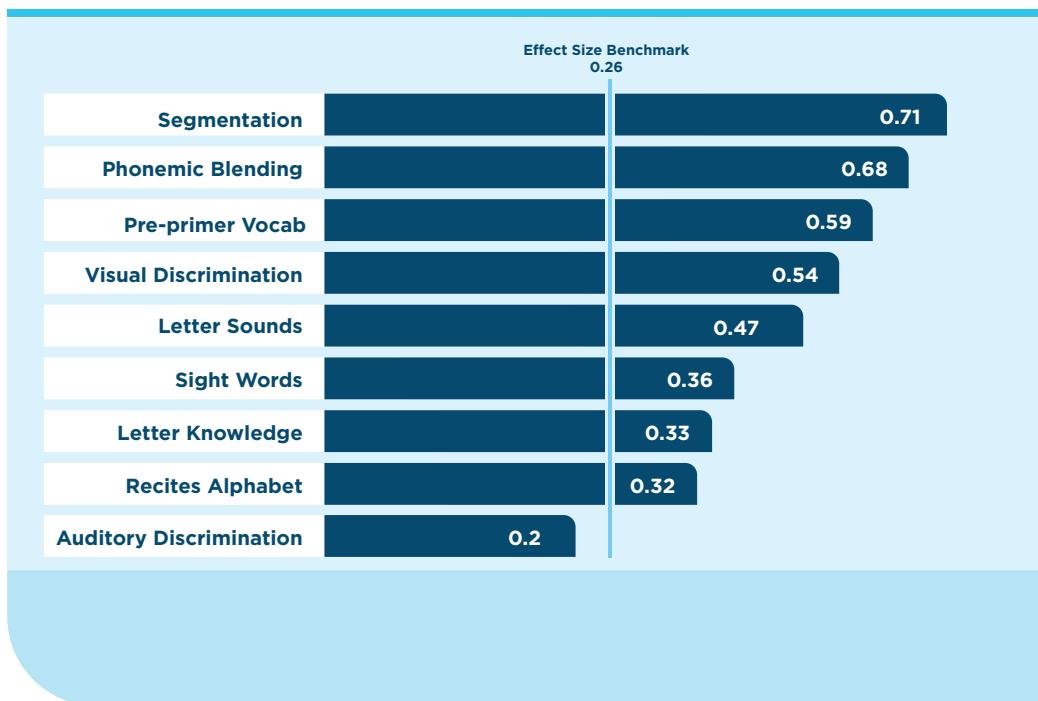
Participation in UPSTART was associated with significant improvement on both phonological awareness strands of the Bader assessment, Phoneme Blending and Phoneme Segmenting. Children participating in UPSTART had a significantly higher growth rate, with a 21-point advantage on Brigance post-test scores compared to non-participating children. Additionally, first grade students who previously participated in UPSTART during pre-kindergarten significantly outperformed their control counterparts on the DIBELS composite at the beginning of first grade. The average beginning of year DIBELS composite score was 10.67 points higher for students who were enrolled in UPSTART compared to students who were not ($ES = 0.22$).

BRIGANCE AND BADER RESULTS

Figure 46: Brigance & Bader Posttest Analysis of Composite Scores



Figure 47: Effect Size Estimates by Magnitude of Effect for Brigance & Bader Subtests



BRIGANCE AND BADER GROWTH RATE RESULTS

Figure 48: Phonological Awareness: Treatment & Control Group Pretest & Posttest Mean Scores

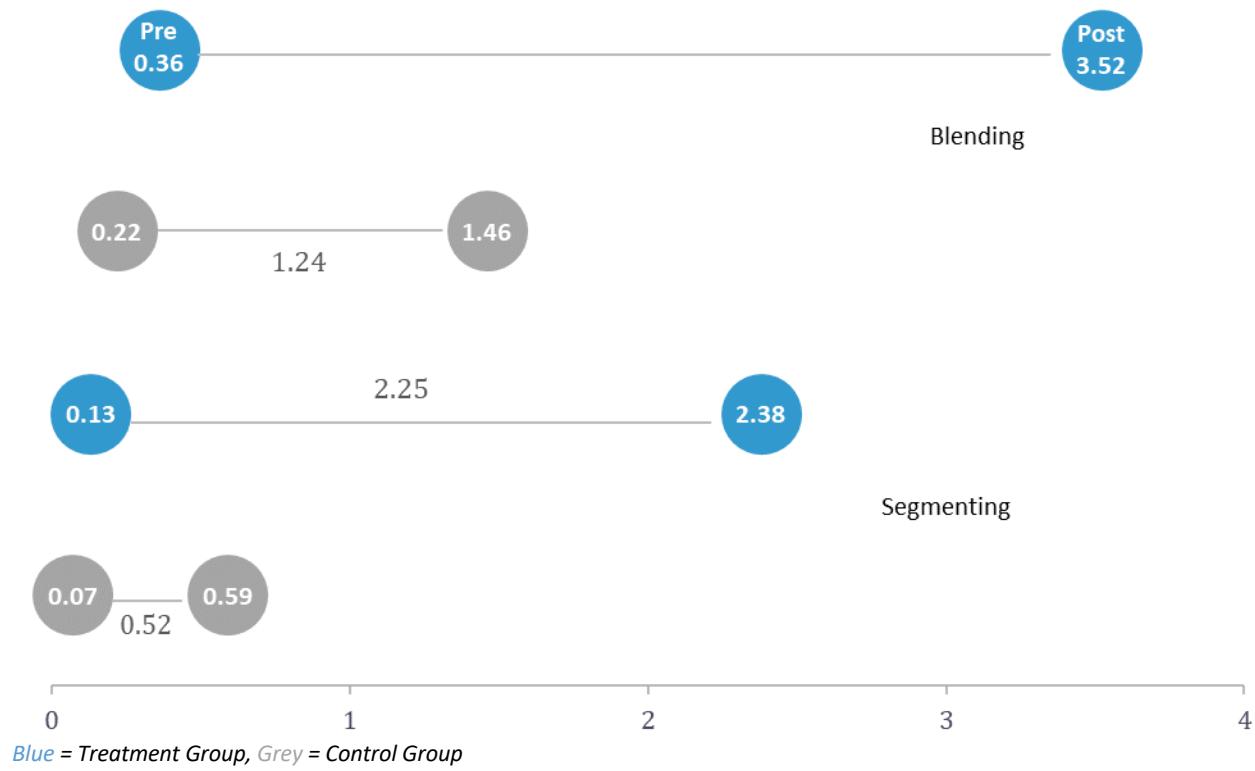


Figure 49: Decoding: Treatment & Control Group Pretest & Posttest Mean Scores

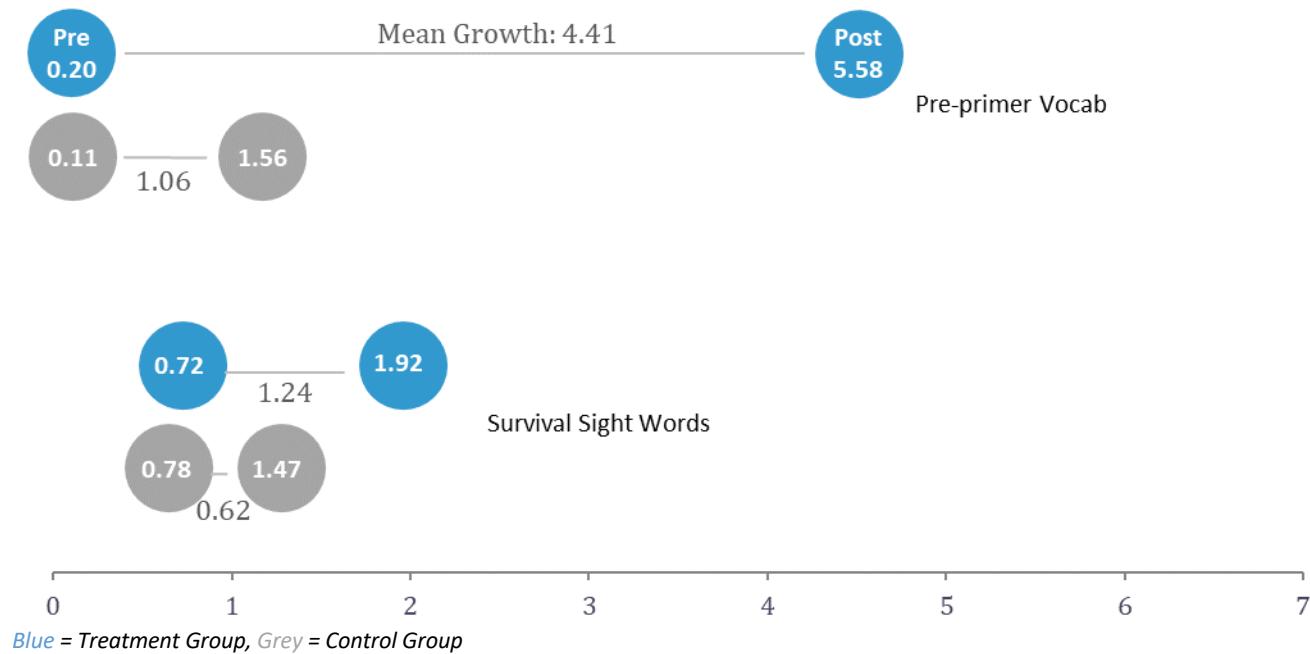


Figure 50: Letter Knowledge: Treatment & Control Group Pretest & Posttest Mean Scores

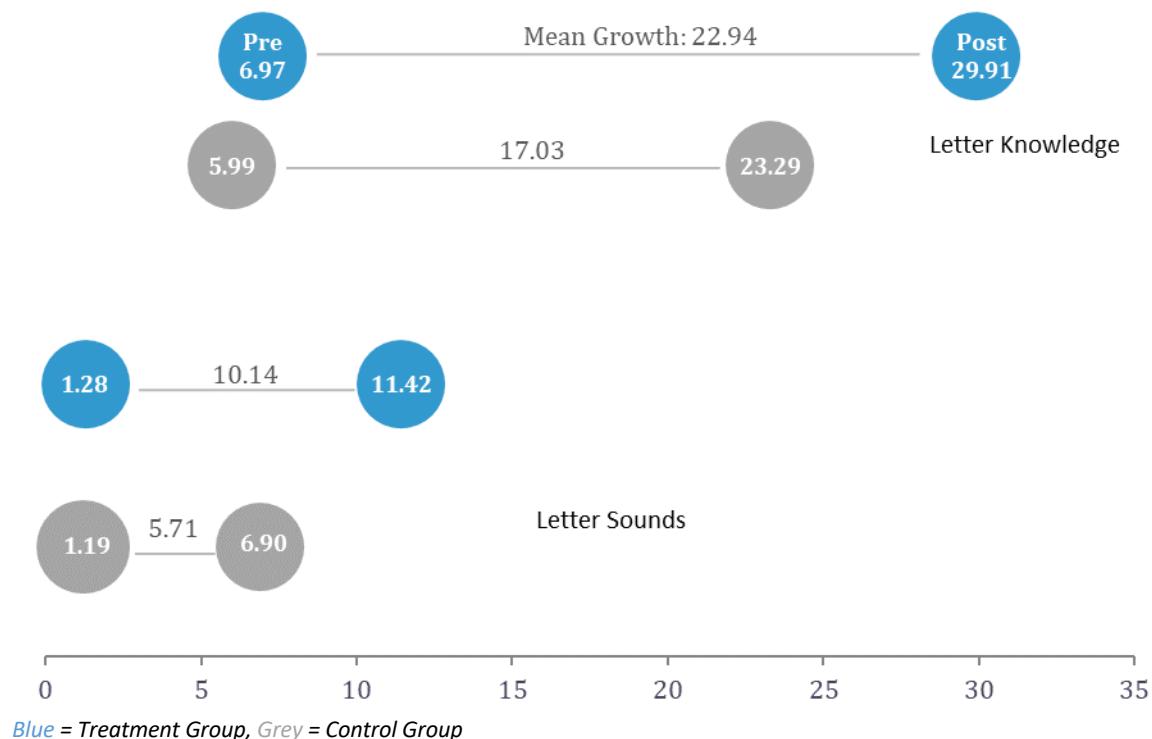
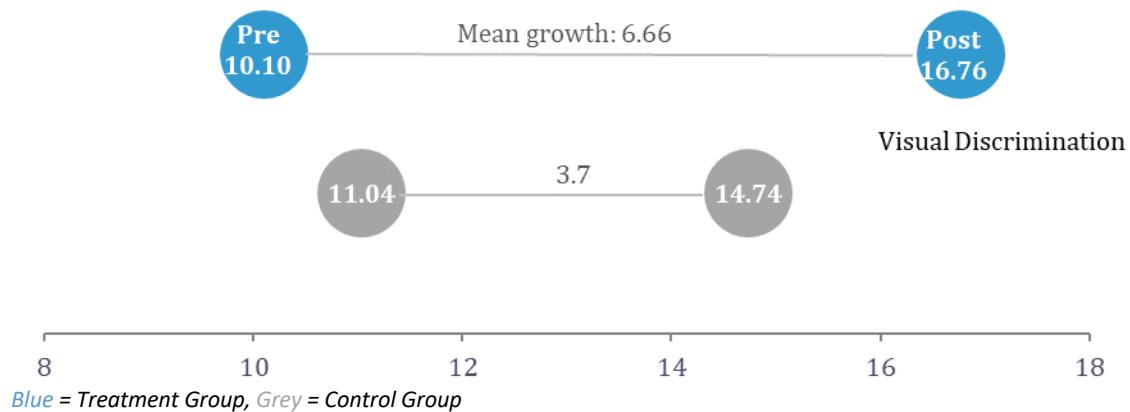


Figure 51: Pre-literacy Discrimination: Treatment & Control Group Pretest & Posttest Mean Scores



Evaluation of Indiana UPSTART

This study investigated the impact of the UPSTART program on four-year-old students in the state of Indiana, measuring the children's literacy, math, and social skills (Evaluation and Training Institute, 2020b). Treatment group students were enrolled in the UPSTART program in 2018-2019, and control group students were not enrolled in the UPSTART program, with a pretest administered during the summer before preschool enrollment and a posttest administered during the summer before kindergarten enrollment. Measures included the Brigance Inventory of Early Development (IED-III) and the Social Skills Improvement System (SSIS).

Treatment group students significantly outperformed control group students across literacy strands, with effect sizes ranging from medium (Visual Discrimination, $ES = 0.29$) to large (Word Recognition, $ES = 0.86$), demonstrating that the UPSTART program can positively impact students' literacy skills before kindergarten entry. While treatment group students did not significantly outperform their control counterparts on the Brigance Math Composite, treatment and control group students performed similarly on social skill scale scores, indicating that students enrolled in UPSTART and their control counterparts experienced similar growth in their social skills over the course of the program.

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 a 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 52: Listening Comprehension Benchmarks at Posttest

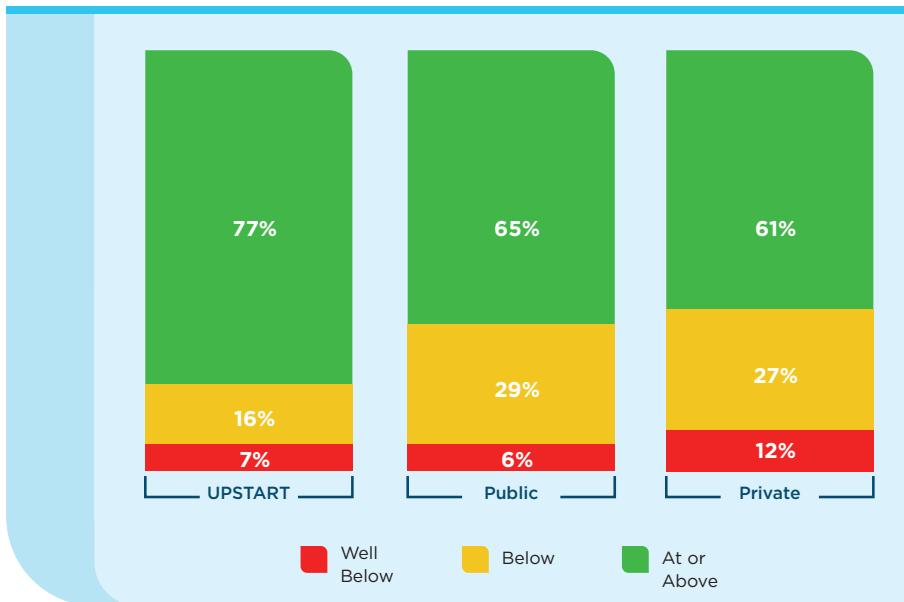
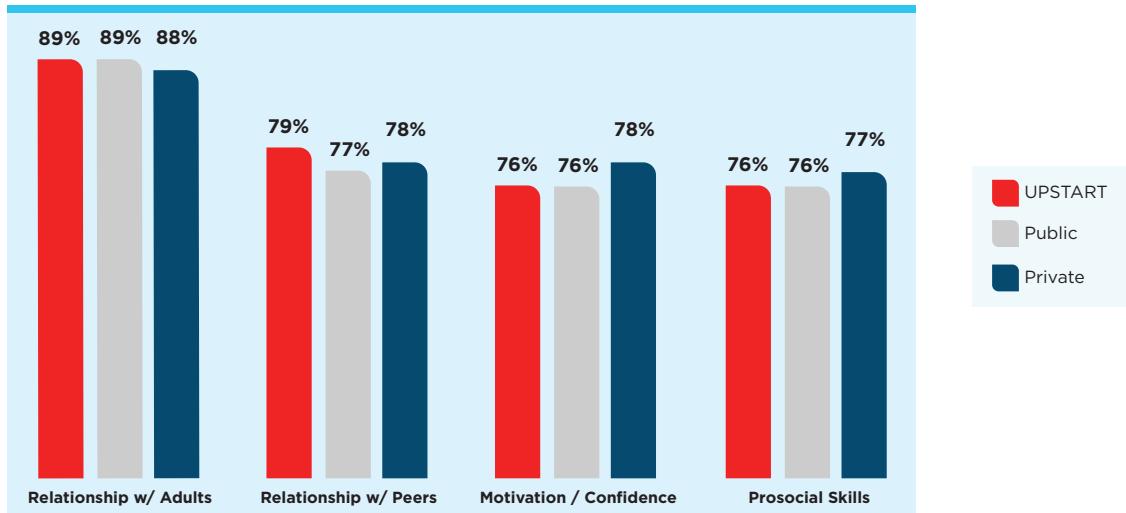


Table 14: Effect Size Estimates by Literacy Domain

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

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

Figure 53: Post-Program Social-Emotional Development Skills by Program Group



High-Quality School Readiness Expansion in Utah, 2018-2019

This study assessed the efficacy of high-quality school readiness programs (Evaluation and Training Institute, 2019b). 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.

Out of the three treatment groups, UPSTART students were the only ones to significantly outperform control students on literacy at the end of program overall Brigance Global Literacy Composite ($ES = .383$), and the only group to outperform control students on key literacy domains, with small to medium effects in phonological awareness ($ES = 0.462$), and small effects in letter knowledge ($ES = 0.270$), as measured by relevant subdomains on the Bader and Brigance, respectively. Readiness to learn was assessed post program using the PELI listening comprehension scale; students from UPSTART had the highest percentage of students with scores at or above benchmark (79% on target) and were the only group to outperform the comparison group (69% on target). By the end of the program year, social emotional development (SED) was similar for all three treatment groups.

Additionally, positive longitudinal results on literacy were found for UPSTART program participation: Students who participated in UPSTART during preschool performed significantly higher than control students on the Acadience Reading composite during the following year at the beginning, middle, and end of kindergarten. Students who participated in UPSTART also achieved significantly higher Acadience Reading composite scores at the beginning of first grade when compared to control students who were matched on demographic characteristics. Among the treatment groups, students who participated in UPSTART were the only ones who showed higher literacy skills in kindergarten and first grade compared to the control group, demonstrating a lasting effect on literacy skills after participating in UPSTART.

Table 15: Effect Size Estimates by Literacy Domain

Literacy Domain	Skills Tested	UPSTART	Public	Private	Instrument
Letter Knowledge	Letter Sounds	0.270*	NS	NS	Brigance
	Recites Alphabet				
	Letter Knowledge				
Listening Comprehension	Inference/Prediction	NS	NS	-0.644**	PELI
	Recollection				
	Cloze				
Phonological Awareness	Rhyme Recognition	0.462*	NS	NS	Bader
	Phonemic Blending				
	Phonemic Segmenting				

Figure 54: Post-Program Social-Emotional Development Skills by Program Group

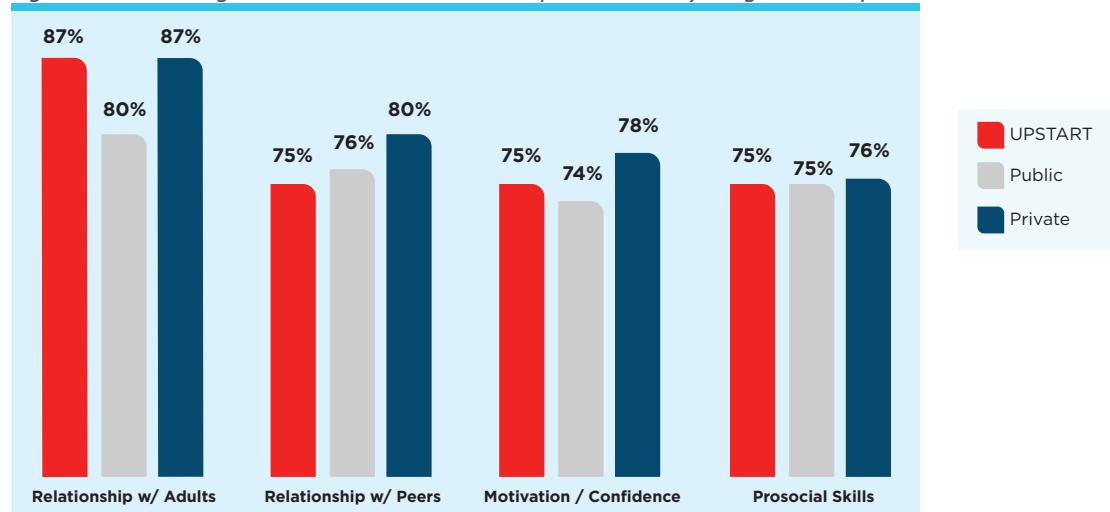
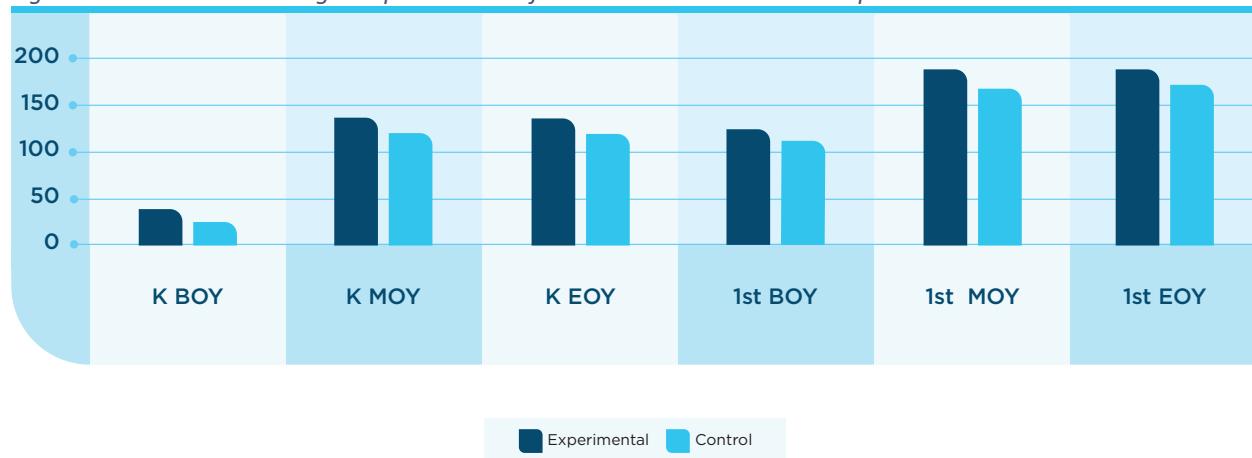


Figure 55: Acadience Reading Composite Scores for UPSTART and Control Groups



Impact of the Upstart Program on Forestalling Summer Learning Loss

This study investigates the Upstart summer program's impact on forestalling summer literacy learning loss in students attending elementary schools in Utah (Hobbs & Overby, 2019b). Treatment students participated in the program during the summer after kindergarten, first grade, and/or second grade. Control students were not enrolled in the Upstart program. The Acadience Reading assessment administered at the end of the year was used as the baseline measure of literacy while the Acadience Reading assessment administered upon return to school at the beginning of the year was used as the outcome measure of literacy. At baseline, the treatment and control groups were equivalent and well balanced across literacy achievement.

Table 16: Baseline Equivalence Assessment for UPSTART Summer Program Participants

Cohort	Measure	Treatment Group			Comparison Group			T-C difference	Standardized difference
		N	Mean	(SD)	N	Mean	(SD)		
1	LNF	297	53.82	12.51	297	53.65	12.36	.17	.01
	PSF	318	59.01	10.52	318	58.87	10.47	.14	.01
	NWF-CLS	300	39.97	15.25	300	40.22	15.34	-.25	-.02
	NWF-WWR	363	3.47	5.96	363	3.39	5.97	.08	.01
	COMP	333	152.22	31.44	333	152.81	31.34	-.59	-.02
2	NWF-CLS	222	90.86	35.29	222	90.68	35.43	.17	.00
	NWF-WWR	190	28.00	14.19	190	28.17	14.12	-.17	-.01
	ORF	251	64.58	27.43	251	64.18	27.29	.4	.01
	COMP	235	215.69	68.23	235	216.09	67.86	-.41	-.01
3	ORF	222	104.54	32.96	222	104.01	32.84	.54	.02
	COMP	207	313.86	69.80	207	314.37	70.02	-.52	.01

Note: Measure names are the following: LNF (Letter Naming Fluency); PSF (Phoneme Segmentation Fluency); NWF-CLS (Nonsense Word Fluency - Correct Letter Sounds); NWF-WWR (Nonsense Word Fluency - Whole Words Read); ORF (Oral Reading Fluency); COMP (Acadience Reading Composite).

Cohort 1

Students in Cohort 1 used the Upstart Summer program during the summer between kindergarten and first grade. Regression analysis revealed that Upstart Summer program participation was a significant predictor in reducing overall literacy learning loss ($p = 0.003$) for students in Cohort 1. Treatment students had a higher average increase than the control students on Reading Composite Scale ($g = 0.22$), NWF-CL ($g = 0.32$), and LNF ($g = 0.17$) test scores.

Table 17: Impact Analysis Results for UPSTART Summer Program Cohort 1 Participants

Measure	Treatment Group			Comparison Group			Impact estimate	Standardized difference	P value
	Unadj N	Model Adjusted Mean	Unadjusted Post-Test SD	Unadj N	Unadjusted Post-Test Mean	Unadjusted Post-Test SD			
COMP Change	333	-22.80	26.14	333	-27.45	23.42	5.17	.21	.003
LNF Change	297	-7.88	10.64	297	-10.70	10.98	1.94	.18	.021
PSF Change	318	-8.76	13.27	317	-11.18	13.26	1.26	.09	.142
NWF-CLS Change	300	-4.48	12.57	300	-7.43	12.75	3.03	.24	.003
NWF-WWR Change	363	0.66	5.61	363	0.18	5.07	0.47	.09	.217

Note: Ns are unadjusted Ns for Treatment and Comparison groups

Measure names are the following: COMP (Acadience Reading Composite Score); LNF (Letter Naming Fluency); PSF (Phoneme Segmentation Fluency); NWF-CLS (Nonsense Word Fluency - Correct Letter Sounds); NWF-WWR (Nonsense Word Fluency - Whole Words Read).

Cohort 2 and Cohort 3

Students in Cohort 2 and Cohort 3 used the Upstart Summer program during the summer between first grade and second grade (Cohort 2) or during the summer between second grade and third grade (Cohort 3). Regression analysis revealed that participation in the Upstart Summer program for Cohort 2 and Cohort 3 was not a significant predictor of forestalling learning loss.

Tier 3: 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 ($N = 1,480$) 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, race/ethnicity, primary home language, and other measures. Waterford was found particularly effective

for MLL students; these groups demonstrated greater gains than the non-MLL group in the comparison schools ($F[1, 1045] = 8.62, p = .003$).

Table 18: MLL & Non-MLL Students on DIBELS Total Reading Score

	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
MLL 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-MLL 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 19: 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								

Measures	N	Covariate			AdjPosttest			F	p
		M	SD	M	SD	ES			
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				
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.

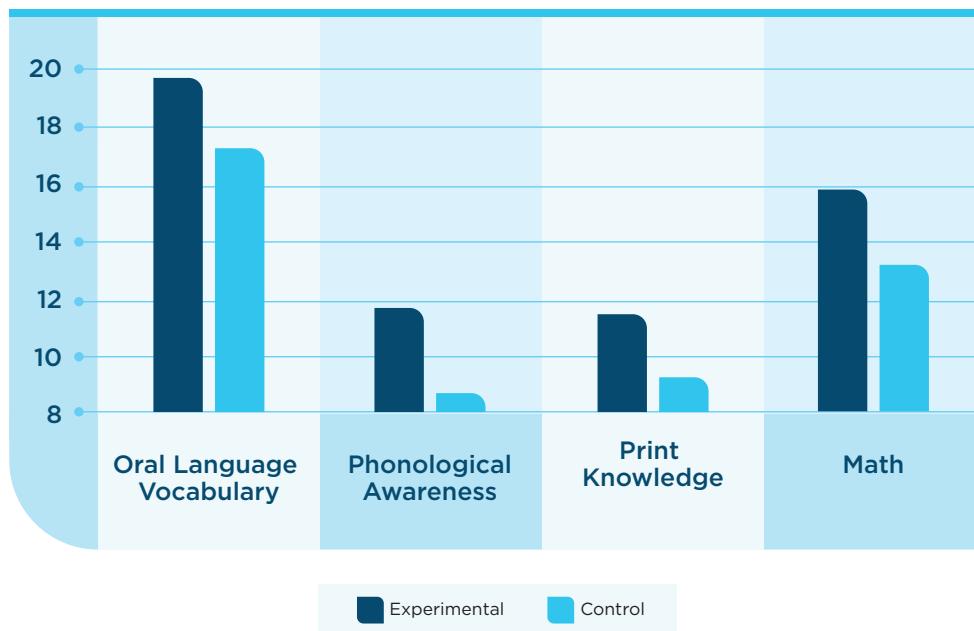
Evaluation of a School District in Florida, 2014-2015

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 reading and math scores of students experiencing poverty (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 17 and Figure 41).

Table 20: 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 56: Spring VPK Sub-Strand Scores, Covarying for Fall



Further analysis was conducted to examine the effects of gender, MLL status, race/ethnicity, and special education services on spring scores. Four separate ANCOVAs 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. The results of the ANCOVAs and the post-hoc analysis demonstrated that the students in the experimental group outperformed the control group on the spring VPK assessment when disaggregated by gender, multi-lingual learner (MLL) status, race/ethnicity, and special education services on all spring strand scores, when covarying for fall scores. Results were significant across all strands for students with and without MLL status, as well as for African American/Black and Latino/a students. The interaction between special education services and treatment was not significant across all strands, which indicates that WEL had a similar impact on all strands for students with and without special education services.

Table 21: 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
MLL Status	1.319	0.251	0.05	0.823	8.447	<0.01	3.81	0.051
Race/Ethnicity	2.43	<0.05	1.928	0.104	2.8	<0.05	2.105	0.076
Special Education Services	2.802	<0.06	2.652	0.071	4.326	<0.05	0.685	0.562

Figure 57: Oral Language Vocabulary Spring Scores by Demographics

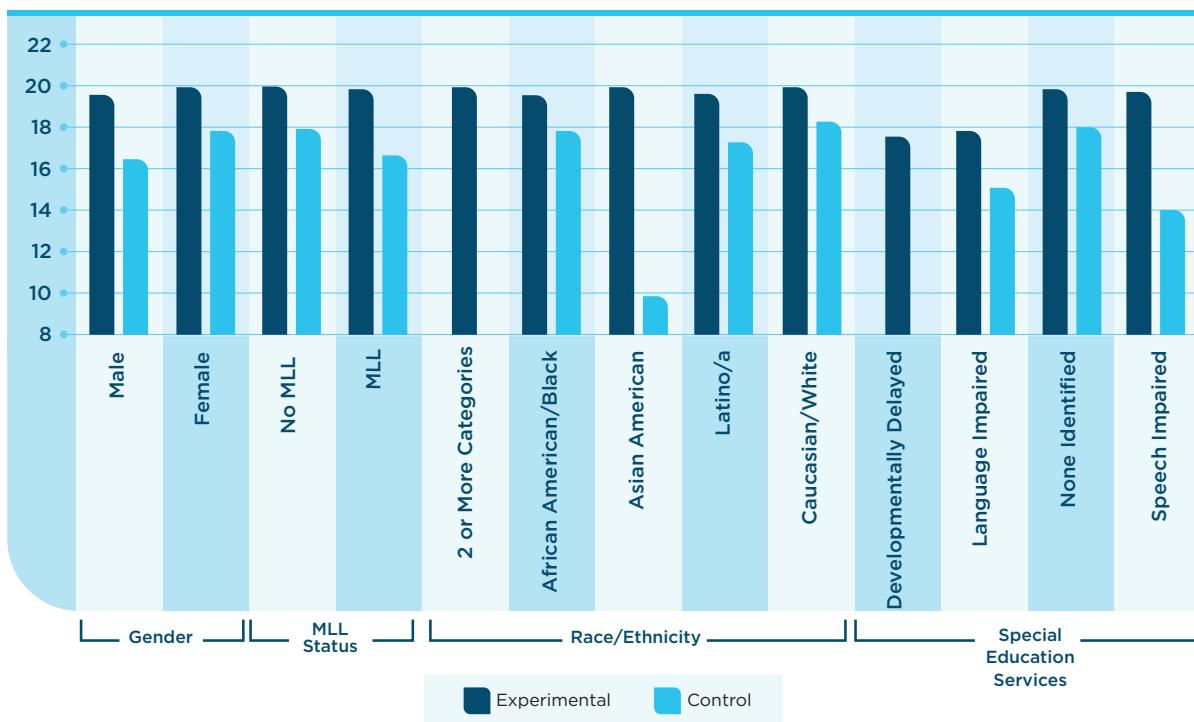


Figure 58: Phonological Awareness Spring Scores by Demographics

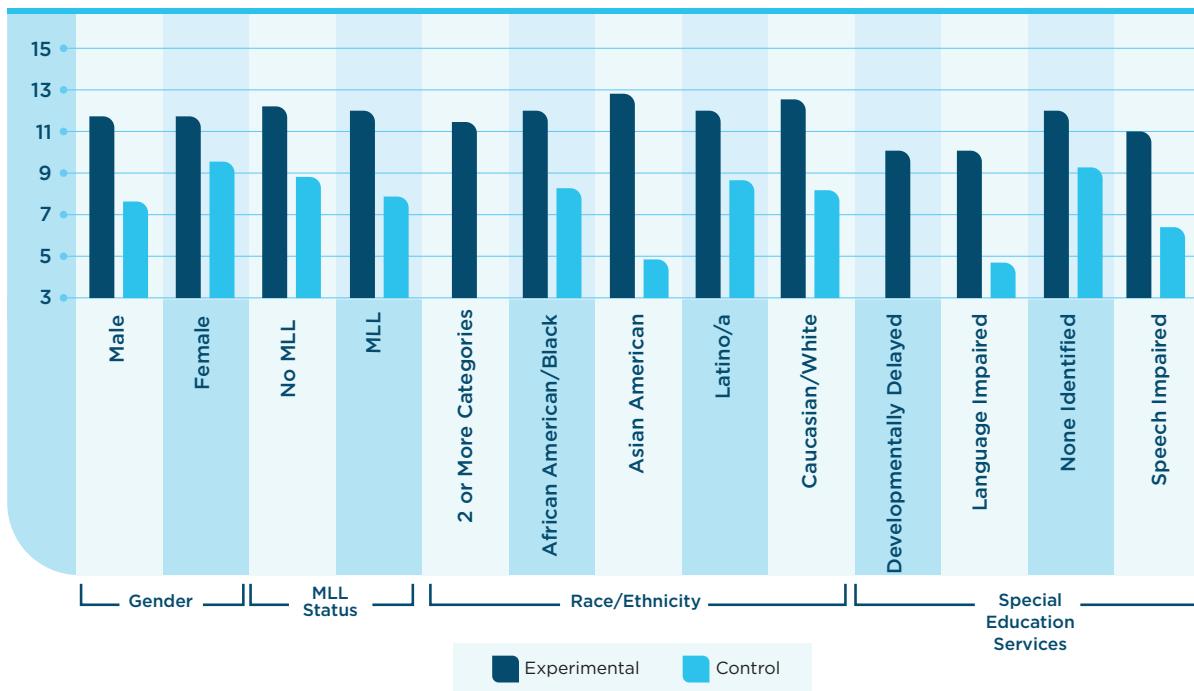


Figure 59: Print Knowledge Spring Scores by Demographics

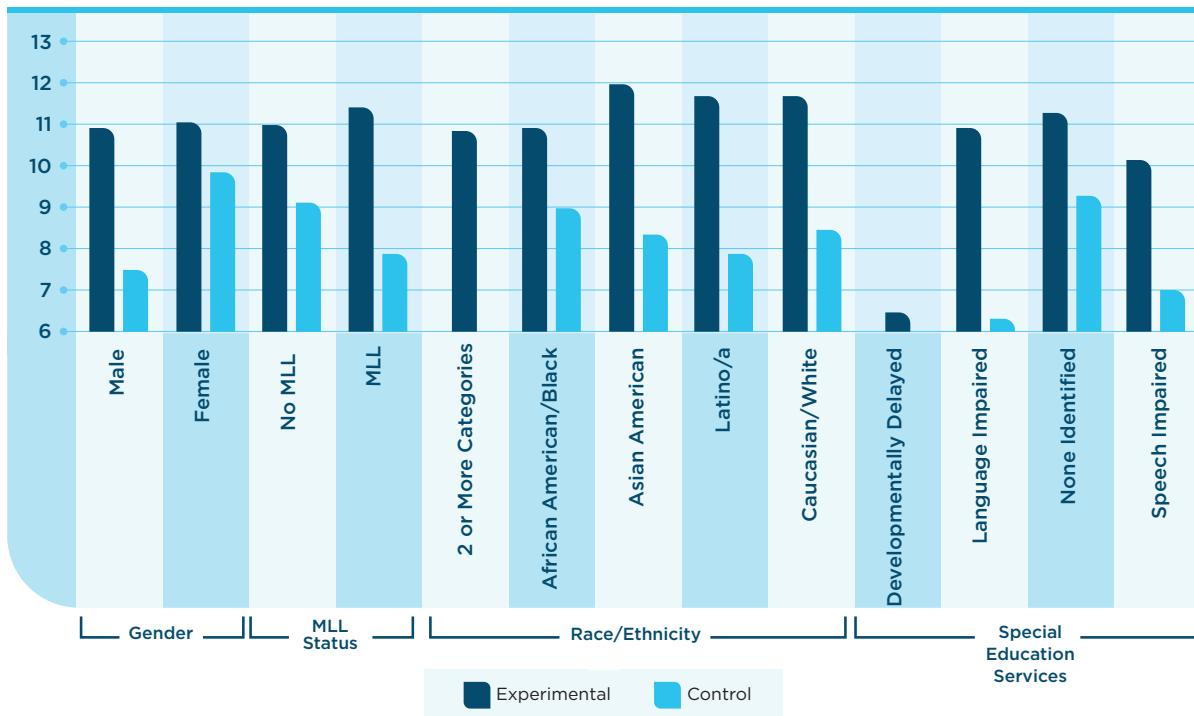
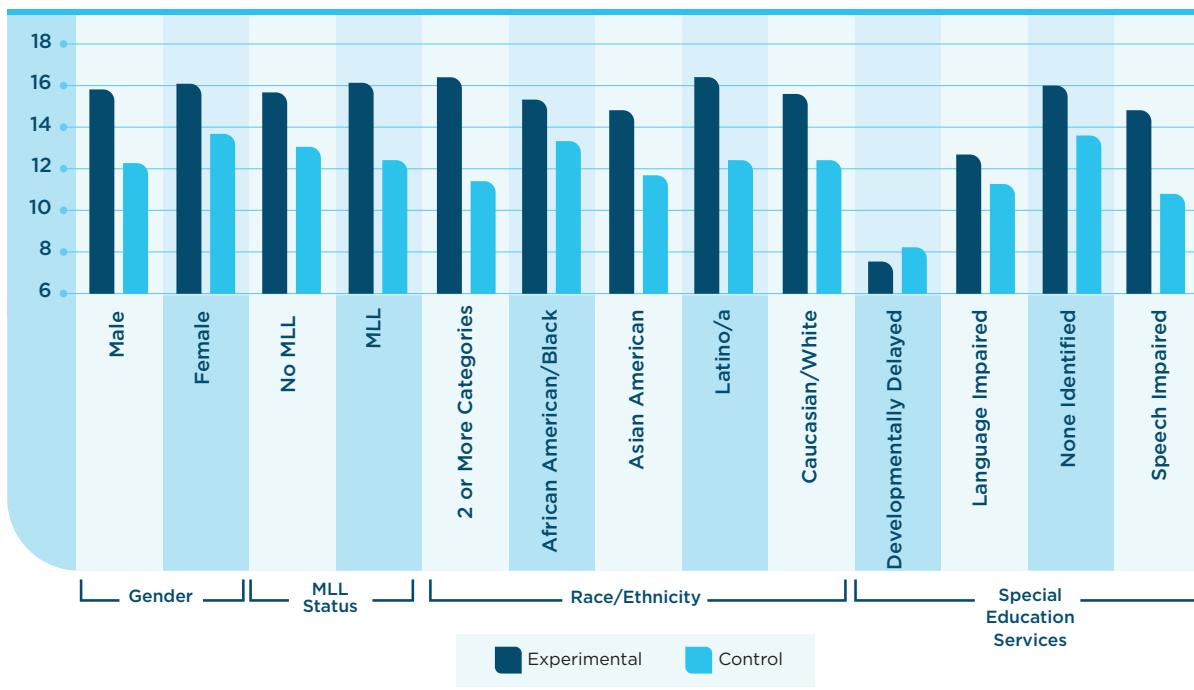


Figure 60: Math Spring Scores by Demographics



Evaluation of the Waterford Early Reading Program in California

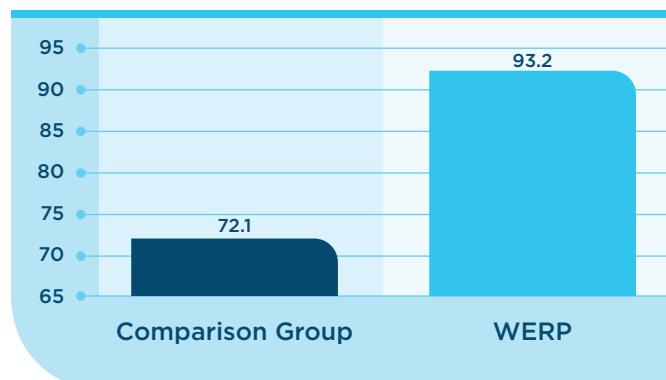
The sample in the current study was composed of 80% Latino/a students and 20% African American/Black students (Shamir & Goethe, 2015). Ninety percent of students qualified for free/reduced lunch. The treatment group consisted of students who used the Waterford Early Reading Program (ERP). The control 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-strands: Letter Sound Fluency (LSF), Letter Name Fluency (LNF), and Nonsense Word Fluency (NWF).

