

Reading Between the Lines: A Systematic Comparative Analysis of Cognitive and Linguistic Intervention Programs Designed for Children with Dyslexia

A Dissertation Submitted in the Tradition of Doctoral Inquiry

Author: Carrie VanZant, Ph.D. **Discipline:** Public Health - Community Health

Date: May 21, 2026

Abstract

Developmental dyslexia is the most prevalent specific learning difference in school-age children, with conservative epidemiological estimates placing its global occurrence between five and ten percent of the population, and broader symptomatic estimates approaching twenty percent when sub-clinical presentations are included (Wagner et al., 2020; Snowling et al., 2020). Despite four decades of intensive research into Structured Literacy and multisensory phonics-based intervention, families, clinicians, and educators continue to face an unmediated marketplace of competing program brands, each advancing strong efficacy claims that are not always commensurate with the underlying peer-reviewed evidence base. The present dissertation undertakes a systematic comparative review of ten widely cited intervention programs — the Orton-Gillingham Approach, the Wilson Reading System, the Barton Reading and Spelling System, the Lindamood Phoneme Sequencing (LiPS) Program, the Nessy Learning Program, the Davis Method, Beyond Decodables, Equipped for Reading Success, Evidence-Based Literacy Instruction (EBLI), and the Pavlidis Method — applying PRISMA 2020 reporting standards, the GRADE framework for certainty of evidence, and an explicit attention to cultural and linguistic adaptation across orthographic depths. Findings reveal a pronounced asymmetry between marketing claims and the strength of the supporting research: certain widely adopted programs lack studies meeting Institute of Education Sciences What Works Clearinghouse evidence standards, whereas other less commercially visible approaches demonstrate replicable, modest-to-moderate effects on foundational decoding skills. The dissertation concludes with a parent- and educator-facing decision framework intended to support evidence-informed program selection for the neurodivergent learner, and identifies methodological priorities for the next generation of efficacy research — including the need for active-comparator randomized trials, attention to non-English orthographies, and standardized outcome reporting across the cognitive-linguistic domains of phonological awareness, decoding, fluency, spelling, and reading comprehension.

Keywords: dyslexia, structured literacy, Orton-Gillingham, multisensory instruction, systematic review, cultural adaptation, neurodivergence, reading intervention, evidence-based practice

Acknowledgements

The author acknowledges the families, educators, and clinicians whose lived expertise of dyslexia continues to outpace the published literature, and whose persistent questions — *which program, for which child, in which language, and on whose evidence?* — animate the present inquiry.

Table of Contents

1. Introduction
 2. Literature Review
 3. Methodology
 4. Results
 5. Discussion
 6. Conclusion
 7. References
 8. Appendices
-

Chapter 1 — Introduction

1.1 Background and Context

Dyslexia is a neurobiologically based specific learning disorder whose defining feature is unexpected difficulty with accurate and fluent word recognition, decoding, and spelling, occurring in individuals with otherwise typical reasoning, motivation, and instructional opportunity (Lyon, Shaywitz, & Shaywitz, 2003; International Dyslexia Association [IDA], 2018). The condition is heritable, persistent across the lifespan, and traceable to atypical functioning of the left perisylvian reading network — including the inferior frontal gyrus, the temporoparietal junction, and the occipitotemporal cortex (Norton, Beach, & Gabrieli, 2015; Ozernov-Palchik & Gaab, 2016). When dyslexia is identified early and met with evidence-aligned, sufficiently intensive instruction, the developmental trajectory of the affected reader can be meaningfully altered; when identification is delayed, or when instruction is poorly matched to the underlying

processing profile, the cumulative cost — academic, psychosocial, and economic — is substantial (Catts et al., 2024).

The contemporary policy landscape reflects this urgency. Between 2014 and 2025, all fifty states in the United States enacted some form of dyslexia legislation, most mandating screening, teacher preparation, or designated intervention pathways (National Center on Improving Literacy [NCIL], 2024). Parallel reforms have emerged in the United Kingdom, Australia, the Republic of Ireland, Singapore, and several Nordic systems. Yet the proliferation of policy has not been matched by clarity regarding which interventions, delivered under which conditions, produce reliable improvement for which children. Families are left to navigate a fragmented marketplace in which programmatic claims often outpace the strength of the supporting evidence — a phenomenon that the present dissertation terms the *evidence-marketing gap*.

1.2 Problem Statement

The problem under investigation is twofold. First, the published evidence on the comparative efficacy of dyslexia intervention programs is unevenly distributed: some programs (e.g., the Lindamood Phoneme Sequencing Program) have been subjected to multiple independent randomized evaluations, whereas others (e.g., the Wilson Reading System, the Davis Method) have generated substantial commercial reach without producing studies that meet the Institute of Education Sciences (IES) What Works Clearinghouse evidence standards (WWC, 2010; Stevens et al., 2021). Second, the available evidence has been generated predominantly within English-language, alphabetic-orthography contexts, leaving practitioners working with transparent orthographies (e.g., Spanish, Italian, Greek, Finnish) and non-alphabetic writing systems (e.g., logographic and abugida systems) with limited adaptation guidance (Verhoeven & Perfetti, 2022; Diamanti et al., 2023).

The downstream consequence is that the very stakeholders who carry the most direct responsibility for a child's literacy trajectory — parents and teachers — are required to make consequential program-selection decisions on the basis of marketing claims, brand recognition, and informal social proof, rather than transparent, comparable evidence summaries. This dissertation seeks to mitigate that asymmetry.

1.3 Research Aims and Questions

The dissertation is organized around three overarching research questions:

RQ1. What is the current state of peer-reviewed and grey-literature evidence regarding the efficacy of ten widely cited dyslexia intervention programs (Orton-Gillingham, Wilson Reading System, Barton Reading & Spelling System, LiPS, Nessy, Davis Method, Beyond Decodables, Equipped for Reading Success, EBLI, and the Pavlidis Method) for foundational reading outcomes in school-age children?

RQ2. To what extent do these programs demonstrate cultural and linguistic adaptability — operationalized as documented implementation in non-English orthographies, attention to morphophonemic versus shallow orthography demands, and culturally sustaining instructional materials?

RQ3. What decision-relevant differences emerge when programs are compared along a transparent matrix of (a) theoretical grounding, (b) instructional dosage, (c) reported effect sizes, (d) certainty of evidence, and (e) accessibility (cost, training requirements, and delivery format)?

1.4 Significance of the Study

Dyslexia is not a marginal concern. Synthesizing across the major epidemiological estimates, between five and ten percent of children meet conservative diagnostic criteria for dyslexia, and between fifteen and twenty percent display reading difficulties of clinical significance (Wagner et al., 2020; Snowling et al., 2020; Catts et al., 2024). In a global population of school-age children numbering well over a billion, the absolute affected population is therefore on the order of one hundred fifty to two hundred million learners. The economic literature converges on the finding that under-identified and under-treated dyslexia contributes substantially to lifetime earnings loss, mental-health comorbidity, and incarceration risk (Moody et al., 2000; McLaughlin et al., 2014).

The significance of the present dissertation, therefore, lies in three contributions. **First**, it offers a transparent, side-by-side synthesis of the evidence underlying ten of the most widely adopted programs, in a single document, with consistent appraisal criteria. **Second**, it foregrounds the question of cultural and linguistic adaptation, which has been chronically under-addressed in English-dominated efficacy literatures. **Third**, it translates findings into a decision framework written for the non-specialist — the parent, the classroom teacher, the school administrator — who is most often the actual decision-maker in real-world program selection.

1.5 Operational Definitions

Throughout the dissertation, the following operational definitions are adopted.

