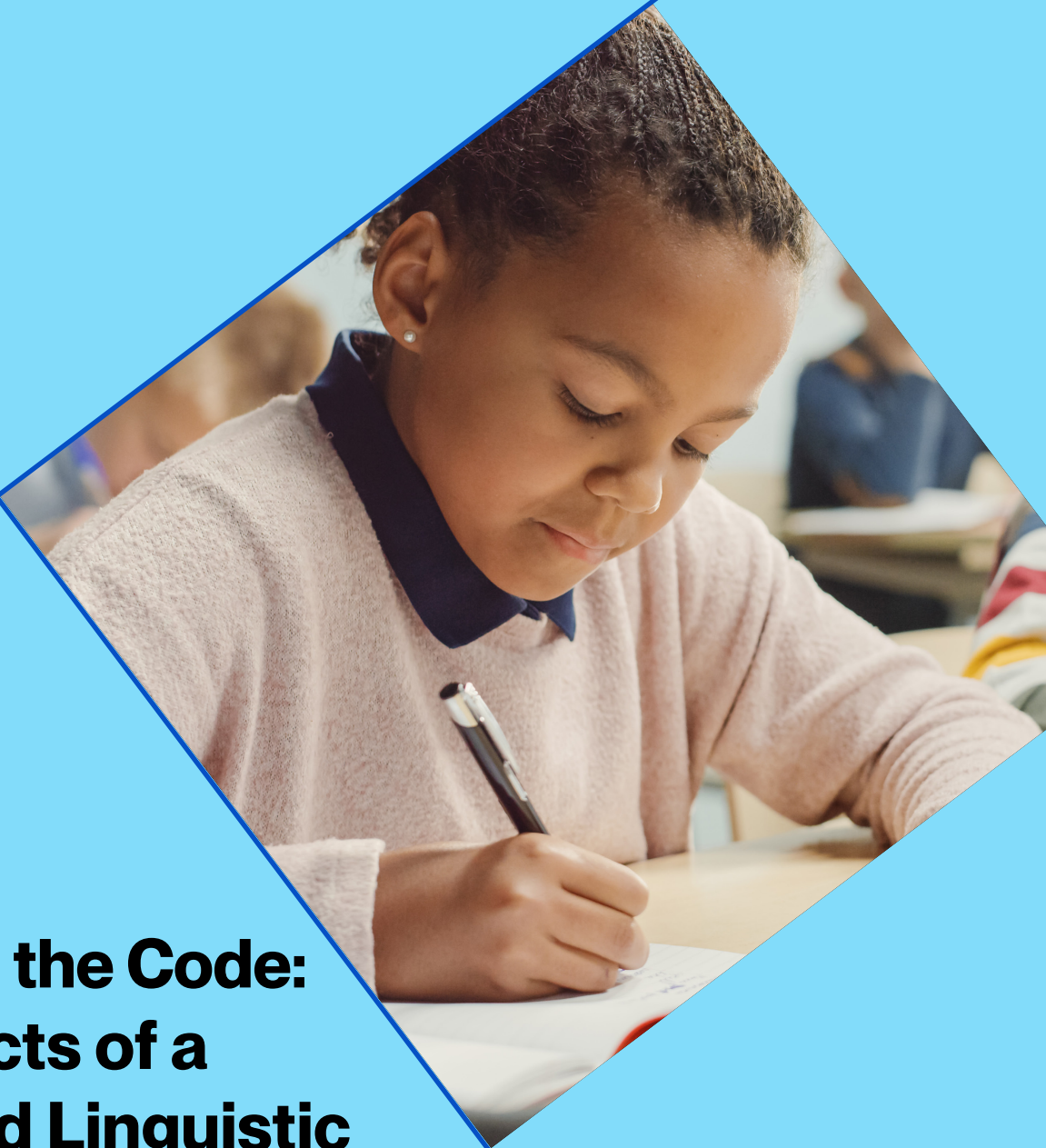




Unlocking the Code: The Impacts of a Structured Linguistic Phonics Program in Grades 4–8

Appendices

Dan Reynolds, Haiwen Wang, Madeline Coole,
Sara Rutherford-Quach, Qican Cao & Katherine Necochea Tinco



August 2026

Authors

Dan Reynolds
Haiwen Wang
Madeline Coole
Sara Rutherford-Quach
Qican Cao
Katherine Necochea Tinco