Figure 61: Mean Scores by AIMSweb Sub-Strands for Kindergarten Students



Figure 62: 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 48-49).

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 ($d = 0.29$).

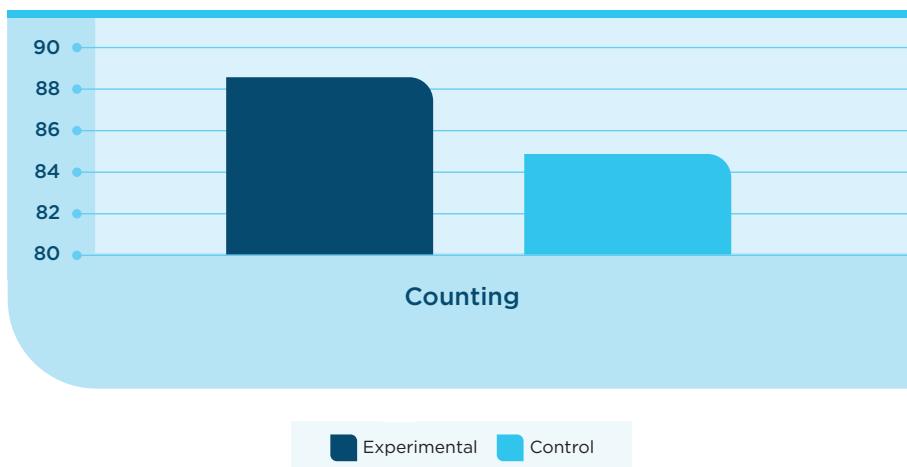
Analysis of Counting end of year scores, while covarying for beginning of year scores, did not reveal a significant difference between groups, $F(1, 168) = 3.43, p = .066$; however, Waterford students had higher end of year scores ($M = 88.64$) than control students ($M = 84.33$).

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



Figure 64: 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 services on Number Identification end of year scores (see Figure 50).

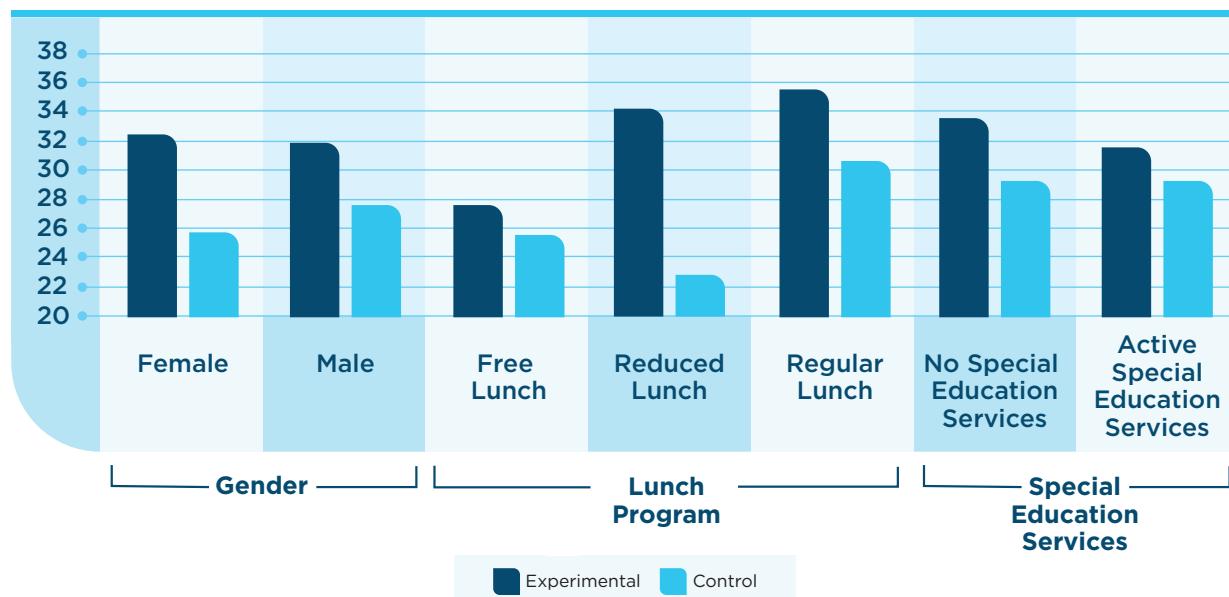
There was no significant interaction between the effects of gender and Waterford curriculum on Number Identification end of year scores, covarying for beginning of year scores, $F(1, 166) = 2.90, p = .091$. Simple effects analysis showed that for females, students in the experimental group significantly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

Figure 65: Number Identification End of Year Scores by Demographics



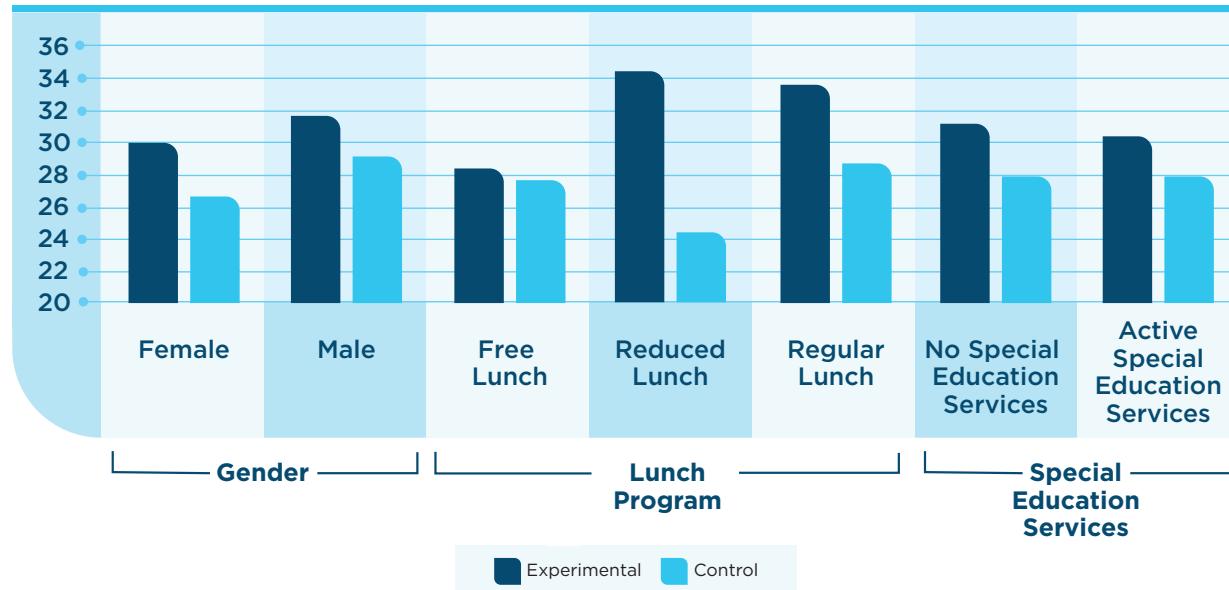
Further analysis was conducted to examine the effects of gender, lunch program, and special education services on Quantity Discrimination end of year scores (see Figure 51).

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 services 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 services and active special education services, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 66: 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 67: First Grade mCLASS: Math End of Year Scores by Sub-Strand

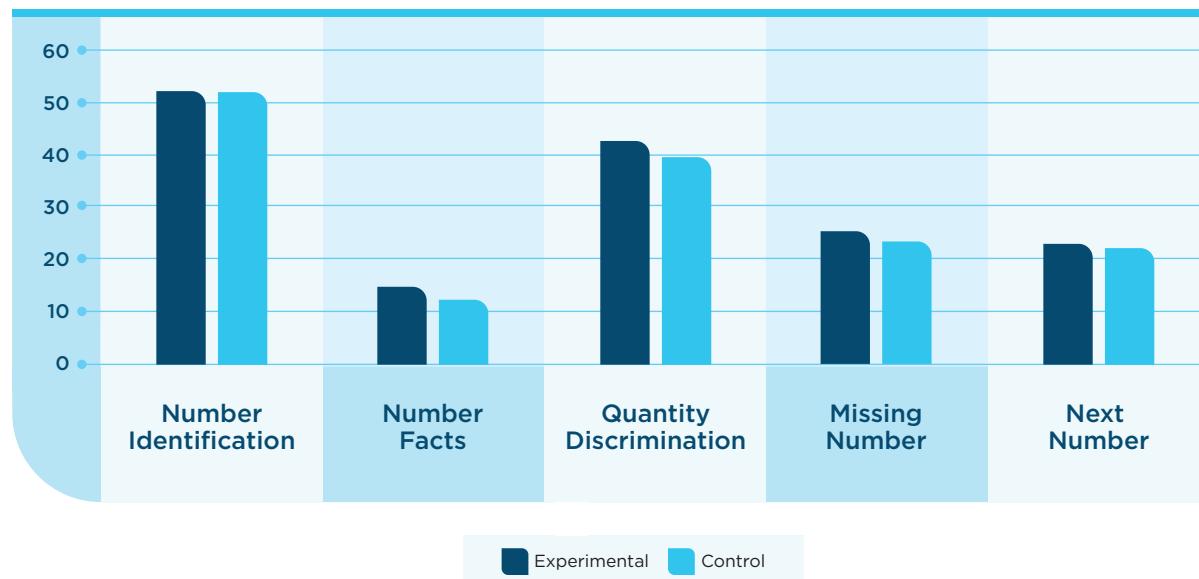
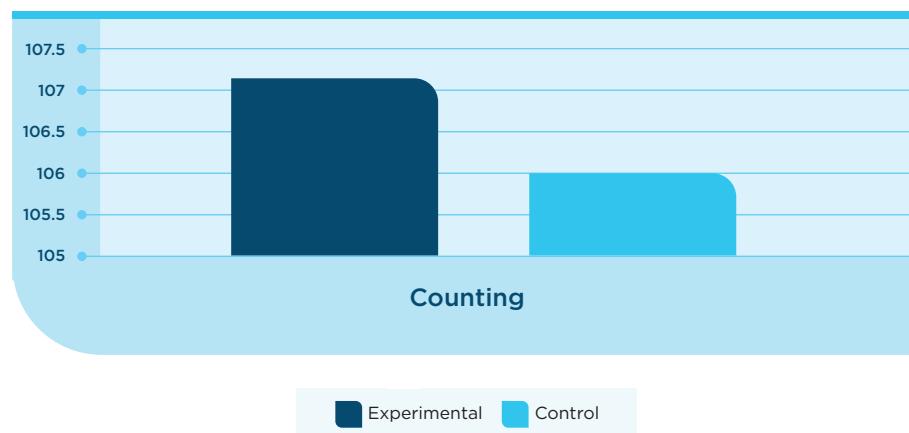


Figure 68: First Grade mCLASS: Math Counting End of Year Scores



GROUP DIFFERENCES BY DEMOGRAPHICS USING ANCOVAs

Further analysis was conducted to examine the effects of gender, lunch program, and special education services 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 services 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 services, the experimental group significantly outperformed the control group. For students with active special education services, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 69: First Grade Number Facts End of Year Scores by Demographics



Further analysis was conducted to examine the effects of gender, Multi-lingual Learner (MLL) status, lunch program, and special education services on end of year Quantity Discrimination scores.

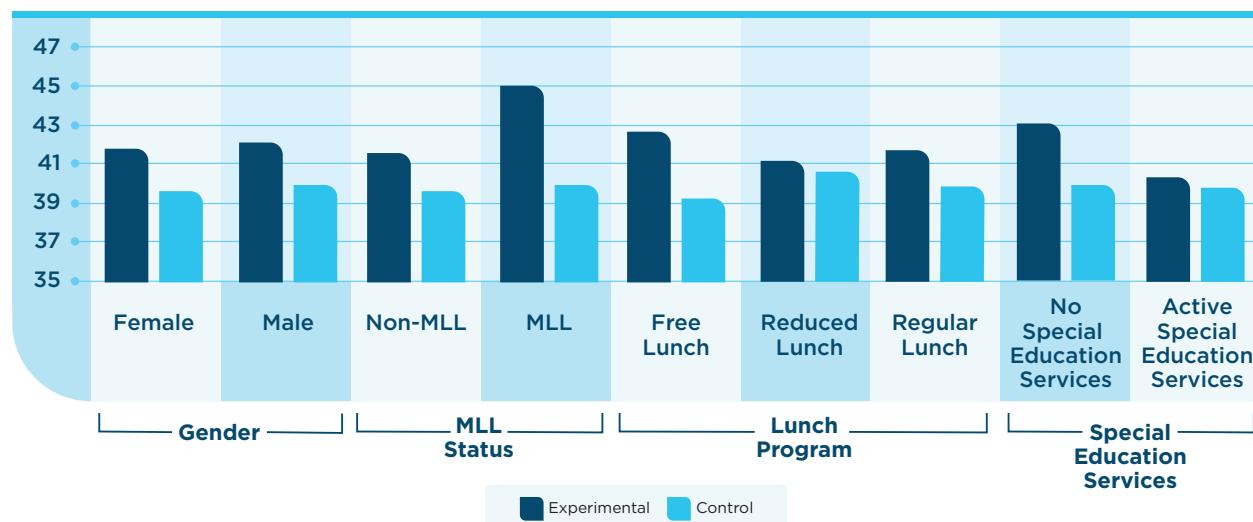
There was no significant interaction between the effects of gender and Waterford curriculum on Quantity Discrimination end of year scores, covarying for beginning of year scores, $F(1, 317) = 0.01, p = .918$. Simple effects analysis showed that for males and females, students' scores in the experimental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

There was no significant interaction between the effects of MLL 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-MLL students' scores in the experimental group were slightly higher than in the control group, approaching significance. MLL 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 services 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 services, the experimental group significantly outperformed the control group. For students with active special education services, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 70: First Grade Quantity Discrimination End of Year Scores by Demographics



Further analysis was conducted to examine the effects of gender, MLL status, lunch program, and special education services 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.

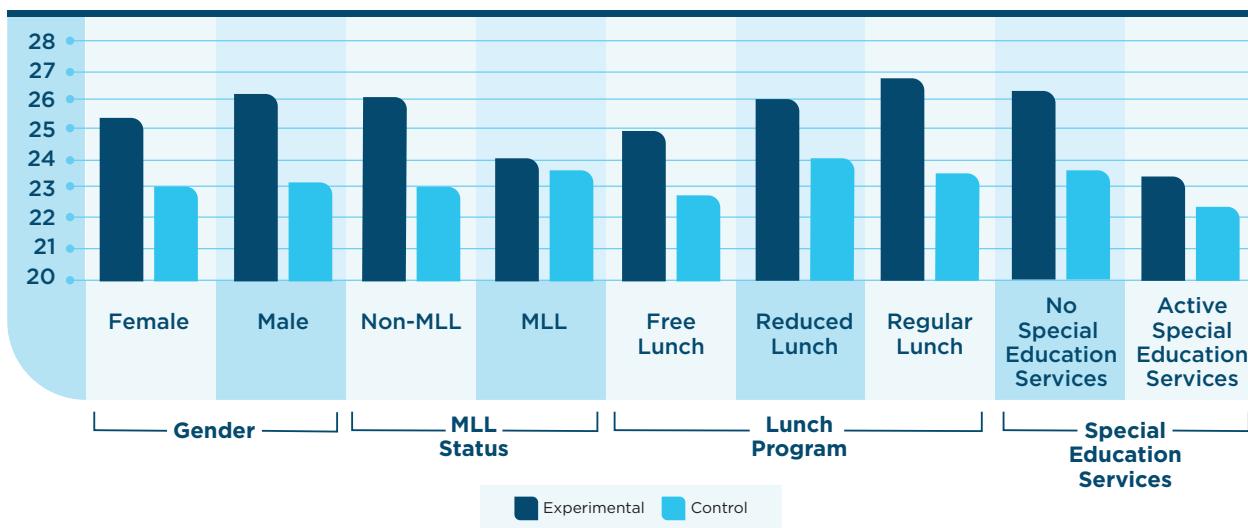
There was no significant interaction between the effects of MLL status and Waterford curriculum on Missing Number end of year scores, covarying for beginning of year scores, $F(1, 317) = 1.47, p = .227$. Simple effects analysis showed that Non-MLL students in the experimental group significantly outperformed students in the control group. MLL 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 services 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 services, the experimental group significantly outperformed the control group. For students with active special education services, scores in the experimental group were slightly higher than in the control group, but the difference was not significant.

Figure 71: First Grade Missing Number End of Year Scores by Demographics



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

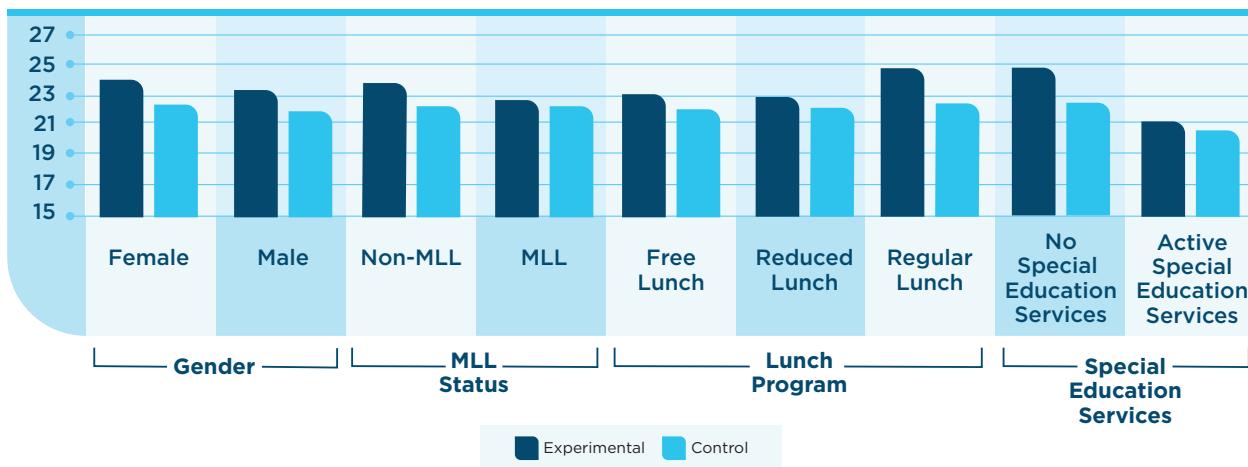
There was no significant interaction between the effects of gender and Waterford curriculum on Next Number end of year scores, covarying for beginning of year scores, $F(1, 317) = 0.07, p = .787$. Simple effects analysis showed that for males, students in the experimental group significantly outperformed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

Figure 72: 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 22: Brigance Posttest Analysis of Treatment-Control Group Differences

<i>Brigance Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SEM</i>	<i>t</i>	<i>Significance</i>
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 23: Bader Posttest Analysis of Treatment-Control Group Differences

<i>Bader Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SEM</i>	<i>t</i>	<i>Significance</i>
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 24: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance	<u>Treatment Group</u>		<u>Control Group</u>		Significance <i>p < .01</i>
	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 25: Treatment-Control Group Differences in Growth Rates on the Bader

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

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.

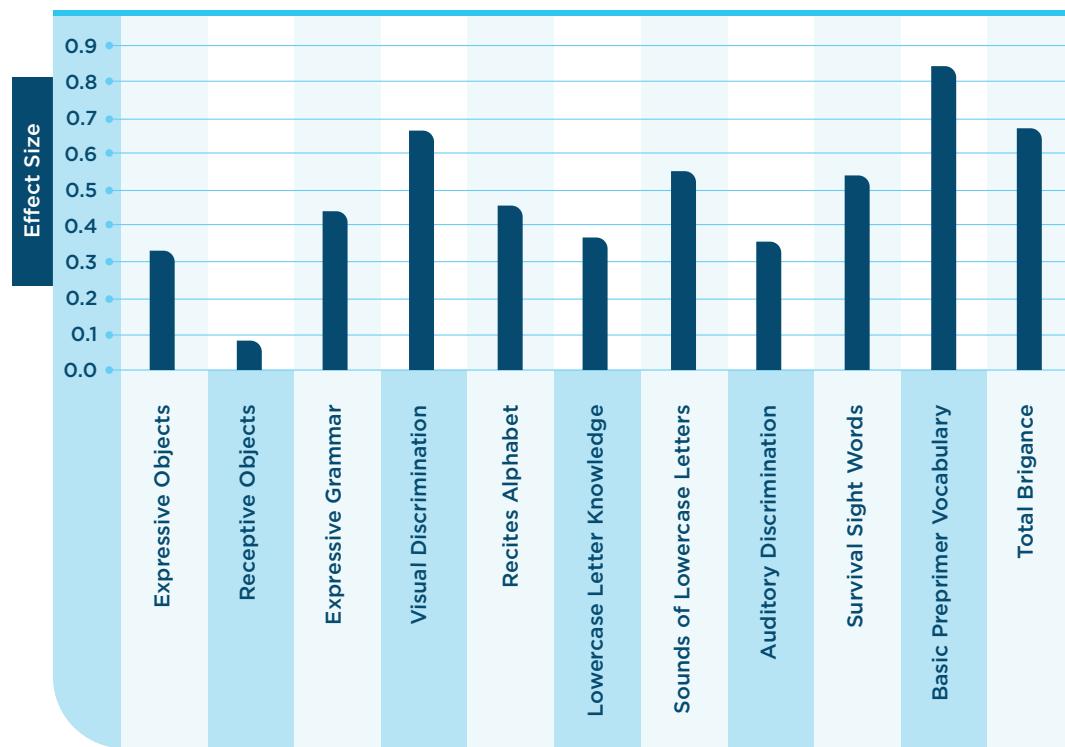
BRIGANCE RESULTS

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

** $p < .01$

Figure 73: UPSTART's Impact on the Brigance in Effect Size Units



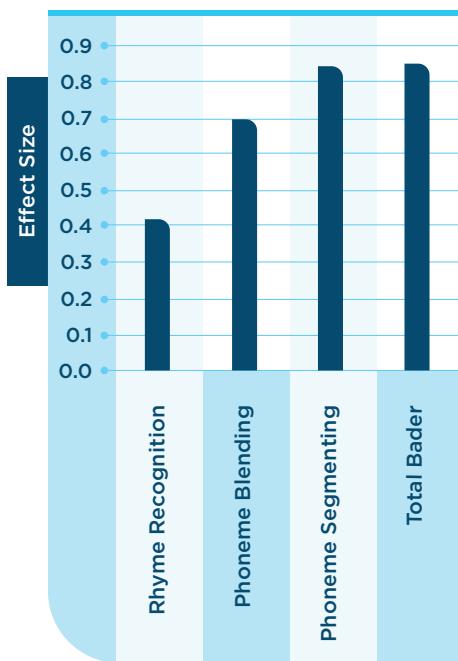
BADER RESULTS

Table 27: Bader Posttest Analysis of Treatment-Control Group Differences

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

** $p < .01$

Figure 74: UPSTART's Impact on the Bader in Effect Size Units



BRIGANCE GROWTH SCORE RESULTS

Table 28: 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 29: Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	<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 30: 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		
Survival Sight Words	Treatment	101	3.703	3.968	1.330	NS

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

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

Table 31: Brigance Effect Size Estimates

<i>Test</i>	<i>Effect Size</i>	<i>Significance</i>	<i>Magnitude of Effect</i>
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 32: Bader Posttest Analysis of Treatment-Control Group Differences

<i>Bader Posttest</i>	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>Significance</i>
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 33: 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 34: 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 35: 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 better than the control group on the Total Brigance and significantly better than the control group on the 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 36: 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	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 37: 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 38: 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 39: 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 40: Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance Test	<u>Control Group</u>		<u>Treatment Group</u>		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 41: Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	<u>Control Group</u>		<u>Treatment Group</u>		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 races/ethnicities benefitted from WEL. African American/Black and Latino/a students in the experimental group scored

higher than those in the control group; however, the difference was significant for only African American/Black students, who used the program significantly more than Latino/a students.

Table 42: Kindergarten DRA Gains & End of Year Scores Covarying For Middle of Year Scores

	Gains					ANCOVA				
	Experimental		Control		p	Experimental		Control		p
M	SD	M	SD		M	SD	M	SD		
Kindergarten Overall	2.41	1.25	1.73	1.38	.00**	4.84	2.01	4.32	2.00	.01*
Gender										
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
Race/Ethnicity										
African American/Black	2.40	1.20	1.20	1.32	.00**	4.82	2.11	3.86	1.60	.01**
Latino/a	2.04	1.31	2.00	0.85	.93	4.55	2.32	4.52	0.99	.94

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

Table 43: First Grade DRA Gains & End of Year Scores Covarying For Middle of Year Scores

	Gains					ANCOVA				
	Experimental		Control		p	Experimental		Control		p
M	SD	M	SD		M	SD	M	SD		
First Grade Overall	3.53	1.33	2.84	1.63	.00**	11.15	1.96	10.02	3.65	.00**
Gender										
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**
Race/Ethnicity										
African American/Black	3.44	1.45	2.60	1.98	.01**	11.04	2.21	9.47	4.18	.00**
Caucasian/White	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 gender and race/ethnicity. These results indicate WEL potentially positively impacts early literacy skills.

KINDERGARTEN

Figure 75: Kindergarten End of Year Scores Covarying for Beginning of Year Scores by Strand

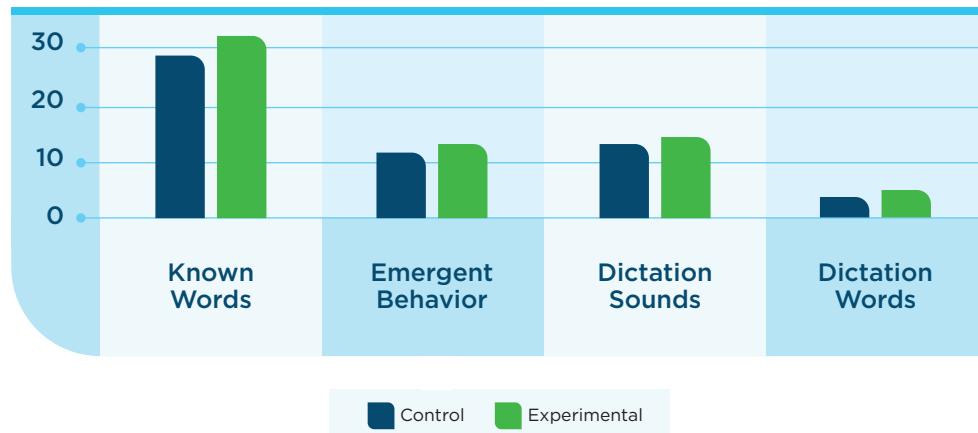


Figure 76: Kindergarten End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics

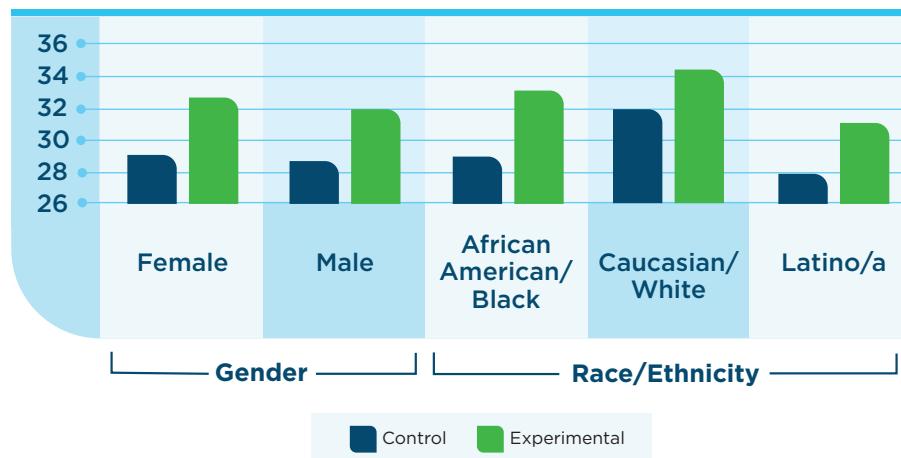


Figure 77: Kindergarten End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics

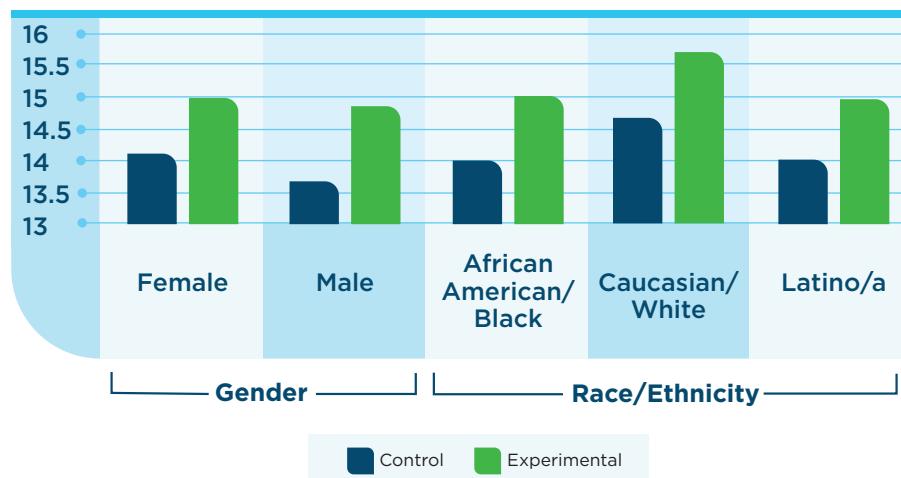


Figure 78: Kindergarten End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

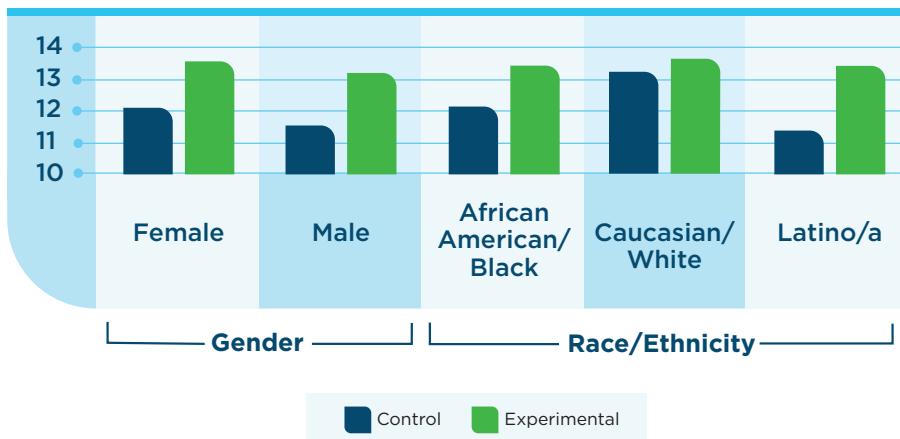
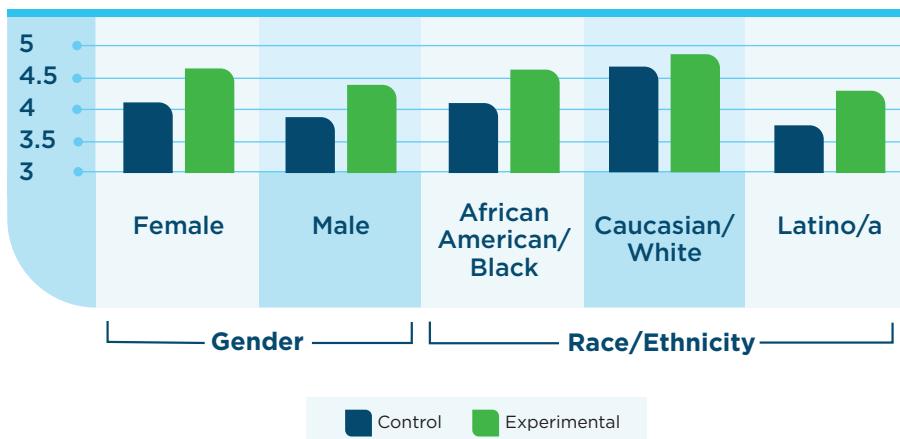


Figure 79: Kindergarten End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



FIRST GRADE

Figure 80: First Grade End of Year Scores Covarying for Beginning of Year Scores by Strand

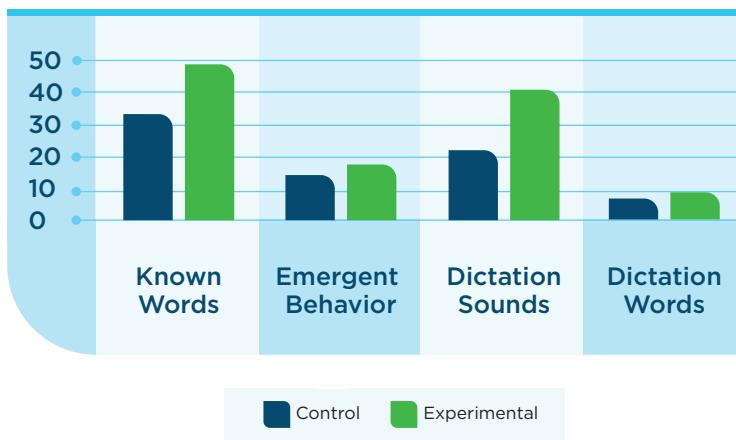


Figure 81: First Grade End of Year Known Words Scores Covarying for Beginning of Year Scores by Demographics

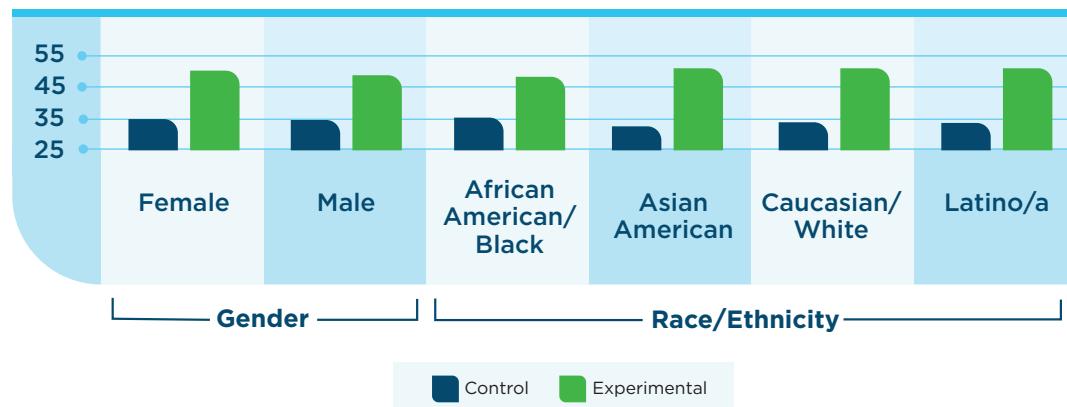


Figure 82: First Grade End of Year Emergent Behavior Scores Covarying for Beginning of Year Scores by Demographics