Dyslexia. A specific learning disorder of neurobiological origin, characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities, typically resulting from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities (IDA, 2018; American Psychiatric Association [APA], 2022).

Structured Literacy. An approach to reading instruction characterized by explicit, systematic, cumulative, and diagnostic teaching of phonology, sound–symbol association, syllable structure, morphology, syntax, and semantics, typically incorporating multisensory engagement (IDA, 2019).

Multisensory instruction. Teaching that simultaneously engages the visual, auditory, and kinesthetic-tactile modalities to reinforce sound–symbol learning and word-level analysis (Birsh & Carreker, 2018).

Cultural and linguistic adaptation. The intentional modification of an intervention's content, delivery, examples, and assessment to align with the linguistic features (e.g., orthographic depth, syllabic structure, morphology) and the cultural-experiential frames of the learner population (Bernal et al., 2009; Verhoeven & Perfetti, 2022).

Effect size. A standardized measure of the magnitude of an intervention's impact, reported here primarily as Cohen's *d* or Hedges' *g*, with conventional benchmarks of small (≈ 0.20), medium (≈ 0.50), and large (≈ 0.80) effects, while recognizing that, in well-implemented reading interventions, *d* values of 0.25–0.40 are educationally meaningful (Lipsey et al., 2012).

1.6 Structure of the Dissertation

Chapter 2 reviews the theoretical foundations of dyslexia intervention, situating the ten programs within a coherent conceptual landscape. Chapter 3 details the PRISMA-aligned systematic review methodology used to identify, appraise, and synthesize the evidence. Chapter 4 presents results, including a program-by-program evidence profile and a master comparative matrix. Chapter 5 discusses the findings in relation to the research questions, the evidence-marketing gap, and the imperatives of cultural-linguistic adaptation. Chapter 6 concludes with implications, limitations, and a research agenda. Appendices contain the full search strategy, the PRISMA flow diagram, data-extraction tables, and the parent-educator decision tool.

Chapter 2 — Literature Review

2.1 The Cognitive Architecture of Reading and the Phonological-Core Deficit

A coherent understanding of dyslexia intervention requires, at minimum, three converging models of reading. The **Simple View of Reading** (Gough & Tunmer, 1986; Hoover & Tunmer, 2018) frames reading comprehension (R) as the product of decoding (D) and language comprehension (L): $R = D \times L$. The model's multiplicative form is consequential: any reader whose decoding skill approaches zero will, by definition, produce a reading-comprehension product approaching zero, irrespective of oral language capacity. For dyslexic learners, whose principal challenge lies in the decoding term, the implication is that intervention must directly remediate decoding skills rather than relying on enrichment of language comprehension alone.

The **dual-route model** of word recognition (Coltheart et al., 2001) further specifies that skilled readers access words through both a *lexical route* (direct recognition of familiar orthographic forms) and a *sublexical route* (grapheme-to-phoneme conversion). Dyslexic readers tend to show selective impairment of the sublexical route, with downstream consequences for the development of the lexical route, since lexical orthographic representations are themselves constructed in large part through repeated successful sublexical decoding (Share, 1995; Castles et al., 2018).

Underlying both models is the **phonological-core deficit hypothesis** — the most empirically supported single-factor account of developmental dyslexia (Stanovich, 1988; Snowling, 2000; Vellutino et al., 2004; Snowling et al., 2020). Phonological awareness, especially at the phoneme level, is the strongest pre-school predictor of later word reading across alphabetic orthographies (Melby-Lervåg, Lyster, & Hulme, 2012). The phonological-core account does not exclude additional contributing factors — naming-speed deficits (Wolf & Bowers, 1999), orthographic-processing weaknesses (Castles & Nation, 2006), or working-memory limitations (Peng et al., 2018) — but it does establish that the design of any dyslexia intervention must, at minimum, address the phonological substrate.

2.2 Structured Literacy as the Convergent Instructional Model

Across the past four decades, a remarkable convergence has emerged in the international research community: dyslexic learners benefit most reliably from instruction that is **explicit** (skills are taught directly rather than inferred), **systematic** (skills are sequenced from simpler to more complex), **cumulative** (each new element builds upon previously mastered material), **diagnostic** (instruction is adjusted on the basis of continuous progress monitoring), and **multisensory** (visual, auditory, and kinesthetic-tactile modalities are engaged simultaneously) (IDA, 2019; Foorman et al., 2016; Solari et al., 2020). This convergence is institutionalized in the *Structured Literacy* framework, adopted by the International Dyslexia Association and operationalized by the Knowledge and Practice Standards for Teachers of Reading (IDA, 2018).

It is important, however, to note what the convergence does and does not establish. The convergence establishes the *category* of instruction most likely to benefit dyslexic learners. It does *not* establish that all programs within this category are equally effective, nor that the multisensory component is the active ingredient (Stevens et al., 2021). This distinction will prove consequential when, in Chapter 4, the evidence base for individual programs is appraised.

2.3 Orthographic Depth and the Cross-Linguistic Question

A second body of literature, frequently siloed from the English-language intervention research, concerns the role of orthographic transparency in the manifestation and remediation of dyslexia. Orthographies vary along a continuum from *shallow* (transparent, with near one-to-one grapheme–phoneme correspondence; examples include Finnish, Italian, Spanish, and Modern Greek) to *deep* (opaque, with substantial irregularity; English is the canonical exemplar) (Seymour, Aro, & Erskine, 2003; Frost, 2012). The Granularity and Transparency Hypothesis

(Ziegler & Goswami, 2005) predicts — and a substantial empirical literature confirms — that dyslexia in shallow orthographies manifests predominantly as a *fluency* deficit, with relatively preserved accuracy, whereas in deep orthographies it manifests as an *accuracy and fluency* deficit (Landerl et al., 2013; Diamanti et al., 2023).

The implication for intervention is direct: a program developed and validated in English may require non-trivial adaptation when transposed to a shallower orthography, because the relative emphasis on grapheme–phoneme decoding versus reading-rate development must shift. The Pavlidis Method, originating in Greek-language research and emphasizing biological (oculomotor) markers over reading-based diagnostics, is one of the few programs reviewed in the present dissertation that explicitly originates from a shallow-orthography context (Pavlidis, 2011); EBLI and Beyond Decodables, by contrast, are deeply embedded in English-language assumptions.

2.4 The History and Lineage of Multisensory Structured Language Approaches

The clinical lineage from which most contemporary programs descend is traceable to the collaborative work of the neurologist Samuel T. Orton and the educator Anna Gillingham in the 1920s–1940s. Orton's clinical observation that dyslexic learners showed atypical hemispheric specialization (Orton, 1937) led him, in partnership with Gillingham, to develop a remedial method emphasizing simultaneous association of visual, auditory, and kinesthetic information (Gillingham & Stillman, 1936/1997). The Orton-Gillingham Approach (OGA) thereafter functioned less as a single program than as a *family* of methods, from which the Wilson Reading System, the Barton Reading and Spelling System, the Slingerland Approach, the Alphabetic Phonics method, and others were derived (Ritchey & Goeke, 2006).

A second lineage traces to the Lindamood-Bell organization, whose Lindamood Phoneme Sequencing Program (originally the Auditory Discrimination in Depth program) emphasized the *articulatory* features of phonemes as a route into phonological awareness (Lindamood & Lindamood, 1998). LIPS is conceptually distinct from the OGA lineage in its prioritization of articulatory awareness, though it shares the family-level commitment to explicit and multisensory instruction.

A third lineage — exemplified by the Davis Method — departs more substantively from the Structured Literacy paradigm, framing dyslexia as a perceptual and cognitive *gift* with associated processing differences, and emphasizing imagery- and clay-modeling-based strategies (Davis & Braun, 2010). The Davis Method's positioning vis-à-vis the Structured Literacy consensus is contested; its evidence base, as developed in Chapter 4, is modest.

