Secondary analysis of the Reading Recovery four-year i3 scale-up.

Robert M. Schwartz
Oakland University

Richard Lomax
The Ohio State University

Paper presented at the Society for the Scientific Study of Reading's Twenty-Fifth Annual Meeting in Brighton, United Kingdom, 2018

Abstract

This paper provides a secondary analysis of the May, Sirinides, Gray, and Goldsworthy (2016) evaluation of the Reading Recovery scale-up. We extend their findings to include analysis of the six subscales of the *Observation Survey (OS) of Early Literacy Achievement* (Clay, 2013) that the What Works Clearinghouse has previously used to access the beginning reading domains of Alphabets, Reading Fluency, and General Reading Ability. We also provide a sub-group analysis of students whose entry scores predict a severe reading difficulty at the end of first grade. This analysis includes the ITBS subscales reported by May et al. for their main analysis. The May et al. design was a randomized controlled trial with independent samples in each of the four years of the study. In total 3444 pairs of first grade students were matched on the fall OS Text Reading Level measure and then randomly assigned to the treatment or control condition. The secondary analysis presents descriptive statistics and Hedges's d effect size calculations for the total sample and sub-group on each of the subscales of the OS. The analysis for the Text Reading Level measure is calculated for both raw scores and scale scores. The scale scores provide an interval measure that is more appropriate for calculations of effect size (D'Agostino, Rodgers, & Mauck, 2018). The results showed medium to large effects on each of the OS measures for the total groups and severely at-risk subgroup. The appendix includes additional tables for each year of the scale-up, demonstrating replication of these findings over the four independent RCTs. The Institute of Education Sciences (IES, 2016) considers its highest evidence level as "an independent evaluation of a fully-developed education intervention with prior evidence of efficacy, when implemented by the end user under routine conditions" (IES, 2016, p.5). The replication of substantial effects on the ITBS and OS subscales demonstrates the ability of the Reading Recovery network to partner with schools to scale-up and implement an effective early intervention.

Purpose

We present a secondary analysis of the May, Sirinides, Gray, and Goldsworthy (2016) evaluation of the Reading Recovery scale-up. Our analysis extends their findings in two ways:

First, we examined the intervention's effects on the six sub-scales of the *Observation Survey (OS) of Early Literacy Achievement* (Clay, 2013). The May et al. analysis only reported gains on the OS total score and Iowa Tests of Basic Skills (ITBS) Word Reading and Comprehension subscales. The OS subscales have previously been used in the *What Works Clearinghouse's intervention report* (WWC, 2013) to assess their beginning reading domains:

Alphabetics:	Letter Identification, Word Test
Reading Fluency:	Text Reading Level (Raw and Scale Scores)
Reading Ability:	Concepts About Print, Hearing and Recording Sounds in Words, Writing Vocabulary

Second, we conducted a sub-group analysis for those students whose entry OS scores predicted a severe reading difficulty at the end of first grade (D'Agostino, Rodgers, & Mauck, 2018). There has been a persistent claim that the Reading Recovery intervention is not effective for the most at-risk beginning readers (Chapman, Greaney, & Tunmer, 2015; Cook, Rodes, & Lipsitz, 2017; Elbaum, 2000; Schwartz et al., 2009). This analysis tested whether the intervention is effective for

students predicted to be classified as reading disabled (D'Agostino, Rodgers, & Mauck, 2018; NCII, 2018).

Methods

The May et al. data includes four large, independent samples of students taught by different teachers from different schools during each of the four years of the scale-up. In total 3444 pairs of first grade students were matched on the fall OS Text Reading Level measure and then randomly assigned to the treatment or control condition. They used a three-level hierarchical linear model with students nested within matched pairs, and matched pairs nested within schools. Given the large sample and effect size in these analyses (see below May et al., Table 2.6, p. 42), for our secondary analyses we present descriptive statistics and Hedges's d effect size calculations consistent with WWC procedures (WWC, 2018, p.13). The analysis for the Text Reading Level measure is calculated for both raw scores and scale scores. The scale scores provide an interval measure that is more appropriate for calculations of effect size (D'Agostino, Rodgers, & Mauck, 2018).