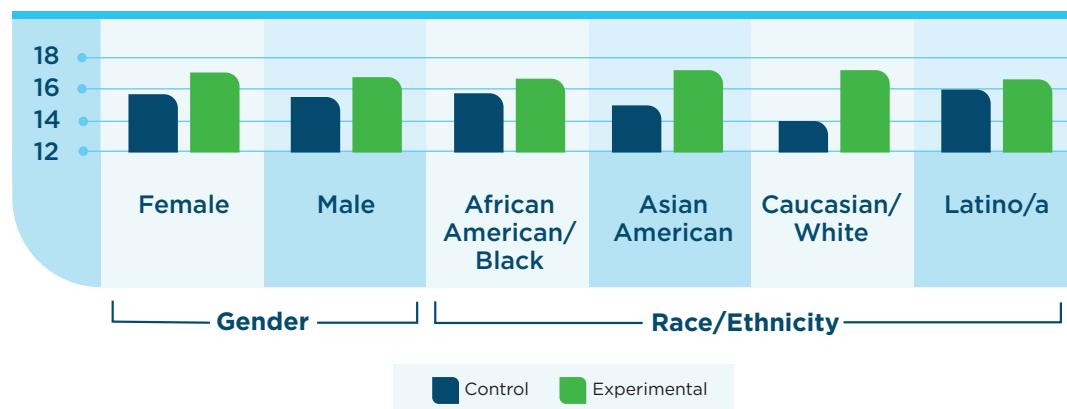


Figure 83: First Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

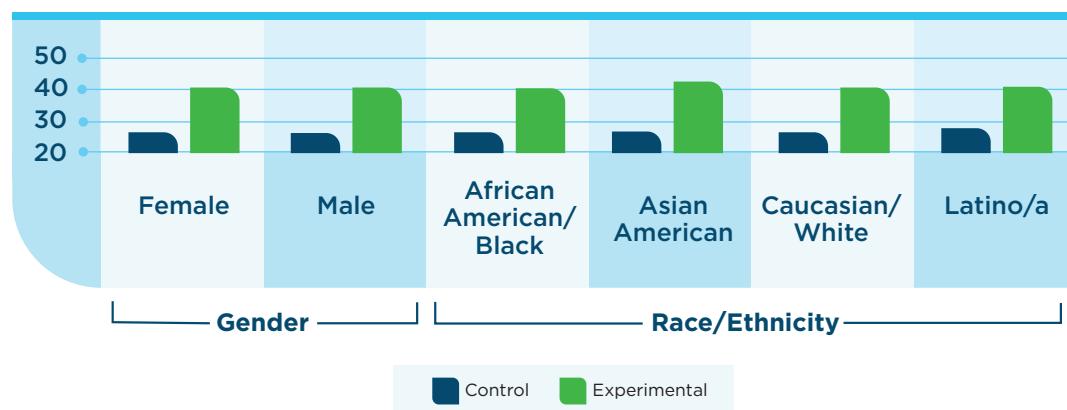
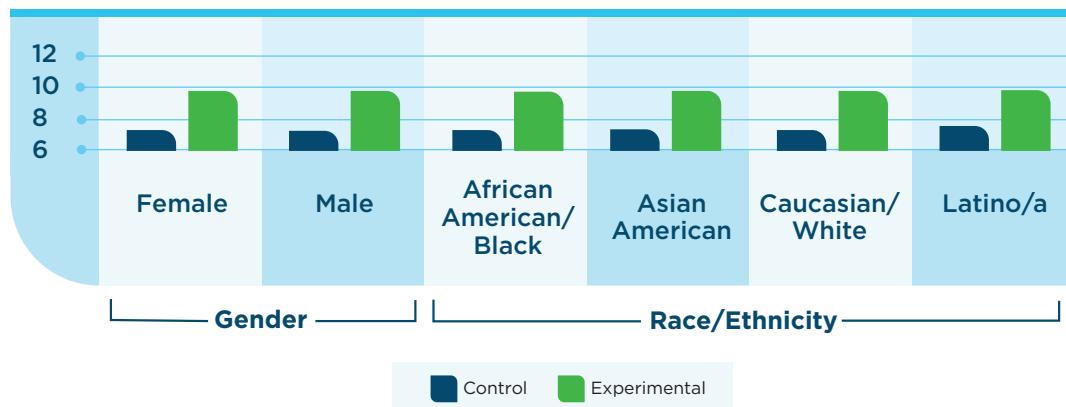


Figure 84: First Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



SECOND GRADE

Figure 85: Second Grade End of Year Scores Covarying for Beginning of Year Scores by Strand

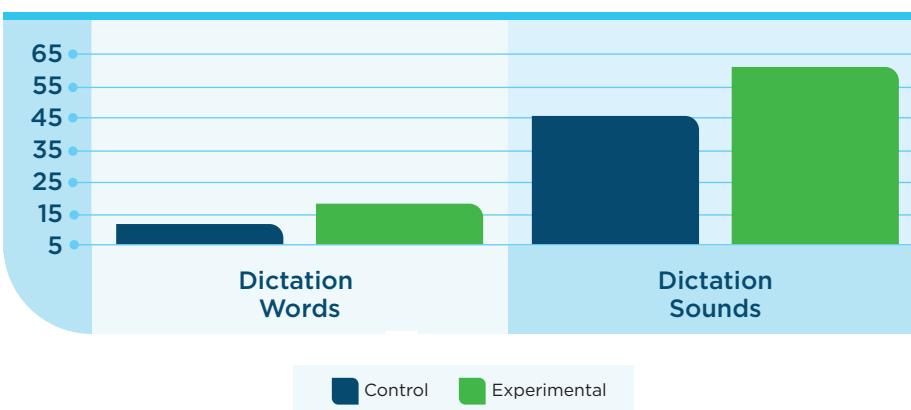


Figure 86: Second Grade End of Year Dictation Sounds Scores Covarying for Beginning of Year Scores by Demographics

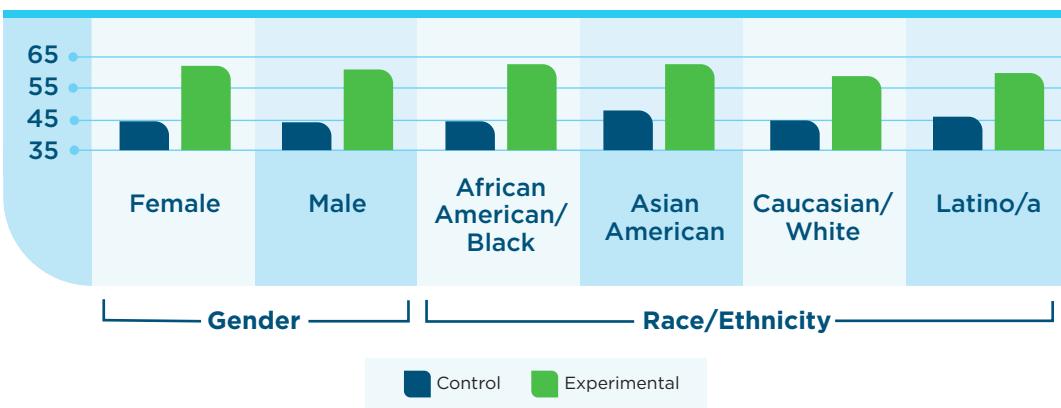
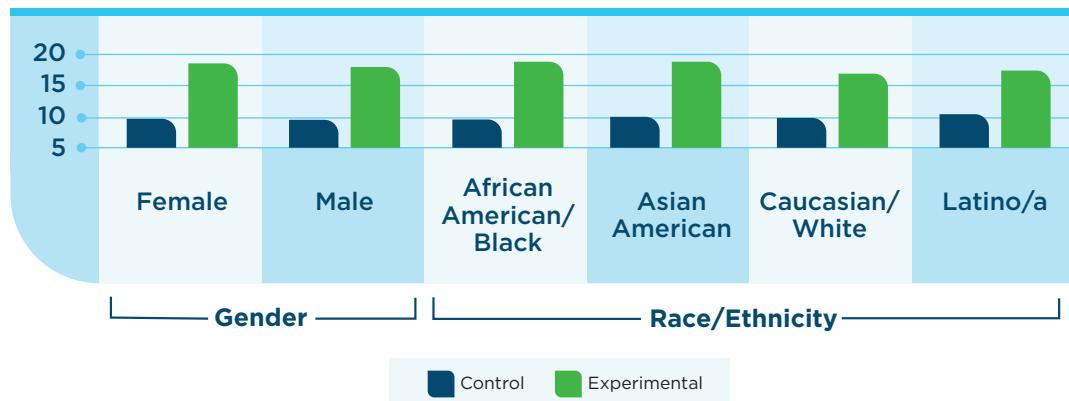


Figure 87: Second Grade End of Year Dictation Words Scores Covarying for Beginning of Year Scores by Demographics



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 88: Kindergarten Beginning & End of Year Overall RIT Scores

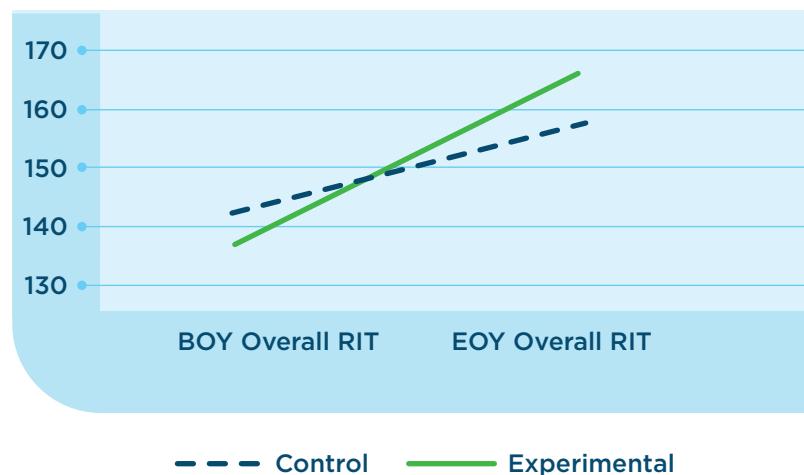
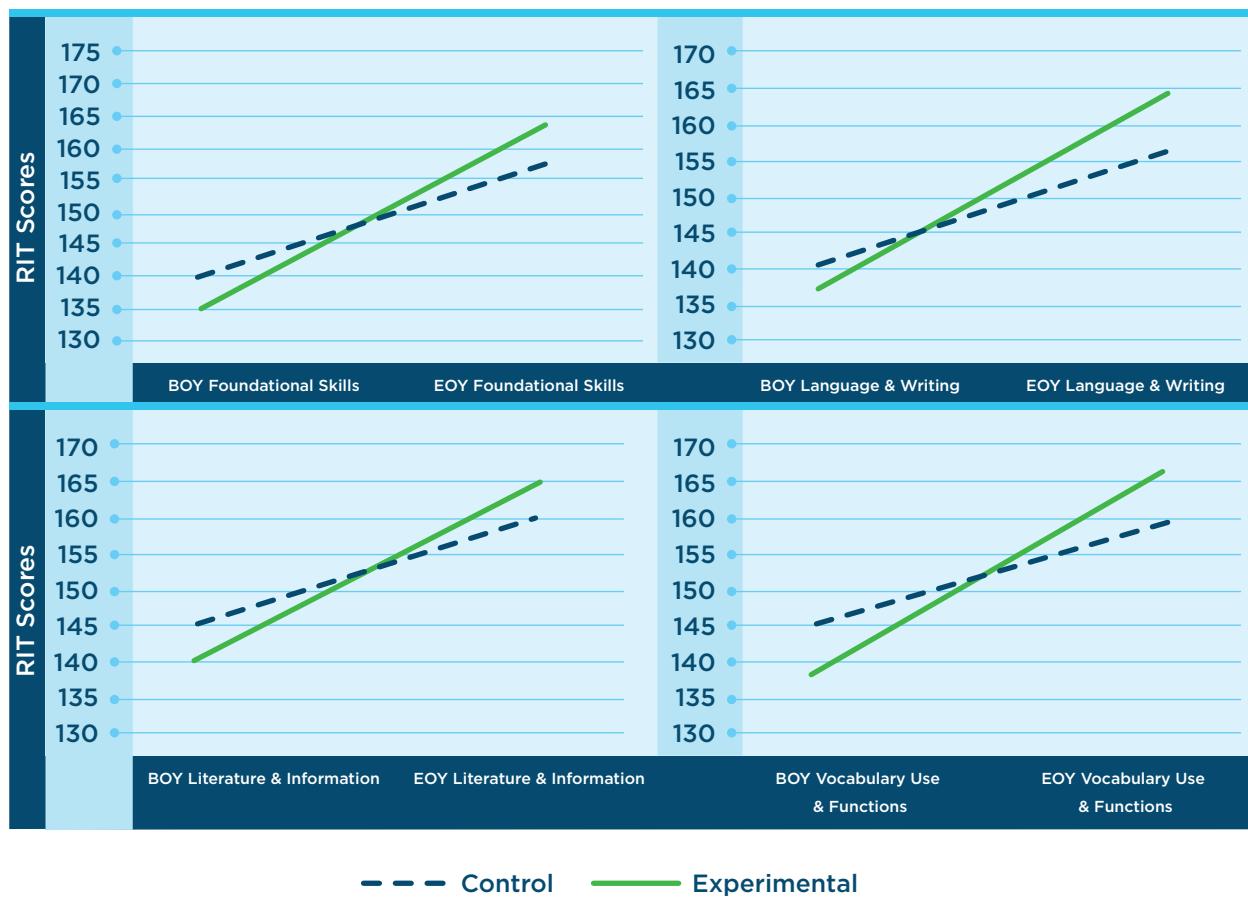
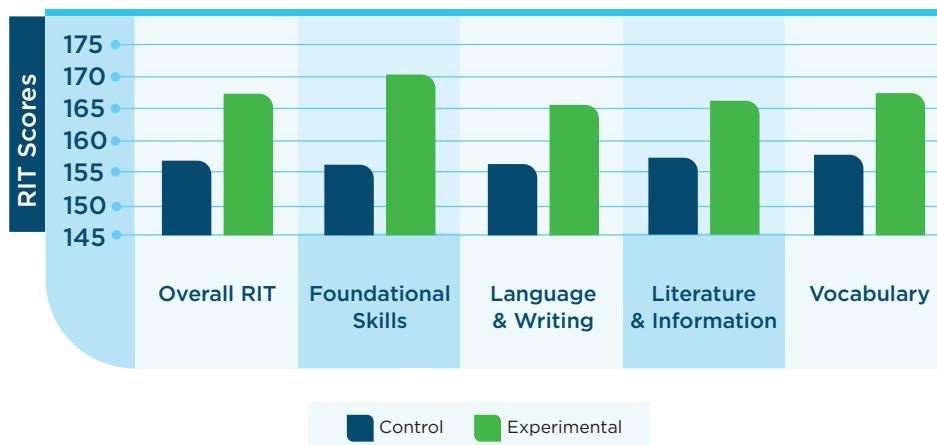


Figure 89: Kindergarten Beginning & End of Year RIT Scores by Strand



GROUP DIFFERENCES USING ANALYSIS OF COVARIANCE (ANCOVA)

Figure 90: Kindergarten End of Year RIT Scores while Covarying for Beginning of Year Scores



FIRST GRADE

GROUP DIFFERENCES USING INDEPENDENT SAMPLES T-TESTS

Figure 91: First Grade Beginning & End of Year RIT Scores

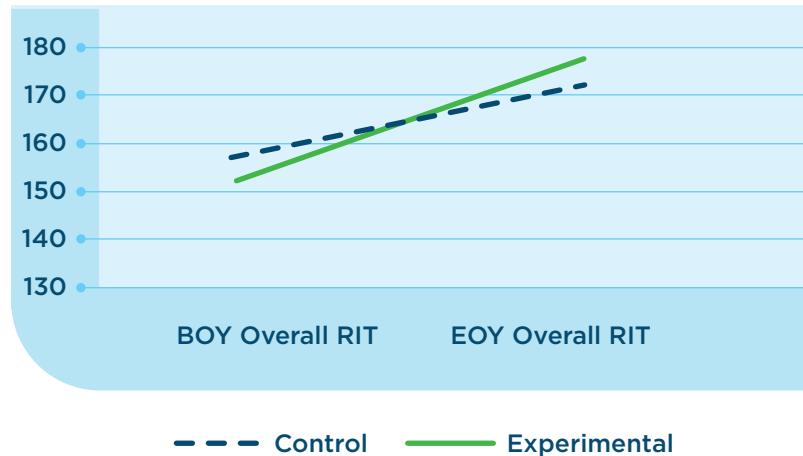
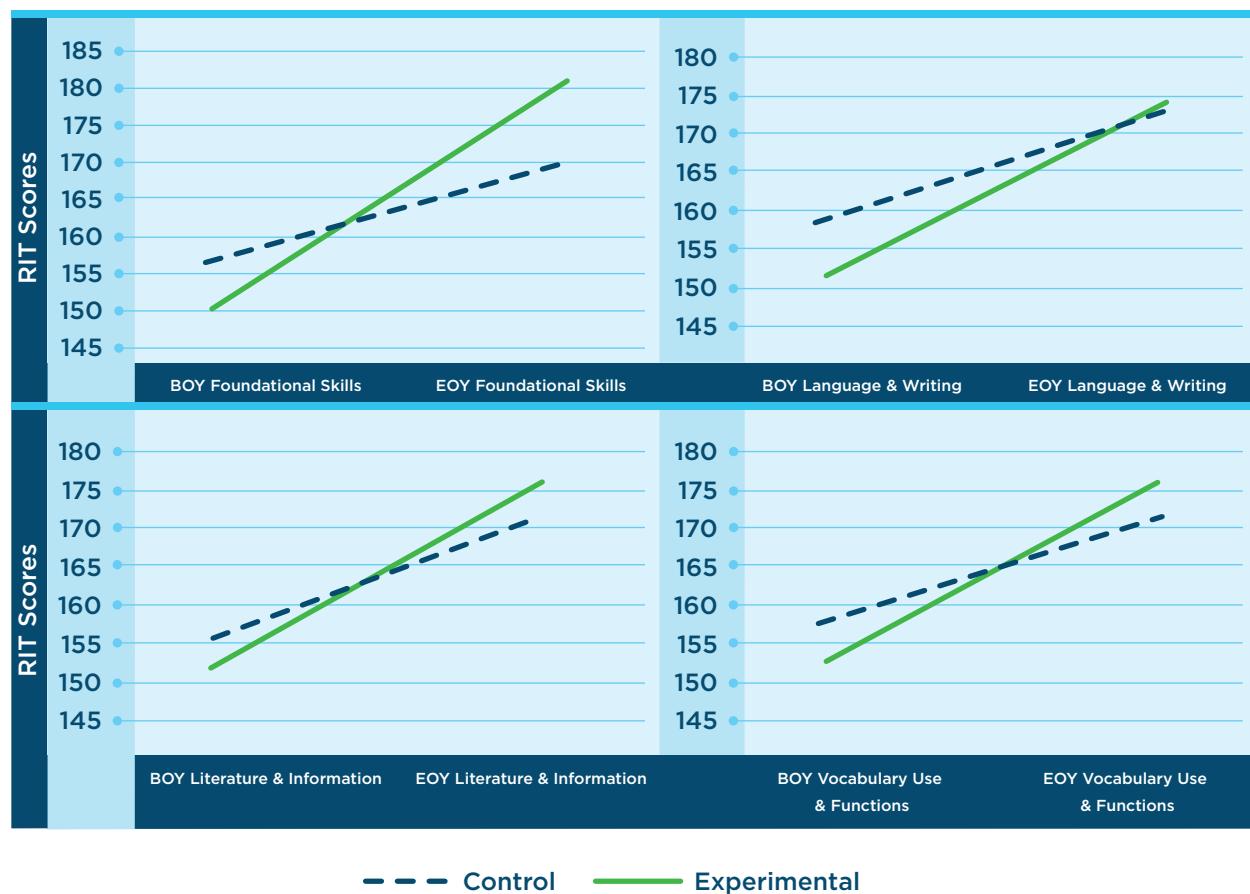


Figure 92: First Grade Beginning & End of Year RIT Scores by Strand



GROUP DIFFERENCES USING ANALYSIS OF COVARIANCE (ANCOVA)

Figure 93: First Grade End of Year RIT Scores while Covarying for Beginning of Year Scores



Evaluation of a School District in South Carolina, 2017-2018

The aim of this study was to investigate the impact of Waterford Reading Academy (WRA) on end of year literacy scores (Shamir, Pocklington, Feehan & Yoder, 2020). Kindergarten students were assigned to the experimental group ($n = 101$) if they used the program for more than 1,000 minutes during the school year; those who were assigned to the control group ($n = 41$) used WRA for less than 300 minutes. Five ANOVAs were conducted to examine the effect of Waterford curriculum on end of year Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) reading scores, while covarying for beginning of year scores. The results showed students in the experimental group scored significantly higher than the control group on Reading RIT (ES = 0.68), Reading Foundations (ES = 0.80), Reading Literature and Nonfiction (ES = 0.69), Reading Vocabulary, (ES = 0.63) and Writing (ES = 0.65) strand scores. The size of the effects ranged from medium to large.

Further analysis was conducted to investigate the effects of Waterford WRA and demographics, including special education services and lunch status, on end of year NWEA MAP scores, while covarying for beginning of year scores. The results indicated that WRA had a beneficial impact on end of year scores across all available demographics, including special education services and subsidized lunch recipiency. For Reading RIT Score, Reading Vocabulary, and Writing, for students with active special education services and no special education services, students in the experimental group significantly outperformed students in the control group. For Reading Foundations and Reading Literature and Nonfiction, for students with no special education services, students in the experimental group significantly outperformed students in the control group; for students with active special education services, students' scores in the experimental group were higher than in the control group, but the difference was not significant. For students with paid lunch status, across strands, students in the experimental group significantly outperformed students in the control group. For students with free/reduced lunch status, across strands except Writing, students in the experimental group significantly outperformed students in the control group. For Writing, scores of students with free/reduced lunch status in the experimental group were higher than in the control group, but the difference was not significant.

Figure 94: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Strand

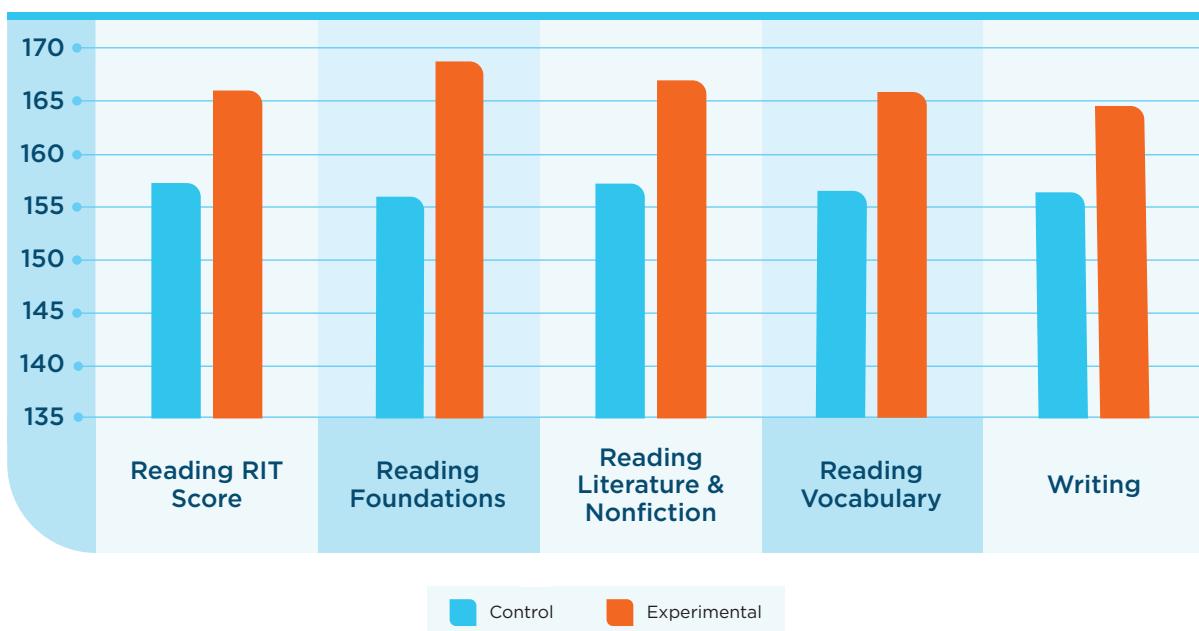


Figure 95: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Special Education Services

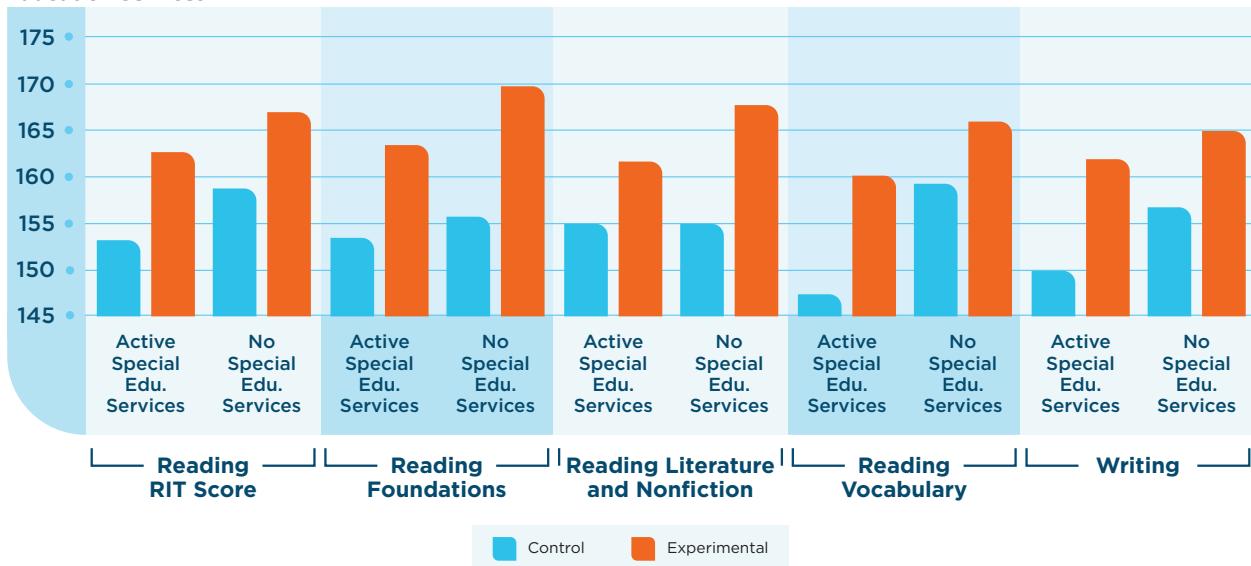
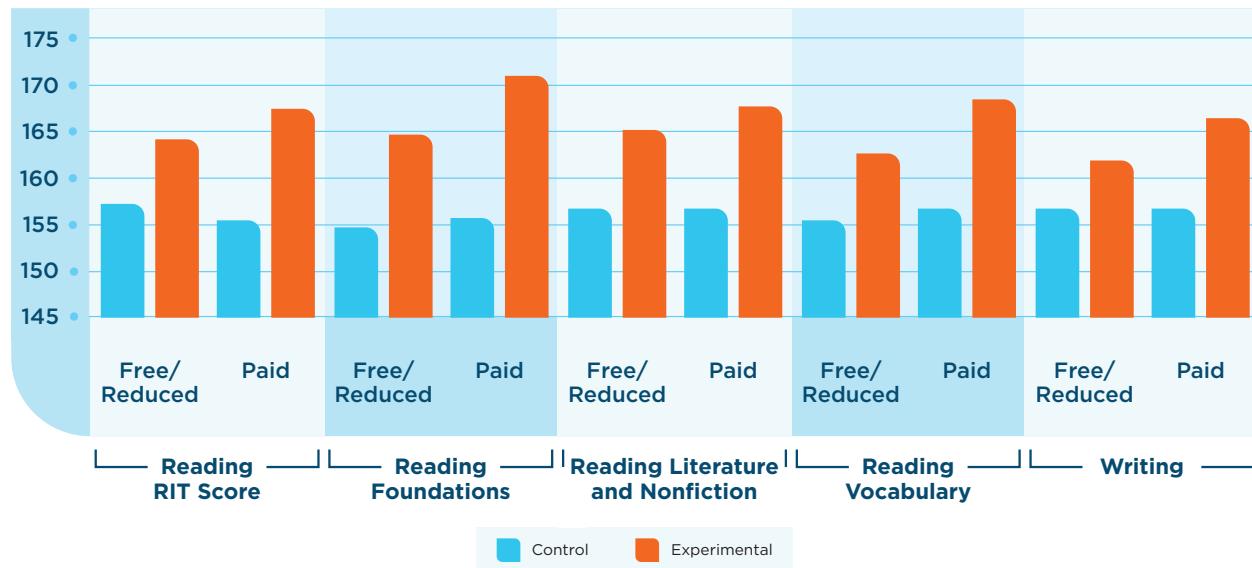


Figure 96: Kindergarten NWEA MAP Reading End of Year Scores Covarying for Beginning of Year Scores by Lunch Status



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 97: Second Grade MAP End of Year Scores by Strand

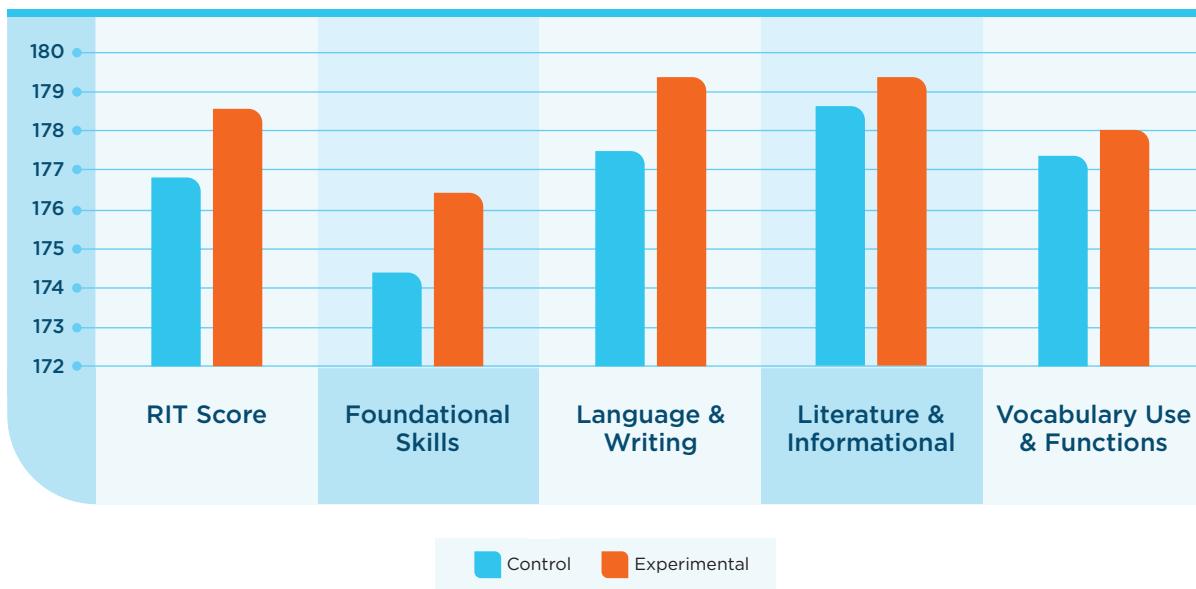


Figure 98: Second Grade MAP End of Year RIT to Reading Scores

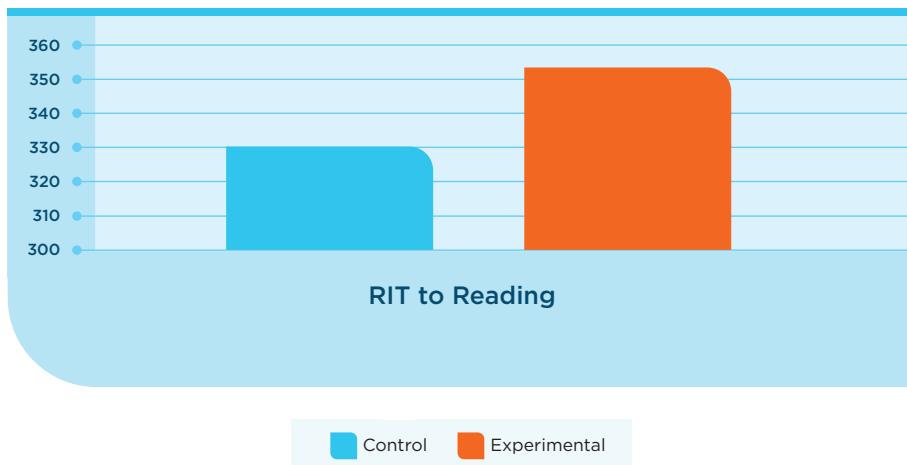


Figure 99: Second Grade MAP End of Year Scores Controlling for Beginning of Year Scores by Strand

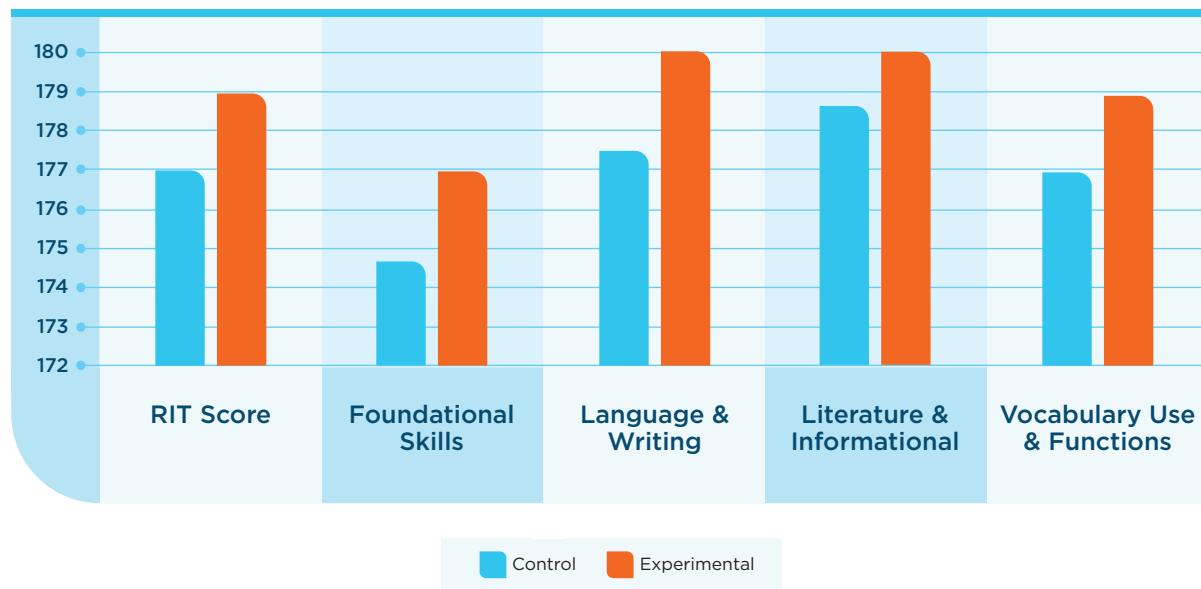
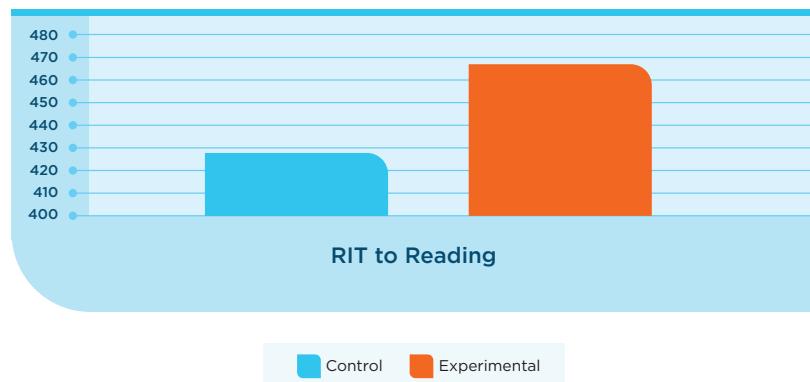


Figure 100: Second Grade MAP End of Year RIT to Reading Scores Controlling for Beginning of Year Scores



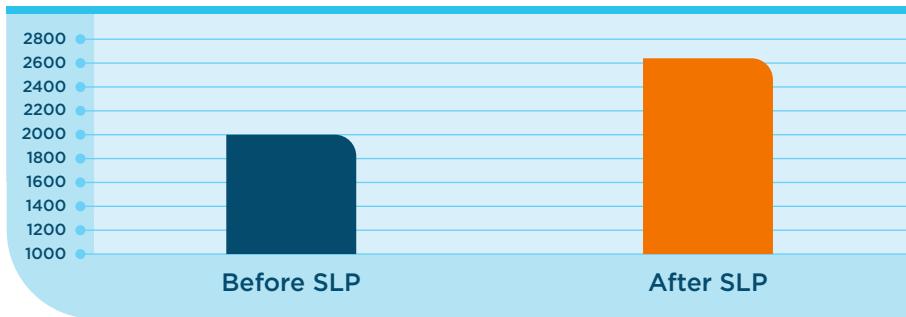
Evaluation of the Efficacy of Waterford Summer Learning Path

This study (Shamir, Ortiz-Wood, Pocklington, & Yoder, 2022) evaluates the efficacy of the Waterford Summer Learning Path (SLP) program, a computer-adaptive reading curriculum that was assigned to four-year-old students ($N = 3,082$) across 11 states during the summer of 2021 before beginning kindergarten. Pre-kindergarten students were expected to use Waterford Reading Academy (WRA) for 20 minutes a day, five days a week. Students were administered the Waterford Assessment of Core Skills (WACS) at the beginning and end of the SLP program to assess the literacy skills of the students before and after their participation in the program.

A paired samples t -test revealed a significant difference between the beginning and end of program WACS scores, $t(1298) = 45.21$, $p < .05$, $d = 1.25$, as students achieved significantly higher WACS scores

after participation in SLP ($M = 2,641.96$, $SD = 578.88$) compared to WACS scores before their participation in SLP ($M = 1,989.85$, $SD = 375.17$).

Figure 101: Average WACS from the Beginning of the SLP Program and the End of the Program



Additionally, mixed design ANOVAs were performed to examine the impacts of demographics and SLP participation on the average end of program WACS scores. There was a significant interaction between the effects of SLP participation and socioeconomic status, $F(1, 1119) = 7.28, p < .05$, with WACS scores being significantly higher at the conclusion of the program than at the beginning. There was also a significant interaction between SLP effects on race/ethnicity, $F(4, 1279) = 5.19, p < .05$. There was no significant interaction between the effects of the SLP program and gender, multi-lingual learner status, prior attendance of other preschool programs, and state. Post-hoc pairwise comparisons showed that on average, the end of program WACS scores were significantly higher than the beginning of the program WACS scores across all races/ethnicities, genders, multi-lingual learner statuses, socioeconomic statuses, prior preschool attendance statuses, and states.