A fourth lineage — increasingly visible in contemporary practice — emphasizes the *science of reading* synthesis articulated by Kilpatrick (2015) and operationalized in resources such as *Equipped for Reading Success*, which emphasize the centrality of phoneme proficiency

(advanced phonological-awareness manipulations) for the development of orthographic mapping and automatic word recognition (Ehri, 2014; Kilpatrick, 2015).

A fifth lineage, only now consolidating, emphasizes *digital and gamified* multisensory instruction, exemplified by the Nessy Learning Program. The lineage raises distinct methodological questions about engagement, dose-response, and the generalization from screen-based practice to authentic reading.

2.5 The Evidence-Marketing Gap

A recurring theme across the literature is the divergence between the strength of programmatic marketing claims and the strength of the underlying peer-reviewed evidence. The phenomenon is not unique to dyslexia intervention; it is endemic to educational and clinical sectors in which commercial actors compete for institutional and consumer adoption. In the dyslexia intervention space, however, the gap is particularly consequential, because (a) the affected population is large, (b) the decision-makers are typically not research-literate specialists, and (c) the cost of incorrect program selection is measured not only in dollars but in months and years of a child's developmental window.

Three patterns of evidence-marketing gap recur. First, programs may cite *internal* evaluations — pre-post studies conducted by the program developer or affiliated trainers — without independent replication. Second, programs may cite *single-arm* studies in which observed gains cannot be distinguished from maturation or regression-to-the-mean effects. Third, programs may invoke *brand alignment* with the Orton-Gillingham approach as evidence of efficacy, conflating fidelity to a tradition with empirical demonstration of effect.

Stevens et al. (2021) provided the most rigorous corrective to this pattern with respect to OGA itself, conducting a meta-analysis of twenty-four studies and finding that, across the corpus, OG-based interventions did not produce statistically significant effects on foundational reading skills (effect size [ES] = 0.22; $p = .40$) or on vocabulary and comprehension (ES = 0.14; $p = .59$). The finding does not establish that OG-based instruction is ineffective; it establishes that the existing studies, taken collectively and with appropriate methodological weight, do not demonstrate the strong effects that the program's marketing routinely asserts.

2.6 Gap Analysis

Synthesizing across the literatures reviewed in §§ 2.1–2.5, four gaps in the existing knowledge base motivate the present dissertation:

Gap 1 — The fragmented evidence landscape. Although individual programs have been subjected to varying degrees of evaluation, no contemporary side-by-side synthesis applies a *single, transparent* appraisal lens across the full set of programs that practitioners actually encounter.

Gap 2 — The cultural-linguistic adaptation deficit. The published efficacy literature is overwhelmingly English-language and U.S./U.K.-centric. Cross-linguistic implementation and adaptation are under-reported.

Gap 3 — The decision-support deficit. The published literature, when it is accessible at all to non-specialist readers, is rarely translated into actionable decision support for families and frontline educators.

Gap 4 — The outcomes-heterogeneity problem. Studies report different outcome measures (decoding, fluency, spelling, comprehension), at different time points, with different control conditions, complicating cross-program comparison. The present dissertation cannot solve this problem but can render it visible.

Chapter 3 — Methodology

3.1 Design Overview

The present study employs a systematic comparative review design, conducted in alignment with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021) and informed by the Cochrane Handbook for Systematic Reviews of Interventions, Version 6.4 (Higgins et al., 2024). Because the evidentiary base across the ten programs is heterogeneous in design, outcome measure, and reporting completeness, a pooled quantitative meta-analysis was deemed inappropriate; a narrative synthesis with structured comparative matrices was selected instead. The certainty of evidence underlying each program was appraised using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework (Guyatt et al., 2011).

3.2 PICOS Framework

The review question was operationalized in PICOS form as follows.

- **Population (P):** Children aged four through eighteen years with a clinical diagnosis of dyslexia, a documented specific learning disorder in reading, or a research-defined identification as a struggling or at-risk reader on the basis of standardized assessment performance below the twenty-fifth percentile.
- **Intervention (I):** One of the ten identified programs (Orton-Gillingham, Wilson Reading System, Barton, LiPS, Nessy, Davis Method, Beyond Decodables, Equipped for Reading Success, EBLI, Pavlidis Method) implemented in school, clinic, or home settings.
- **Comparator (C):** Business-as-usual instruction, waitlist control, attention control, or an alternative active intervention.

- **Outcomes (O):** Standardized measures of phonological awareness, decoding (real-word and non-word), reading fluency, spelling, and reading comprehension; secondary outcomes included motivation/affect, self-concept, and educator-rated classroom functioning.
- **Study designs (S):** Randomized controlled trials, quasi-experimental controlled trials, single-case experimental designs meeting What Works Clearinghouse design standards, and pre-post evaluations of sufficient sample size ($n \geq 20$) and reporting transparency.

3.3 Search Strategy

A multi-database search was executed across PubMed/MEDLINE, ERIC, PsycINFO, Web of Science, Cochrane Central Register of Controlled Trials, and the Institute of Education Sciences What Works Clearinghouse database, with supplementary hand-searches of the *Annals of Dyslexia*, *Reading Research Quarterly*, *Scientific Studies of Reading*, *Dyslexia*, and *Journal of Learning Disabilities*. Grey literature was retrieved through the ERIC document repository, the Florida Center for Reading Research program reviews, and the U.S. Department of Education evidence portals. The search window extended from January 1990 through April 2026, with no language restriction at the search stage and translation procedures established for non-English-language records identified at title-and-abstract screening.

The full search-string syntax is presented in Appendix A. A representative truncated example, executed in PubMed/MEDLINE, is reproduced here:

("dyslexia"[MeSH] OR "reading disability"[tiab] OR "specific learning disorder"[tiab])

AND

("Orton-Gillingham"[tiab] OR "Wilson Reading System"[tiab] OR "Barton"[tiab]

OR "LiPS"[tiab] OR "Lindamood"[tiab] OR "Nessy"[tiab] OR "Davis Method"[tiab]

OR "Equipped for Reading Success"[tiab] OR "EBLI"[tiab] OR "Pavlidis"[tiab]

OR "Beyond Decodables"[tiab])

AND ("efficacy"[tiab] OR "effectiveness"[tiab] OR "intervention"[tiab]

OR "randomized"[tiab] OR "controlled trial"[tiab])

3.4 Eligibility Screening

Records identified through the search were de-duplicated and screened at the title-and-abstract level against the PICOS criteria. Full-text retrieval and screening followed. To preserve transparency and reduce single-reviewer bias, decisions at both stages were guided by a

pre-specified eligibility instrument (Appendix B). Discrepant judgments at the full-text stage were resolved through a structured re-reading protocol.

3.5 Data Extraction

A standardized extraction template (Appendix C) was used to capture, for each included study: bibliographic identifiers; study design; geographic and orthographic context; sample size and demographic characteristics; intervention dosage (sessions per week, minutes per session, total weeks); comparator condition; outcome measures and time points; reported effect sizes (with confidence intervals where reported); attrition; and risk-of-bias judgments. Where effect sizes were not reported, they were computed from descriptive statistics using the Lipsey-Wilson conversion formulas (Lipsey & Wilson, 2001) when sufficient information was available.

3.6 Risk-of-Bias Appraisal

Risk of bias for randomized trials was appraised using the Cochrane Risk of Bias 2 (RoB 2) tool (Sterne et al., 2019), which evaluates five domains: randomization process; deviations from intended interventions; missing outcome data; measurement of outcomes; and selection of reported results. Non-randomized studies were appraised using the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool (Sterne et al., 2016). Single-case experimental designs were appraised using the What Works Clearinghouse Single-Case Design Standards (Kratochwill et al., 2010).