Table 2.6. Impact Estimates on ITBS Subscales and OS Total Scores

Mid-Year Outcomes	Treatment Group (*n*=3444)	Control Group (*n*=3444)	Difference	Glass's Δ[d]	Cohen's *d*[e]
ITBS Reading Words Scale Scores					
Adjusted Mean	140.55	136.98	+3.57		
(Standard Error)	(0.19)	(0.17)	(0.20)	+0.43	+0.35
Mean Percentile Rank[a]	43	27	+16		
ITBS Comprehension Scale Scores					
Adjusted Mean	139.82	135.92	+3.90		
(Standard Error)	(0.21)	(0.18)	(0.21)	+0.43	+0.38
Mean Percentile Rank[a]	39	23	+16		
OS Total Raw Scores[bc]					
Adjusted Mean	495.37	451.88	+43.49		
(Standard Error)	(0.76)	(0.79)	(0.95)	+0.89	+0.99
Mean Percentile Rank	31	7	+24		

[a] Percentile ranks based on ITBS Grade 1 mid-year norms (Hoover et al., 2006).
[b] Percentile ranks based on U.S. Norms for OS Mid-Year (D'Agostino, et. al. 2012).
[c] treatment *n*=3371; control *n*=3322.
[d] Control SD: ITBS-C SD=8.98; ITBS-RW SD=8.23; OS-T SD=49.43
[e] Population SD, Level 6, Fall: ITBS-C SD=10.2; ITBS-RW SD=10.2; OS-T SD=43.96

Results

Observation Survey Subscale Analysis

Table 1 shows pre-test, post-test means, standard deviations, and effect size calculations for the Treatment (T) and Control (C) groups on the Observation Survey subscales pooled across the four year i3 study (n = 3439 per group). After the intervention the comparison of the treatment to control groups shows medium to large effects on each of Observation Survey subscales. These results were replicated in each of the four independent, Randomized Controlled Trials (RCT) conducted in the four-year scale-up evaluation (see Appendix, Tables 1A to 6A).

Sub-Group Analysis: Students Predicted to Need Intensive Intervention

Table 2 shows pre-test, post-test means, standard deviations, and effect size calculations for ITBS and Observation Survey measures pooled across the four year i3 study for students predicted to need intensive intervention (OS Fall Total Score < 419, n = 2712 per group). After the intervention the comparison of the treatment to control groups shows medium to large effects on each of Observation Survey and ITBS subscales. These results were replicated in each of the four independent, RCT conducted in the four-year scale-up evaluation (see Appendix, Tables 7A to 12A).

Conclusions

The Institute of Education Sciences (IES, 2016) considers its highest evidence level as "an independent evaluation of a fully-developed education intervention with prior evidence of efficacy, when implemented by the end user under routine conditions" (IES, 2016, p.5). The replication of substantial effects on the ITBS and OS subscales demonstrates the ability of the Reading Recovery network to partner with schools to scale-up and implement an effective early intervention. The sub-group analysis disconfirms claims that the intervention is not effective for the lowest performing first grade readers. The May et al. (2016) evaluation and this secondary analysis provide a model for educational effectiveness evidence and research-based early literacy interventions.

References

Chapman, J. W., Greaney, K. T., & Tunmer, W. E. (2015). Is Reading Recovery an effective early literacy intervention programme for children who most need literacy supports? In W. E. Tunmer & J. W. Chapman (Eds.), *Excellence and equity in literacy instruction: The case of New Zealand* (pp. 41-70). Basingstoke, UK: Palgrave Macmillan.

Clay, M. M. (2013). *An observation survey of early literacy achievement.* (3rd ed.). Portsmouth, NH: Heinemann.