Figure 102: Before and After SLP WACS Scores by Demographic Factors

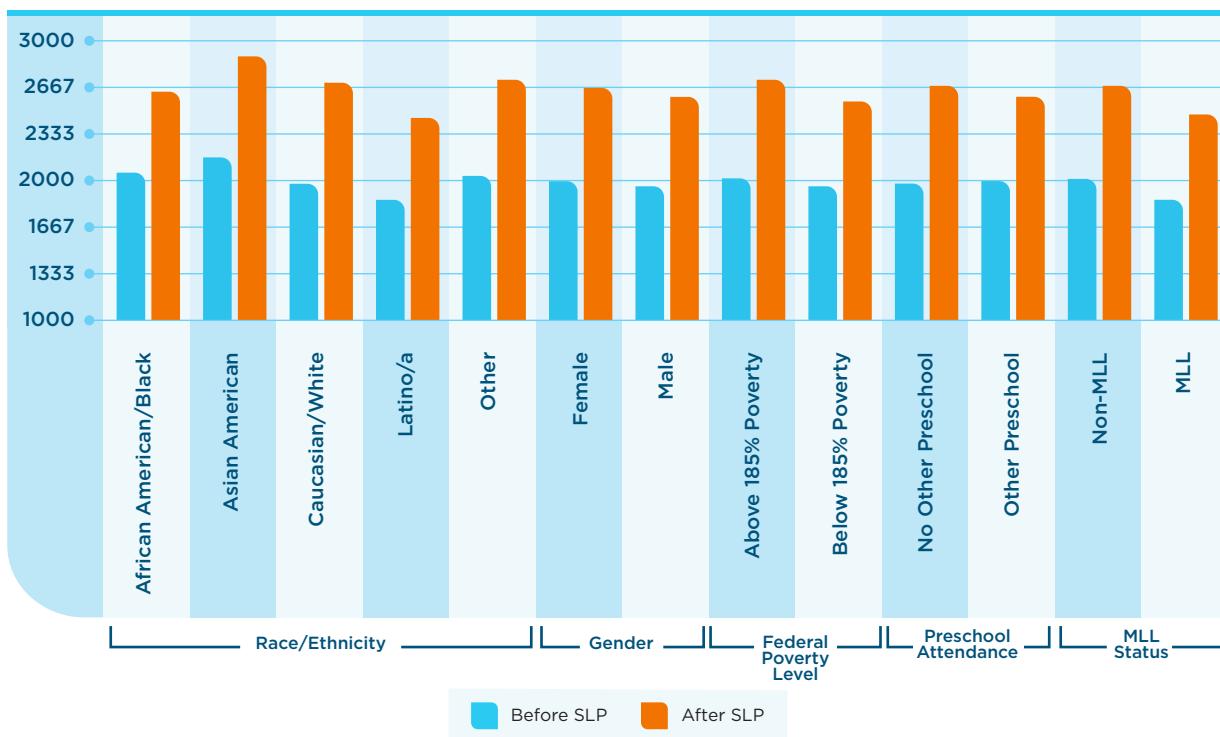
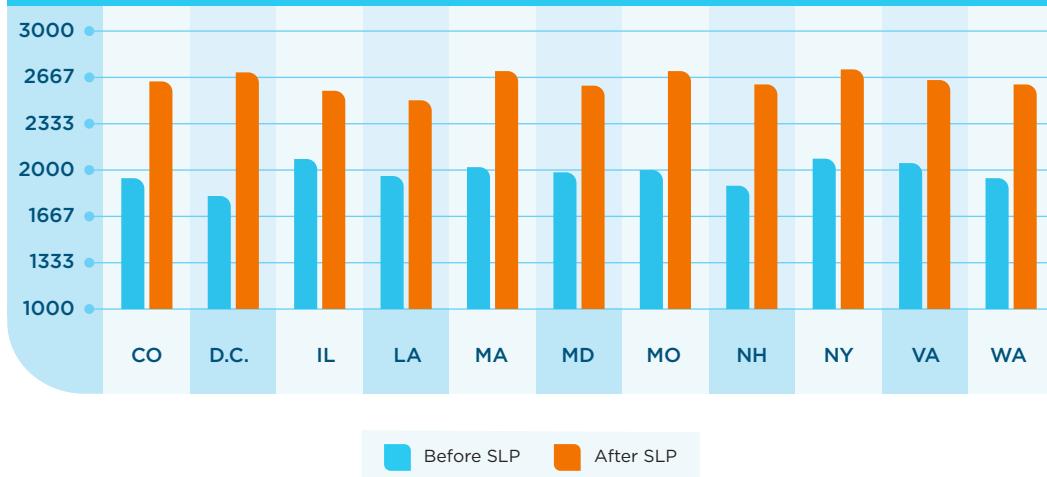


Figure 103: Before and After SLP WACS Scores by State



Evaluation of Usage Fidelity in the Waterford Summer Learning Path

This study (Shamir, Pocklington, & Yoder, 2023) examined the degree to which fidelity of usage moderates the efficacy of the Waterford Summer Learning Path (SLP) program. Pre-kindergarten students ($N = 3,082$) were expected to use Waterford Reading Academy (WRA) for 20 minutes a day, five days per week. The experimental group included students who used WRA for at least 1,300 minutes during the SLP program. The control group included students who used WRA for less than 750 minutes

during the SLP program. The Waterford Assessment of Core Skills (WACS) was administered at the beginning and end of the program to assess the literacy skills of the students before and after their participation in the program. There was no significant difference in beginning of program scores between groups, $t(1, 657) = 0.99, p = .324$. There was a significant difference in end of program scores between groups, $t(1, 659) = 4.01, p < .05$, due to the higher end of program scores made by experimental students ($M = 2,701.57$) than by control students ($M = 2,525.28$). Effect size ($d = 0.31$).

Five two-way ANOVAs were performed to examine the impact of demographics and SLP on average end of the program scores. There were no significant interactions for the effects of the SLP program and race/ethnicity, gender, poverty, prior attendance of another preschool program, or multi-lingual learner status. For all comparisons between the experimental and control group, across all demographics, students in the experimental group consistently outperformed students in the control group.

Figure 104: Beginning and End of Program Assessment Scores

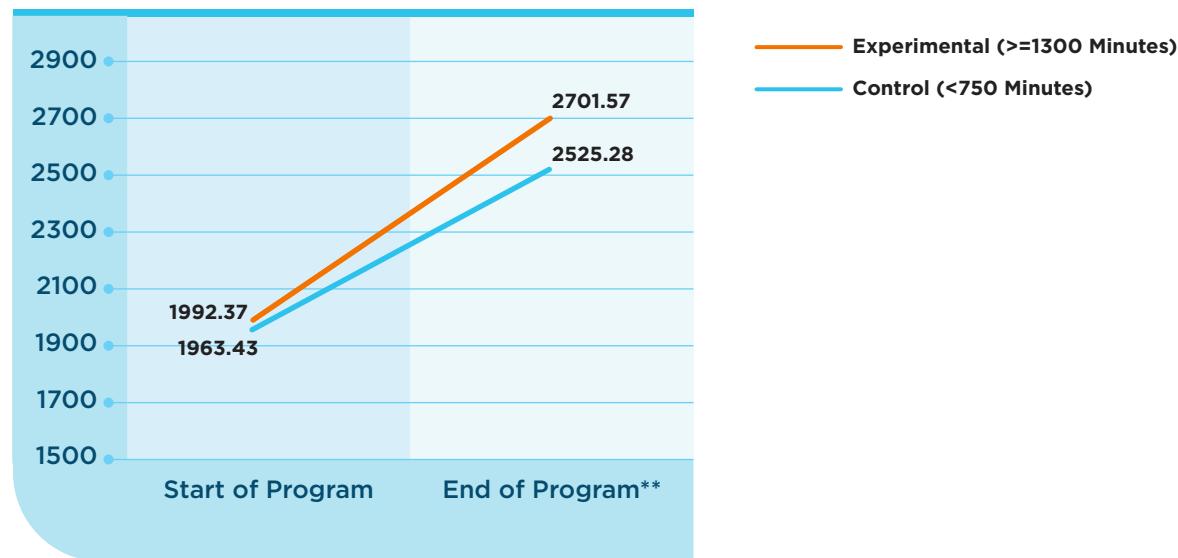


Table 44: Assessment Scores by Demographics

	Control (<750 Minutes)			Experimental (>= 1300 Minutes)			<i>P</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	
Overall							
Start of Program	338	1963.43	395.049	323	1992.37	359.469	0.32
End of Program	338	2525.28	558.221	323	2701.57	571.615	0.00**
Race/Ethnicity							
African American/Black	99	2490.64	557.96	76	2693.76	628.56	0.02*
Asian American	14	2633.86	530.15	34	2947.62	517.52	0.08
Caucasian/White	81	2569.42	575.90	90	2775.11	497.28	0.02*
Latino/a	93	2401.33	505.29	82	2487.70	536.14	0.30
Other	46	2726.09	597.37	38	2792.03	606.47	0.59
Gender							
Female	185	2550.70	572.00	142	2664.25	558.41	0.07
Male	153	2494.54	541.36	181	2730.85	581.63	0.00**
Experiencing Poverty							
No	106	2590.00	485.30	122	2829.25	496.95	0.00**

	Control (<750 Minutes)			Experimental (>= 1300 Minutes)			P
	n	M	SD	n	M	SD	
Yes	170	2470.42	565.78	171	2575.65	572.90	0.07
Preschool Attendance							
No Other Preschool	197	2525.07	540.74	207	2739.92	588.60	0.00**
Other Preschool	141	2525.57	583.72	116	2633.14	535.65	0.13
MLL							
No	268	2541.80	557.49	243	2761.15	569.19	0.00**
Yes	65	2447.55	566.26	77	2517.32	539.28	0.46

Evaluation of the Waterford Early Math and Science Program in Illinois, 2021-2022

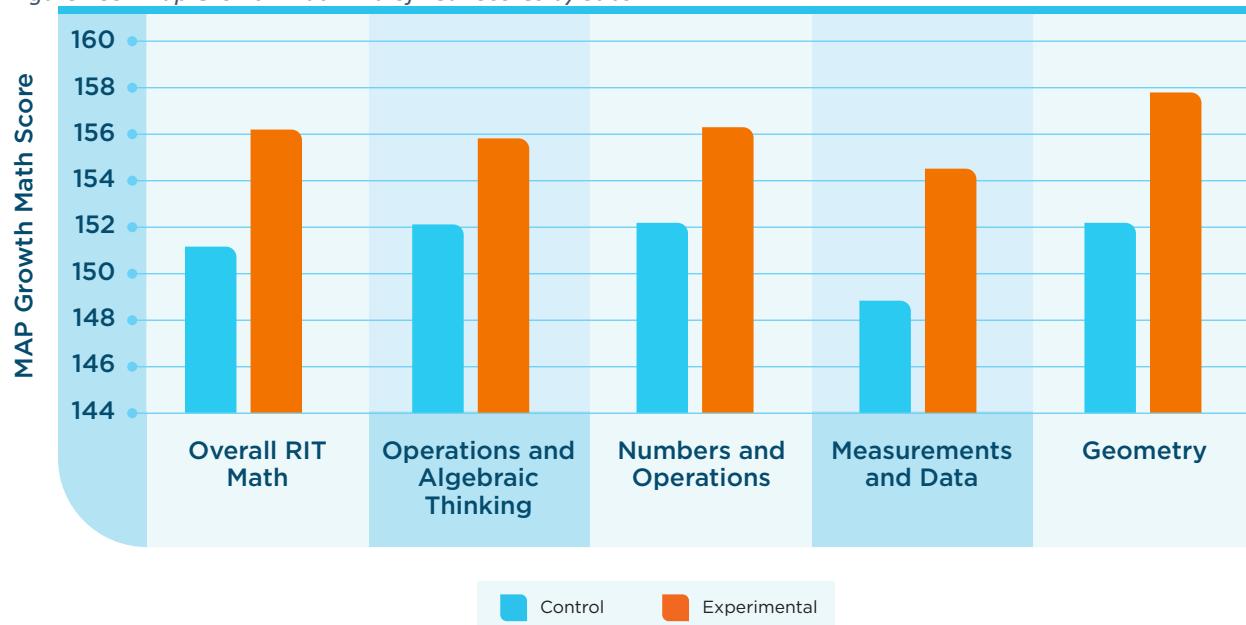
This study investigated the impact of Waterford Reading Academy (WRA) on end of year math skills for kindergarten students ($N = 559$) enrolled in a public school district in Illinois (Shamir, Yoder, Pocklington, Wang, Greene, 2023). Students who used the program for more than 1,500 minutes ($n = 459$) were assigned to the experimental group and students who used the program for less than 800 minutes ($n = 100$) were assigned to the control group. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) Growth assessment was administered at the beginning and end of the program year to measure Overall RIT Math, Operations and Algebraic Thinking, Number and Operations, Measurement and Data, and Geometry on a standardized Rasch Unit (RIT) scale. Baseline equivalence was established using independent samples t -tests.

Table 45: MAP Growth Math Beginning of Year Scores by Subskill

	Experimental		Control		t	p
	n	M	n	M		
Overall RIT Math	439	137.60	65	136.77	-0.60	.546
Operations and Algebraic Thinking	432	136.39	65	135.22	-0.67	.505
Numbers and Operations	438	136.48	65	137.37	0.57	.567
Measurement and Data	439	139.06	65	137.66	-0.89	.374
Geometry	438	139.08	65	136.74	-1.35	.177

Independent samples t -tests were used to examine posttest group differences finding that students in the experimental group scored significantly higher than the control group on Overall RIT Math ($d = 0.38$), Operations and Algebraic Thinking ($d = 0.25$), Numbers and Operations ($d = 0.30$), Measurement and Data ($d = 0.37$), and Geometry ($d = 0.35$).

Figure 105: Map Growth Math End of Year Scores by Subskill



Five two-way ANOVAs were conducted to examine the effects of WRA and special education services on scores. Scores for students with special education services were significantly higher in the experimental group than the control group for all subskills. All students in the experimental group achieved higher scores across all subskills compared to the control group, showing that all students benefited from using WRA. Overall, these results indicate that CAI had a meaningful impact on early mathematical skills across all the students in the sample.

Figure 106: Overall RIT Math End of Year Scores by Special Education Services

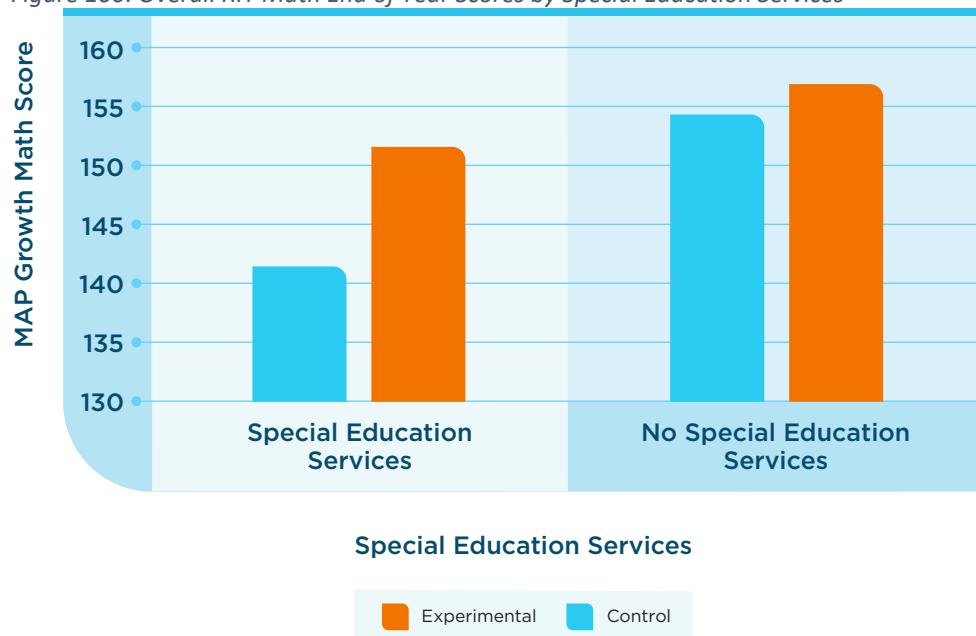


Figure 107: Operations and Algebraic Thinking End of Year Scores by Special Education Services

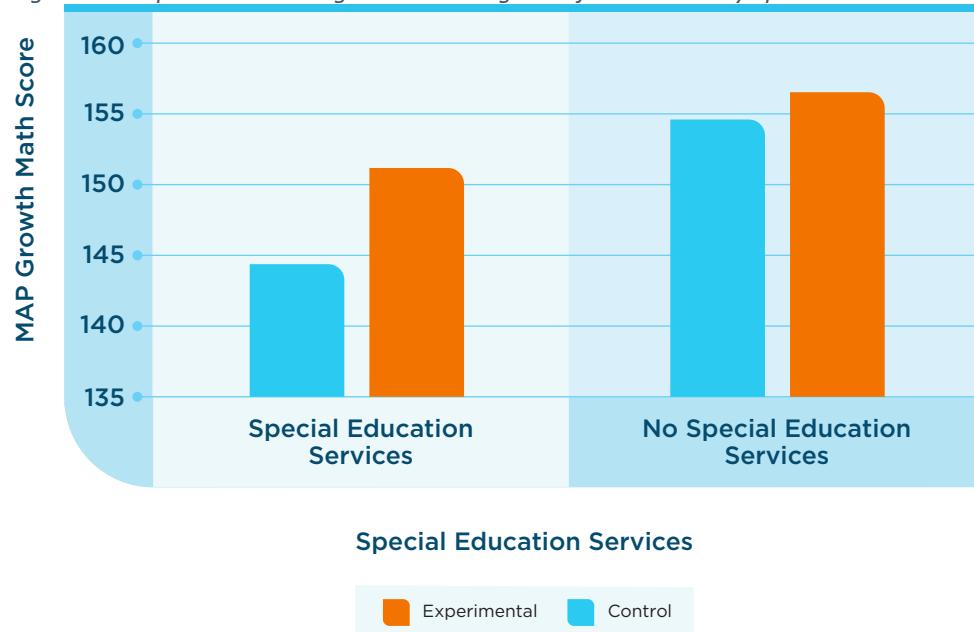


Figure 108: Number and Operations End of Year Scores by Special Education Services

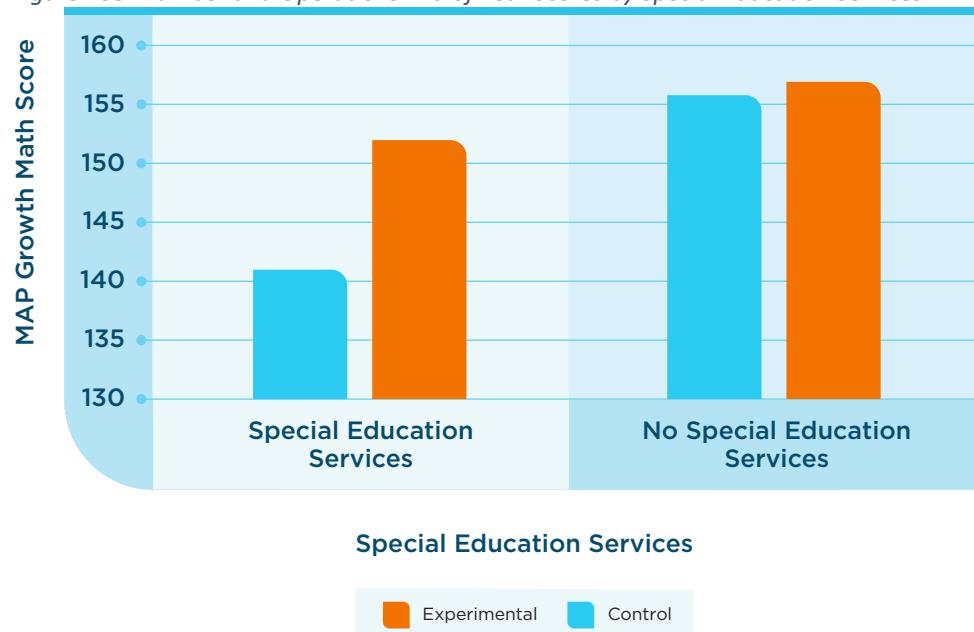


Figure 109: Measurement and Data End of Year Scores by Special Education Services

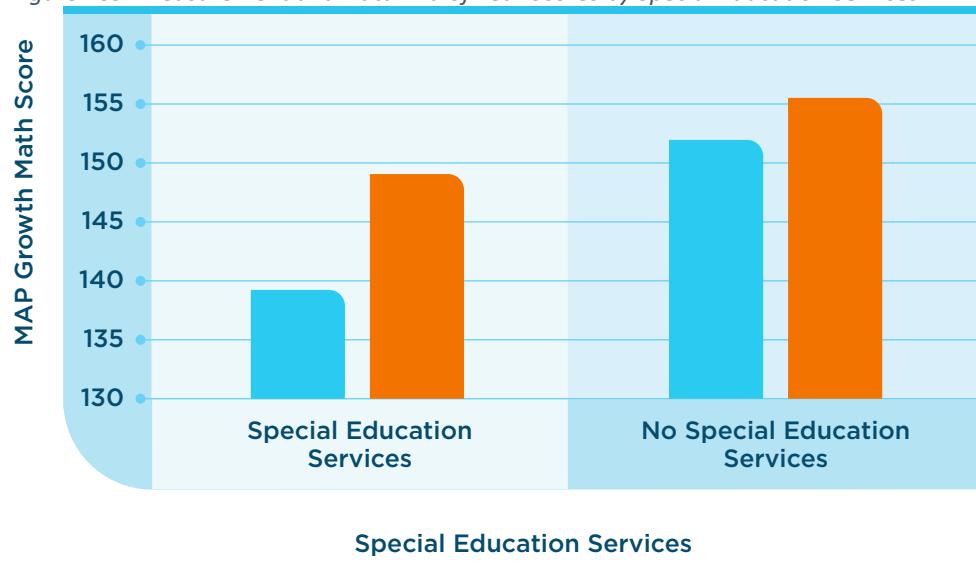
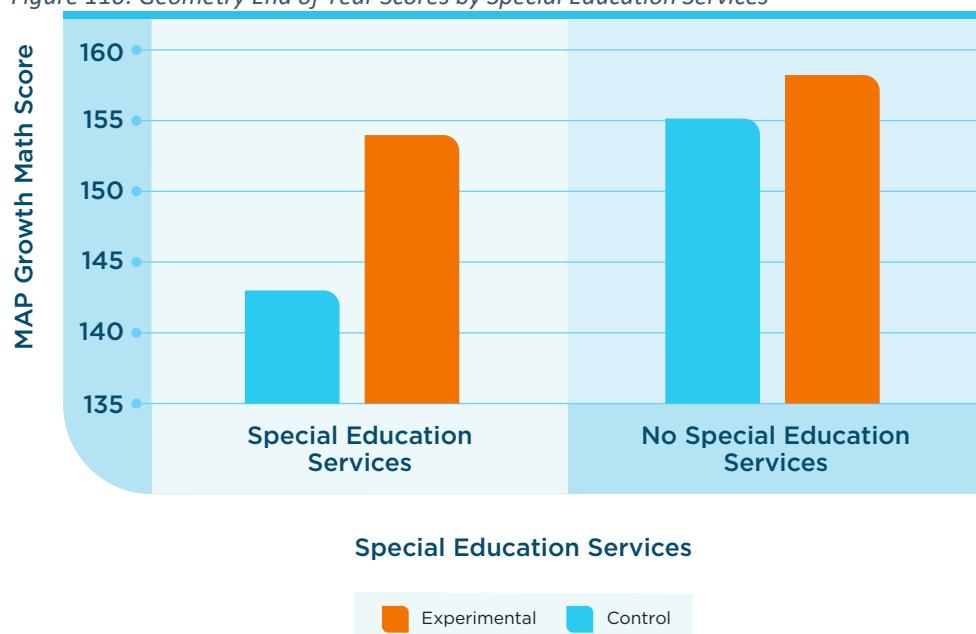


Figure 110: Geometry End of Year Scores by Special Education Services



Evaluation of the Waterford Reading Academy in Illinois

This study (Shamir, Yoder, & Pocklington, 2023a) investigates the impact of the Waterford Reading Academy (WRA) on kindergarten students' early literacy skills. Kindergarten students ($N = 606$) enrolled in a public school district in Illinois during the 2021-2022 school year were assessed at the beginning and end of the school year using the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) Growth assessment. The experimental group ($n = 540$) consisted of students who used

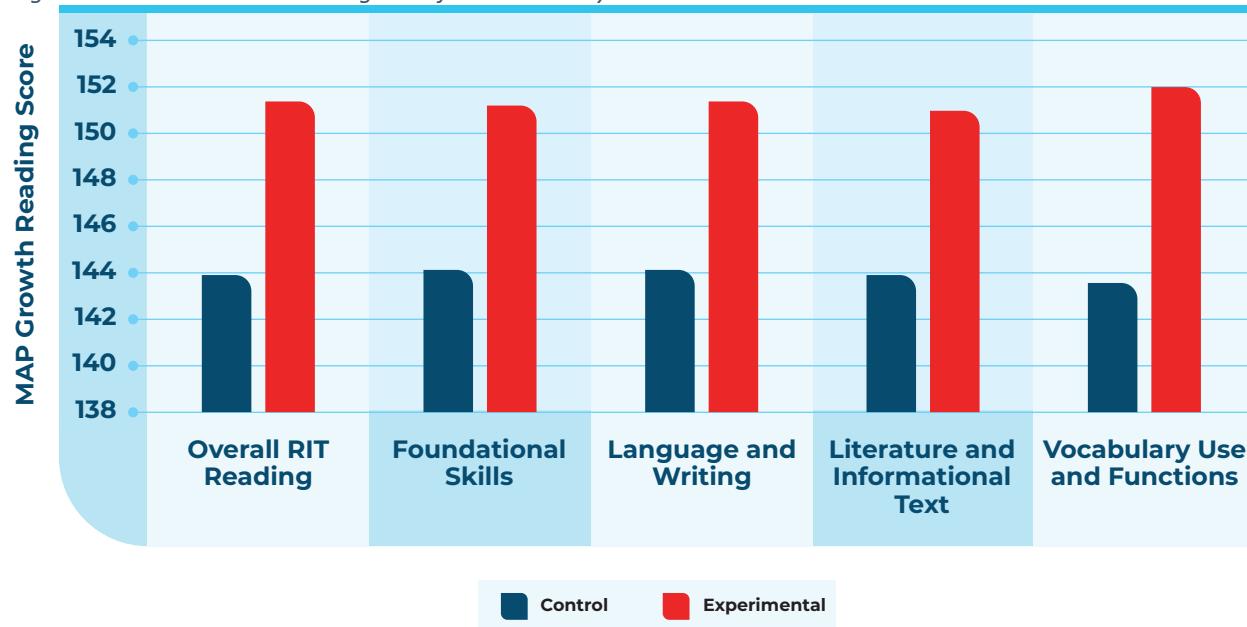
the computer adaptive instruction program for more than 1,500 minutes. The control group ($n = 66$) consisted of students who used the computer-adaptive instruction program for less than 800 minutes.

Baseline equivalence was established between the experimental and control groups using independent samples t -tests. Across all subskills, scores were not significantly different between experimental and control groups at the beginning of the school year. Independent samples t -tests examining group differences between experimental and control groups at the end of the school year revealed significantly higher end of year scores for students in the experimental group on Overall RIT Reading ($ES = 0.61$), Foundational Skills ($ES = 0.52$), Language and Writing ($ES = 0.56$), Literature and Informational Text ($ES = 0.50$), and Vocabulary Use and Functions ($ES = 0.54$).

Table 46: MAP Growth Reading Beginning of Year Scores by Subskill

Subskill	Experimental		Control		<i>t</i>	<i>p</i>
	<i>N</i>	<i>M</i>	<i>N</i>	<i>M</i>		
Overall RIT Reading	511	134.57	34	134.50	-0.03	.977
Foundational Skills	509	131.33	34	131.44	0.06	.954
Language and Writing	511	135.29	34	137.44	0.86	.393
Literature and Informational Text	510	136.76	34	135.21	-0.78	.435
Vocabulary Use and Functions	511	134.99	34	133.91	-0.42	.679

Figure 111: MAP Growth Reading End of Year Scores by Subskill



Separate ANOVAs were conducted for each of the subskills to examine the effects of WRA and demographics on end of year scores. The results of the ANOVAs and the post-hoc analyses demonstrated that there were no significant interactions between the effects of WRA and race/ethnicity.

or special education services. The students in the experimental group significantly outperformed the control group students on the end of year assessment, across all races/ethnicities and special education services, in all subskills other than the Literature and Informational Text subskill, where students with special education services in the experimental group scored higher than the control group, but the difference was not significant.

Figure 112: Overall RIT Reading End of Year Scores by Demographics

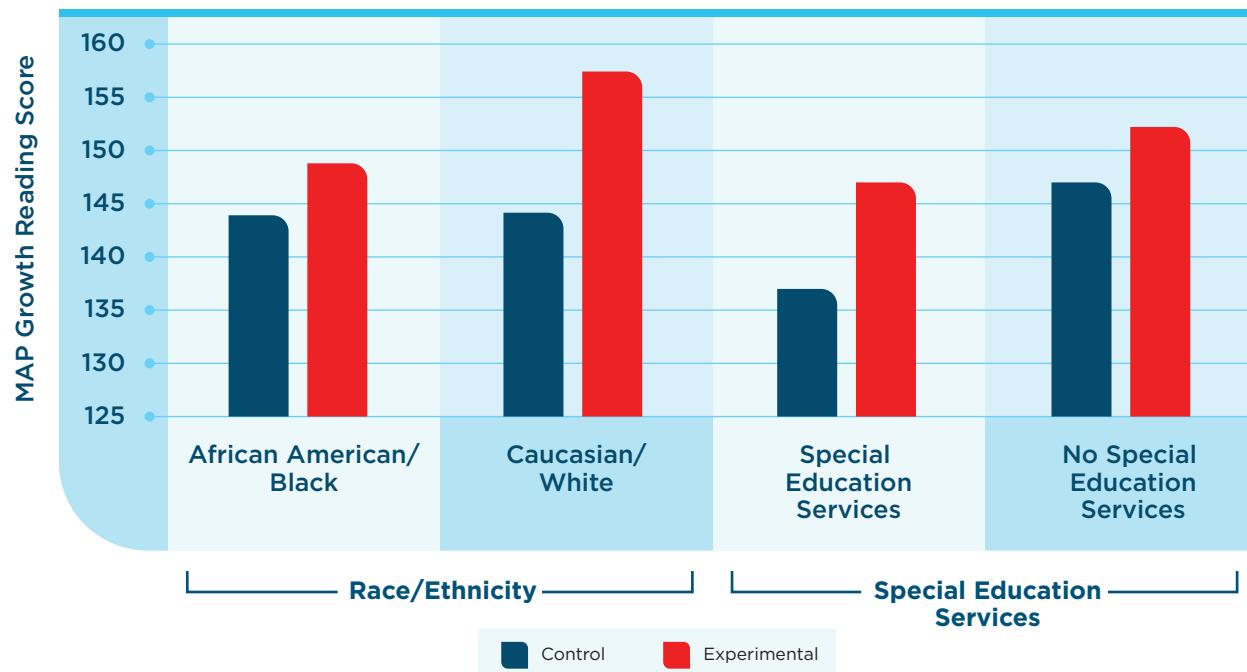


Figure 113: Foundational Skills End of Year Scores by Demographics

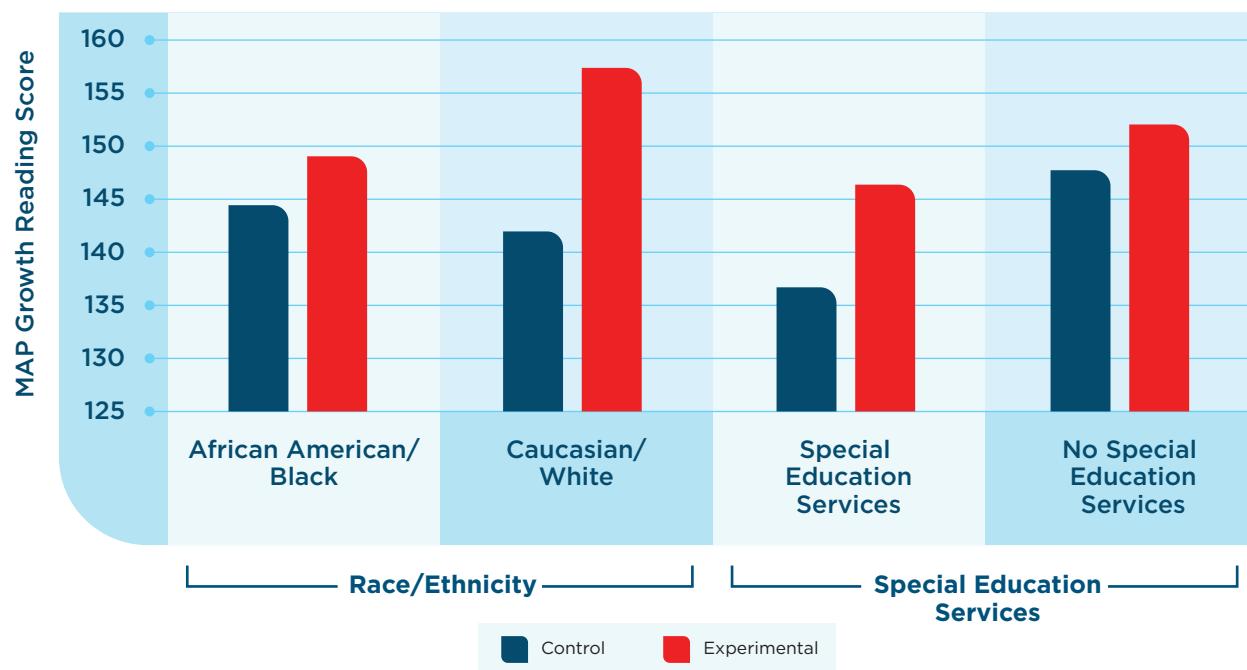


Figure 114: Language and Writing End of Year Scores by Demographics

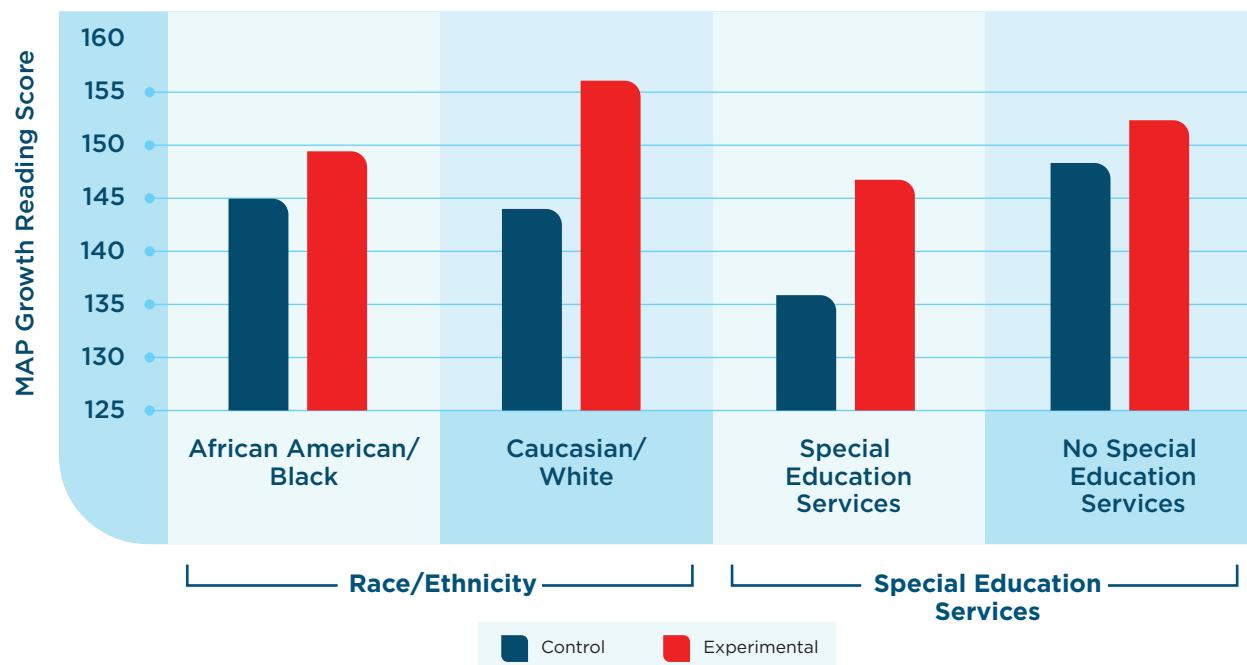


Figure 115: Literature and Informational Text End of Year Scores by Demographics

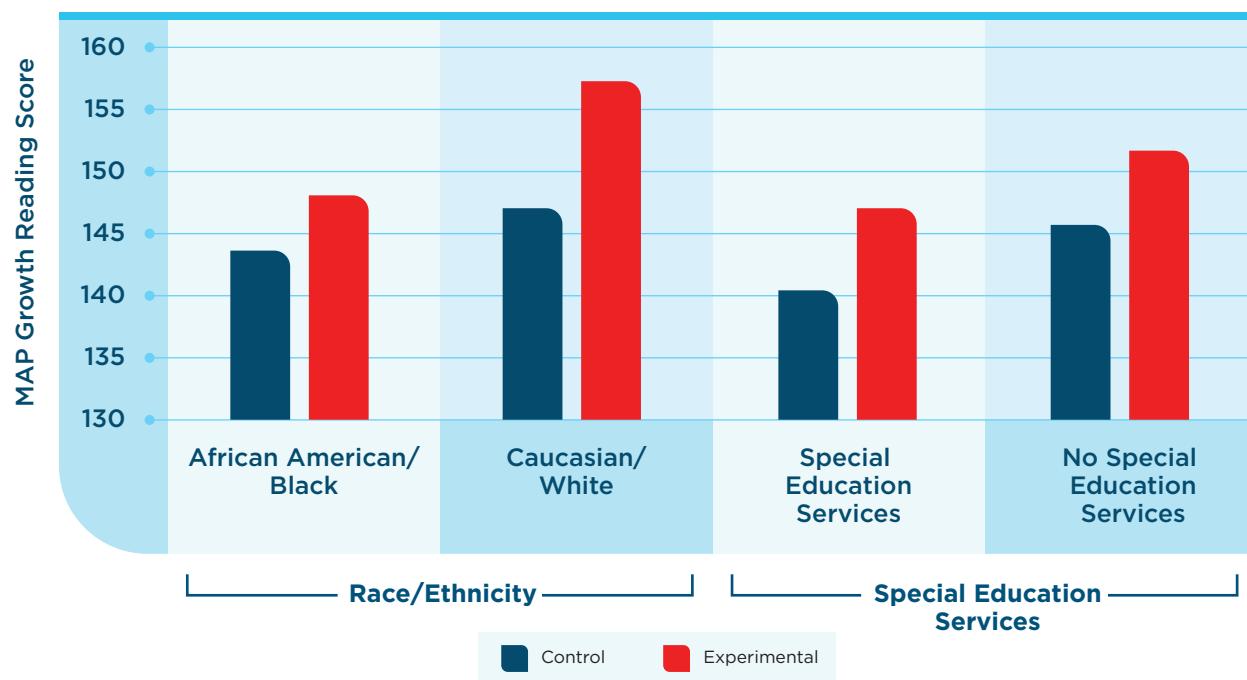
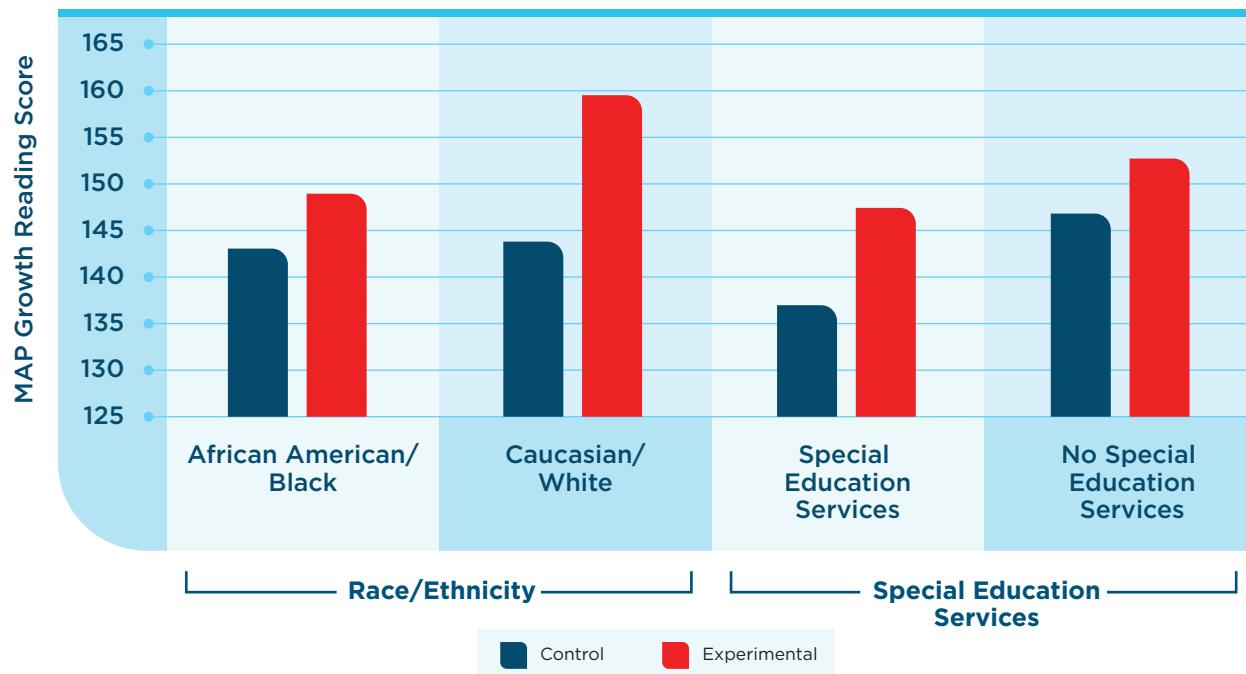


Figure 116: Vocabulary Use and Functions End of Year Scores by Demographics



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 Multi-lingual Learner (MLL) populations. These early results were the first to demonstrate the particular effectiveness of the software for MLL students: In both of the New York schools (PS 43 in the Bronx and PS 1 in Manhattan), where largely MLL classes used the software, these classes performed better on posttests than the non-MLL 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 MLL populations: In a matched sample comparison, in which the majority of students were Multi-lingual Learners, students who used Waterford software achieved gains almost twice as high as students who did not use Waterford (Electronic Education, 2002). Another study in Maryland (at Glenridge Elementary School) produced similar results: MLL 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 MLL students and other students.