Suggested citation

Reynolds, D., Wang, H., Coole, M., Rutherford-Quach, S., Cao, Q., & Necochea Tinco, K. (2026). *Unlocking the Code: The Impacts of a Structured Linguistic Phonics Program in Grades 4–8. Appendices*. [Appendices to research report]. SRI. <https://www.sri.com/education-learning/project/evaluation-of-structured-linguistic-phonics-instruction-in-grades-4-8/>

Acknowledgments

The SRI research team would like to thank the district leaders, school leaders, and teachers for welcoming us into their schools, implementing high-quality literacy instruction, and partnering in advancing research. We appreciate the time and input from these great teams of educators. At SRI, we would also like to thank Charles Harding and Jennifer Medeiros for their support with improving the writing quality and graphic design. In addition, Andrew Pratulon's support in adapting the spelling assessments for remote, digital administration was invaluable, and Carrie Parker was instrumental in leading this project at an early stage. We would like to thank Maya Brunton and May Farenthold at the Stanford University Reading and Dyslexia Research Program for their support in administering the ROAR. Finally, this study would not be possible without the support of Rebecca Sutherland and the Reading Reimagined program at the Advanced Education Research and Development Fund (AERDF).

Appendix A. Technical Details

Sample Attrition

Exhibit A1 shows the number of students assigned to study teachers at the start of the fall semester and the number with outcome test scores who were therefore included in the impact analysis, presented separately for the structured linguistic phonics (SLP) and business-as-usual (BAU) groups. According to What Works Clearinghouse (WWC, 2022) standards, attrition in randomized controlled trials (RCTs) must be evaluated in terms of both the overall rate and the difference between treatment and comparison groups, and must be assessed at both the teacher and student levels for a teacher-level RCT. All recruited teachers participated in the study and are included in the impact analyses, except for one teacher assigned to the SLP (treatment) group in District A who withdrew because she was unable to complete the training course. In District A, where we conducted an RCT, overall attrition and differential attrition fell within the WWC's optimistic boundary for study samples for all outcomes. Because our overall analysis combined data from District A (with an RCT) and District B (with a quasi-experimental design), the findings from this combined analysis have the potential to meet WWC standards with reservations.

Exhibit A1. Student Numbers on Roster and in Final Analytic Samples

District/Group		ROAR Word		ROAR Sentence		Spelling	
	On roster	Analytic sample	Attrition rate	Analytic sample	Attrition rate	Analytic sample	Attrition rate
District A							
<i>SLP</i>	148	120	18.9%	122	17.6%	115	22.3%
<i>BAU</i>	138	110	20.3%	109	21.0%	96	30.4%
<i>Total</i>	286	230	19.6%	231	19.2%	211	26.2%
District B							
<i>SLP</i>	144	121	16.0%	116	19.4%	107	25.7%
<i>BAU</i>	85	80	5.9%	80	5.9%	70	17.6%
<i>Total</i>	229	201	12.2%	196	14.4%	177	22.7%

Baseline Equivalence Between SLP and BAU Students Across Cohorts

We examined baseline equivalence on each outcome measure. Rapid Online Assessment of Reading (ROAR) scores are age-standardized by the publisher, with a mean of 100 and a standard deviation of 15 (Yeatman et al., 2021). We scaled Elementary and Upper-Level Spelling Inventory (ESI and USI; Bear et al., 2023) item scores respectively using a two-parameter item response theory (IRT) model that accounts for student ability and item difficulty, and then standardized to a z-score with a mean of 0 and a standard deviation of 1.

As shown in Exhibit A2, all standardized differences between the SLP and BAU groups in the baseline measures were below 0.25, establishing baseline equivalence. Following WWC (2022) standards, we

assessed equivalence among students with no missing baseline or outcome data. In the impact analysis, for each outcome, baseline values were imputed for students without the baseline score using a logistic regression model that included the SLP treatment indicator, outcome score, and all covariates specified in the final impact model. Together with student background indicators, the imputed baseline score was adjusted statistically in the impact model to improve precision of the impact estimate.