Cook, P., Rodes, D. R., & Lipsitz, K. L. (2017). The Reading Wars and Reading Recovery: What Educators, Families, and Taxpayers Should Know. *Learning Disabilities: A Multidisciplinary Journal, 22*(2), 12-23. doi:https://doi.org/10.18666/LDMJ-2017-V22-I2-8391

D'Agostino, J. V., Rodgers, E., & Mauck, S. (2018). Addressing Inadequacies of the observation survey of early literacy achievement. *Reading Research Quarterly, 53*(1), 51-69. doi:10.1002/rrq.181

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure?: A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*(4), 605-619.

Institute of Education Sciences. (2016). *Building evidence: What comes after an efficacy study?* Retrieved from https:// ies.ed.gov/ncer/whatsnew/techworkinggroup/pdf/ BuildingEvidenceTWG.pdf

May, H., Sirinides, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. Retrieved from Philadelphia: http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale (see also Sirinides, P., Gray, A., & May, H. (201X). The Impacts of Reading Recovery at Scale: Results From the 4-Year i3 External Evaluation. Educational Evaluation and Policy Analysis, XX(X), 1-20. DOI: 10.3102/0162373718764828)

National Center on Intensive Intervention (NCII, 2018) https://charts.intensiveintervention.org/chart/academic-screening/observation-survey-early-literacy-achievement#title

Schwartz, R. M., Hobsbaum, A., Briggs, C., & Scull, J. (2009). Reading Recovery and evidence-based practice: A response to Reynolds and Wheldall (2007). *International Journal of Disability, Development and Education, 56*(1), 5-15.

What Works Clearinghouse, (2013). Reading Recovery Intervention Report. Retrieved from https://ies.ed.gov/ncee/wwc/Intervention/209

What Works Clearinghouse, (2018). Procedures Handbook, Version 4, https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf .

Table 1: Pre-test, Post-test Means, Standard Deviations, and Effect Size Calculations for the Treatment (T) and Control (C) Groups on the Observation Survey Sub-Scales Pooled Across the Four Year i3 Study (n = 3439 per group).

OS Measure	Letter ID		Word Test		CAP		HRSW		WV		TRL Scale Score	
Treatment/Control	T	C	T	C	T	C	T	C	T	C	T	C
Pre-test Mean	46.4	46.1	3.2	3.1	11.7	11.6	18	17.7	9	8.9	307.1	305.8
(Standard Deviation)	(8.1)	(8.5)	(3.1)	(3)	(3.5)	(3.6)	(9.7)	(9.7)	(6.2)	(6.3)	(76.5)	(76.8)
Post-test Mean	52.4	51.1	14.7	10.3	18.3	15.7	33.4	29.6	38.8	27.1	489.7	431
(Standard Deviation)	(2.9)	(4.9)	(4.6)	(5.3)	(3.2)	(3.3)	(5)	(7.4)	(14.1)	(12.9)	(44.3)	(70.7)
Effect Size	+0.32		+0.89		+0.80		+0.60		+0.87		+0.99	

Table 2: Pre-test, Post-test Means, Standard Deviations, and Effect Size Calculations for ITBS and Observation Survey Measures Pooled Across the Four Year i3 Study for Students Predicted to Need Intensive Intervention (OS Fall Total Score < 419, n = 2712 per group).