Later Effectiveness Studies

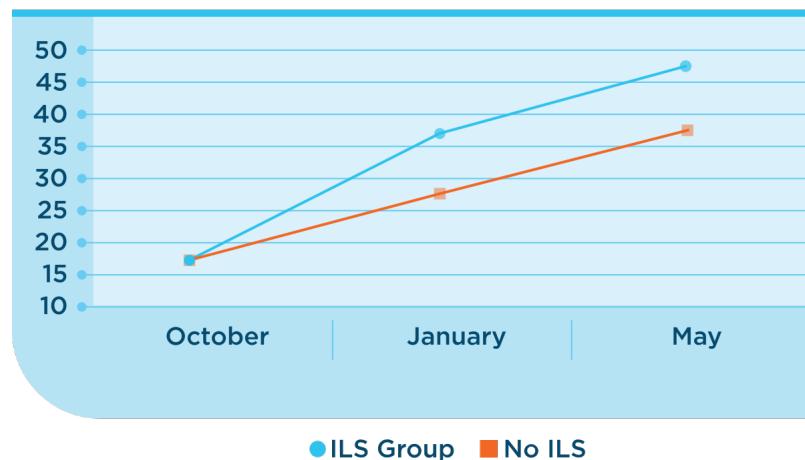
Cassady and Smith (2003) published the first of their two Waterford-related studies in 2003. An Indiana school implemented Waterford software in its kindergarten classes to work in conjunction with existing literacy instruction (ILS group); the evaluation used the Phonological Abilities Tests (PAT) at 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 47: Performance Measure Means at the Beginning, Middle, & End of the Year

	October <i>M</i> (<i>SD</i>)	January <i>M</i> (<i>SD</i>)	May <i>M</i> (<i>SD</i>)
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 117: Kindergarten Students' Phonological Awareness Growth

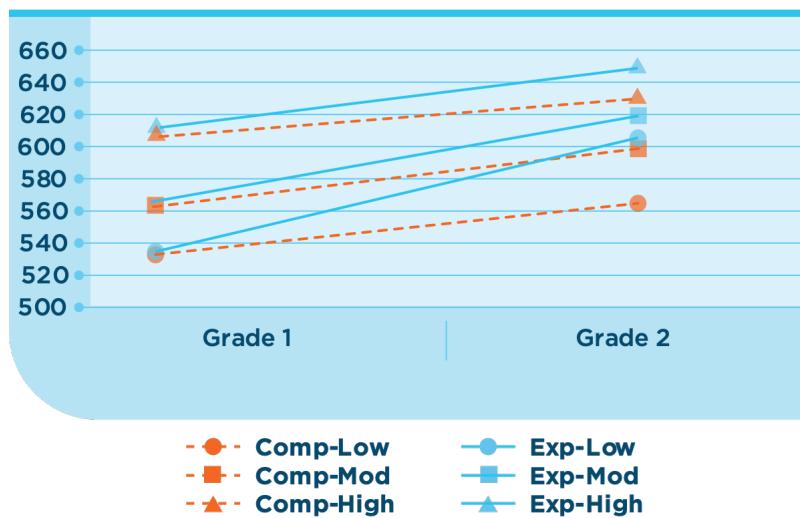


Cassady and Smith (2005) examined the effect on reading achievement gains during the first grade year. Again, students who used the software experienced significantly greater reading skill gains on a standardized test (the CTBS Terra Nova) than the comparison group, $F(1, 91) = 10.61, p < .01, \eta^2 = .10$. 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 48: 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)	561.82 (27.77)	559.62 (8.14)	599.35 (32.14)	612.00 (26.68)	633.40 (33.46)
	<i>n</i> = 11		<i>n</i> = 26		<i>n</i> = 10	
	535.75 (16.09)	610.83 (26.29)	561.35 (8.93)	617.95 (36.67)	608.50 (22.06)	649.71 (38.33)
Experimental	<i>n</i> = 12		<i>n</i> = 20		<i>n</i> = 14	

Figure 118: 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 American/Black students 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 119: Mean Gains by STAR Sub-Strands for First Sample Group



Figure 120: Mean Gains by STAR Sub-Strands for Second Sample Group



Figure 121: 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 122: Mean Reading Gains by DRA Overall Kindergarten Scores



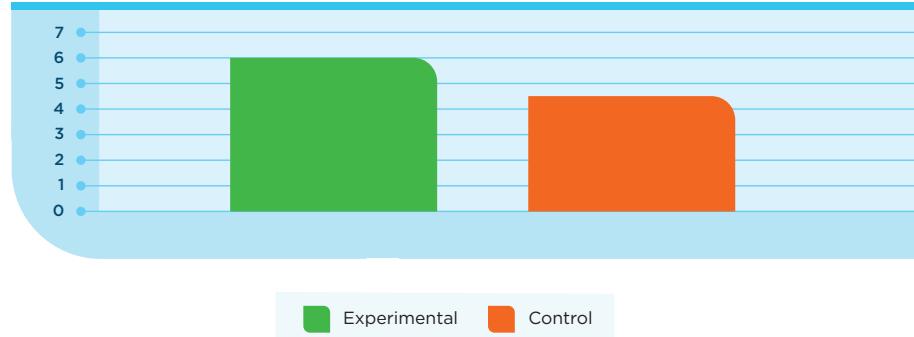
Evaluation of the Waterford Early Reading Program in South Carolina

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.

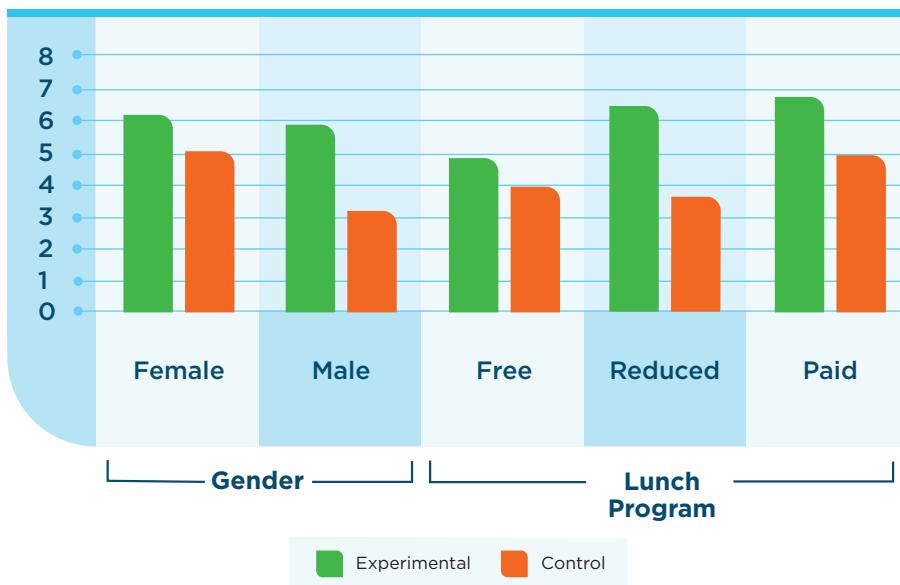
An independent samples t -test examining group differences in DRA kindergarten end of year scores between the experimental group and the control group was conducted. Analysis of end of year scores revealed a significant difference between groups due to higher end of year scores made by experimental students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

Figure 123: Kindergarten DRA End of Year Scores



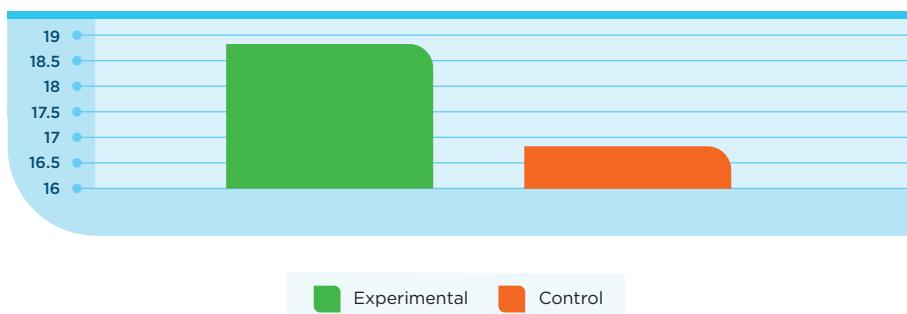
Further analysis was conducted to examine the effects of gender and subsidized lunch on DRA end of year scores, covarying for DRA beginning of year scores. Two separate two-way ANCOVAs were conducted to examine the effect of Waterford curriculum and demographics on DRA end of year scores, covarying for beginning of year DRA scores. The ANCOVAs showed that the interaction between lunch status and treatment was not significant across all strands, which indicates that WEL had a similar impact on end of year scores for students with free, reduced, and paid lunch status.

Figure 124: Kindergarten DRA End of Year Scores by Demographics



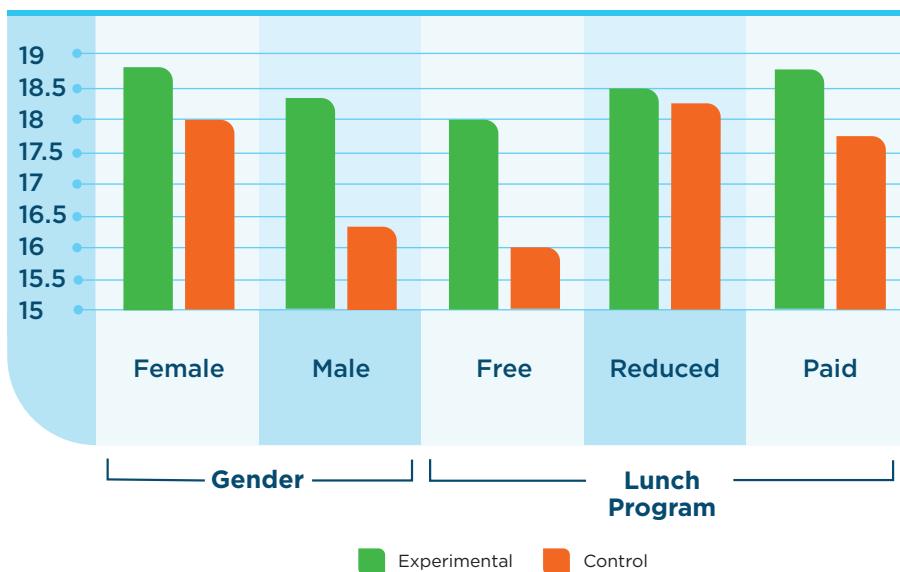
For first grade, analysis of DRA end of year scores, while covarying for DRA beginning of year scores, revealed a significant difference between groups due to higher end of year scores made by experimental students than by control students.

Figure 125: 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 126: 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, $t(1, 2257) = -2.57, p < .05$. Additionally, two-way ANOVAs were conducted to examine effects of the Waterford curriculum and demographics on VLT performance, indicating that the Waterford program benefited kindergarten, first grade and second grade students across demographics. The interaction between treatment and race/ethnicity, MLL status, and special education services was not significant, which indicates that the Waterford reading program had a similar impact on VLT scores for African American/Black, Caucasian/White, and Latino/a students, as well as students with MLL status, non-MLL status, special education services, and no special education services.

Figure 127: End of Year VLT Scores by Grade



Figure 128: Kindergarten End of Year VLT Scores by Demographics

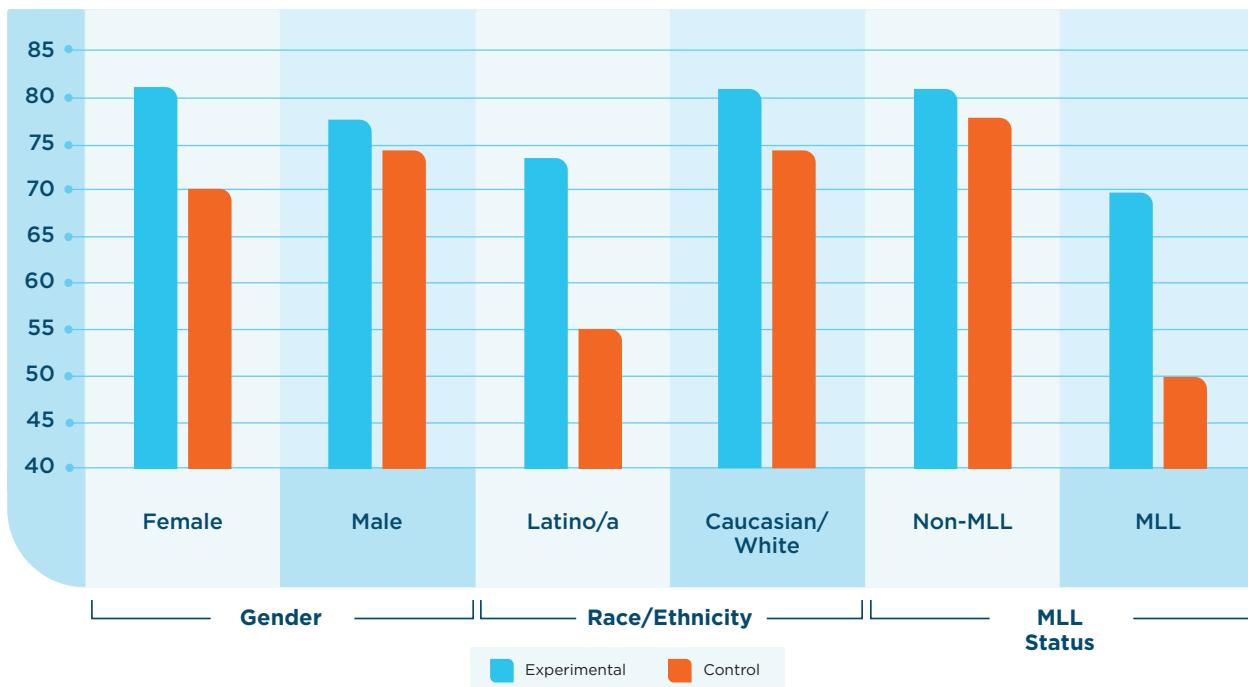


Figure 129: First Grade End of Year VLT Scores by Demographics

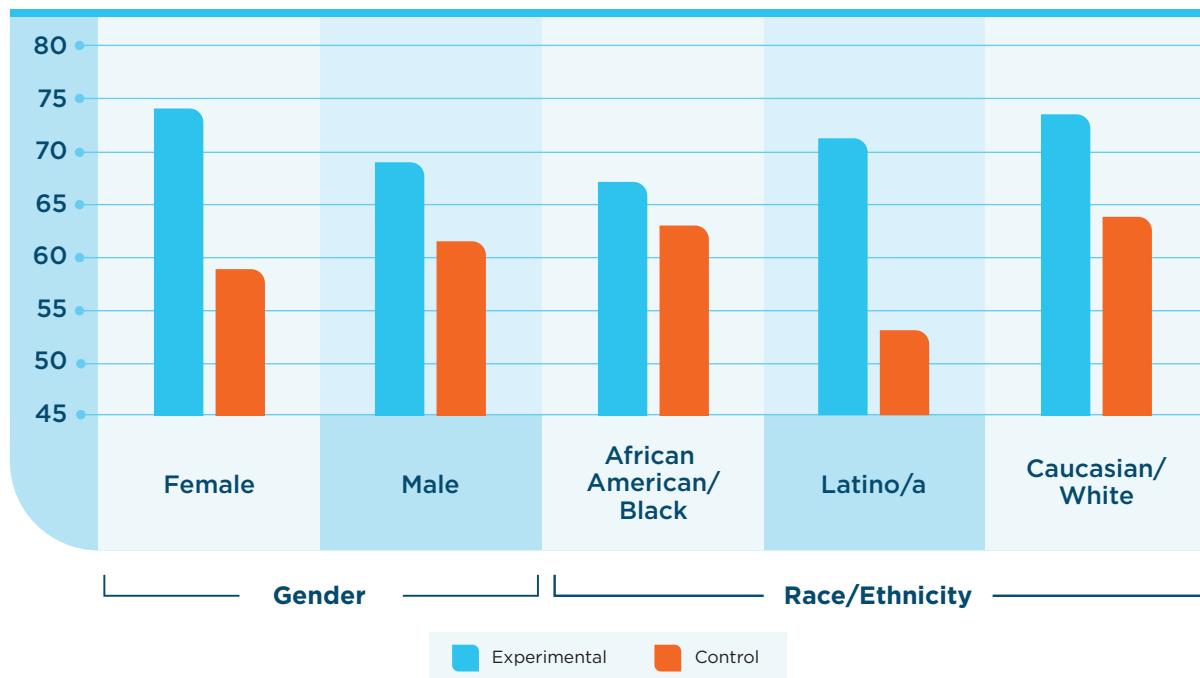
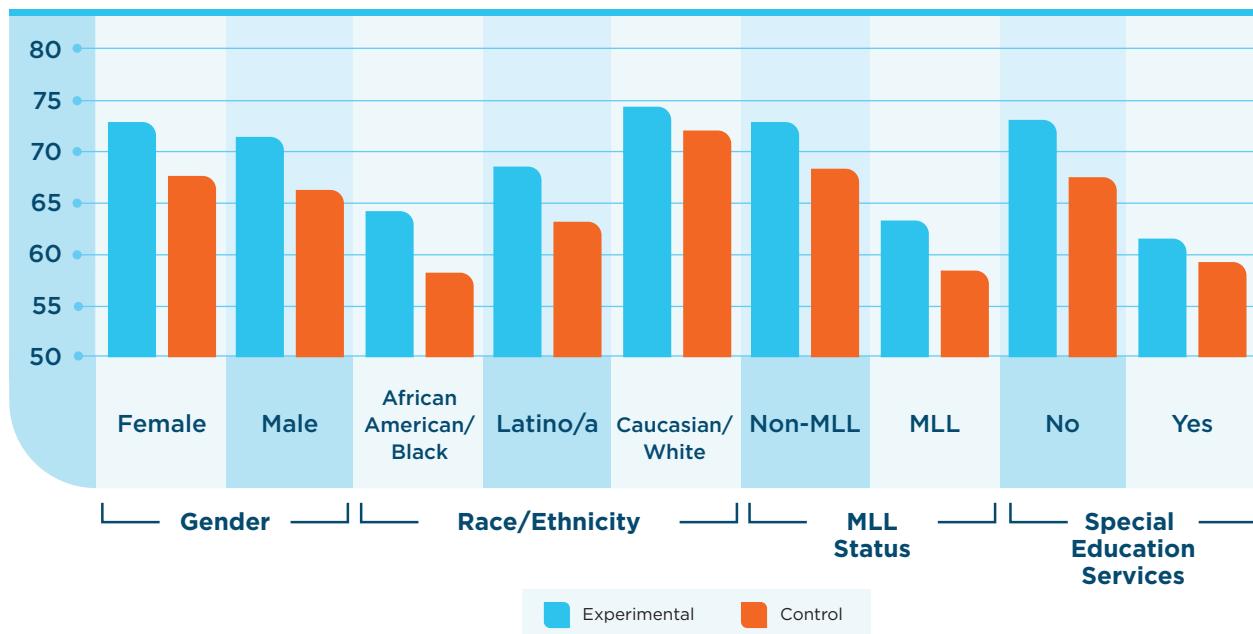


Figure 130: Second Grade End of Year VLT Scores by Demographics



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.

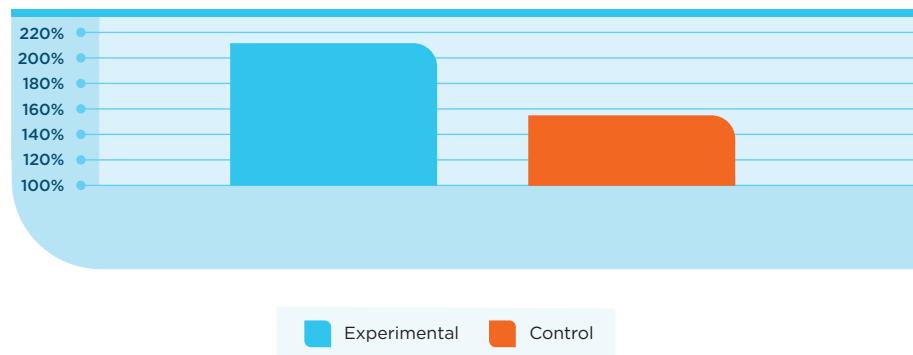
FIRST GRADE

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 131: 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 services 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 Services

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

Figure 132: First Grade DRA Percent Gains by Demographics



SECOND GRADE

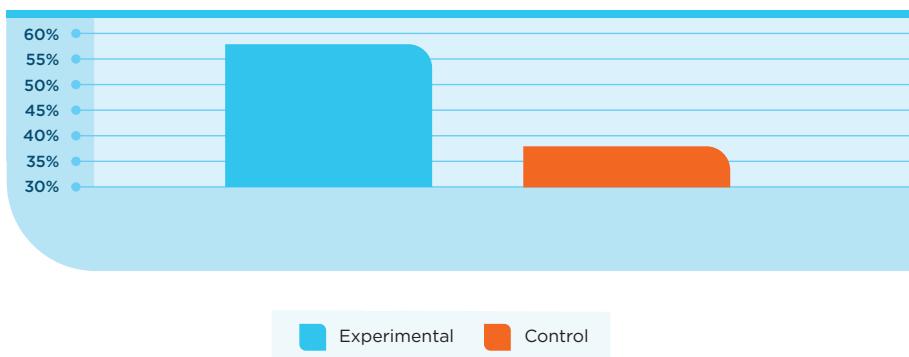
The experimental group for second grade ($N = 70$) consisted of students that used the Waterford Early Reading Program during the 2015-2016 school year. The control group ($N = 407$) consisted of students that did not use the Waterford Early Reading Program during the 2015-2016 school year.

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 133: Second Grade DRA Percent Gains



GROUP DIFFERENCES BY DEMOGRAPHICS USING Two-WAY ANOVAs

Further analysis was conducted to examine the effects of gender, race/ethnicity, lunch program, MLL status, and special education services 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.

Race/Ethnicity

There was a significant interaction between the effects of race/ethnicity and Waterford curriculum on DRA percent gains, $F(4, 464) = 2.74, p < .05$. Simple effects analysis showed that Multiracial and Caucasian/White students' percent gains in the experimental group were significantly higher than in the control group. Asian American and African American/Black 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.

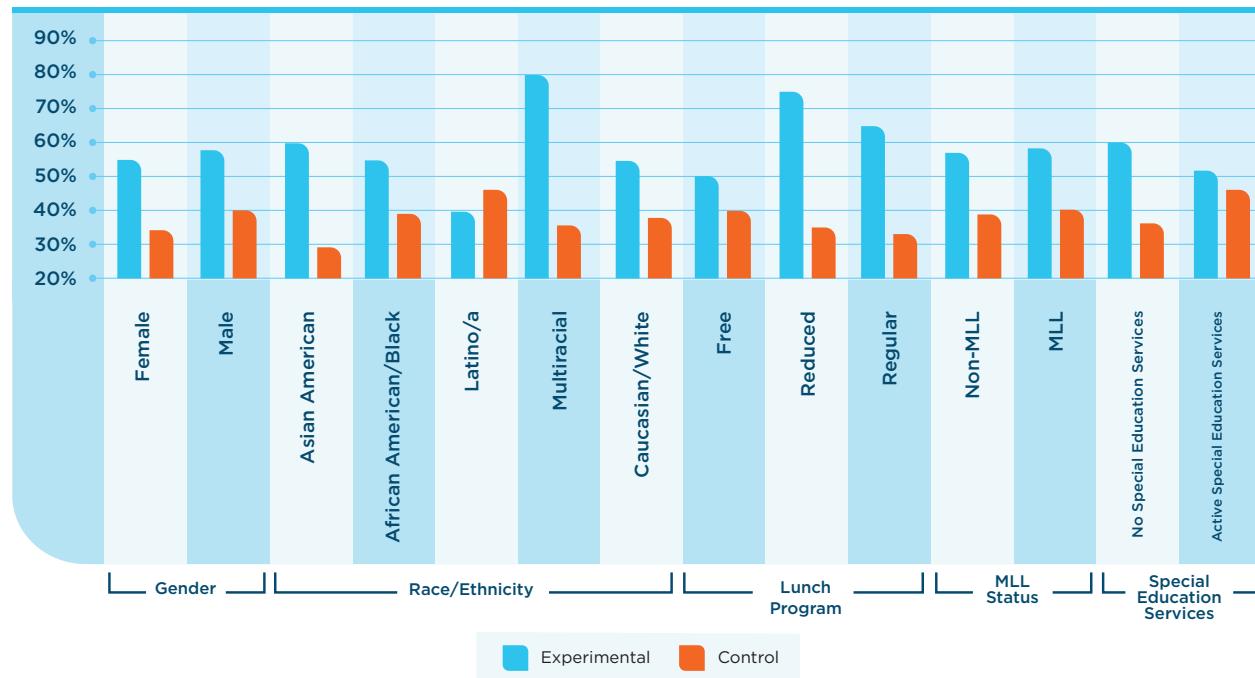
Multi-lingual Learner (MLL) Status

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

Special Education Services

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

Figure 134: 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 multi-lingual learners (MLL) (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 MLL students in the experimental group outperformed students in the control group on standardized literacy assessments (Tables 43-48).

DISTRICT 1

Table 49: 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 50: District 1 Overall Kindergarten TPRI Gains by MLL 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 51: 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 52: District 1 Overall Second Grade TPRI Gains by MLL 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 53: 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 54: District 2 VLT End of Year Scores by MLL 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 students experiencing poverty (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 students experiencing poverty in the experimental group outperformed students in the control group on multiple reading and math assessments (Tables 49-56). ANCOVAs showed that the interaction between lunch status and treatment was not significant across all strands, which indicates that WEL had a similar impact on end of year scores for students with free and paid lunch status, across both districts and grades.

DISTRICT 1

Table 55: 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*
Experiencing Poverty	4.85	2.82	381	4.00	4.08	10	.44

Table 56: 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**
Experiencing Poverty	18.02	5.48	377	16.01	5.09	29	.01**

DISTRICT 2

Table 57: 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 58: 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 59: 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 60: 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 61: 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 62: 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 services 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 services, but also the control group of students without active special education services.

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

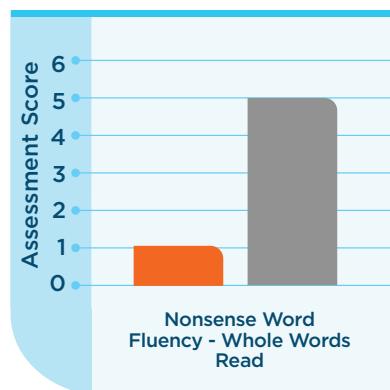


Figure 136: mCLASS: DIBELS Next & Math Scores by Strand by Special Education Control—Letter Naming Fluency & Correct Letter Sounds

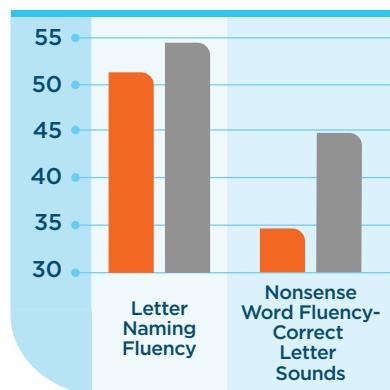


Figure 137: mCLASS DIBELS Next and Math Scores by Strand by Special Education Control—Number Identification & Quantity Discrimination



Figure 138: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Whole Words Read

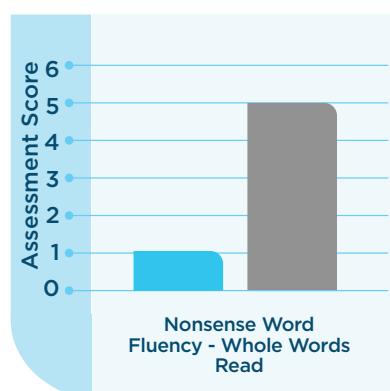


Figure 139: mCLASS DIBELS Next & Math Scores by Strand by Non-Special Education Control—Letter Naming Fluency & Correct Letter Sounds

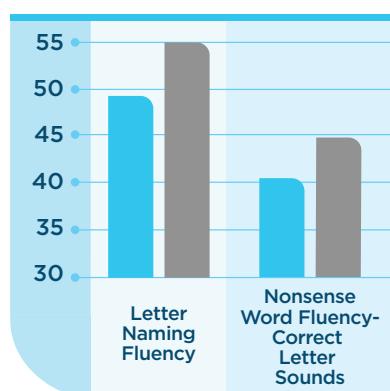


Figure 140: mCLASS DIBELS Next and Math Scores by Strand by Non-Special Education Control—Number Identification & Quantity Discrimination



■ Non-Special Education Control ■ Special Education Experimental

DISTRICT 2

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

Figure 141: TPRI Gain Scores by Strand by Special Education Control

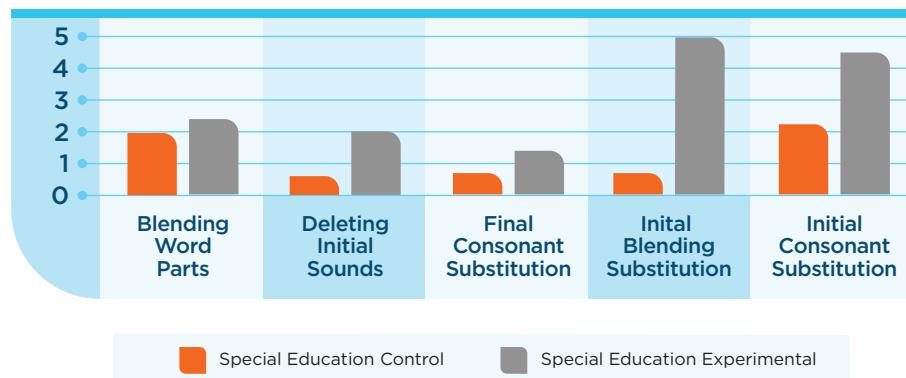
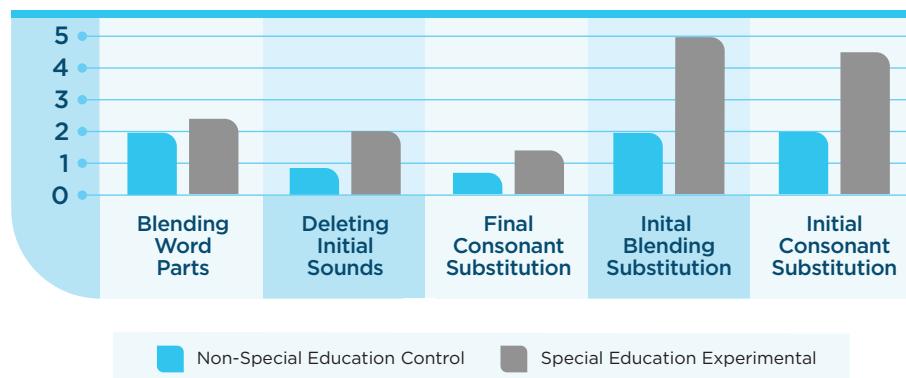


Figure 142: TPRI Gain Scores by Strand by Non-Special Education Control



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 special education services, Multi-lingual Learner (MLL) status, and race/ethnicity.

KINDERGARTEN

Figure 143: Kindergarten TPRI Gains by Substrand

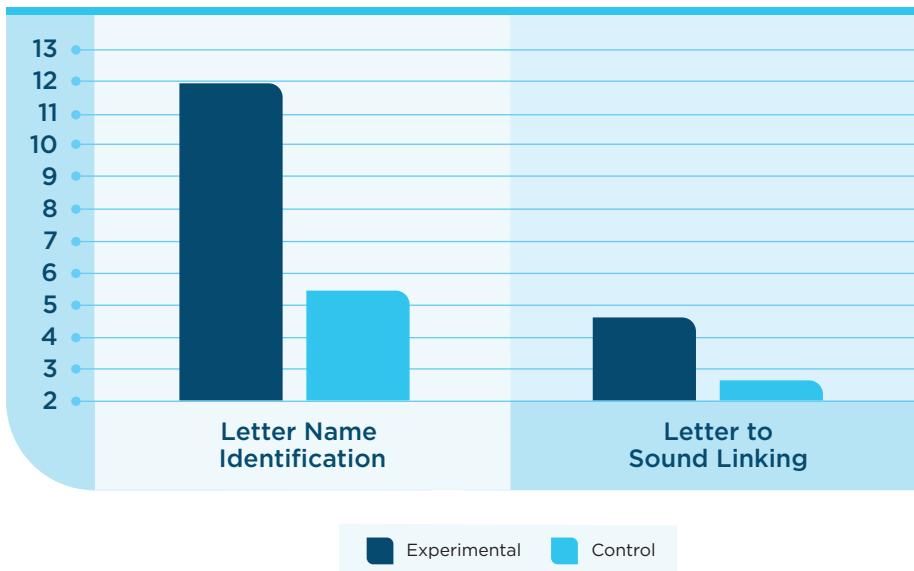


Figure 144: Kindergarten TPRI Gains by Substrand

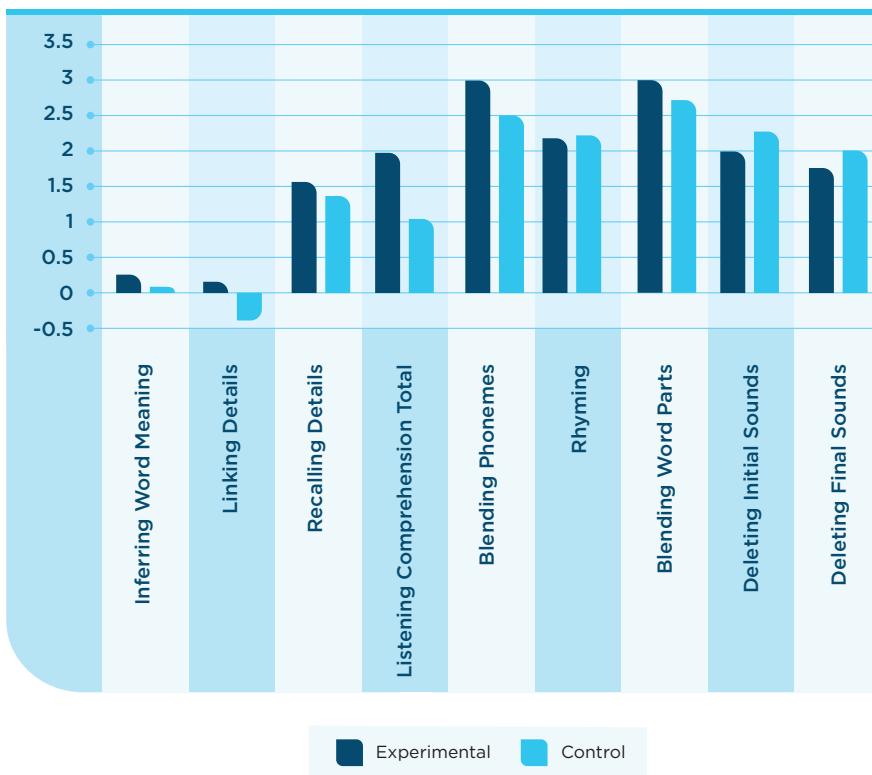


Table 63: Kindergarten TPRI Gains by Special Education Services

Kindergarten	Active Special Education Services				No Special Education Services			
	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 64: Kindergarten TPRI Gains by MLL Status

Kindergarten	MLL				Non-MLL			
	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 65: Kindergarten TPRI Gains by Race/Ethnicity

Kindergarten	African American/Black				Latino/a				Caucasian/White			
	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 145: First Grade TPRI Gains by Substrand

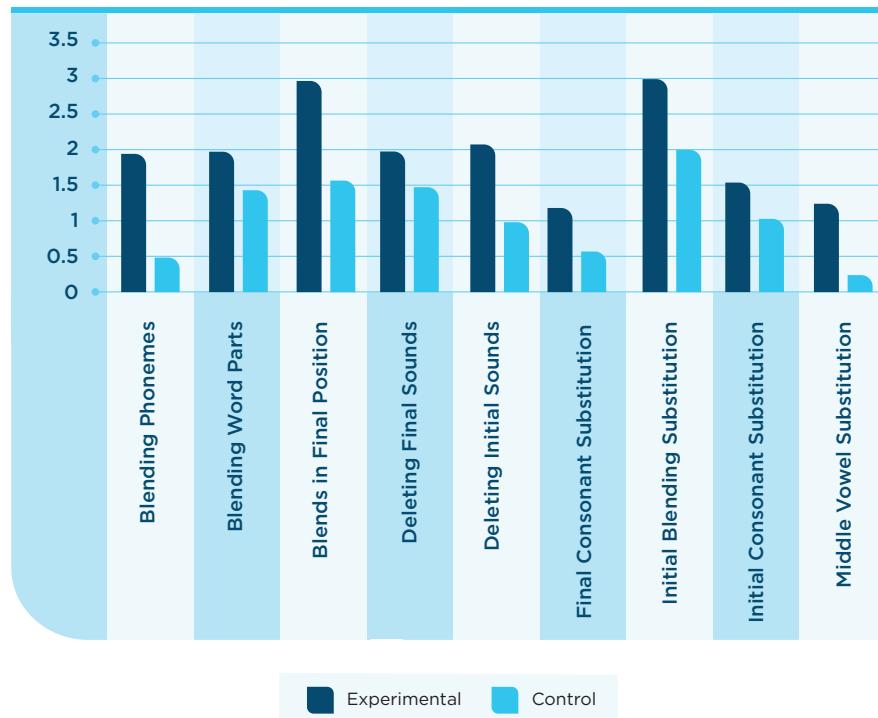


Table 66: First Grade TPRI Gains by Special Education Services

First Grade	Active Special Education Services				No Special Education Services			
	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 67: First Grade TPRI Gains by MLL Status

First Grade	MLL				Non-MLL			
	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 68: First Grade TPRI Gains by Race/Ethnicity

First Grade	African American/Black				Latino/a				Caucasian/White			
	Treatment		Control		Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
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.