Exhibit A2. Baseline Equivalence, Analytic Sample Combining All Three Cohorts

Group	ROAR Word	ROAR Sentence	Spelling
SLP			
<i>Mean</i>	95.66	84.52	-0.06
<i>SD</i>	13.01	15.40	0.98
<i>N</i>	241	238	222
BAU			
<i>Mean</i>	96.89	87.39	-0.04
<i>SD</i>	12.76	15.05	0.99
<i>N</i>	190	189	166
Standardized difference	-0.10	-0.19	-0.02

It is notable that for the two standardized measures, students at baseline were much stronger at ROAR Word (just under the standardized mean of 100) than at ROAR Sentence (about 1 standard deviation below the mean of 100). This suggests that students in both conditions were stronger decoders than they were efficient readers of basic sentences.

Study Outcome Measures

ROAR Word. This single-word verification measure assesses students' ability to determine if a letter string is a nonword (e.g., *concentent*) or a word (e.g., *overwhelm*). Accurate word recognition helps determine if students can match letter strings to their knowledge of words, and the ability to do so effortlessly and automatically is a predictor of students' reading development. The SLP approach systematically teaches students to match the 175 most common spellings of English words to the 44 most common sounds, which may improve students' single-word recognition.

ROAR Sentence. This sentence verification measure assesses students' ability to quickly read a sentence (e.g., "All cars are yellow") and note if it is true or false. SLP's systematic approach may improve students' ability to quickly read sentences with basic comprehension.

Spelling. These spelling assessments measure single word spelling. Students hear a spoken word, then hear the word in a sentence, then hear it spoken again, and then spell the word. The ESI has 25 items ranging from simple one-syllable word (e.g. *bed* and *ship*) to complex multisyllabic words (e.g. *civilize* and *opposition*). The USI has 32 items ranging in difficulty from *switch* and *smudge up* to *irresponsible* and *succession*.

Appendix B. District A-Only Impact Analysis

While our primary analysis combined samples from both districts to increase statistical power and generalizability, we also conducted a separate impact analysis for District A as a smaller-scale randomized controlled trial (RCT). This analysis included only 23 District A teachers, which limited statistical power and reduced the ability to detect the impacts of structured linguistic phonics (SLP) with precision. Nonetheless, the District A-only analysis serves as an important validity check, providing additional evidence on whether the combined-district findings are consistent with results from the smaller-scale RCT study.

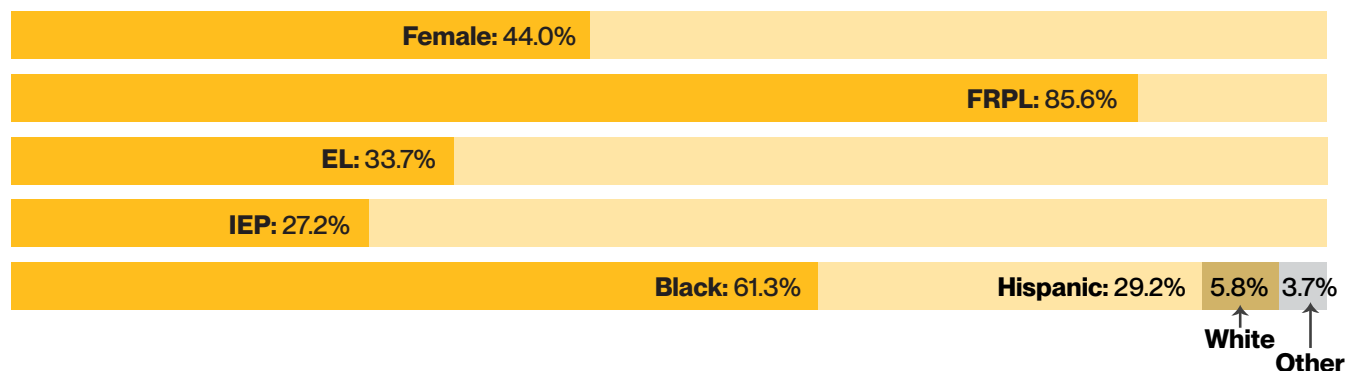
Sample Attrition

In District A, where we conducted the RCT, the overall and differential attrition for all outcome samples fell within the What Works Clearinghouse (WWC, 2022) optimistic boundary (Exhibit A1). Therefore, the findings for this analysis have the potential to meet WWC standards without reservations.