OS Measure	Letter ID		Word Test		CAP		HRSW		WV		TRL	
Treatment/Control	T	C	T	C	T	C	T	C	T	C	T	C
Pre-test Mean	45.6	45.1	2.5	2.4	11.2	11.1	16.0	15.6	7.4	7.3	0.8	0.8
(Standard Deviation)	(8.5)	(8.9)	(2.4)	(2.3)	(3.4)	(3.5)	(8.9)	(9)	(4.5)	(4.7)	(1)	(1.1)
Post-test Mean	52.2	50.9	14.2	9.5	18	15.4	32.9	28.9	37.0	25.4	9.8	4.7
(Standard Deviation)	(3.2)	(5.0)	(4.7)	(5.1)	(3.2)	(3.3)	(5.3)	(7.5)	(13.6)	(12)	(4.7)	(3.6)
Effect Size	+0.31		+0.96		+0.80		+0.62		+0.90		+1.22	

Measure	OS Total Score		ITBS Word		ITBS Comprehension		OS TRL Scale Score	
Treatment/Control	T	C	T	C	T	C	T	C
Pre-test Mean	360.4	358.9					295	293.5
(Standard Deviation)	(32.0)	(32.4)					(71.9)	(72.1)
Post-test Mean	490.7	444.9	139.0	134.9	139.8	135.9	484.5	424.3
(Standard Deviation)	(44.2)	(47.4)	(9.2)	(8.5)	(8.6)	(7.6)	(45.8)	(70.4)
Effect Size	+1.00		+0.46		+0.48		+1.01	

Appendix A: Tables by Year for Full Group and Subgroup Predicted to Need Intensive Intervention

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
TRL pre-test										
Mean	1	1	1	0.9	1	1	1.1	1.1	1	1
(Standard Deviation)	(1.3)	(1.1)	(1.2)	(1.1)	(1.2)	(1.1)	(1.2)	(1.3)	(1.2)	(1.2)
TRL post-test										
Mean	10.6	5.2	10.3	5.1	10.4	5.4	10.5	5.2	10.5	5.2
(Standard Deviation)	(4.8)	(3.7)	(4.7)	(4.2)	(4.9)	(3.9)	(4.9)	(4)	(4.9)	(4)
TRL pre-test scale score										
Mean	304.4	302.2	302.1	300.4	308.1	306.2	309.7	309.4	307.1	305.8
(Standard Deviation)	(74.9)	(75.4)	(75.3)	(75.2)	(75.5)	(76.3)	(77.9)	(78.2)	(76.5)	(76.8)
TRL post-test scale score										
Mean	492	435.5	488.7	425	489.5	433.4	489.6	431.3	489.7	431
(Standard Deviation)	(39.5)	(64.5)	(44.7)	(77.1)	(44.5)	(69.5)	(45.4)	(69.7)	(44.3)	(70.7)

Raw Score Effect Size	+1.46	+1.24	+1.28	+1.33	+1.33
Scale Score Effect Size	+0.88	+0.83	+0.81	+0.84	+0.83

Table 1A: Effect Size Calculations for Text Reading Level Raw Scores and Scale Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
LID pre-test										
Mean	46.7	46.6	46.4	46	46.7	46.5	46.1	45.7	46.4	46.1
(Standard Deviation)	(6.7)	(7.2)	(7.5)	(8.5)	(7.8)	(7.9)	(8.8)	(9.2)	(8.1)	(8.5)
LID post-test score										
Mean	52.4	51.4	52.3	51	52.5	51.4	52.3	50.8	52.4	51.1
(Standard Deviation)	(2.6)	(4.6)	(3.2)	(4.7)	(2.5)	(3.8)	(3.1)	(5.5)	(2.9)	(4.9)

Effect Size	.22	.28	.29	.27	.26

Table 2A: Effect Size Calculations for Letter Identification Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
OWT pre-test										
Mean	3	2.8	3.1	3	3.3	3.2	3.3	3.2	3.2	3.1
(Standard Deviation)	(3.1)	(2.8)	(3)	(3)	(3)	(3)	(3.1)	(3)	(3.1)	(3)
OWT post-test score										
Mean	14.7	10.4	14.9	10.4	14.8	10.5	14.6	10	14.7	10.3
(Standard Deviation)	(4.4)	(4.9)	(4.5)	(5.5)	(4.6)	(5.3)	(4.6)	(5.3)	(4.6)	(5.3)