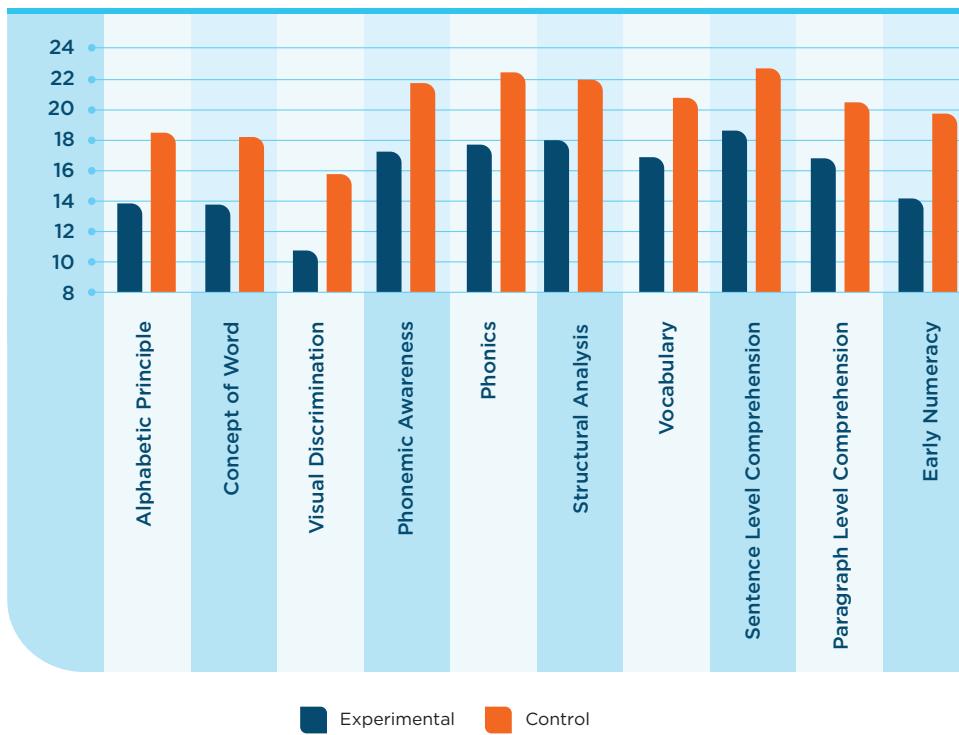
STAR EARLY LITERACY

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

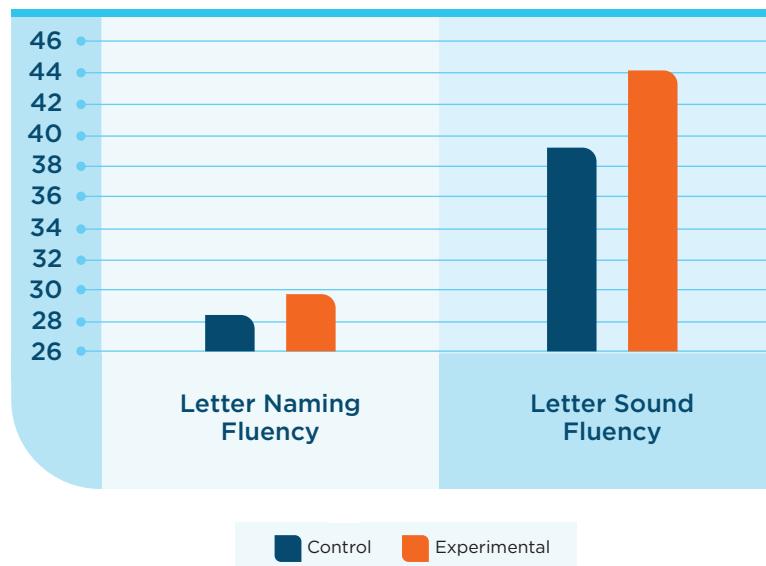


Figure 147: STAR Early Literacy Gains from Beginning of Year to End of Year by Strand



IRI

Figure 148: 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/White students' scores are included in the tables.

DISTRICT 1

KINDERGARTEN

Table 69: District 1 – Kindergarten TPRI Gains by Race/Ethnicity

Kindergarten	African American/Black				Latino/a				Caucasian/White			
	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	2.31	1.49	1.17	1.54	2.04	1.65	1.21	1.70	1.55	1.44	0.85	1.30
- Total Score.												

FIRST GRADE

Table 70: District 1 – First Grade TPRI Gains by Race/Ethnicity

First Grade	<u>African American/Black</u>				<u>Latino/a</u>				<u>Caucasian/White</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
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 71: District 2 – Second Grade DRA Percent Gains by Race/Ethnicity

Second Grade	<u>African American/Black</u>				<u>Multiracial</u>				<u>Caucasian/White</u>			
	<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>		<u>Treatment</u>		<u>Control</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
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 72: District 3 – Kindergarten VLT End of Year Scores by Race/Ethnicity

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

FIRST GRADE

Table 73: District 3 – First Grade VLT End of Year Scores by Race/Ethnicity

First Grade	African American/Black				Latino/a				Caucasian/White			
	Treatment		Control		Treatment		Control		Treatment		Control	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
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 who were experiencing poverty 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 multi-lingual learner (MLL) status, race/ethnicity, and experiencing poverty. 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 multi-lingual learner (MLL) students, students of minority races/ethnicities, and students with free lunch status. Students with MLL status, free lunch status, as well as Latino/a students in the experimental group significantly outperformed their control group counterparts across all strands. African American/Black students in the experimental group also significantly outperformed African American/Black students in the control group on all strands except for Concepts of Print.

Figure 149: KRT Overall End of Year Scores by Demographics

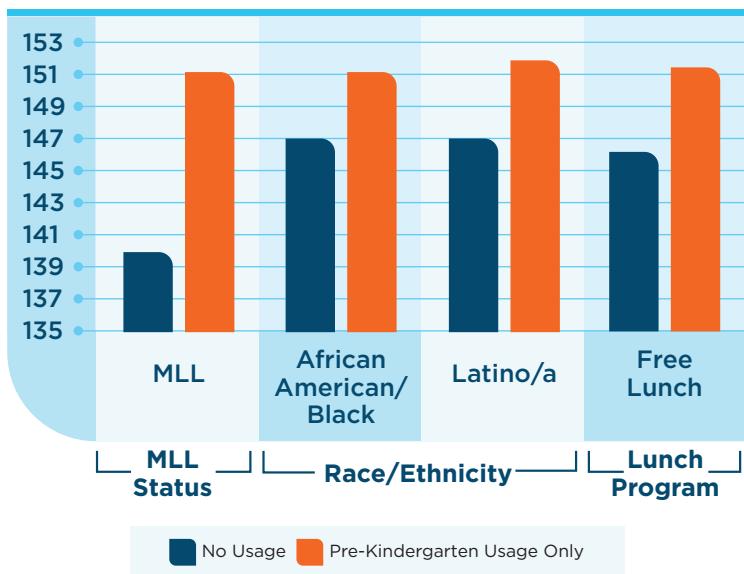


Figure 150: KRT Letter ID & Sounds End of Year Scores by Demographics

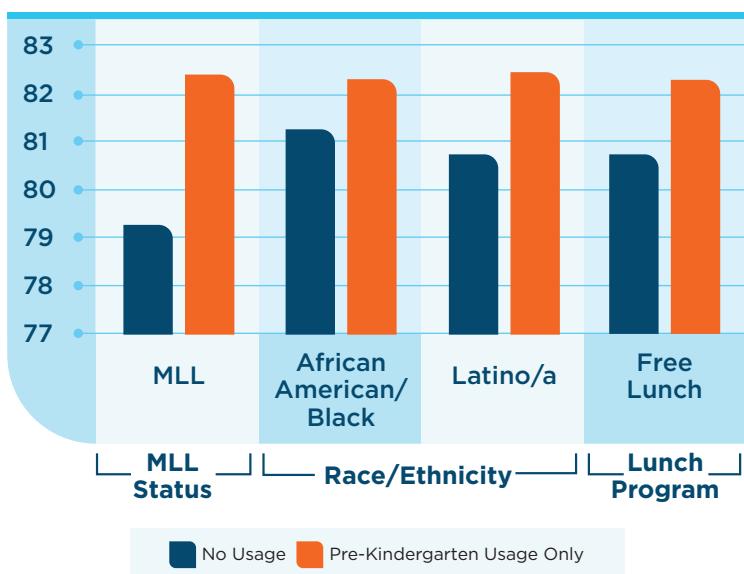


Figure 151: KRT Concepts of Print End of Year Scores by Demographics

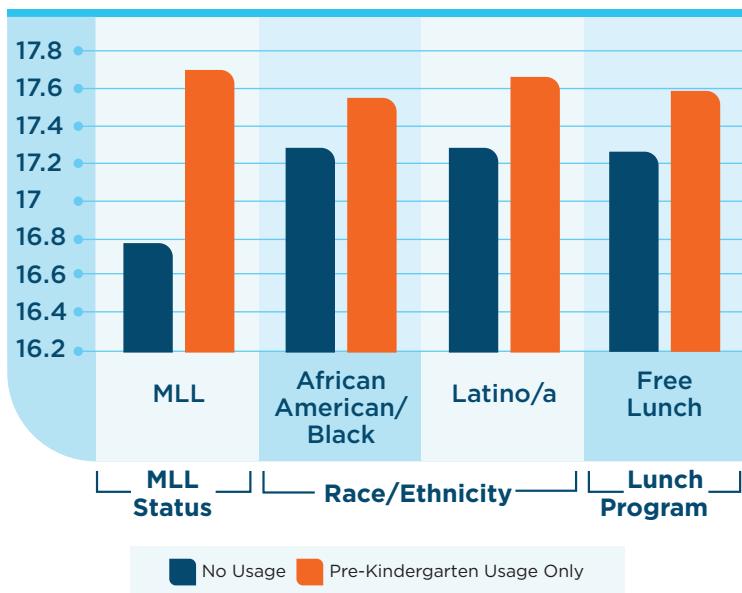


Figure 152: KRT Phonemic Awareness End of Year Scores by Demographics

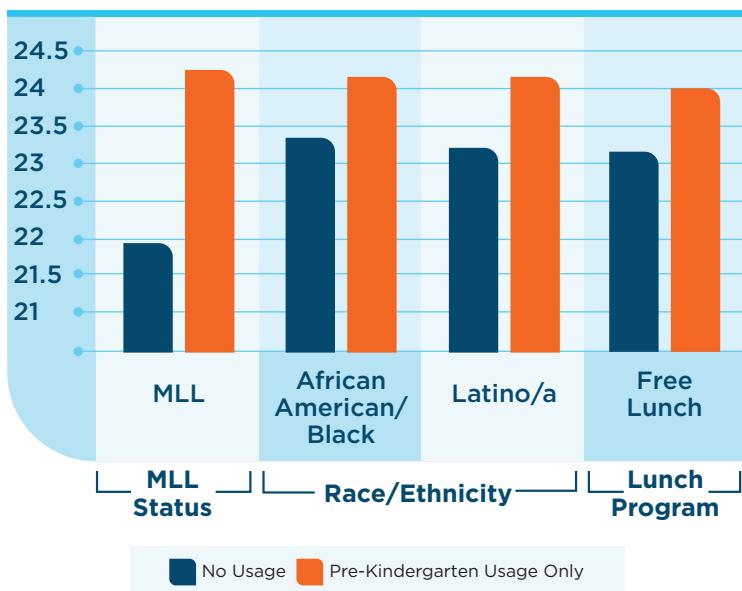


Table 74: End of Kindergarten Scores by Strand & Demographics

	Experimental			Control			<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>		
Overall								
MLL Status	151.43	15.98	111	140.34	32.27	2627	-6.75	.000**
African American/Black	151.67	18.16	100	147.25	26.37	2528	-2.34	.021*
Latino/a	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**
Letter ID and Sounds								
MLL Status	82.50	2.30	111	79.23	12.14	2613	-10.13	.000**
African American/Black	82.14	3.24	100	81.16	7.27	2509	-2.76	.006**
Latino/a	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**
Concepts of Print								
MLL Status	17.69	1.32	111	16.78	3.17	2612	-6.55	.000**
African American/Black	17.53	2.04	100	17.35	2.18	2506	-0.81	.421
Latino/a	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*
Phonemic Awareness								
MLL Status	24.14	3.45	111	22.06	6.32	2623	-5.95	.000**
African American/Black	24.10	3.17	100	23.39	4.72	2521	-2.15	.033*
Latino/a	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 69-72).

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

Table 75: End of First Grade DRA Scores

		Treatment			Control			<i>P</i>
		<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
K and 1 st 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 76: End of Second Grade DRA Scores

		Treatment			Control			<i>p</i>
		<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
K and 1 st Usage	Cohort 2	28.11	5.74	1235	24.88	7.61	281	.00**
	K Only Usage	25.47	9.19	49	24.88	7.61	281	.63
1 st 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 77: End of First Grade High Usage DRA Scores

		Treatment			Control			<i>p</i>
		<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
K and 1 st 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 78: End of Second Grade High Usage DRA Scores

		Treatment			Control			<i>p</i>
		<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
K and 1 st Usage	Cohort 2	28.56	5.29	967	24.88	7.61	281	.00**
	K Only Usage	27.70	8.00	33	24.88	7.61	281	.05*
1 st 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 153: Three Years of Usage vs. No Usage End of Year Scores by Strand

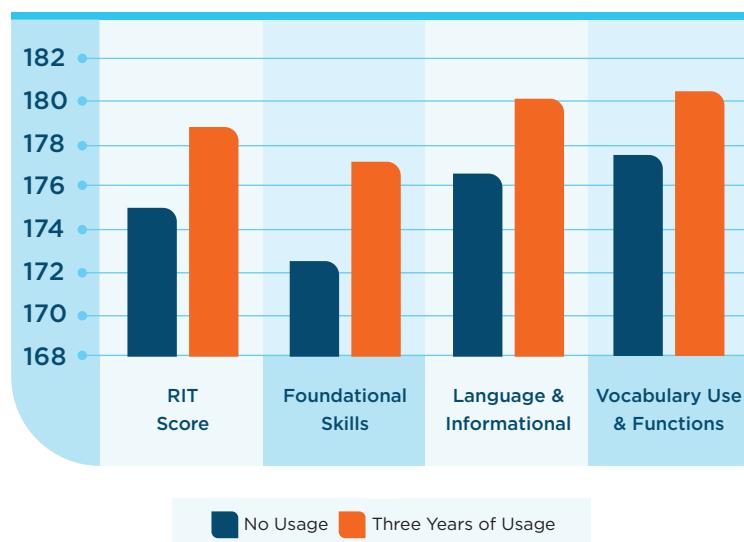
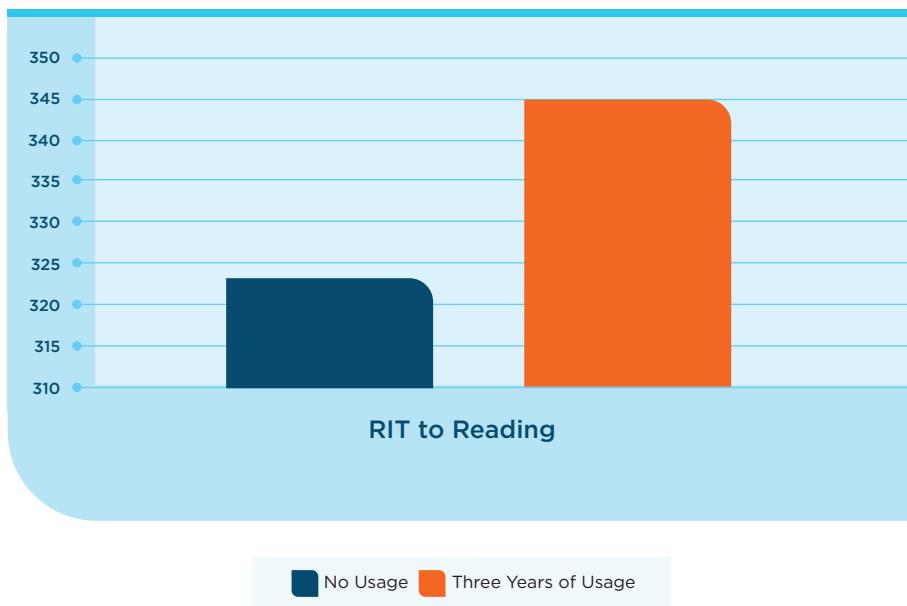


Figure 154: Three Years of Usage vs. No Usage RIT to Reading End of Year Scores



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

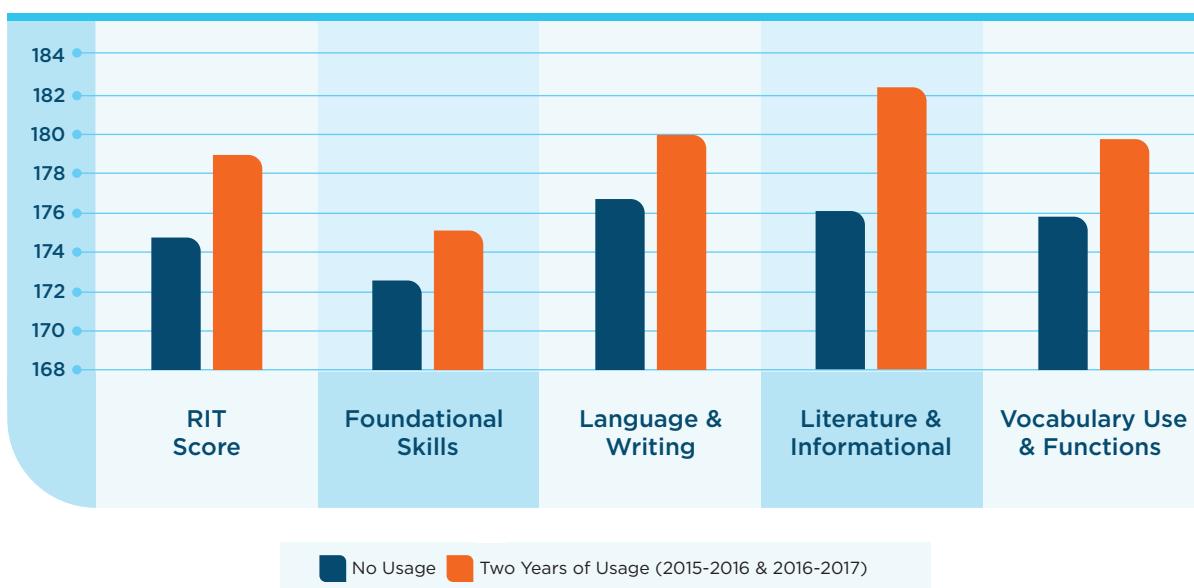
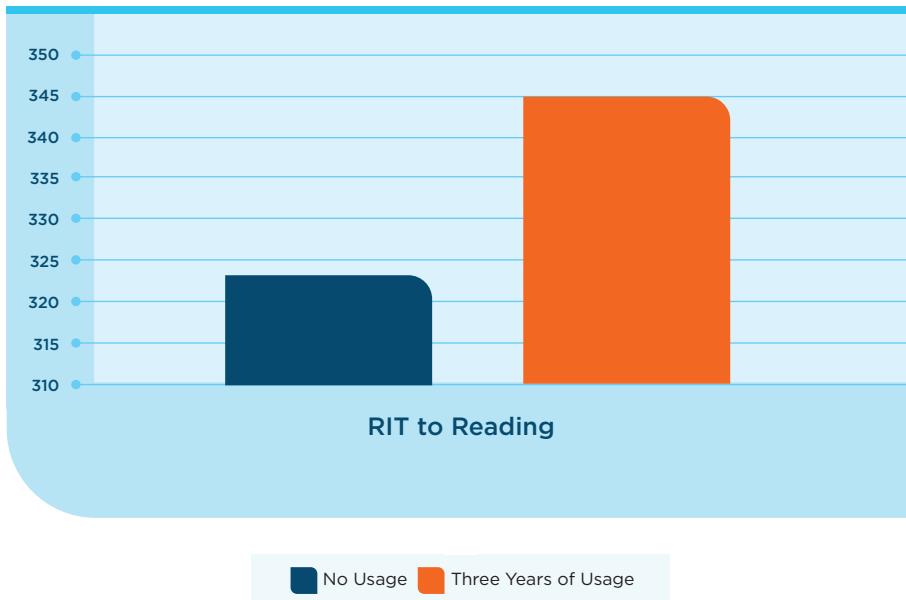


Figure 156: Two Years of Usage vs. No Usage RIT to Reading End of Year Scores



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

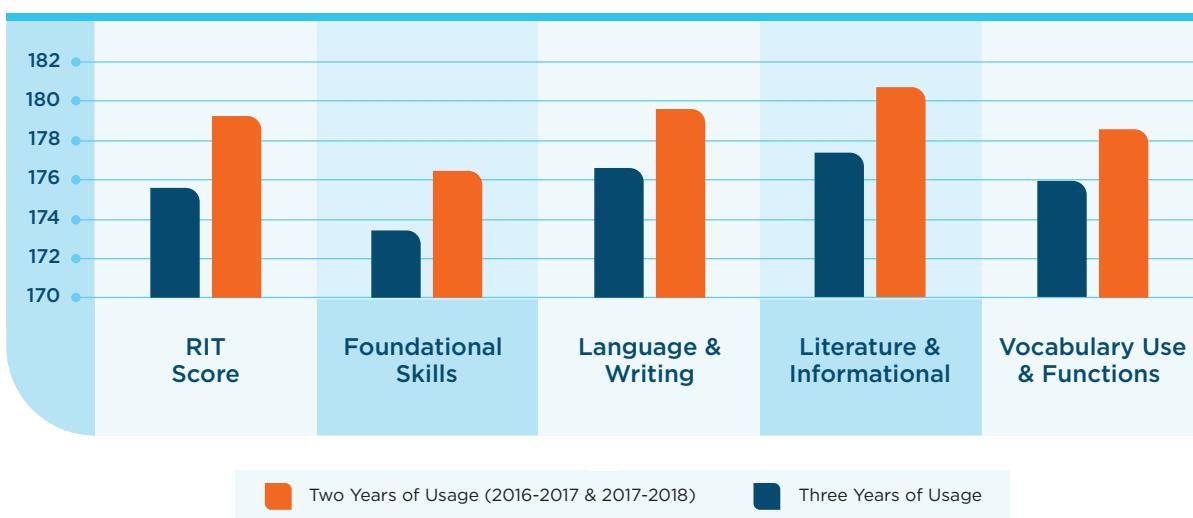
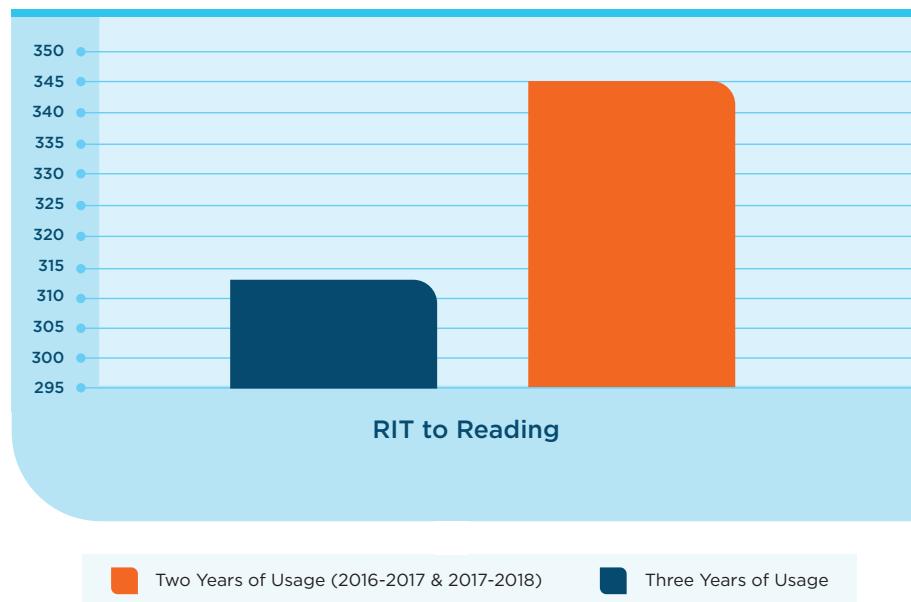


Figure 158: Three Years of Usage vs. Two Years of Usage RIT to Reading End of Year Scores



Longitudinal Evaluation of Three School Districts, 2015-2019

The following longitudinal study (Shamir, Pocklington, Yoder & Feehan, 2020) assessed the efficacy of Waterford Early Learning (WEL) for second grade students across three school districts: District 1 used an assessment that measured literacy skills, including Dictation Sounds and Dictation words; District 2 used a reading assessment that was part of a universal reading screener; and District 3 used the Developmental Reading Assessment (DRA). Across all districts, kindergarten students were expected to use WEL for 15 minutes per day five days per week during their kindergarten year, and for 30 minutes per day, five days per week during first and second grade.

The study tested three hypotheses: (1) Students who used CAI for two or three years would have higher literacy assessment scores than students who did not use CAI, demonstrating an overall effect of CAI on literacy scores. (2) Students who used CAI for one or two years, when assessed one year after use of CAI has stopped, would have higher literacy assessment scores than students who did not use CAI, demonstrating a long-term effect of CAI on literacy scores. (3) Students who began using CAI during kindergarten would have higher literacy assessment scores than students who did not begin using CAI until first grade, demonstrating an early effect of CAI on literacy scores. Experimental and control groups were formed for each district based on the different number of school years in which students had usage between kindergarten, first grade, and second grade.

All three hypotheses were supported across the different districts, demonstrating overall, long-term, and early effects of using WEL. For the analysis of overall effects, in the two districts where overall effects were assessed, students who used WEL for all three years significantly outperformed control students, who did not use WEL, on all strands assessed. For the analysis of long-term effects, students who used WEL prior to second grade outperformed control students at the end of second grade on all strands assessed and in all three districts. For the analysis of early effects, across all districts, students

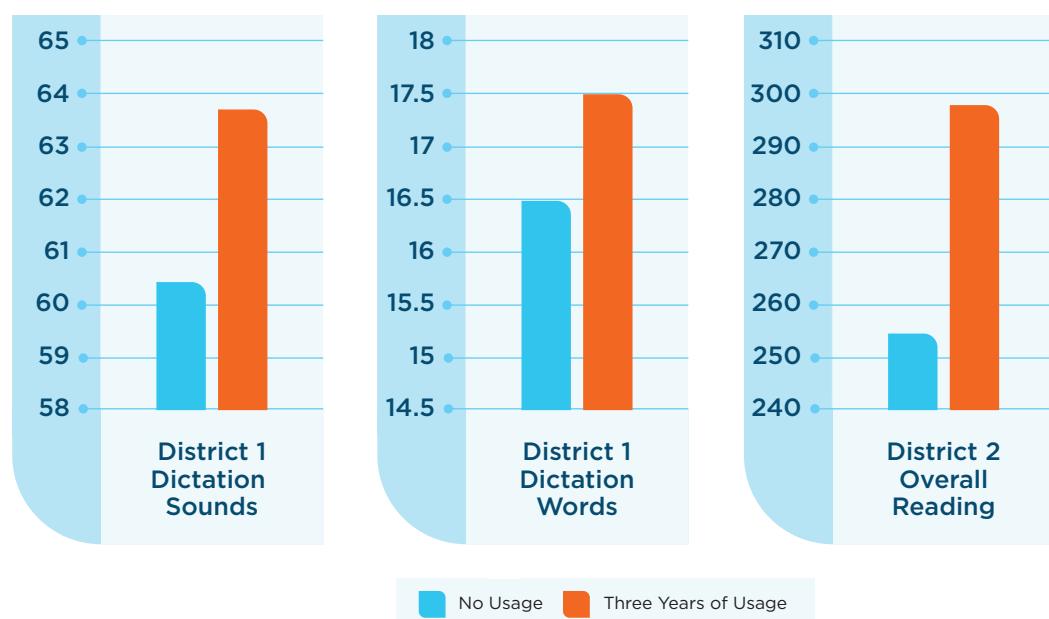
who started using WEL during kindergarten and consistently used for two or three years, significantly outperformed students who started using the program one year later on all strands assessed.

Table 79: Participants by District and Usage Group Comparisons

	District 1	District 2	District 3	Usage Group Comparisons
Total 2nd Grade Students	5,002	1,796	1,849	
3 Years of Usage Experimental	2,108	585	-	Overall Effects, Early Effects
K & 1st Grade Usage Experimental	288	478	1,228	Long-Term Effects, Early Effects
No Usage Control	593	142	287	Overall Effects, Long-Term Effects
No K Usage Control	729	137	278	Early Effects

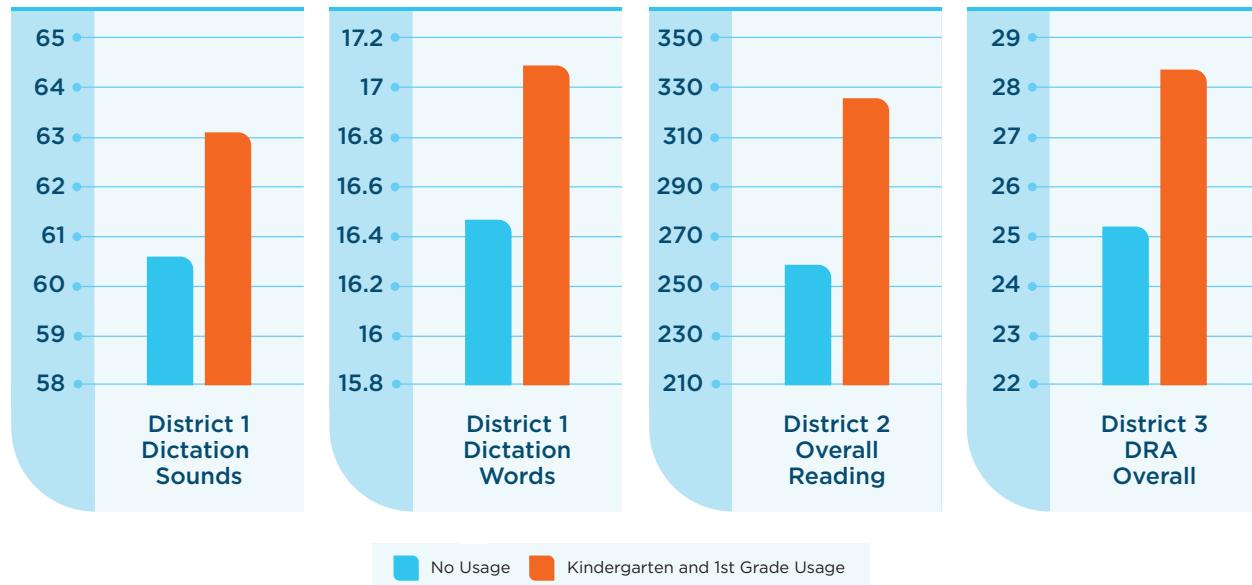
OVERALL EFFECTS

Figure 159: Overall Effects - Second Grade Scores



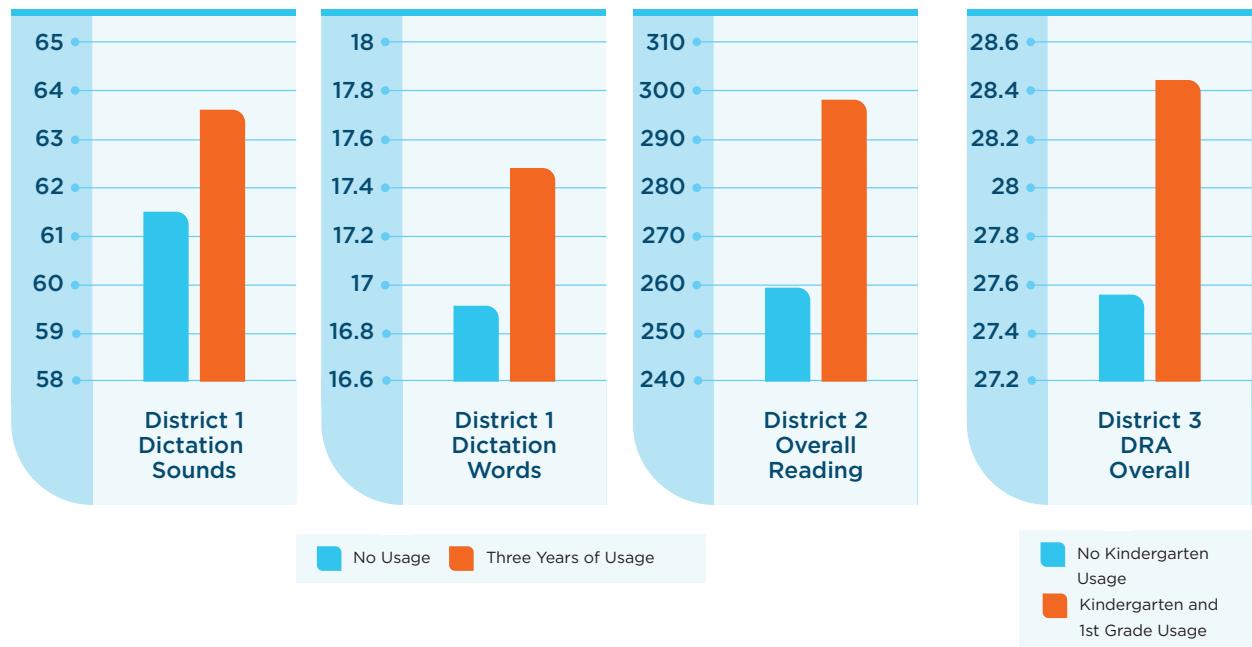
LONG-TERM EFFECTS

Figure 160: Long-Term Effects - End of Second Grade Scores



EARLY EFFECTS

Figure 161: Early Effects - End of Second Grade Scores



Longitudinal Evaluation of Waterford Upstart in Mississippi, 2018–2020

The aim of this longitudinal study was to investigate the long-lasting effects of the Waterford Upstart program (Evaluation and Training Institute, 2020c). This study consisted of 112 kindergarten students from Mississippi who participated in the Waterford Upstart program during the 2018–2019 school year, while in pre-kindergarten. The majority of the students (96%) were African American/Black, and almost half of the students (41%) were below 100% of the poverty level. Students were assigned to either the treatment group ($N = 50$), which used the reading program, or the comparison group ($N = 62$), which used the math program. At the end of their kindergarten year, students' literacy and math skills were assessed using the Kaufman Test of Educational Achievement Third Edition (KTEA-3).

Students in the treatment group significantly outperformed students in the comparison group on all literacy subskills, with small to medium effect sizes, including letter-word recognition ($ES = 0.41$), reading comprehension ($ES = 0.39$), and phonological processing ($ES = 0.38$). Additionally, students in the comparison group who were part of the math program performed significantly higher than the treatment group on the math concepts and applications subskill ($ES = 0.50$). These results demonstrate meaningful long-term effects for students who were part of Waterford Upstart during preschool, which lasted through the end of kindergarten, approximately one year after they had participated in the program.