Student Demographic and Academic Background

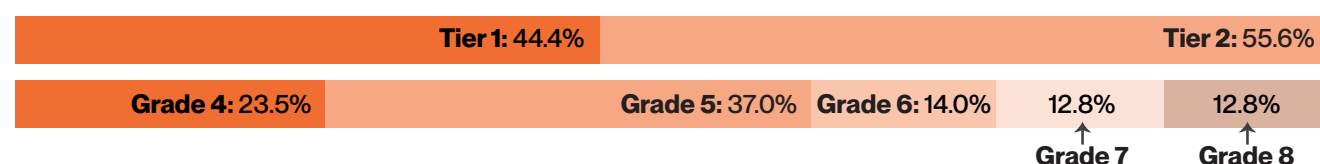
To provide some context of the study's student sample, we examined the demographic and academic characteristics of students included in the analysis. In the overall sample for District A, the vast majority of students were eligible for free or reduced-price lunch, about a third were English learners, and most identified as Black or Hispanic (Exhibit B1). There were more students in grade 4 and 5 than in grades 6–8 (Exhibit B2).

Exhibit B1. Demographic Information, District A Analytic Sample (n = 243)



Note. FRPL = free or reduced-price lunch; EL = English learner; IEP = Individualized Education Program.

Exhibit B2. School Background Information, District A Analytic Sample (n = 243)



Baseline Equivalence

As shown in Exhibit B3, standardized differences between the SLP (treatment) and business-as-usual (BAU) groups in the baseline measures were below 0.25 for all outcomes. We adjusted these baseline scores in the impact models to achieve better precision.

Exhibit B3. Baseline Equivalence, District A Analytic Sample

	ROAR Word	ROAR Sentence	Spelling
SLP			
<i>Mean</i>	92.11	80.02	-0.20
<i>SD</i>	12.26	13.67	0.94
<i>N</i>	120	122	115
BAU			
<i>Mean</i>	94.30	83.17	-0.09
<i>SD</i>	11.78	13.61	0.90
<i>N</i>	110	109	96
Standardized difference	-0.18	-0.23	-0.12

District A-Only Impact Findings

Exhibit B4 reports estimated impacts and effect sizes for the overall sample and separately for Tier I and Tier II instruction for District A. The study detected a statistically significant and positive overall impact on ROAR Word (effect size [ES] = 0.25, $p = .014$; improvement index = 9.9) and ROAR Sentence (ES = 0.25, $p = .020$; improvement index = 9.9), with no significant impact on the spelling assessments.

When we examine the two instructional modes separately, estimated impacts were positive and statistically significant on ROAR Word (ES = 0.27, $p = .043$; improvement index = 10.6) for Tier II but not for Tier I students. Conversely, impacts were positive and statistically significant on ROAR Sentence (ES = 0.57, $p = .000$; improvement index = 21.6) for Tier I but not for Tier II students. Moreover, the difference between the two tiers (treatment-by-tier interaction) was not statistically significant for ROAR Word but was statistically significant for ROAR Sentence, suggesting the impact on ROAR Sentence may be higher for Tier I versus Tier II students on ROAR Sentence but not on ROAR Word. Given the relatively small sample of teachers and students, these results should be considered suggestive rather than definitive.

Exhibit B4. Impact Estimates, District A Analytic Sample

Assessment	Estimated Impact	Standard Error	<i>p</i> Value	Student <i>N</i>	Teacher <i>N</i>	Effect Size
ROAR Word						
Overall impact	3.37*	1.37	.014	230	23	0.25*
Tier I	2.99	2.02	.140	106	5	0.24
Tier II	3.66*	1.81	.043	124	18	0.27*
Difference (Tier II – Tier I)	0.67	2.68	.802	230	23	0.03
ROAR Sentence						
Overall impact	3.64*	1.56	.020	231	23	0.25*
Tier I	6.96***	1.86	.000	104	5	0.57***
Tier II	1.74	1.66	.294	127	18	0.11
Difference (Tier II–Tier I)	-5.22	2.42	.031	231	23	-0.46
Spelling (ESI/USI)						
Overall impact	-0.10	0.07	.146	211	23	-0.11
Tier I	-0.08	0.10	.413	100	5	-0.09
Tier II	-0.12	0.09	.214	111	18	-0.12
Difference (Tier II – Tier I)	-0.03	0.14	.809	211	23	-0.02

Note. ESI = Elementary Spelling Inventory; USI = Upper-Level Spelling Inventory.

p* < .05. **p* < .001.