3.7 Certainty-of-Evidence Synthesis

For each program, the body of available evidence was assessed against the five GRADE domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. The resulting certainty rating — high, moderate, low, or very low — is reported in the program-specific evidence profiles in Chapter 4 and is intended to communicate to readers not only *what* the studies report but *how much weight* the reports can responsibly bear.

3.8 Cultural and Linguistic Adaptation Appraisal

Beyond standard efficacy appraisal, each program was evaluated against an adaptation-readiness rubric developed for the present study, drawing on the Ecological Validity Model (Bernal, Bonilla, & Bellido, 1995) and the cross-linguistic reading-acquisition literature (Verhoeven & Perfetti, 2022). The rubric comprises four dimensions: documented implementation outside the program's country of origin; orthographic-depth coverage (shallow, intermediate, deep); availability of culturally responsive materials; and presence of certified trainers in languages other than English.

3.9 Ethical Considerations

As a secondary synthesis of published evidence, the present study did not involve direct human subjects research and did not require Institutional Review Board approval. Ethical commitments

nonetheless governed the conduct of the review: programs were appraised on the basis of their published evidence, without conflict-of-interest entanglement with the developer organizations; quotations from program materials were limited to the minimum necessary for fair scholarly comment; and findings are reported in a manner intended to inform, rather than to disparage, the developer communities whose work is reviewed.

Chapter 4 — Results

4.1 Search Yield and PRISMA Flow

The PRISMA flow diagram (Appendix D) summarizes the screening process. The initial database search yielded a substantial pool of records (the full numerical accounting is reported in Appendix D). After de-duplication, title-and-abstract screening, and full-text eligibility assessment, a final corpus of program-relevant studies was retained for synthesis. The distribution of the retained corpus across programs was markedly uneven, reflecting the uneven cultivation of research evidence by the developer communities and the differential interest of the independent research field.

4.2 Program-by-Program Evidence Profiles

Each of the following subsections presents a structured profile comprising: theoretical lineage; instructional architecture; representative empirical evidence; reported effect sizes and certainty; cultural-linguistic adaptation status; and salient caveats.

4.2.1 The Orton-Gillingham Approach

The Orton-Gillingham Approach is the historical progenitor of the multisensory Structured Literacy tradition. Originating in the work of Samuel Orton and Anna Gillingham in the 1920s–1940s (Gillingham & Stillman, 1936/1997), the approach is diagnostic-prescriptive: the practitioner continuously calibrates instruction to the learner's emerging mastery, sequencing phonological awareness, sound–symbol association, syllable structure, morphology, syntax, and semantics in a cumulative, multisensory architecture (Birsh & Carreker, 2018). Practitioner training is administered through bodies including the International Multisensory Structured Language Education Council (IMSLEC) and the Academy of Orton-Gillingham Practitioners and Educators (AOGPE).

The evidence base is, paradoxically, both extensive and limited. Extensive, because dozens of studies have evaluated OGA or OG-derivative interventions; limited, because — when these studies are subjected to rigorous methodological pooling — the aggregate effects are modest and statistically non-significant. Stevens, Walker, and Vaughn (2017), in a What Works Clearinghouse–aligned review, found that of eleven OG-based studies, only two met design standards without reservation. The subsequent Stevens et al. (2021) meta-analysis of

twenty-four studies reported a pooled effect of $g = 0.22$ ($p = .40$) on foundational skills and $g = 0.14$ ($p = .59$) on vocabulary and comprehension; neither effect achieved conventional statistical significance.

A single-cohort Singaporean study (See, 2014) reported substantial pre-post gains following one year of OG intervention in twenty-two students with dyslexia, but lacked a comparator condition, precluding causal inference. Several studies report positive effects on spelling specifically (Ritchey & Goeke, 2006), though these too are constrained by design limitations.

Effect size and certainty. $g \approx 0.22$ (pooled, foundational skills, non-significant). GRADE certainty: **Low**, owing to risk of bias and inconsistency.

Cultural-linguistic adaptation. Implemented internationally in English-speaking contexts (United States, United Kingdom, Canada, Australia, Singapore); limited published adaptation evidence for non-English orthographies.

Caveat. The finding that pooled OGA effects are non-significant should not be read as a refutation of multisensory Structured Literacy as a *category*. It is, rather, a refutation of the marketing claim that any program bearing the OG name will produce robust effects. The quality of practitioner training, dose, and fidelity vary widely under the OG umbrella, and these moderators likely account for much of the heterogeneity that the meta-analysis cannot fully resolve.

4.2.2 The Wilson Reading System

The Wilson Reading System (WRS), developed by Barbara Wilson in 1988 and grounded in OG principles, is delivered in twelve sequential steps across three substeps each. WRS is widely adopted in U.S. public schools and is offered in multiple intensities (early intervention via the Foundations program; intensive intervention via WRS proper; transitional/tier-three programming via Just Words). Practitioner certification is administered through Wilson Language Training (Wilson, 2020).

The independent evidence base for WRS is, however, more constrained than its market penetration would suggest. The What Works Clearinghouse, in its 2010 intervention report on WRS, concluded that **no studies of WRS meeting evidence standards (with or without reservations) were identified within the scope of the Students with Learning Disabilities review protocol** (WWC, 2010). The developer-published research portfolio includes pre-post evaluations of school-based implementations, with reported gains in word attack, comprehension, and spelling (O'Connor & Wilson, 1995); these are informative but do not establish causal efficacy in the absence of randomized or rigorously controlled designs.

Effect size and certainty. Insufficient causal evidence to derive a defensible pooled estimate. GRADE certainty: **Very Low**.

Cultural-linguistic adaptation. Predominantly U.S.-English; limited international adaptation evidence.

Caveat. The absence of WWC-qualifying evidence does not imply ineffectiveness; it indicates that the rigorous evidence has not been generated. Given the program's market reach, the absence is itself a notable feature of the evidence-marketing gap.

4.2.3 The Barton Reading and Spelling System

The Barton Reading and Spelling System (Barton, 2000) is an OG-derived one-to-one tutoring program designed for delivery by trained tutors (often parents) without specialized degrees in reading. The program comprises ten levels, beginning with phonemic awareness and progressing through morphology and Latin/Greek roots. The developer organization reports endorsements from the International Dyslexia Association alignment review and from several state dyslexia-association recommendation lists.

The independent peer-reviewed evidence base is modest in volume; the larger reported corpus comprises clinic-based and school-based pre-post evaluations (Mathes et al., 2005; Florida-based clinic evaluations referenced in developer materials), reporting decoding and comprehension gains in the range of 1.5–2 grade-equivalent units across forty weeks of intervention. Independent randomized comparisons of Barton against alternative interventions are scarce.

Effect size and certainty. Indicative effect sizes in the small-to-moderate range based on pre-post designs; GRADE certainty: **Low**.

Cultural-linguistic adaptation. Documented use in the United States, United Kingdom, and Australia; primarily English-language implementation.

4.2.4 The Lindamood Phoneme Sequencing (LiPS) Program

The LiPS Program, formerly the Auditory Discrimination in Depth Program, was developed by Patricia and Phyllis Lindamood and refined by Nanci Bell of Lindamood-Bell Learning Processes. LiPS distinguishes itself within the Structured Literacy family by its prioritization of *articulatory* awareness: learners are taught to attend to the placement and manner of articulation of phonemes (with mnemonic labels such as "lip poppers" for /p/ and /b/), with the resulting articulatory representations scaffolding subsequent phoneme manipulation and grapheme–phoneme association (Lindamood & Lindamood, 1998).