Effect Size	+0.88	+0.82	+0.81	+0.87	+0.83

Table 3A: Effect Size Calculations for Ohio Word Test (OWT) Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
CAP pre-test										
Mean	11.5	11.5	11.4	11.5	11.9	11.8	11.7	11.5	11.7	11.6
(Standard Deviation)	(3.4)	(3.5)	(3.5)	(3.6)	(3.4)	(3.4)	(3.6)	(3.7)	(3.5)	(3.6)
CAP post-test score										
Mean	18.3	15.6	18.2	15.6	18.5	15.8	18.3	15.7	18.3	15.7
(Standard Deviation)	(2.9)	(3.3)	(3.2)	(3.4)	(3.1)	(3.4)	(3.2)	(3.3)	(3.2)	(3.3)

Effect Size	+0.79	+0.76	+0.79	+0.79	+0.79

Table 4A: Effect Size Calculations for Concept About Print (CAP) Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
HRSW pre-test										
Mean	17.6	17	17	17	18.5	18.2	18.3	17.9	18	17.7
(Standard Deviation)	(9.2)	(9.5)	(9.4)	(9.6)	(9.8)	(9.7)	(10)	(9.9)	(9.7)	(9.7)
HRSW post-test score										
Mean	33.6	30.1	33.5	29.2	33.5	30	33.2	29.4	33.4	29.6
(Standard Deviation)	(4.5)	(7.1)	(4.8)	(7.9)	(4.9)	(6.8)	(5.3)	(7.4)	(5)	(7.4)

Effect Size	+0.49	+0.54	+0.52	+0.51	+0.51

Table 5A: Effect Size Calculations for Hearing and Recording Sounds in Words (HRSW) Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	429	429	725	725	855	855	1430	1430	3439	3439
WV pre-test										
Mean	8.5	8.5	8.1	8.4	9.2	9.2	9.5	9.2	9	8.9
(Standard Deviation)	(5.6)	(6.1)	(5.5)	(5.8)	(6.2)	(6.5)	(6.6)	(6.5)	(6.2)	(6.3)
WV post-test score										
Mean	37.7	27.7	38.1	26.8	39.1	27.7	39.3	26.8	38.8	27.1
(Standard Deviation)	(12.9)	(12.5)	(13.2)	(13.3)	(14.4)	(12.6)	(14.6)	(12.9)	(14.1)	(12.9)

Effect Size	+0.80	+0.85	+0.90	+0.97	+0.91

Table 6A: Effect Size Calculations for Writing Vocabulary (WV) Scores across the Four Year i3 Study

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
TRL pre-test										
Mean	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
(Standard Deviation)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1.2)	(1)	(1.1)
TRL post-test										
Mean	10.2	4.9	9.9	4.6	9.7	4.9	9.7	4.6	9.8	4.7
(Standard Deviation)	(4.8)	(3.3)	(4.6)	(3.7)	(4.7)	(3.6)	(4.8)	(3.5)	(4.7)	(3.6)
TRL pre-test scale score										
Mean	298.1	294.2	290.9	290	296.5	294.1	295.2	294.7	295	293.5
(Standard Deviation)	(70.7)	(70.8)	(71.1)	(70.2)	(72.2)	(72.4)	(72.6)	(73.2)	(71.9)	(72.1)
TRL post-test scale score										
Mean	488.8	432.1	485	417.2	484	427.4	483.3	423.7	484.5	424.3
(Standard Deviation)	(40.9)	(63.3)	(46.2)	(77.2)	(45.9)	(68.3)	(46.8)	(69.6)	(45.8)	(70.4)

Raw Score Effect Size	+1.61	+1.43	+1.33	+1.46	+1.42
Scale Score Effect Size	+0.90	+0.88	+0.83	+0.86	+0.86