Appendix C. Spelling Feature Score Analysis

We further examined intervention effects on phonetic and orthographic subcomponents emphasized by the structured linguistic phonics (SLP) approach, focusing on the Within-Word Pattern and Letter Name spelling stages. These measures were operationalized using feature-level data from the Elementary and Upper-Level Spelling Inventories (ESI and USI) for Cohort 1 students only. All scoring guidelines come directly from the ESI and USI instruments (Bear et al., 2023)

The Within-Word Pattern spelling feature scoring measure was derived from ESI features assessing common long-vowel patterns and other vowel patterns, and from USI features assessing vowels, blends and digraphs, and complex consonants. For example, a student who spelled *trane* when asked to spell *train* would receive one of two possible points: one point for spelling the blend *tr* but no points for the incorrect spelling of the vowel sound /a/.

The Letter Name spelling feature scoring measure was derived exclusively from ESI features, including final consonants, short vowels, digraphs, and consonant blends. For example, a student who spelled *chip* when asked to spell *ship* would receive one out of two points on the Letter-Name feature score: one point for correctly spelling the vowel sound /i/ but no points for misspelling the consonant digraph *sh*.

We estimated feature scores for ESI and USI separately using item response theory (IRT), standardized to *z-scores* ($M = 0$, $SD = 1$) and averaged within construct to produce composite Letter Name and Within-Word Pattern scores.

Baseline Equivalence

As shown in Exhibit C1, the standardized difference between the SLP (treatment) and business-as-usual (BAU) groups in the baseline Within-Word Pattern measure was below 0.25, establishing baseline equivalence. However, the difference in the baseline Letter Name measure is larger than .25. Following What Works Clearinghouse (WWC, 2022) standards, we used propensity score weighting to assign weights to students in the BAU group based on their baseline scores. After weighting, the standardized difference in Letter Name was reduced to 0.04, meeting WWC (2022) equivalence standards. The same weights were used in the subsequent impact model.

Exhibit C1. Baseline Equivalence of Spelling Feature Scores, Cohort 1 Analytic Sample Combining Both Districts

Group	Within-Word Pattern	Letter Name (grades 4 & 5 only)	Letter Name (grades 4 & 5 only, weighted)
SLP			
Mean	-0.15	-0.21	-0.21
SD	1.05	1.02	1.02
N	172	137	137
BAU			
Mean	0.05	0.17	-0.25
SD	0.88	0.76	1.08
N	125	71	71
Standardized difference	-0.20	-0.40	0.04

Impact Findings

Exhibit C2 shows estimated impacts and effect sizes for the overall sample for both measures, and separately for Tier I and Tier II instruction, for Within-Word Pattern. We did not conduct separate analyses by instructional tier for Letter Name because only elementary-grade students were included in the analysis and the Tier II sample size was insufficient to yield reliable estimates. No statistically significant impacts were detected for either measure. These findings are consistent with the previously reported overall spelling outcome analysis and with the randomized controlled trial (RCT) analysis using District A data only (see Appendix B).

Exhibit C2. Impact Estimates for Spelling Feature Scores, Cohort 1 Analytic Sample Combining Both Districts

Measure	Estimated Impact	Standard Error	p Value	Student N	Teacher N	Effect Size
Within-Word Pattern						
Overall impact	-0.04	0.09	.65	309	18	-0.04
Tier I	-0.14	0.12	.26	188	10	-0.15
Tier II	0.07	0.13	.60	121	8	0.07
Difference (Tier I – Tier II)	0.20	0.17	.24	309	18	0.21
Letter Name (grades 4 & 5 only)						
Overall impact	-0.02	0.13	.89	220	16	-0.02



SRI Education, a division of SRI, is helping federal and state agencies, school districts, major foundations, nonprofit organizations, and international and commercial clients tackle some of the most complex issues in education to help students succeed. Our mission is **to reduce barriers and optimize outcomes for all children, youth, and families**. We do this by conducting high-quality research, supporting use of data and evidence, helping to strengthen state and local systems, and developing tools that improve teaching and accelerate and deepen learning. Our work covers a range of topics, including early learning and development, student behavior and well-being, teaching quality, digital learning, STEM and computer science, literacy and language arts, and college and career pathways.

SRI is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society.

Silicon Valley

(SRI Headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000
education@sri.com

Washington, DC

1100 Wilson Boulevard
Suite 2700
Arlington, VA 22209
+1.703.524.2053
www.sri.com/education-learning/

©2026 SRI International. SRI International is a registered trademark, and SRI Education is a trademark of SRI International. All other trademarks are the property of their respective owners.

STAY CONNECTED