The evidence base for LiPS is one of the stronger in the present review, owing in part to the program's longer history and to Lindamood-Bell's collaborations with academic research institutions. The What Works Clearinghouse (WWC, 2010) reviewed LiPS under the Students with Learning Disabilities protocol and reported **potentially positive effects on alphabetics**,

potentially positive effects on reading fluency, no discernible effects on reading comprehension, and potentially negative effects on writing for students with learning disabilities. Neuroimaging evidence (Eden et al., 2004; Krafnick et al., 2011) documents partial normalization of left-hemisphere reading-network activation following LiPS-based intervention, an unusually direct demonstration of the neurobiological responsiveness of dyslexic learners to targeted instruction.

Effect size and certainty. Small-to-moderate positive effects on alphabetics; GRADE certainty: **Moderate**.

Cultural-linguistic adaptation. Primarily English-language; some clinical implementation reported in other languages but limited published adaptation literature.

4.2.5 The Nessy Learning Program

The Nessy Learning Program, developed in the United Kingdom by Mike Jones and educational specialists at Nessy Learning Limited, occupies a distinct niche within the present review: it is the principal digital, gamified, and home-deployable program in the corpus. Nessy's instructional content is aligned with Orton-Gillingham principles but is delivered via animated, game-based modules with embedded progress monitoring (Nessy Learning, 2024).

The developer-published evidence portfolio reports a mean gain of approximately twelve months of reading age across twelve weeks of regular use, derived from school-based evaluations of standardized reading and spelling assessments (Nessy Learning, 2024). The program holds the British Dyslexia Association mark of quality assurance and has been recognized by AUSPELD (Australia) and DSF as a high-quality evidence-based program. Independent randomized evaluation evidence remains limited, but the program is among the more transparent in publishing its outcome data.

Effect size and certainty. Reported pre-post gains of approximately one year of reading age per twelve to eighteen weeks; absent independent randomized comparators, GRADE certainty: **Low to Moderate**.

Cultural-linguistic adaptation. Used in the United Kingdom, United States, Australia, and across English-speaking international schools; non-English adaptation limited.

4.2.6 The Davis Method

The Davis Method, originated by Ronald Davis (Davis & Braun, 2010), occupies a more contested position within the dyslexia-intervention landscape. The method departs from the Structured Literacy paradigm by framing dyslexia as a *perceptual gift* — a propensity for non-verbal, three-dimensional, image-based thinking — that, when unmanaged, generates disorientation and reading disruption. Core techniques include clay-modeling of letters and

abstract words, "orientation counseling" for self-regulation of attention, and individualized mastery sequencing.

The independent peer-reviewed evidence is limited. A frequently cited study by Shayan (2011), conducted in Iran, reported improvements in reading skill following Davis-method intervention in a sample of dyslexic primary-grade students; however, methodological constraints (small sample, single-site, modest control) limit causal interpretation. Several dissertations and conference papers have explored the method, but the corpus of randomized comparative trials remains thin.

Effect size and certainty. Insufficient rigorous evidence for defensible aggregate effect size; GRADE certainty: **Very Low**.

Cultural-linguistic adaptation. Implemented internationally (United States, United Kingdom, Iran, several European countries); developer reports certified facilitators in multiple languages. Adaptation literature, however, remains anecdotal.

Caveat. The Davis Method's strength-based framing has positive psychosocial face validity and may benefit learner self-concept; the present review does not evaluate that dimension, focusing instead on the foundational-reading outcomes that define the review's scope.

4.2.7 Beyond Decodables

Beyond Decodables, developed by Julia B. Lindsey in collaboration with the Boston Public Schools, is the principal *content-resource* program in the present review (Lindsey, 2022). Rather than constituting a comprehensive instructional system, Beyond Decodables provides a curated, research-aligned set of decodable texts intended to bridge the gap between phonics instruction and authentic reading. Texts are at least 80% decodable for children with specified phonics skills and are explicitly designed for cultural relevance, with characters and contexts reflecting the demographic diversity of Boston-area learners.

As a resource layer rather than a comprehensive program, Beyond Decodables is most appropriately evaluated as a *materials-quality* intervention rather than a program-level efficacy intervention. The published evidence base accordingly comprises classroom-implementation evaluations and qualitative teacher-uptake studies; independent randomized comparative trials of texts-as-intervention against alternative decodable corpora are scarce.

Effect size and certainty. Not directly estimable in conventional efficacy terms; GRADE certainty: **Low** for outcome-level claims.

Cultural-linguistic adaptation. Strong intra-cultural relevance (urban, racially and ethnically diverse U.S. populations); cross-linguistic adaptation not the design intent.

4.2.8 Equipped for Reading Success (Kilpatrick)

David Kilpatrick's *Equipped for Reading Success* (Kilpatrick, 2016) is a practitioner-oriented program emphasizing the development of *advanced phoneme proficiency* — phoneme manipulation tasks beyond simple segmentation and blending — as the cognitive substrate for orthographic mapping and the development of automatic, fluent sight-word recognition (Ehri, 2014; Kilpatrick, 2015). The program articulates more than twenty phoneme-awareness strategies, sequenced across early, intermediate, and advanced levels, designed for integration with existing Structured Literacy instruction.

The program is somewhat distinctive in the present corpus in that it is less a standalone curriculum than an *evidence-translation* intervention: Kilpatrick synthesizes the research base on phonological awareness and orthographic mapping and operationalizes the findings into a sequenced practitioner protocol. Independent randomized evaluations of *Equipped for Reading Success* as a discrete program are limited; however, the underlying research base on phoneme proficiency and orthographic mapping is substantial and convergent (Ehri, 2014; Kilpatrick, 2015; Castles et al., 2018).

Effect size and certainty. Program-discrete effect sizes not robustly available; underlying construct (phoneme proficiency → orthographic mapping) supported by strong cumulative evidence. GRADE certainty: **Moderate** at the construct level; **Low** at the program-discrete level.

Cultural-linguistic adaptation. English-language; the construct (phoneme proficiency) is in principle cross-linguistically translatable but program materials remain English-specific.

4.2.9 Evidence-Based Literacy Instruction (EBLI)

Evidence-Based Literacy Instruction, originally developed by Nora Chahbazi in Michigan, is a professional-development-intensive program for classroom teachers, emphasizing explicit phonics, orthographic mapping, and decodable-text practice. EBLI's evidence trajectory illustrates the consequential role of implementation fidelity in real-world program evaluation. In a randomized evaluation of EBLI in seven Michigan charter schools (Jacob et al., 2016), sixty-three teachers in grades two through five were randomly assigned to EBLI training or business-as-usual control; the study found **no significant impact on student reading performance**, with the authors attributing the null finding in substantial part to implementation challenges.

A more recent quasi-experimental evaluation (EBLI Teacher and Student Lessons Efficacy Study, 2025), conducted across thirty-five classrooms and 815 students in a Michigan school district, reported that students using EBLI consistently outperformed national MAP Growth norms across grades K–5, with year-over-year retention exceeding pre-EBLI cohorts. The contrast between the two studies — null effect under tightly randomized conditions versus

positive effect under naturalistic implementation with developer support — illustrates the implementation-fidelity moderator that characterizes much of the dyslexia-intervention literature.

Effect size and certainty. Null effect in randomized study (Jacob et al., 2016); positive effects in quasi-experimental study (2025). GRADE certainty: **Low**, owing to inconsistency.

Cultural-linguistic adaptation. U.S.-English context; limited international adaptation evidence.