Figure 162: KTEA-3 End of Kindergarten Literacy Scores for Treatment and Comparison Groups

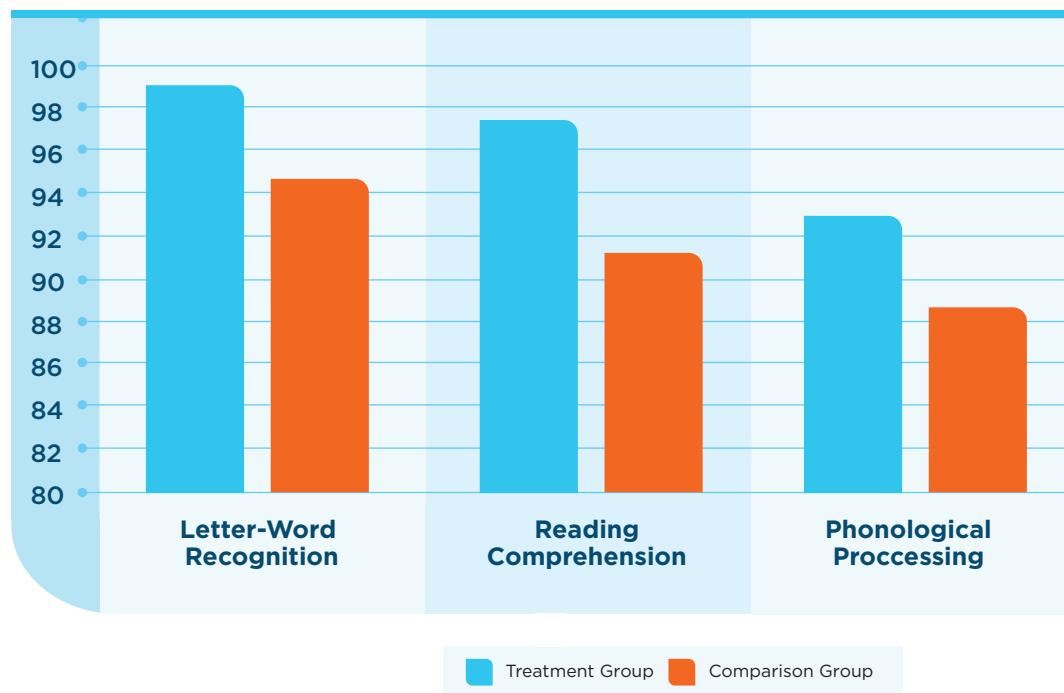
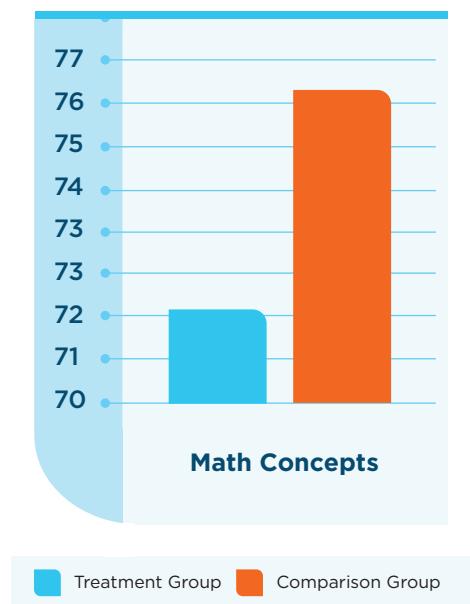


Figure 163: KTEA-3 End of Kindergarten Math Scores for Treatment and Comparison Groups



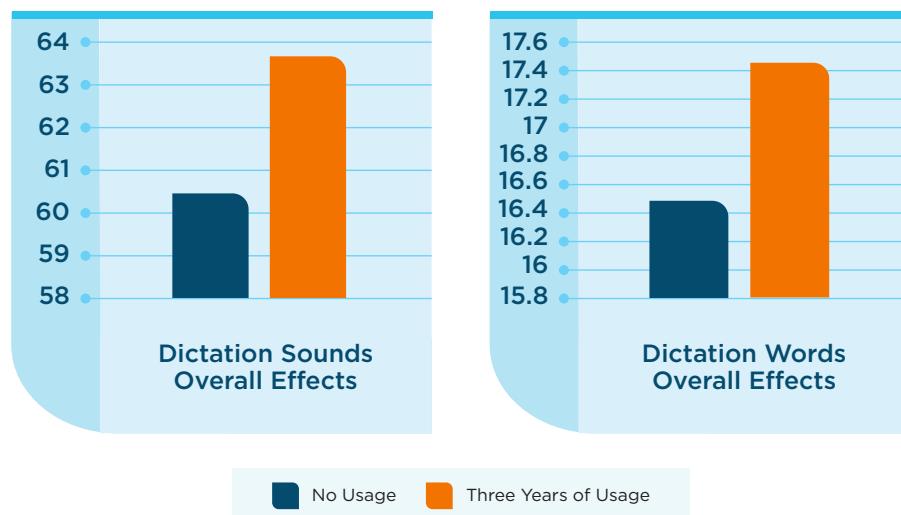
Longitudinal Evaluation of Waterford Early Learning in Maryland, 2015-2018

The following longitudinal study (Shamir, Yoder, Pocklington, Feehan, & Ortiz-Wood, 2022) assessed the efficacy of Waterford Early Learning (WEL), for students who used the software for one, two, or three years. Participants consisted of second grade students ($N = 5,002$) enrolled in a large suburban public school district in Maryland. A District-Administered Literacy Assessment was administered to all students at the end of the 2017-2018 school year, and independent samples t -tests were conducted to examine group differences in end of second grade scores between experimental and control groups. The study reports findings across three distinct comparisons: Overall effects, long-term effects, and early effects.

THREE YEARS OF USAGE VS. NO USAGE: OVERALL EFFECTS

For the overall effects, students in the experimental group used WEL during kindergarten, first grade, and second grade, while students in the control group did not use WEL during all three school years. Usage, in this context, was defined as using WEL for more than 100 minutes within a school year. Participants were said to have not used WEL if they used it for less than 100 minutes. Students in the experimental group significantly outperformed students in the control group on both literacy subskills: Dictation Sounds, $t(1, 673) = -3.97, p < .01, (d = 0.19)$ and Dictation Words, $t(1, 729) = -3.10, p < .01, (d = 0.15)$.

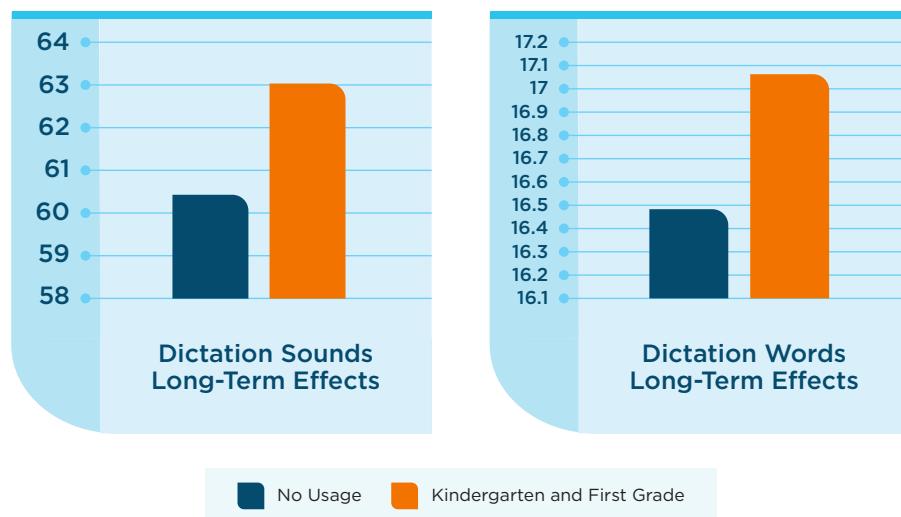
Figure 164: Three Years of Usage Vs. No Usage End of Second Grade Scores



KINDERGARTEN AND FIRST GRADE USAGE VS. NO USAGE: LONG-TERM EFFECTS

To examine the long-term effects of WEL usage, students in the experimental group used WEL during kindergarten and first grade but did not use the program during second grade. Meanwhile, students in the control group did not use WEL during all three school years. Experimental group students had significantly higher Dictation Sound scores than control group students, $t(1, 702) = -2.30, p < .05, d = 0.17$. Experimental group students scored higher than control group students on Dictation Words, but the difference was not significant.

Figure 165: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores

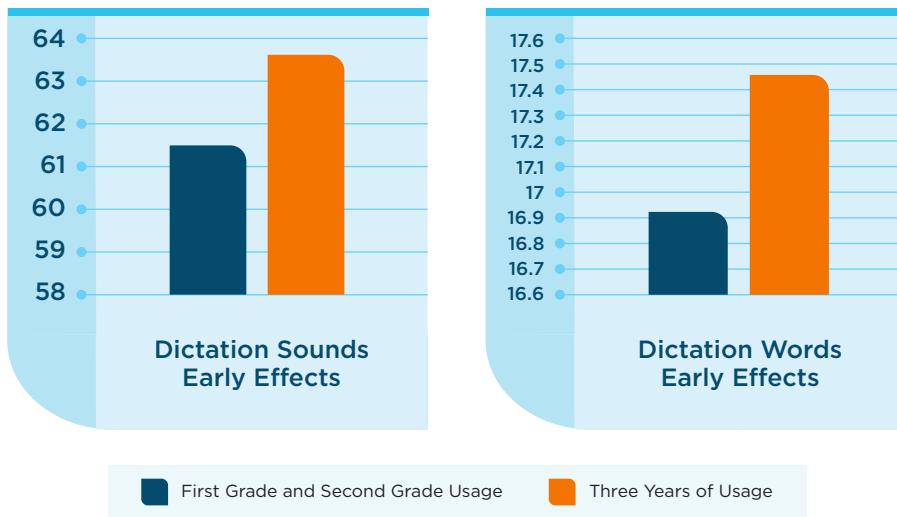


THREE YEARS OF USAGE VS. FIRST GRADE AND SECOND GRADE USAGE: EARLY EFFECTS

For the early effects, students in the experimental group used WEL during kindergarten, first grade, and second grade, while students in the control group used WEL during first grade and second grade.

Students in the experimental group significantly outperformed students in the control group on both literacy subskills: Dictation Sounds, $t(1, 1029) = -3.44, p < .01, (d = 0.15)$ and Dictation Words, $t(1, 1126) = -2.10, p < .05, (d = 0.09)$.

Figure 166: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores



Longitudinal Evaluation of Waterford Early Learning in Texas, 2018-2020

The following longitudinal study (Shamir, Pocklington, Yoder, Feehan, & Ortiz-Wood, 2022) assessed the efficacy of Waterford Early Learning (WEL), a computer-adaptive program that was assigned to students in a small city public school district in Texas during the 2018-2019 school year. A District-Administered Literacy Assessment was administered to all students at the end of the school year and independent samples t -tests were conducted to examine group differences in end of second grade scores between experimental and control groups. Two-way ANOVAs were conducted to examine the effects of WEL and demographics on end of second grade scores.

THREE YEARS OF USAGE VS. NO USAGE: OVERALL EFFECTS

For the overall effects, students in the experimental group used WEL during kindergarten, first grade, and second grade, while students in the control group did not use WEL during all three school years. Usage, in this context, was defined as using WEL for more than 100 minutes within a school year. Participants were said to have not used WEL if they used it for less than 100 minutes. Students in the experimental group significantly outperformed students in the control group on assessment scores at the end of second grade, $t(1,725) = -3.25, p < 0.01, d = 0.30$ scores ($d = 0.30$). Across most demographics, students in the experimental group outperformed students in the control group. Simple effects analysis showed that females, students with no special education services, students who were not experiencing poverty, multi-lingual learner students, and Latino/a students in the experimental group significantly outperformed students in the control group.

Figure 167: Three Years of Usage Vs. No Usage End of Second Grade Scores

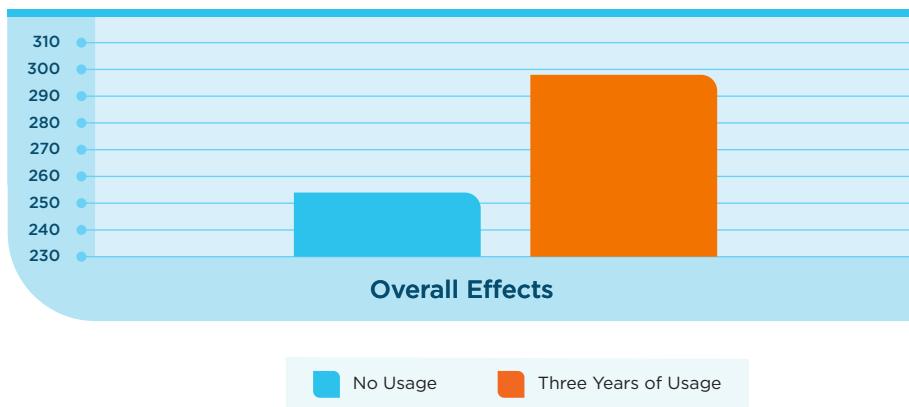
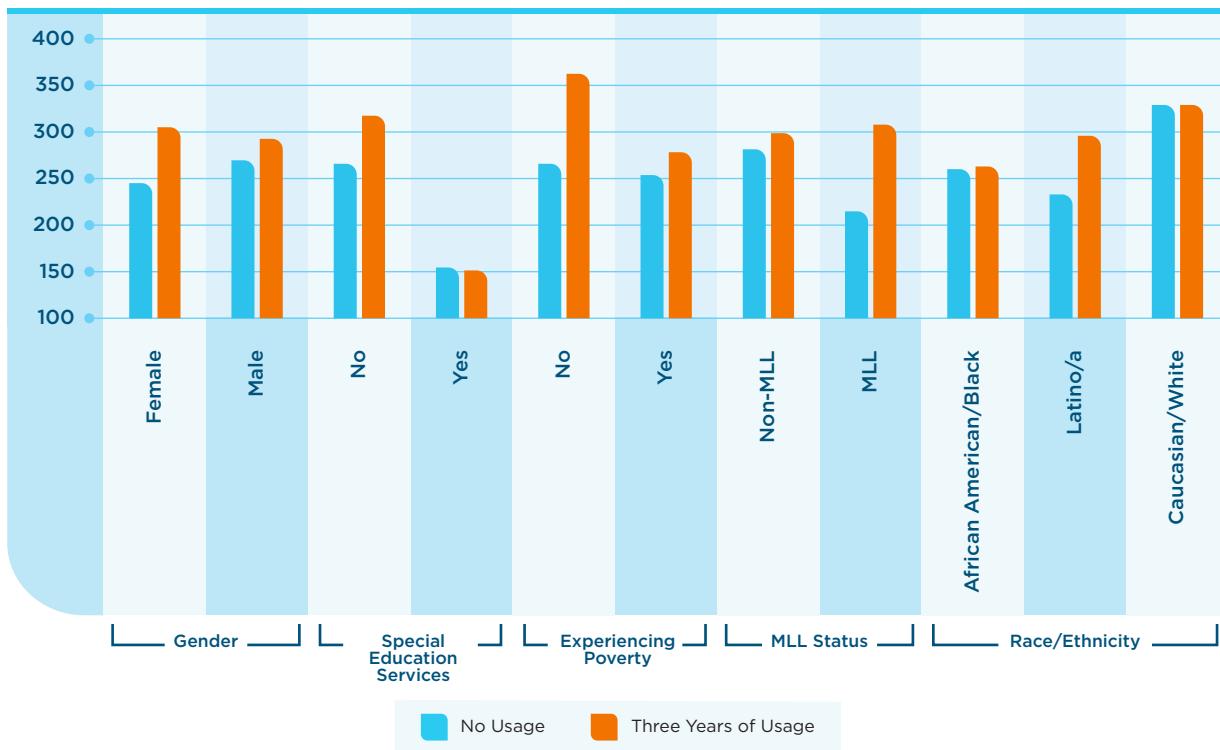


Figure 168: Three Years of Usage Vs. No Usage End of Second Grade Scores by Demographics



KINDERGARTEN AND FIRST GRADE USAGE VS. NO USAGE: LONG-TERM EFFECTS

To examine the long-term effects of WEL usage, students in the experimental group used WEL during kindergarten and first grade, while students in the control group did not use WEL during all three school years. Students in the experimental group significantly outperformed students in the control group on end of second grade scores, $t(1,618) = -5.01, p < 0.01, (d = 0.48)$. Across all demographics, students in the experimental group outperformed students in the control group. Simple effects analysis showed that males and females, students with no special education services, students experiencing poverty and students not experiencing poverty, students with multi-lingual learner status and monolingual students,

and Latino/a students in the experimental group significantly outperformed students in the control group.

Figure 169: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores

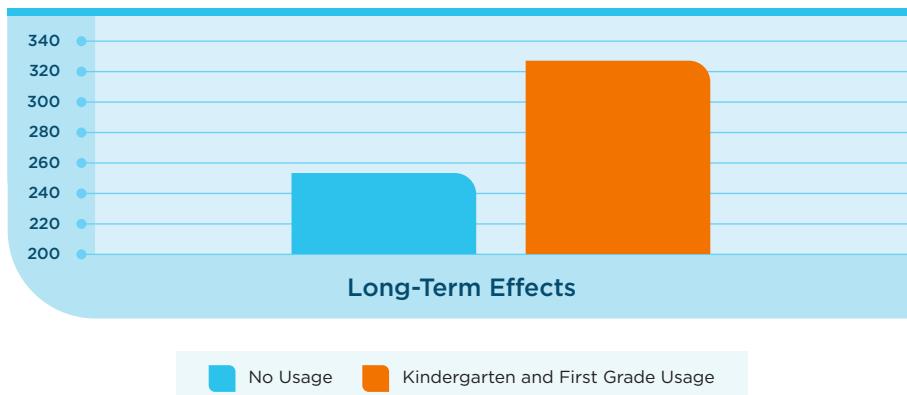
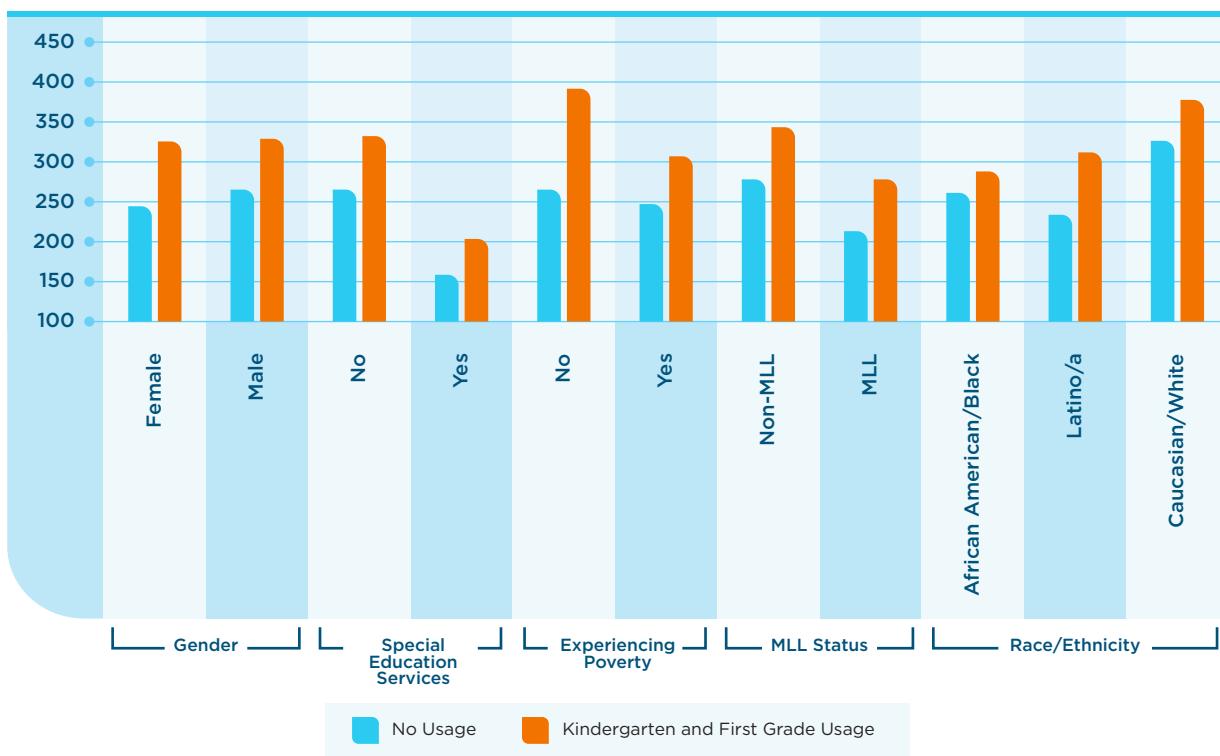


Figure 170: Kindergarten and First Grade Usage Vs. No Usage End of Second Grade Scores by Demographics



THREE YEARS OF USAGE VS. FIRST GRADE AND SECOND GRADE USAGE: EARLY EFFECTS

To examine early effects of WEL usage, students in the experimental group used WEL during kindergarten, first grade, and second grade, while students in the control group used WEL during first grade and second grade. Students in the experimental group significantly outperformed students in the control group on end of second grade scores, $t(1,720) = -2.77, p < 0.01, (d = 0.26)$. Across all demographics, students in the experimental group outperformed students in the control group. Simple

effects analysis showed that males, students with no special education services, students experiencing poverty and students not experiencing poverty, students who are not multi-lingual learners, and Caucasian/White students in the experimental group significantly outperformed students in the control group.

Figure 171: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores

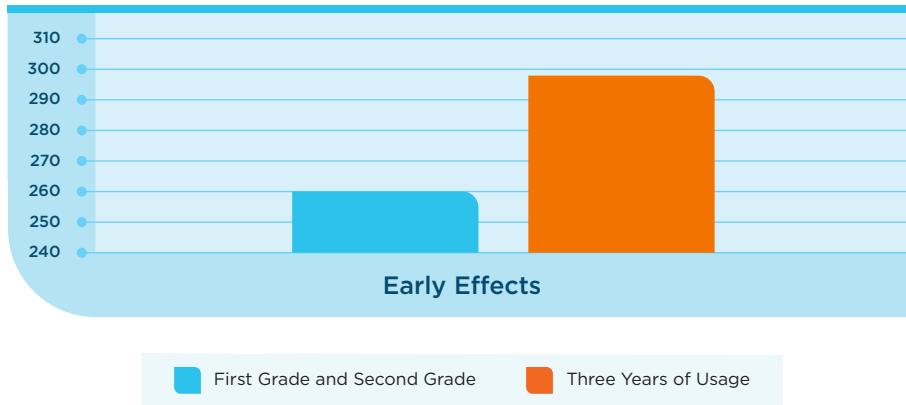
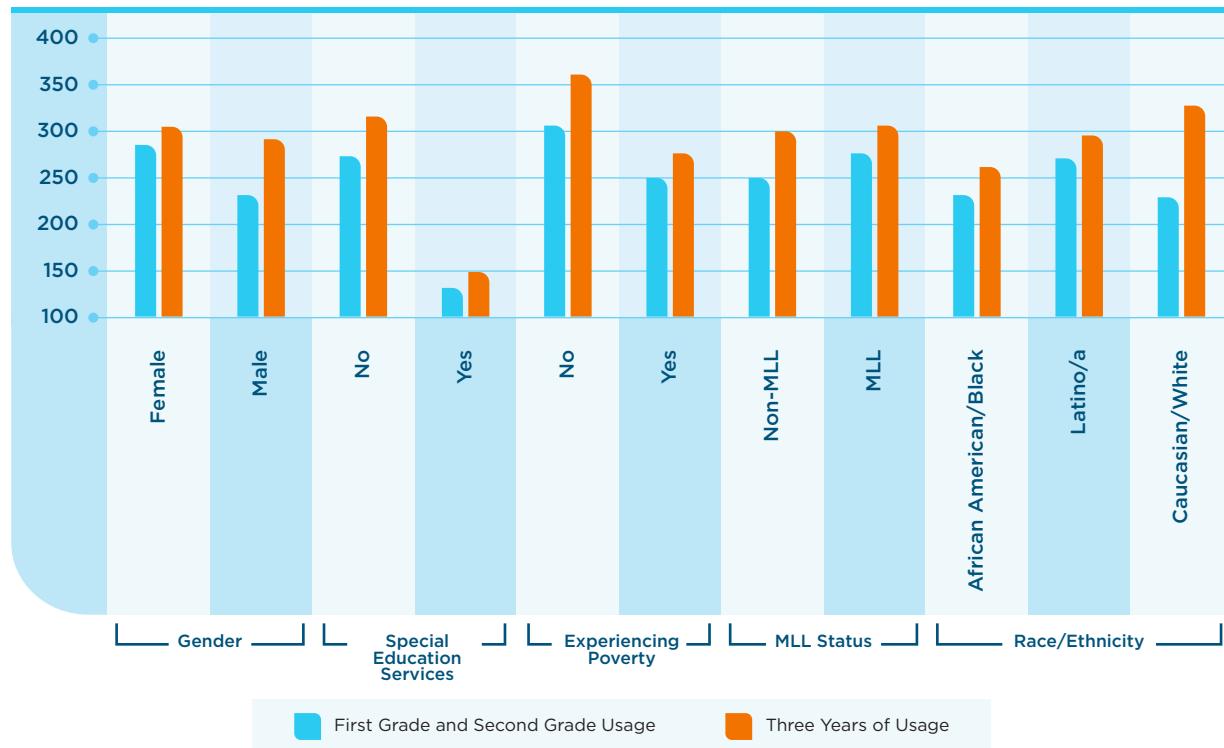


Figure 172: Three Years of Usage Vs. First Grade and Second Grade Usage End of Second Grade Scores by Demographics



Longitudinal Evaluation of a School District in Texas, 2019-2022

This study investigated the long-term effects of using Waterford Reading Academy (WRA) across two cohorts (Shamir, Yoder, & Pocklington, 2023b). First grade students ($N = 1,368$) in a school district in Texas were assigned to experimental and control groups based on their total time spent using WRA during the previous school year when they were in kindergarten. For the 2020-2021 cohort, students with at least 1,000 minutes during kindergarten in 2019-2020 were included in the experimental group ($n = 534$), while students with less than 500 minutes were included in the control group ($n = 115$). For the 2021-2022 cohort, students with at least 1,000 minutes during kindergarten in 2020-2021 were included in the experimental group ($n = 592$), while students with less than 500 minutes were included in the control group ($n = 127$). The TPRI literacy assessment was administered to students in each cohort at the end of their school year in first grade.

Independent samples t -tests were conducted to examine end-of-first-grade scores for each cohort and a combined sample of both years. Results revealed that the experimental group significantly outperformed the control group on TPRI Word Reading for the 2020-2021 cohort ($d = 0.59$), the 2021-2022 cohort ($d = 0.65$), and the combined sample ($d = 0.62$). Two-way ANOVAs were also conducted to examine the long-term effects of WRA and demographics on end-of-first-grade Word Reading scores, which revealed that all students across race/ethnicity, special education services, experiencing poverty, and multi-lingual learner status, benefited from the use of WRA at the end of first grade, one year after they had stopped using the software. This demonstrates the long-term impact of WRA on literacy skills of early learners.

Figure 173: 2020-2021 End of First Grade TPRI Word Reading Scores



Figure 174: 2020-2021 End of First Grade TPRI Word Reading Scores by Demographics

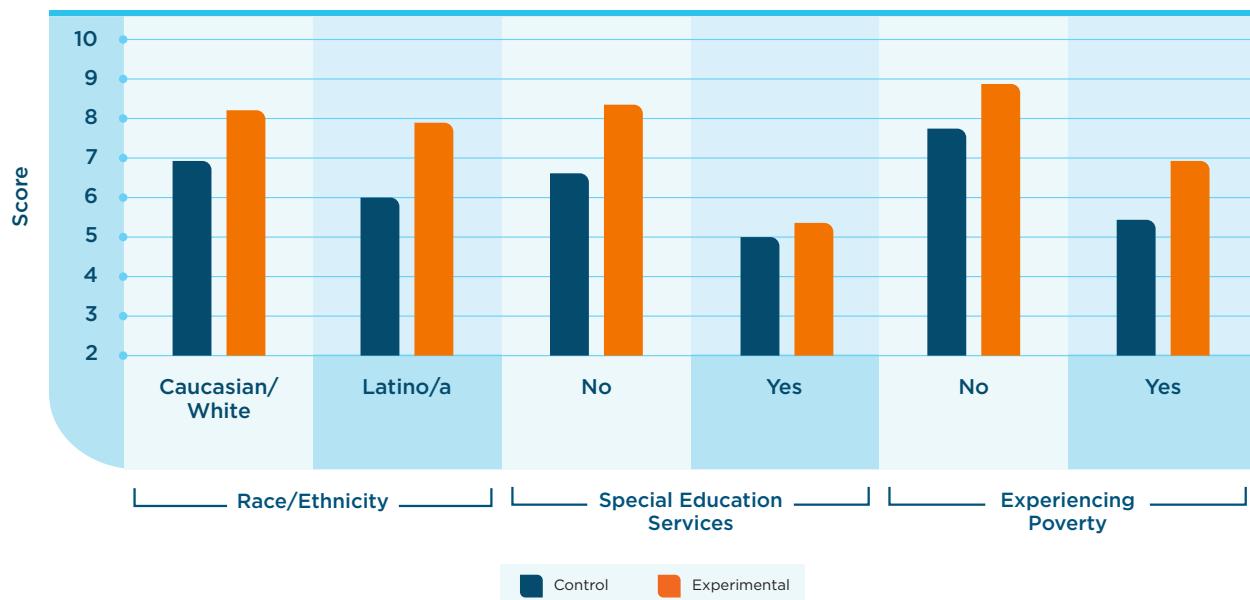


Figure 175: 2021-2022 End of First Grade TPRI Word Reading Scores

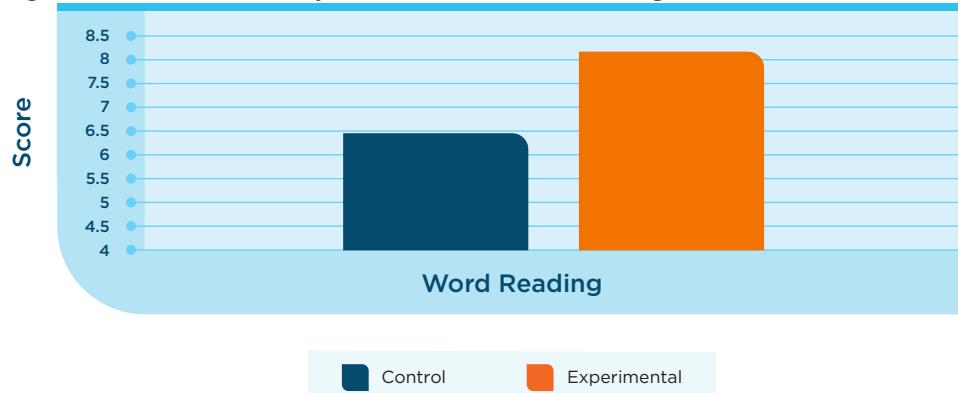


Figure 176: 2021-2022 End of First Grade TPRI Word Reading Scores by Demographics

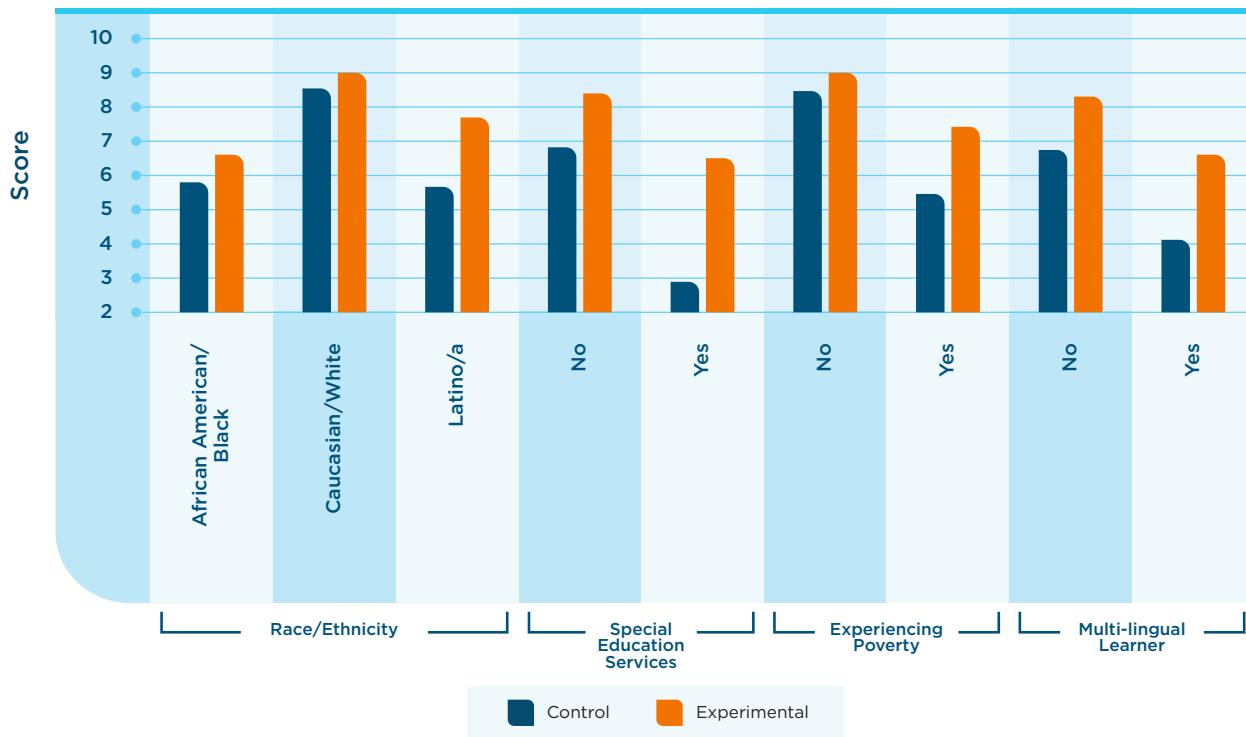
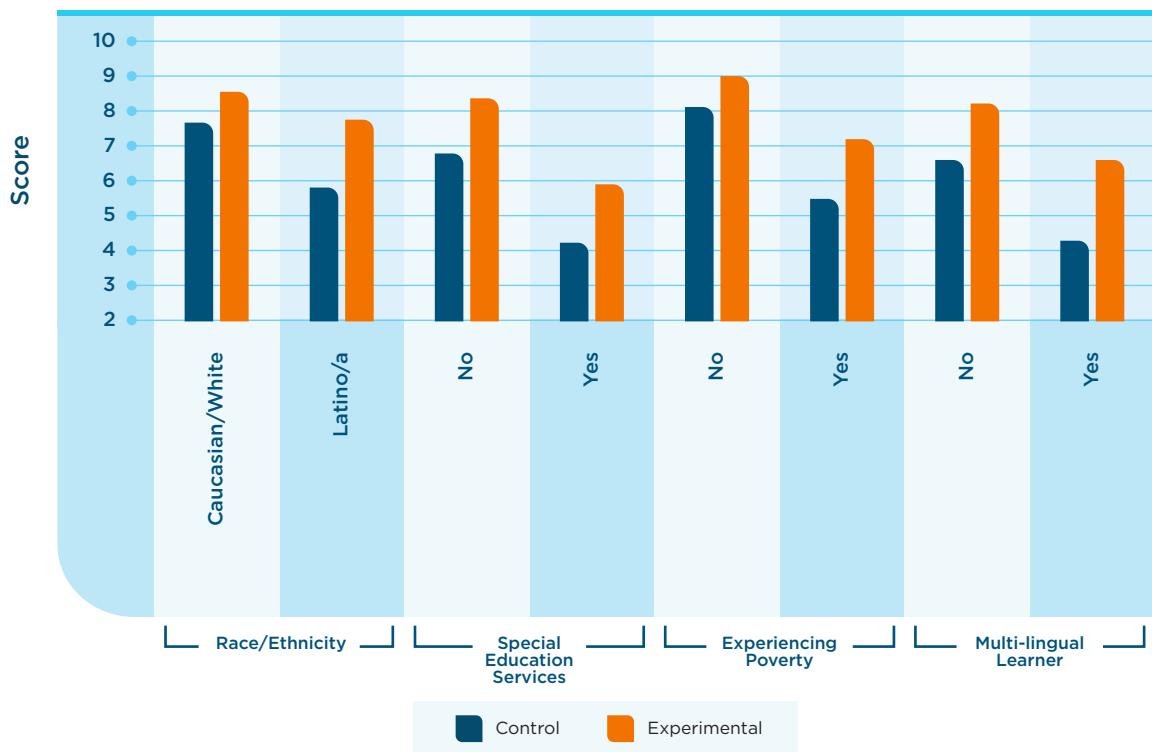


Figure 177: Combined End of First Grade TPRI Word Reading Scores



Figure 178: Combined End of First Grade TPRI Word Reading Scores by Demographics



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

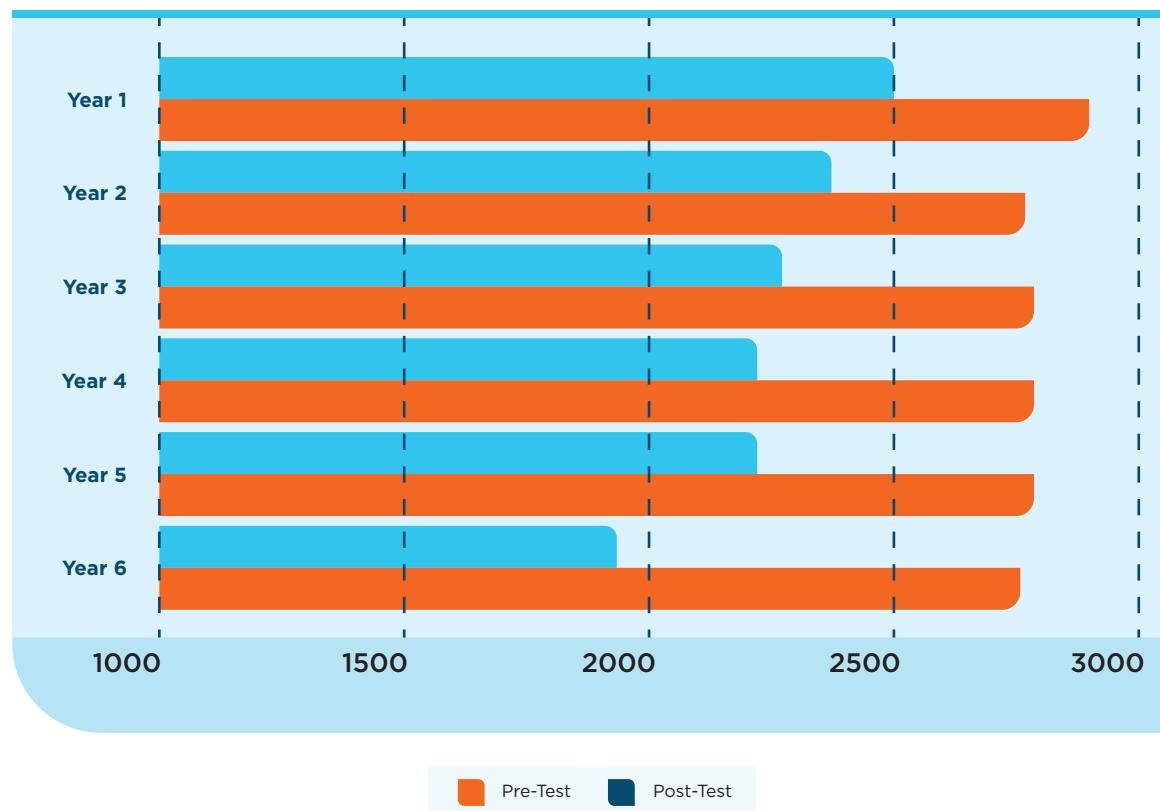
WACS results are provided for years five and six (Shamir, Miner, Izzo, Pocklington, Feehan, & Yoder, 2018). Results for years one through four are very similar and can be provided upon request.

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 133 shows the improvement in academic achievement as measured by WACS pretest and posttest scores.

Table 80: 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 179: 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 races/ethnicities, and from upper and lower SES levels, indicating that the UPSTART program closes the achievement gap between groups of students (Miner, 2014).