Table 7A: Effect Size Calculations for Text Reading Level Raw Scores and Scale Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
LID pre-test										
Mean	46.2	45.9	45.8	45.2	45.9	45.6	45	44.5	45.6	45.1
(Standard Deviation)	(6.7)	(7.5)	(7.7)	(8.6)	(8.1)	(8.2)	(9.5)	(9.7)	(8.5)	(8.9)
LID post-test score										
Mean	52.3	51.2	52.1	50.7	52.3	51.3	52.1	50.6	52.2	50.9
(Standard Deviation)	(2.8)	(5.1)	(3.4)	(5)	(2.7)	(3.5)	(3.4)	(5.6)	(3.2)	(5)

Effect Size	+0.22	+0.52	+0.28	+0.27	+0.26

Table 8A: Effect Size Calculations for Letter Identification Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
OWT pre-test										
Mean	2.4	2.3	2.5	2.5	2.5	2.5	2.5	2.4	2.5	2.4
(Standard Deviation)	(2.4)	(2.2)	(2.4)	(2.4)	(2.3)	(2.4)	(2.4)	(2.3)	(2.4)	(2.3)
OWT post-test score										
Mean	14.4	9.9	14.5	9.8	14.1	9.7	14	9.2	14.2	9.5
(Standard Deviation)	(4.4)	(4.8)	(4.6)	(5.3)	(4.8)	(5.1)	(4.7)	(5.1)	(4.7)	(5.1)

Effect Size	+0.94	+0.89	+0.86	+0.94	+0.92

Table 9A: Effect Size Calculations for Ohio Word Test (OWT) Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
CAP pre-test										
Mean	11.2	10.9	11.1	11	11.5	11.4	11.1	10.9	11.2	11.1
(Standard Deviation)	(3.3)	(3.4)	(3.5)	(3.5)	(3.3)	(3.3)	(3.5)	(3.6)	(3.4)	(3.5)
CAP post-test score										
Mean	18	15.4	18	15.3	18.2	15.5	17.9	15.3	18	15.4
(Standard Deviation)	(2.9)	(3.2)	(3.2)	(3.4)	(3.2)	(3.3)	(3.3)	(3.3)	(3.2)	(3.3)

Effect Size	+0.81	+0.81	+0.82	+0.79	+0.79

Table 10A: Effect Size Calculations for Concept About Print (CAP) Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
HRSW pre-test										
Mean	16.4	15.2	15.4	15.2	16.2	16.1	16	15.5	16	15.6
(Standard Deviation)	(8.6)	(8.6)	(8.6)	(8.9)	(8.9)	(9.1)	(9)	(9.1)	(8.9)	(9)
HRSW post-test score										
Mean	33.3	29.7	33.2	28.6	33	29.3	32.7	28.5	32.9	28.9
(Standard Deviation)	(4.8)	(7.1)	(4.9)	(8)	(5.2)	(6.8)	(5.6)	(7.6)	(5.3)	(7.5)

Effect Size	+0.51	+0.58	+0.54	+0.55	+0.53

Table 11A: Effect Size Calculations for Hearing and Recording Sounds in Words (HRSW) Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)

	2011-12		2012-13		2013-14		2014-15		Pooled	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
N	336	336	588	588	679	679	1109	1109	2712	2712
WV pre-test										
Mean	7.4	7.1	7	7.1	7.4	7.4	7.5	7.4	7.4	7.3
(Standard Deviation)	(4.2)	(4.5)	(4.4)	(4.4)	(4.7)	(4.8)	(4.6)	(4.8)	(4.5)	(4.7)
WV post-test score										
Mean	37	26.7	37	25.5	36.8	25.9	37.1	24.8	37	25.4
(Standard Deviation)	(13)	(11.9)	(12.8)	(12.8)	(13.8)	(11.4)	(14.1)	(11.8)	(13.6)	(12)

Effect Size	+0.87	+0.90	+0.96	+1.04	+0.97

Table 12A: Effect Size Calculations for Writing Vocabulary (WV) Scores across the Four Year i3 Study for Students Predicted to need Intensive Intervention (OS Total Score < 419)