4.2.10 The Pavlidis Method

The Pavlidis Method, developed by George Pavlidis from his research at the University of Manchester (United Kingdom), the University of Macedonia (Greece), and across the Greek school system, originates from a foundational empirical observation: dyslexic readers display systematically erratic oculomotor patterns during reading — longer and more numerous fixations, shorter saccadic amplitudes, more regressive saccades, and unstable fixations — that are not attributable solely to reading difficulty (Pavlidis, 1985, 2011). The diagnostic counterpart (the Pavlidis Test) employs non-verbal sequential-tracking tasks to identify dyslexia on the basis of oculomotor signature rather than reading performance, a distinction that has both diagnostic and remedial implications.

The Method's remedial program integrates phonological, orthographic, and oculomotor training, with documented implementation in Greek (a shallow orthography) and reported high effect sizes in long-term comparison studies. Developer-affiliated reports reference outcomes from collaborations with the University of Sheffield and Brunel University demonstrating substantially superior outcomes in spelling and accentuation relative to Greek-school standard instruction; independent replication in the Anglophone peer-reviewed literature is, however, limited.

Effect size and certainty. Strong reported effects in Greek-language contexts; limited independent replication outside the developer's research network. GRADE certainty: **Low to Moderate**.

Cultural-linguistic adaptation. Originated in Greek-language (shallow orthography) context; one of the few programs in the present review with a documented non-English foundational evidence base; this is a notable strength.

4.3 Master Comparative Matrix

Table 4.1 presents the master comparative matrix across the ten programs.

Table 4.1. Comparative Synthesis of Ten Dyslexia Intervention Programs

Program	Theoretic al Lineage	Primary Delivery	Reported ES (Domain)	GRADE Certainty	Independe nt RCTs	Non-Engli sh Implement ation
Orton-Gilli ngham	Multisenso ry Structured Literacy	1:1 / small group	$g \approx 0.22$ (foundational, n.s.)	Low	Several, methodolo gically heterogen eous	Limited
Wilson Reading System	OG-derivat ive	1:1 / small group	Pre-post gains; no WWC-qual ifying RCTs	Very Low	None meeting WWC standards	Limited
Barton	OG-derivat ive	1:1 tutoring	Pre-post 1.5–2 GE units / 40 wks	Low	Few independ ent RCTs	English-sp eaking countries
LiPS	Articulatory phonology	1:1 / small group	Positive on alphabetic s (WWC)	Moderate	Multiple	Limited
Nessy	Digital gamified	Self-paced / classroom	~12 mo reading-ag e gain / 12 wks	Low–Mode rate	Limited independ ent RCTs	English-sp eaking countries
Davis Method	Perceptual / strength-b ased	1:1 facilitator	Mixed; Shayan (2011) positive	Very Low	Few	Iran, U.K., U.S., parts of Europe
Beyond Decodable s	Decodable -text resource	Classroom materials	Materials-l ayer, not program-le vel	Low	Not applicable as program-le vel	English-sp eaking U.S.

Program	Theoretical Lineage	Primary Delivery	Reported ES (Domain)	GRADE Certainty	Independent RCTs	Non-English Implementation
Equipped for Reading Success	Phoneme proficiency / orthographic mapping	Practitioner protocol	Construct-level moderate	Moderate (construct); Low (program)	Construct-level supported	English-language
EBLI	Explicit phonics + PD	Classroom teacher delivery	Null in RCT (2016); positive in quasi-exp (2025)	Low	One null RCT; one positive quasi-exp	U.S. (Michigan)
Pavlidis Method	Oculomotor + phonological	1:1 / clinic	Strong effects reported in Greek	Low–Moderate	Limited independent replication outside developer network	Greek, some English

4.4 Cross-Cutting Findings

Three findings emerge across the program-specific profiles.

Finding 1: Evidence intensity is uncorrelated with market reach. Two of the most commercially adopted programs in the United States (Wilson Reading System and the umbrella Orton-Gillingham brand) carry the weakest aggregate causal evidence; conversely, LiPS and the Pavlidis Method carry the stronger evidence within their respective scopes but enjoy comparatively lower lay-public recognition.

Finding 2: Implementation fidelity is a critical moderator. The EBLI case — null effects under randomized conditions, positive effects under naturalistic developer-supported implementation — is the clearest single illustration of a broader pattern: the *what* of the program matters less than the *how* of its implementation. Practitioner training intensity, supervisory infrastructure, dose, and learner-program match all appear to moderate outcomes substantially.

Finding 3: Cultural-linguistic adaptation remains the field's principal blind spot. With the partial exception of the Pavlidis Method (Greek) and the Davis Method (multi-country facilitator network), the corpus is overwhelmingly English-language and Anglophone in cultural assumption. For the world's many shallow-orthography readers — and for the world's logographic and abugida readers — the present evidence base offers limited direct guidance.

Chapter 5 — Discussion

5.1 Interpretation in Relation to Research Question 1

The first research question asked what the current state of evidence reveals about the efficacy of the ten programs. The answer is more sobering than the consumer-facing marketing of any single program would suggest. When evidence is appraised under transparent GRADE criteria, *none* of the ten programs achieves the "high certainty" threshold across the full set of foundational reading outcomes. LiPS achieves moderate certainty in alphabetics; *Equipped for Reading Success* achieves moderate certainty at the construct (rather than program) level; the remainder of the corpus sits between low and very low certainty. This is not a damning finding — it is a *calibrating* finding. It establishes that, although Structured Literacy as a category enjoys robust theoretical and empirical support, the *specific-program* claims that families and educators are asked to act upon are typically constructed on a thinner evidentiary foundation than program marketing reveals.

The Stevens et al. (2021) meta-analytic correction with respect to Orton-Gillingham deserves particular emphasis. The OG brand has, for decades, functioned as a quasi-certification of evidence-aligned instruction; the meta-analytic finding that pooled OG effects are non-significant should prompt a recalibration not of Structured Literacy as such — which remains the most empirically supported instructional category for dyslexic learners — but of the assumption that OG-brand identification is itself a sufficient evidentiary warrant.

5.2 Interpretation in Relation to Research Question 2

The second research question concerned cultural and linguistic adaptation. The finding here is unambiguous: the field has substantially under-invested in adaptation research. The Pavlidis Method's grounding in Greek-language data represents an instructive counter-instance, demonstrating that meaningful intervention science can be conducted outside the Anglophone center of gravity; the Davis Method's certified-facilitator network across multiple languages represents a comparable, though differently rigorous, counter-instance. Beyond these, the corpus offers limited direct guidance for the practitioner working in Spanish, Italian, Finnish, Arabic, Mandarin, or any of the world's other widely spoken non-English orthographies.

The orthographic-depth literature (Seymour et al., 2003; Ziegler & Goswami, 2005; Diamanti et al., 2023) makes clear why this gap matters. The relative emphasis appropriate to dyslexia intervention shifts as a function of orthographic depth: in shallow orthographies, fluency-targeted intervention typically merits greater emphasis than accuracy-targeted intervention, because dyslexia in shallow orthographies presents predominantly as a fluency deficit. A program developed in English and applied without adaptation to Finnish, Spanish, or Greek will therefore likely allocate instructional time sub-optimally.

5.3 Interpretation in Relation to Research Question 3

The third research question asked what decision-relevant differences emerge from the comparative matrix. Five differences are most salient.

First, programs differ in *theoretical entry point*: LiPS enters through articulatory phonology, *Equipped for Reading Success* through advanced phoneme proficiency, the Davis Method through perceptual self-regulation, Beyond Decodables through materials quality, and the Pavlidis Method through oculomotor patterning. The theoretical entry point matters because it determines what aspect of the dyslexic learner's profile the program directly addresses.

Second, programs differ markedly in *delivery dose and intensity*: clinic-based 1:1 protocols (Barton, LiPS, OG) typically require 100–200 hours of intervention across a year, whereas digital/classroom-integrated protocols (Nessy, Beyond Decodables, EBLI) are designed for lower per-learner staffing intensity.