Figure 180: Year 5 WACS Reading Scores, Grouped by Subtests

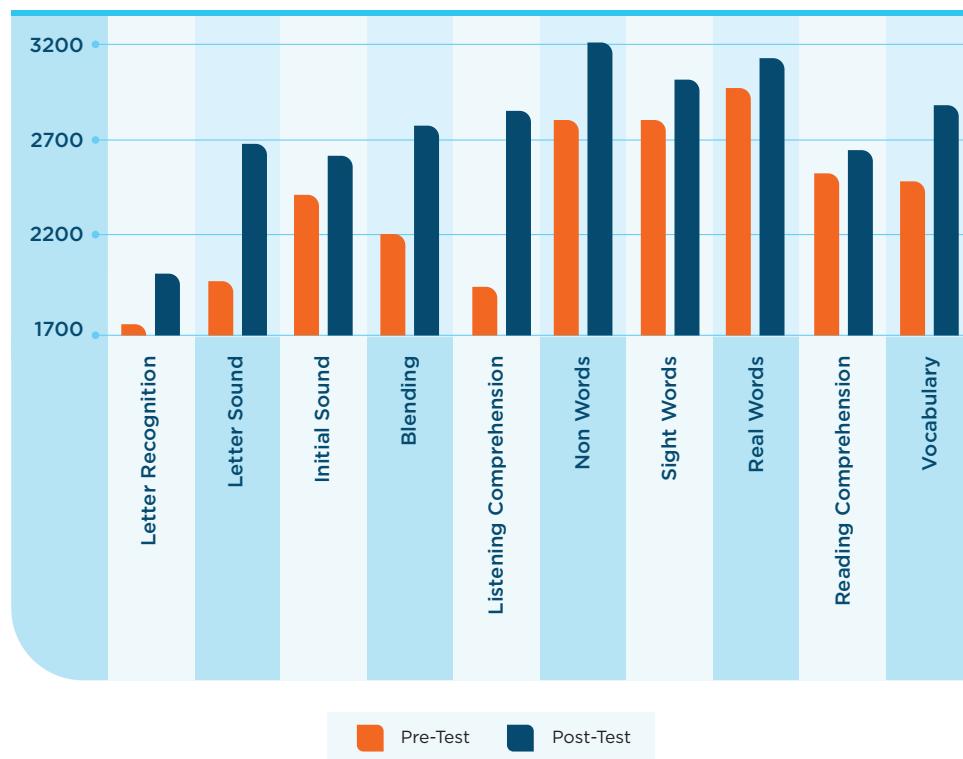


Table 81: WACS Year 5 Posttests

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 181: WACS Reading Score Gains Grouped by Race/Ethnicity

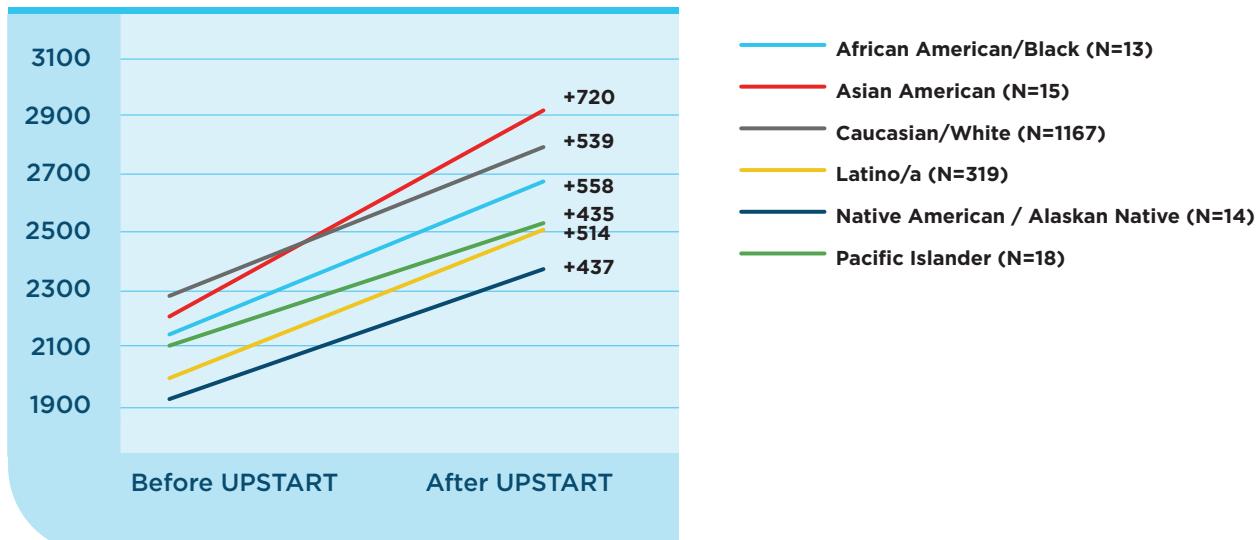


Figure 182: WACS Reading Score Gains Grouped by Other Preschool Attendance

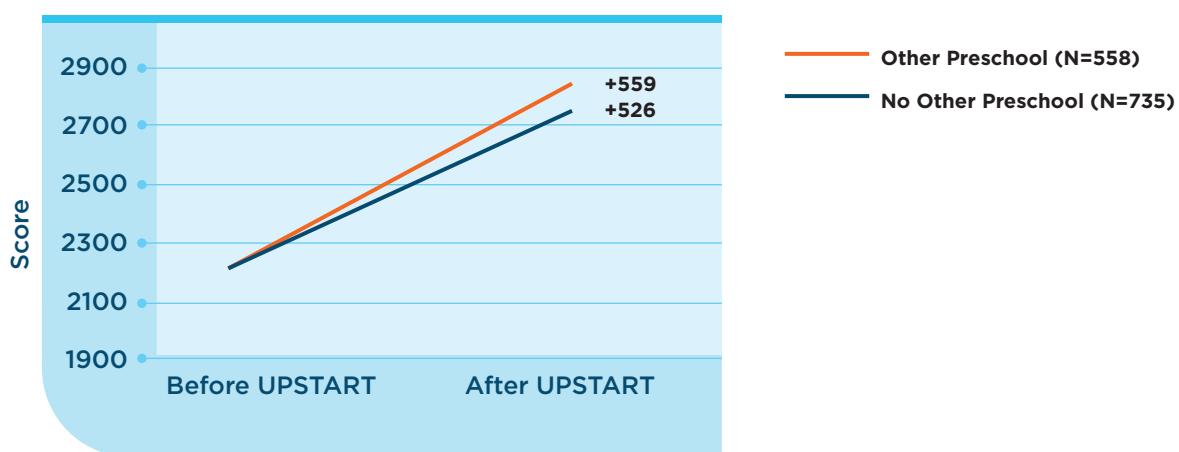


Figure 183: WACS Reading Score Gains Grouped by SES

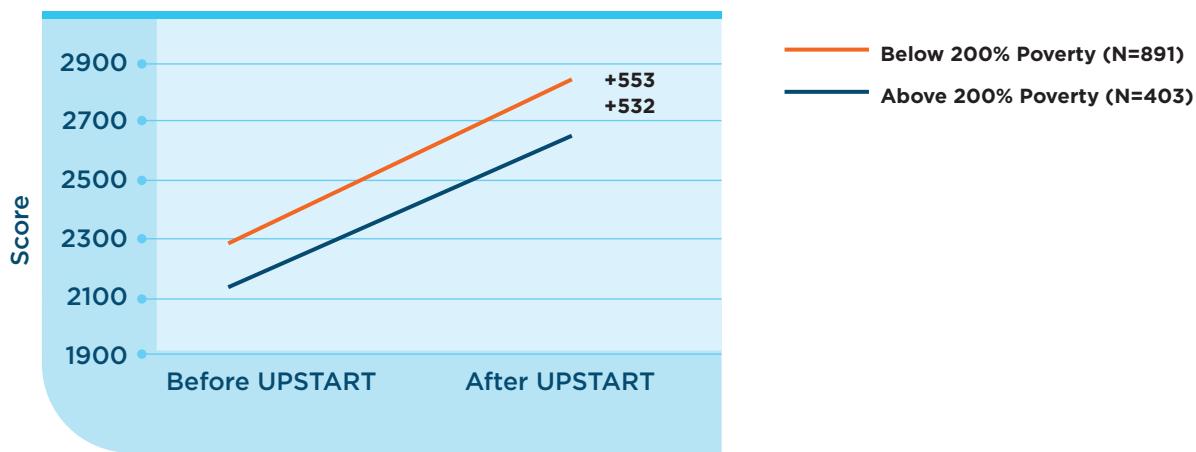


Figure 184: Year 6 WACS Reading Scores, Grouped by Subtests

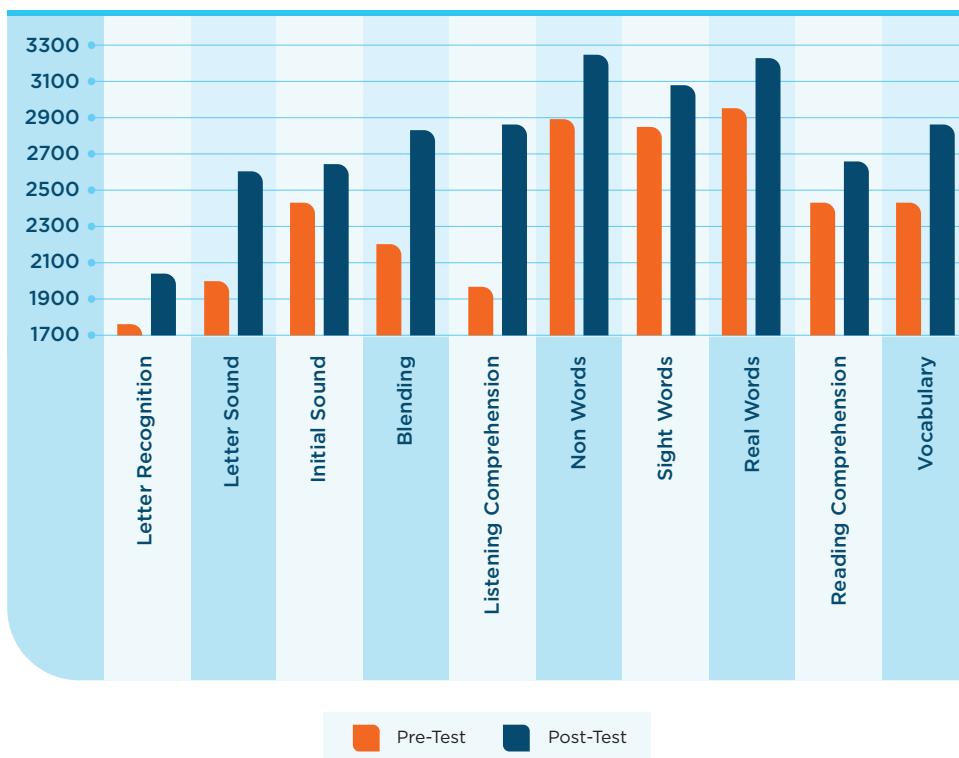


Table 82: WACS Year 6 Posttests

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 185: WACS Reading Score Gains Grouped by Race/Ethnicity

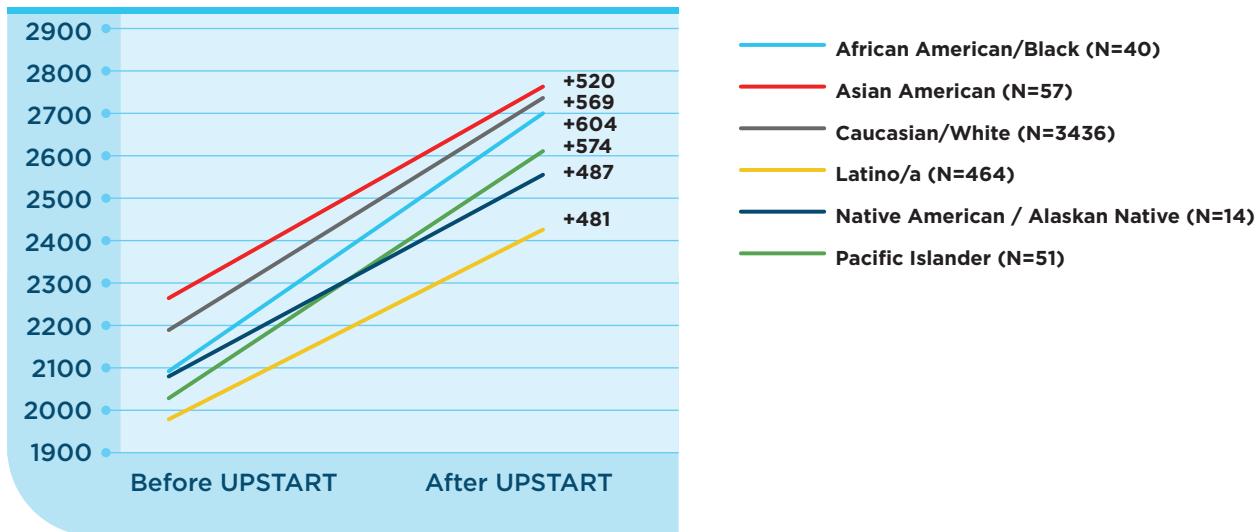


Figure 186: WACS Reading Score Gains Grouped by Other Preschool Attendance

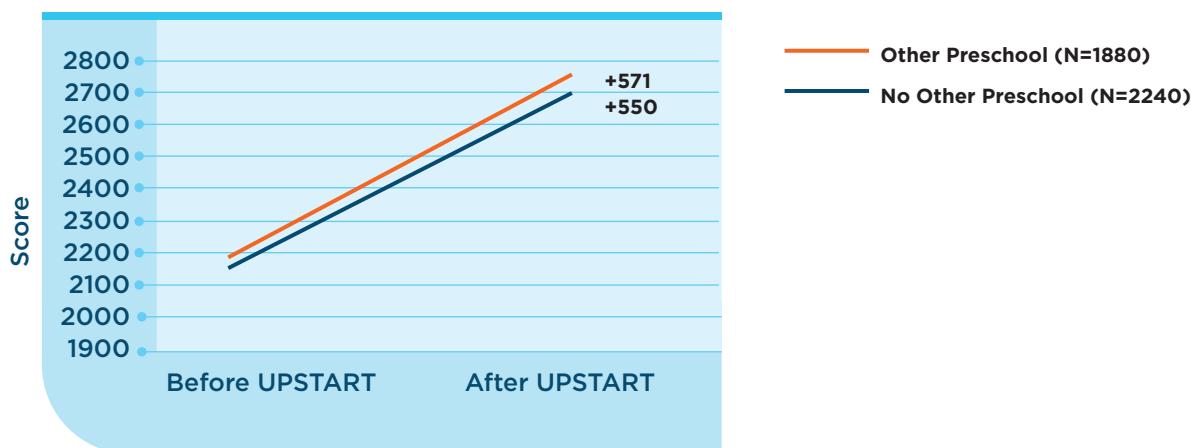
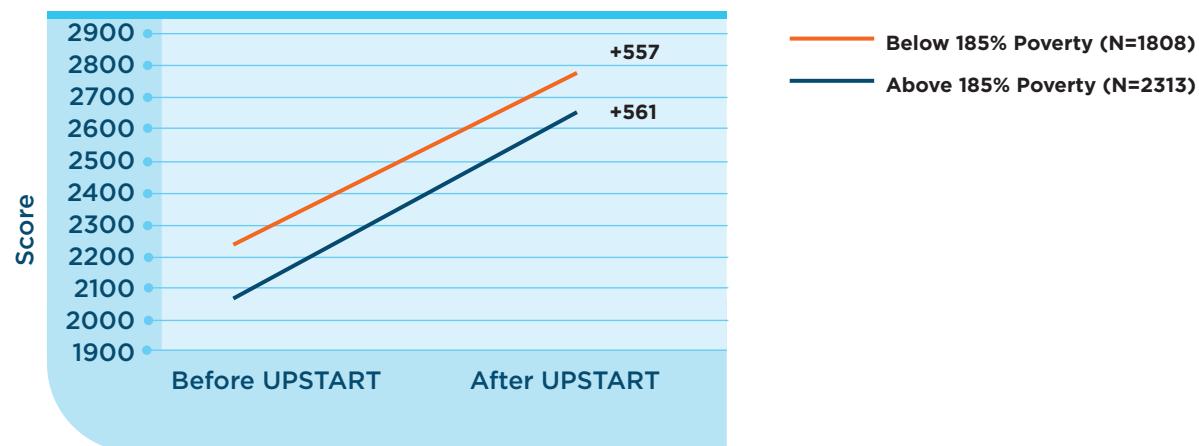


Figure 187: WACS Reading Score Gains Grouped by SES



Figure 188: 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 189: UPSTART Students & State Average DIBELS Scores, Grades 1 through 3

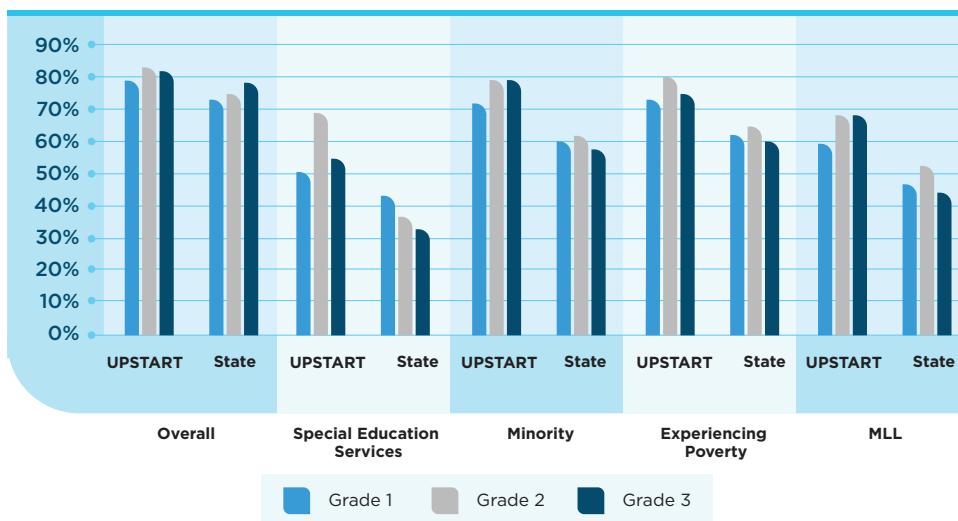


Figure 190: UPSTART Students & State Average SAGE Overall Scores, Grades 3 & 4

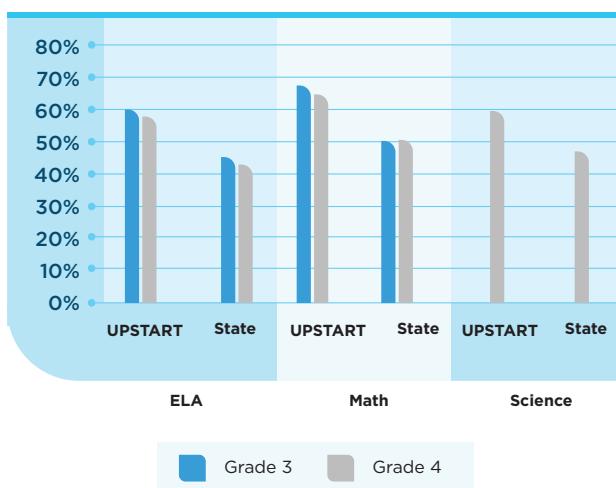


Figure 191: UPSTART Students & State Average SAGE SPED Scores, Grades 3 & 4

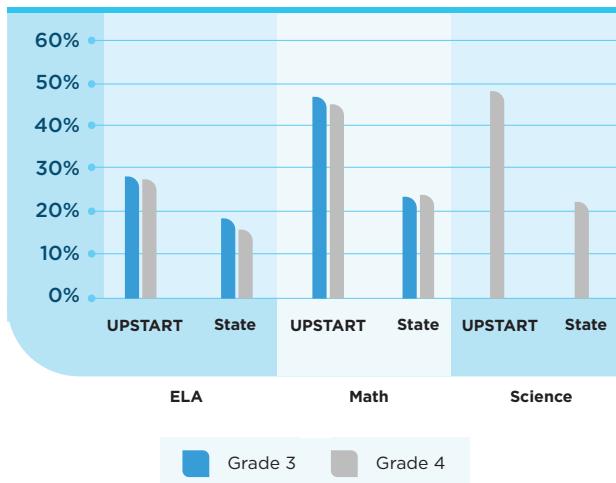


Figure 192: UPSTART Students & State Average SAGE Minority Scores, Grades 3 & 4

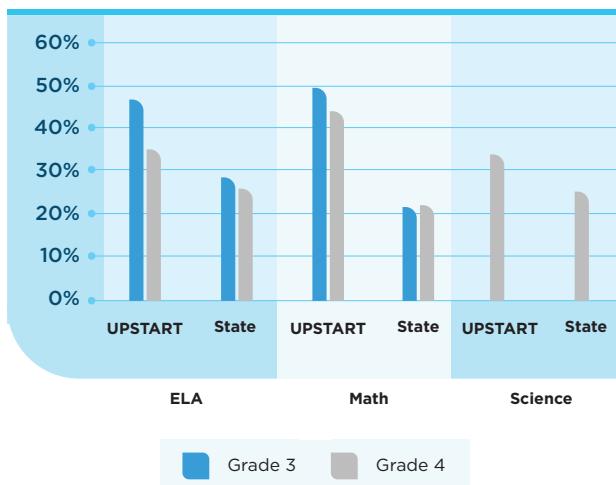


Figure 193: UPSTART Students & State Average SAGE Low Income Scores, Grades 3 & 4

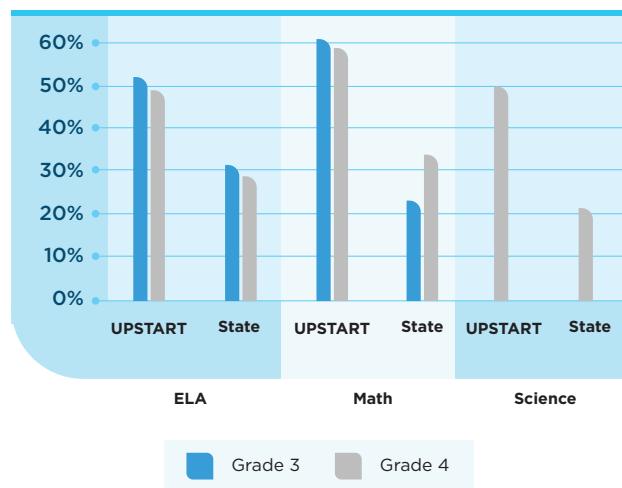
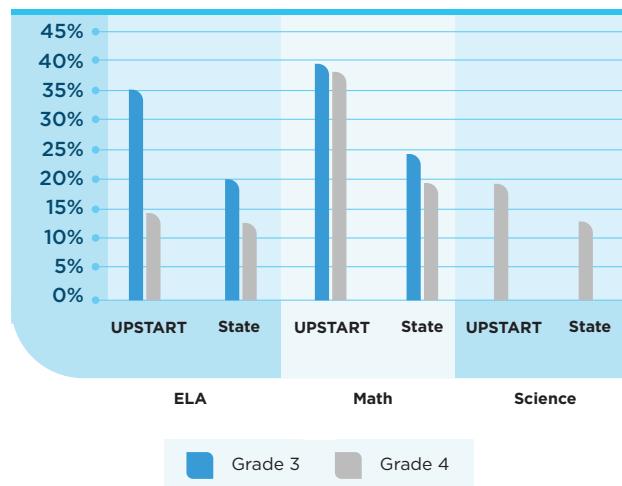


Figure 194: UPSTART Students & State Average SAGE MLL Scores, Grades 3 & 4



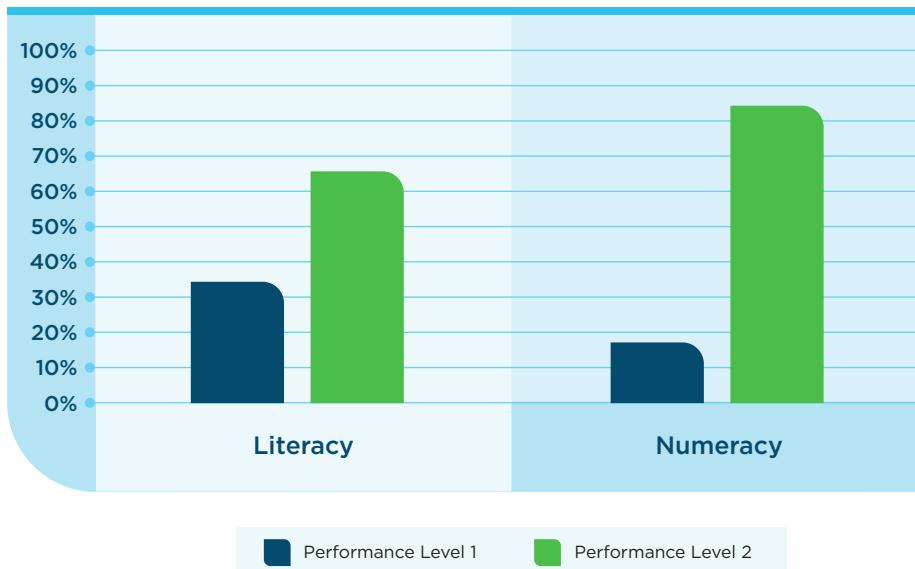
YEAR 13

Waterford Institute enrolled 13,404 preschoolers in its thirteenth year of operation during the 2021-22 program year (Evaluation and Training Institute, 2022b). Students were assessed using Utah's Pre-Kindergarten Entry and Exit Profile (PEEP). Students using the program grew from 63% at benchmark in literacy and 82% at benchmark in numeracy prior to the start of the program to 90% at benchmark for literacy and 93% at benchmark for numeracy prior to entering kindergarten. Students in higher tiers of parent support had a larger proportion of students with higher levels of literacy at the end of the program. 75% of students who were classified as 'in need of support' at the start of the program achieved the highest level of performance at the end of the program.

Comparable results were also found when examining students from the previous year. Most (77%) of Cohort 12 Upstart students ended the program at the requisite performance benchmark. At the completion of their kindergarten year, the majority of Cohort 12 students displayed sufficient

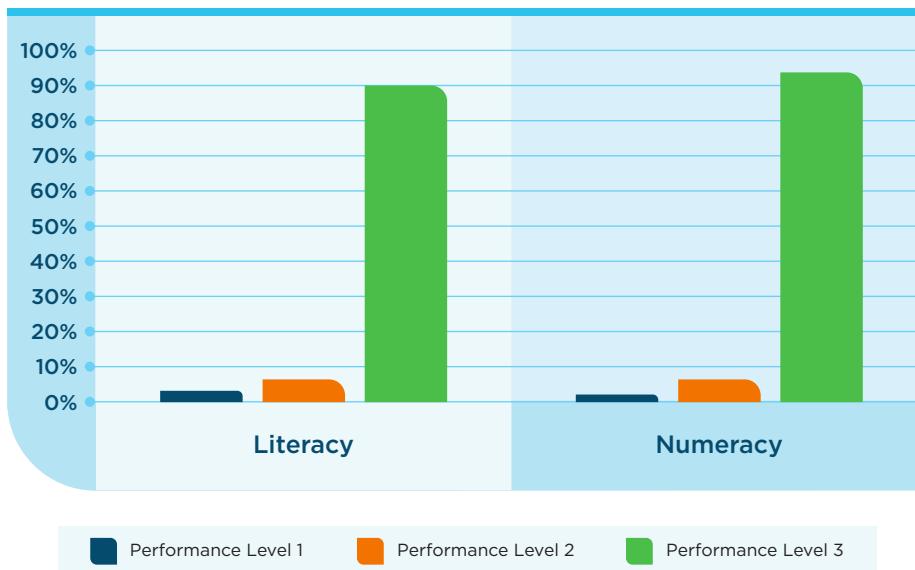
knowledge and skills (73%) as measured by KEEP Exit. The majority of the Cohort 12 Upstart sample entered kindergarten at Level 3 displaying sufficient numeracy knowledge and skills (87%). At the end of kindergarten, Cohort 12 students maintained those proficiency levels (89%).

Figure 195: Upstart Student Proficiency Level at PEEP Entry



Literacy and Numeracy- n=352

Figure 196: Upstart Student Performance Level at PEEP Exit

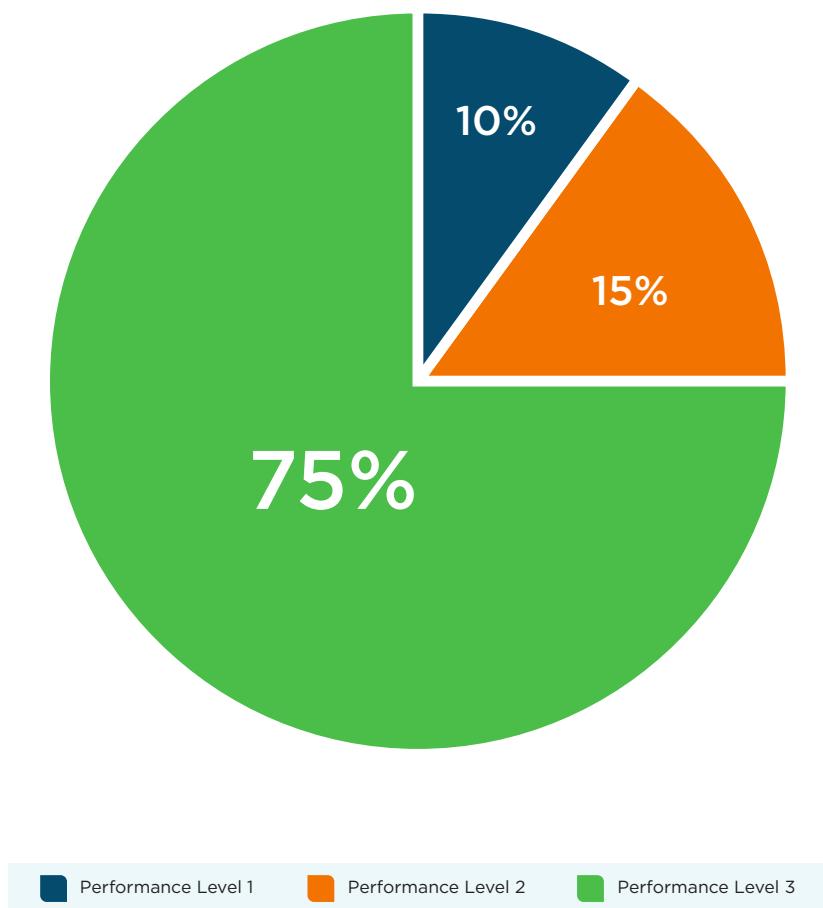


Literacy and Numeracy- n=321

Table 83: PEEP Exit Raw Scores for Literacy and Numeracy by Support Tier

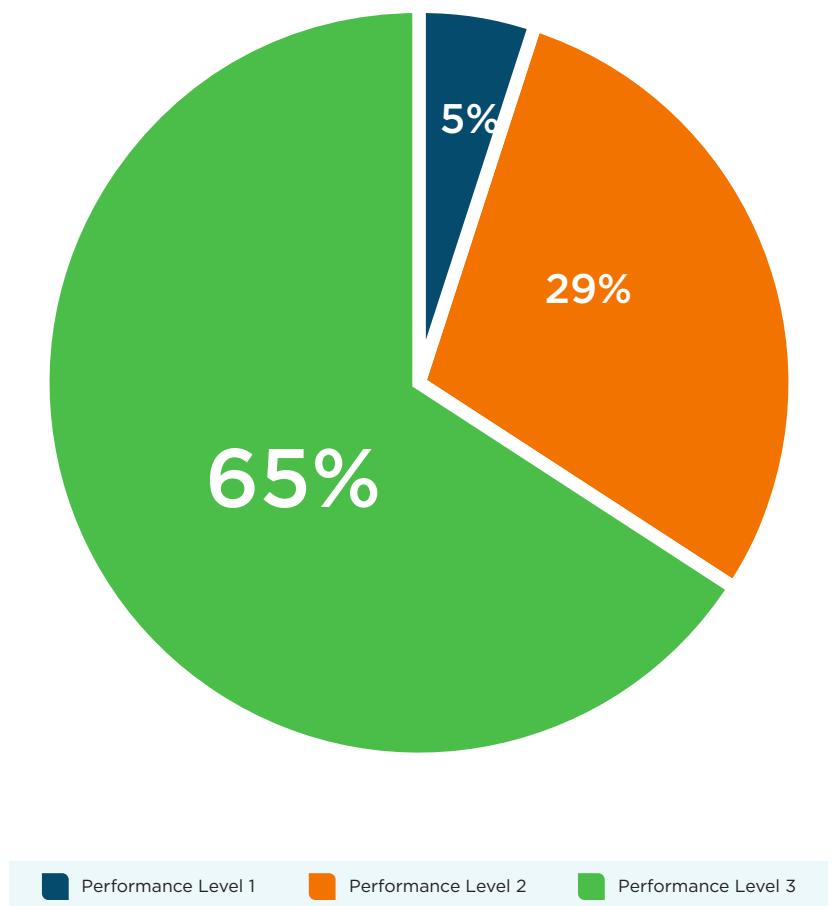
Domain	Tier	N	% at Level 3
Literacy	1	109	72%
	2	111	87%
	3	135	81%
Numeracy	1	109	80%
	2	111	86%
	3	135	82%

Figure 197: Kindergarten Readiness Among Those “Needing Support” at Entry for Literacy



Students with “support needed” for literacy at PEEP Entry ($n=111$)

Figure 198: Kindergarten Readiness Among Those “Needing Support” at Entry for Numeracy



Students with “support needed” for literacy at PEEP Entry (n=55)

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.
- Evaluation and Training Institute. (2019a). UPSTART Program Evaluation: Year 9 Program Results.
- Evaluation and Training Institute. (2019b). Utah High-Quality School Readiness Expansion (HQSR-E) Program Evaluation.
- Evaluation and Training Institute. (2020a). UPSTART Program Evaluation: Year 10 Program Results.
- Evaluation and Training Institute. (2020b). Indiana UPSTART Evaluation.
- Evaluation and Training Institute. (2020c). Mississippi UPSTART and Head Start Evaluation Kindergarten Impact Study.

Evaluation and Training Institute (2020d). EIR Year 1 summary outcomes and fidelity results

Evaluation and Training Institute (2021). EIR Great Plains Task Force Grant Year 2 Summary Report

Evaluation and Training Institute (2022a). EIR Year 3 Summary: Implementation and Preliminary Outcomes

Evaluation and Training Institute (2022b). UPSTART Program Evaluation: Year 13 Program Results

Evaluation and Training Institute (2023a). Improving early math skills: UPSTART program effects from pre-kindergarten to kindergarten.

Evaluation and Training Institute (2023b). Preparing Students for Kindergarten: Waterford Upstart Summer Learning Path Program Effects.

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., & Coordt, A. (2021). Impact of Upstart participation on Nevada pre-kindergarteners' school readiness.

Hobbs, L. J., & Overby, M. (2019a). Impact of UPSTART Reading Participation on Preschooler's Early Literacy Skills. Retrieved from
https://docs.wixstatic.com/ugd/796969_313351e185694f4c84ab8cc1eae59dfa.pdf

Hobbs, L. J., & Overby, M. (2019b). Impact of the UPSTART Program on Forestalling Summer Learning Loss. In ERIC. <https://eric.ed.gov/?id=ED605302>

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). Dosage effects of CAI on literacy skills. In *Smart Education and e-Learning 2019* (pp. 149-158). Springer, Singapore.

Shamir, H., Feehan, K., Pocklington, D., & Yoder, E. (2019b). Effects of long-term and early use of CAI on students' literacy skills. In *EdMedia+ Innovate Learning* (pp. 1406-1411). Association for the Advancement of Computing in Education (AACE).

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. 1425-1434). 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 9th 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 9th 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 11th 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 11th 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). Can CAI improve reading achievement in young ELL students? In J. Vopava (Ed.), *Proceedings of the Multidisciplinary Academic Conference on Education, Teaching, and E-Learning (MAC-ETEL 2018)* (pp. 229-237). Prague, Czech Republic: Academic Conferences Association.

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*, 5(3), 191-197.

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*, 5(3), 220-226.

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2019c). Game-based learning for young learners. *International Journal of Learning and Teaching*, 5(3), 206-212.

Shamir, H., Pocklington, D., Feehan, K., & Yoder, E. (2020). Using CAI with Fidelity: Impacts on Literacy Skills of Kindergarten Students Across Demographics. In D. Schmidt-Crawford (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 832-837). Online: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Pocklington, D., Yoder, E., & Feehan, K. (2019). Reinforcing second grade literacy skills using a computer-adaptive reading program. In *EdMedia+ Innovate Learning* (pp. 1412-1417). Association for the Advancement of Computing in Education (AACE).

Shamir, H., Pocklington, D., Yoder, E., & Feehan, K. (2020). Long-Term Effects of CAI in Early Education: Evidence from Three Districts. In D. Schmidt-Crawford (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 826-831). Online: Association for the Advancement of Computing in Education (AACE).

Shamir, H., Pocklington, D., & Yoder, E. (2023). Using computer-assisted instruction to fidelity to combat learning loss in early learners. *Proceedings of American Educational Research Association (AERA) 2023*.

Shamir, H., Pocklington, D., Yoder, E., Feehan, K., & Ortiz-Wood, C. (2022). Long-Term and Early Effects of Computer-Assisted Instruction in Low Socioeconomic Status Students. In *Smart Education and e-Learning-Smart Pedagogy* (pp. 77-86). Singapore: Springer Nature Singapore

- Shamir, H., Ortiz-Wood, C., Pocklington, D. & Yoder, E. (2022). Can Children Achieve Literacy Gains During the Summer?. In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 711-718). Online: Association for the Advancement of Computing in Education (AACE).
- Shamir, H., Yoder, E., Feehan, K., & Pocklington, D. (2019). Randomized controlled trial of kindergarten students using literacy technology. In *Smart Education and eLearning 2019* (pp. 243-252). Springer, Singapore.
- 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.
- Shamir, H., Yoder, E., Pocklington, D., Feehan, K., & Ortiz-Wood, C. (2022). Long-Term and Early Effects of Computer-Assisted Instruction in Early Childhood. In *Smart Education and e-Learning-Smart Pedagogy* (pp. 109-117). Singapore: Springer Nature Singapore
- Shamir, H., Yoder, E., & Pocklington, D. (2023a). Using CAI to Provide Early Literacy Instruction for All Learners. In *Smart Education and eLearning 2023* (In Press).
- Shamir, H., Yoder, E., & Pocklington, D. (2023b). Usage matters: Longitudinal benefits of computer-assisted instruction for early learners. To be presented at EDULEARN23 (2023).
- Shamir, H., Yoder, E., Pocklington, D., Wang, C., & Greene, E. (2023). Using technology to teach young learners math skills. To be presented at EDULEARN23 (2023).
- Suddreth, D., Thronsen, 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.

Young, J. W., & Tracey, D. H. (1999). An evaluation of the Waterford Early Reading Program: Newark, New Jersey. Research Compendium, 1998.