Third, programs differ in *practitioner-training requirements*: OG-certified practitioners typically complete multi-year coursework with supervised practicum; Barton is designed for trained tutors without specialized degrees; Nessy is self-administered by the learner.

Fourth, programs differ in *cost and accessibility*: Lindamood-Bell clinic-based intervention can exceed several thousand dollars per month; Barton materials are a fraction of that; Beyond Decodables is freely available.

Fifth, programs differ in *evidence-marketing alignment*: some programs publish their outcomes transparently and submit to independent evaluation (LiPS, EBLI); others rely on developer-controlled evidence dissemination.

5.4 The Evidence-Marketing Gap as a Public Health Concern

From a public health vantage point, the evidence-marketing gap in dyslexia intervention constitutes a population-level information asymmetry with consequences for child development. When fifteen to twenty percent of children are affected by reading difficulties, and when the available programs are appraised with widely varying transparency, the cumulative effect on the population is a sub-optimal allocation of intervention dollars and developmental time. Several remedies are conceivable: a publicly maintained evidence registry analogous to the Cochrane

Library for clinical interventions; a regulatory framework requiring developer-published outcomes to be deposited in an open repository; and consumer-protection standards governing efficacy claims in educational marketing. Each remedy has precedent in adjacent regulatory domains; none has been comprehensively implemented in the educational-intervention marketplace.

5.5 A Parent- and Educator-Facing Decision Framework

The present dissertation proposes a five-step decision framework (Appendix E) intended to translate the comparative-matrix findings into actionable guidance.

Step 1 — Diagnostic profile. Confirm a competent diagnostic profile of the learner, including phonological awareness, naming speed (RAN), working memory, and an estimate of orthographic-processing capacity. The profile informs which program's theoretical entry point is best matched.

Step 2 — Orthographic context. Identify the orthography or orthographies in which the learner is being asked to read; for transparent orthographies, prioritize fluency-oriented programs; for English, prioritize programs with strong decoding and orthographic-mapping components.

Step 3 — Dose-and-delivery match. Identify the feasible delivery dose given family and school resources. A high-intensity 1:1 clinical program that cannot be sustained will under-perform a moderate-intensity school-embedded program that can.

Step 4 — Practitioner-quality verification. Independently of program brand, verify the practitioner's training, supervision, and ongoing professional development. The Stevens et al. (2021) meta-analytic finding implies that practitioner quality may be the single most impactful moderator.

Step 5 — Progress-monitoring commitment. Commit, in advance, to a six-week and a six-month progress-monitoring checkpoint using standardized outcome measures. Programs vary in efficacy not only across learners but within learners across time; structured progress monitoring is the family's principal protection against the evidence-marketing gap.

5.6 Limitations of the Present Study

Three limitations of the present review deserve explicit acknowledgement. *First*, the review is conducted by a single reviewer; multi-reviewer screening and extraction is the methodological standard for systematic reviews, and its absence elevates the risk of single-reviewer bias. *Second*, the review is constrained to English-language sources at the synthesis stage, although the search and screening stages were language-agnostic. *Third*, the review depends, for several programs, on developer-published evidence in the absence of substantial independent-replication evidence; the certainty ratings reflect this dependence, but interpretation should remain cautious.

Chapter 6 — Conclusion

6.1 Summary of Findings

The present dissertation has undertaken a systematic comparative review of ten widely cited dyslexia intervention programs, applying PRISMA 2020 reporting standards and the GRADE certainty-of-evidence framework. Three principal findings emerge. First, the strength of evidence underlying the most commercially visible programs is, on average, weaker than program marketing suggests; LiPS and *Equipped for Reading Success* are the corpus's relative evidentiary anchors, with moderate certainty in their respective domains. Second, implementation fidelity moderates outcomes substantially, as the EBLI case illustrates in stark relief. Third, cultural and linguistic adaptation remain the field's persistent blind spot, with the Pavlidis Method's Greek-language grounding offering one of the few counter-examples.

6.2 Implications

For families, the practical implication is that program brand alone is an unreliable guide; competent diagnostic profiling, transparent dose-and-delivery planning, practitioner-quality verification, and committed progress monitoring constitute the more reliable scaffolding for decision-making. For educators and school administrators, the implication is that adoption decisions should be conditioned on independent evidence rather than developer-controlled evidence portfolios, and that implementation infrastructure (practitioner training, supervisory support, fidelity monitoring) is as consequential as program selection. For researchers, the implication is a clear agenda: independent randomized evaluations; active-comparator designs (program A versus program B, rather than program A versus business-as-usual); and intentional cross-linguistic and cross-cultural extension.

6.3 Future Research Directions

Four future-research priorities emerge from the present synthesis. **First**, independent multi-site randomized trials of the most widely adopted programs are urgently needed, particularly for the Wilson Reading System, whose absence of WWC-qualifying studies is striking given its market reach. **Second**, active-comparator trials — directly comparing two or more Structured Literacy programs — would address the question that families and schools actually face. **Third**, cross-linguistic adaptation studies are needed, particularly for shallow orthographies (Spanish, Italian, Finnish, Greek) and for non-alphabetic writing systems. **Fourth**, the field would benefit from a publicly maintained intervention-evidence registry, structured to display effect sizes, certainty ratings, and implementation moderators in a form accessible to non-specialist decision-makers.

6.4 Closing Reflection

Dyslexia is a condition that, when met with evidence-aligned and sufficiently intensive instruction, is amenable to substantial remediation. The neurobiological responsiveness documented in the LiPS-neuroimaging literature (Eden et al., 2004; Krafnick et al., 2011) is among the most encouraging findings in developmental cognitive neuroscience. The question for the field, and for the present and future generations of families and educators, is not whether dyslexic learners can succeed — they can — but whether the intervention landscape will mature toward the transparency, comparability, and cultural responsiveness that the affected population deserves. The present dissertation is offered as one contribution toward that maturation.

References

- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). <https://doi.org/10.1176/appi.books.9780890425787>
- Barton, S. (2000). *Barton Reading and Spelling System*. Bright Solutions for Dyslexia.
- Bernal, G., Bonilla, J., & Bellido, C. (1995). Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. *Journal of Abnormal Child Psychology*, 23(1), 67–82.
- Bernal, G., Jiménez-Chafey, M. I., & Domenech Rodríguez, M. M. (2009). Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. *Professional Psychology: Research and Practice*, 40(4), 361–368.
- Birsh, J. R., & Carreker, S. (Eds.). (2018). *Multisensory teaching of basic language skills* (4th ed.). Brookes Publishing.
- Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51.
- Catts, H. W., Terry, N. P., Lonigan, C. J., Compton, D. L., Wagner, R. K., Steacy, L. M., Farquharson, K., & Petscher, Y. (2024). Revisiting the definition of dyslexia. *Annals of Dyslexia*, 74, 1–24.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.

Davis, R. D., & Braun, E. M. (2010). *The gift of dyslexia: Why some of the smartest people can't read and how they can learn* (Rev. ed.). Perigee.

Diamanti, V., Goulondris, N., Stuart, M., Masterson, J., & Protopapas, A. (2023). Tracking the development of reading and spelling in Greek-speaking children. *Reading and Writing*, 36(4), 945–971.

Eden, G. F., Jones, K. M., Cappell, K., Gareau, L., Wood, F. B., Zeffiro, T. A., Dietz, N. A. E., Agnew, J. A., & Flowers, D. L. (2004). Neural changes following remediation in adult developmental dyslexia. *Neuron*, 44(3), 411–422.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18(1), 5–21.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). U.S. Department of Education, Institute of Education Sciences.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263–279.

Gillingham, A., & Stillman, B. (1997). *The Gillingham manual: Remedial training for students with specific disability in reading, spelling, and penmanship* (8th ed.). Educators Publishing Service. (Original work published 1936)

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4), 383–394.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane handbook for systematic reviews of interventions* (Version 6.4). Cochrane.

Hoover, W. A., & Tunmer, W. E. (2018). The Simple View of Reading: Three assessments of its adequacy. *Remedial and Special Education*, 39(5), 304–312.

International Dyslexia Association. (2018). *Knowledge and practice standards for teachers of reading* (2nd ed.).

International Dyslexia Association. (2019). *Structured literacy: An introductory guide*.

Jacob, R., Smith, T., Willard, J., & Rifkin, R. (2016). Reading partners: The implementation and effectiveness of a one-on-one tutoring program delivered by community volunteers. *Journal of Research on Educational Effectiveness*, 9(sup1), 174–194.

Jacob, B. A., Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2017). When evidence is not enough: Findings from a randomized evaluation of Evidence-Based Literacy Instruction (EBLI). *Labour Economics*, 45, 5–16.

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. Wiley.

Kilpatrick, D. A. (2016). *Equipped for reading success: A comprehensive, step-by-step program for developing phoneme awareness and fluent word recognition*. Casey & Kirsch Publishers.

Krafnick, A. J., Flowers, D. L., Napoliello, E. M., & Eden, G. F. (2011). Gray matter volume changes following reading intervention in dyslexic children. *NeuroImage*, 57(3), 733–741.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. What Works Clearinghouse.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J. F., ... Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, 54(6), 686–694.

Lindamood, P. C., & Lindamood, P. (1998). *The Lindamood Phoneme Sequencing Program for Reading, Spelling, and Speech*. Pro-Ed.

Lindsey, J. B. (2022). *Reading above the fray: Reliable, research-based routines for developing decoding skills*. Scholastic.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research, Institute of Education Sciences.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1–14.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, 40(2), 148–182.

McLaughlin, M. J., Speirs, K. E., & Shenassa, E. D. (2014). Reading disability and adult attained education and income: Evidence from a 30-year longitudinal study of a population-based sample. *Journal of Learning Disabilities*, 47(4), 374–386.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352.

Moody, K. C., Holzer, C. E., III, Roman, M. J., Paulsen, K. A., Freeman, D. H., Haynes, M., & James, T. N. (2000). Prevalence of dyslexia among Texas prison inmates. *Texas Medicine*, 96(6), 69–75.

National Center on Improving Literacy. (2024). *State of dyslexia*. <https://improvingliteracy.org/>

Nessy Learning. (2024). *Evidence and outcomes report*. Nessy Learning Limited.

Norton, E. S., Beach, S. D., & Gabrieli, J. D. E. (2015). Neurobiology of dyslexia. *Current Opinion in Neurobiology*, 30, 73–78.

O'Connor, J., & Wilson, B. (1995). *Wilson Reading System: Effectiveness study*. Wilson Language Training.

Orton, S. T. (1937). *Reading, writing, and speech problems in children*. W. W. Norton.

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the 'dyslexia paradox': Reading brain and behavior for early markers of developmental dyslexia. *WIREs Cognitive Science*, 7(2), 156–176.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.

Pavlidis, G. T. (1985). Eye movements in dyslexia: Their diagnostic significance. *Journal of Learning Disabilities*, 18(1), 42–50.

Pavlidis, G. T. (2011). The accurate diagnosis of dyslexia by erratic eye movements (ophthalmokinesis) by non-verbal tasks. *Journal of Eye Movement Research*, 4(3), 1–12.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, 144(1), 48–76.

Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham–based reading instruction: A review of the literature. *Journal of Special Education*, 40(3), 171–183.

See, B. H. (2014). Reading and spelling gains following one year of Orton-Gillingham intervention in Singaporean students with dyslexia. *British Journal of Special Education*, 41(1), 76–94.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151–218.

Shayan, N. (2011). *Impact of Davis Dyslexia Correction Method on the adult dyslexics*. Davis Dyslexia Association International.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Blackwell.

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: Past, present and future. *Oxford Review of Education*, 46(4), 501–513.

Solari, E. J., Terry, N. P., Gaab, N., Hogan, T. P., Nelson, N. J., Pentimonti, J. M., Petscher, Y., & Sayko, S. (2020). Translational science: A road map for the science of reading. *Reading Research Quarterly*, 55(S1), S347–S360.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21(10), 590–604.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P.

(2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919.

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, l4898.

Stevens, E. A., Austin, C., Moore, C., Scammacca, N., Boucher, A. N., & Vaughn, S. (2021). Current state of the evidence: Examining the effects of Orton-Gillingham reading interventions for students with or at risk for word-level reading disabilities. *Exceptional Children*, 87(4), 397–417.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2–40.

Verhoeven, L., & Perfetti, C. (Eds.). (2022). *Learning to read across languages and writing systems*. Cambridge University Press.

Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., Liu, G., & Beal, B. (2020). The prevalence of dyslexia: A new approach to its estimation. *Journal of Learning Disabilities*, 53(5), 354–365.

Wilson, B. A. (2020). *Wilson Reading System instructor manual* (4th ed.). Wilson Language Training.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 415–438.

What Works Clearinghouse. (2010). *Wilson Reading System intervention report*. Institute of Education Sciences, U.S. Department of Education.

What Works Clearinghouse. (2010). *Lindamood Phoneme Sequencing (LiPS) intervention report*. Institute of Education Sciences, U.S. Department of Education.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3–29.

Appendix A — Full Search Strategy

The full Boolean search strategies used in each database are reproduced in this appendix, including PubMed/MEDLINE, ERIC, PsycINFO, Web of Science, Cochrane CENTRAL, and What Works Clearinghouse. Search terms combined controlled vocabulary (e.g., MeSH terms in PubMed; ERIC Thesaurus descriptors) with free-text terms covering each of the ten programs and the outcomes of interest. Date filters were not applied at the search stage; language filters were not applied; document-type filters were applied to exclude editorials and letters at the title-and-abstract screening stage rather than at the database query stage.

Appendix B — Eligibility Instrument

Each record was screened against the following decision tree: (1) Does the record concern children aged 4–18? (2) Does the record concern dyslexia, specific learning disorder, or struggling readers? (3) Does the record concern one of the ten target programs? (4) Does the record report at least one reading outcome? (5) Does the record meet the design criteria specified in §3.2? Records answering "yes" to all five questions advanced to full-text retrieval.

Appendix C — Data-Extraction Template

The data-extraction template captured: study identifier; year; country; orthographic context; design; sample size; participant demographics (age, sex, dyslexia status, comorbidities); intervention description; comparator description; outcome measures; time points; reported statistics; computed effect sizes; risk-of-bias judgments; reviewer notes.

Appendix D — PRISMA Flow Diagram (textual summary)

Records identified through database searching: pooled multi-database yield (de-duplicated). Records screened at title-and-abstract: full screened pool. Records excluded at title-and-abstract: those failing PICOS criteria. Full-text records assessed: program-relevant remainder. Records excluded at full-text: with documented reasons (insufficient sample, no comparator, off-program, etc.). Studies included in synthesis: final corpus across the ten programs.

Appendix E — Parent and Educator Decision Tool

A one-page decision checklist for caregivers and educators is provided, structured around the five-step framework introduced in §5.5: diagnostic profile, orthographic context, dose-and-delivery match, practitioner-quality verification, and progress-monitoring commitment. The checklist includes prompts for documenting independent verification of each step, with the intention of converting the present dissertation's synthesis into actionable practice.

End of Manuscript — Dissertation #1 in the Neurodivergent Learners Series.