The Jaw as a Cross-Script Anchor: An Embodied, Grain-Size-Sensitive Model for Literacy Acquisition across Writing Systems

Abstract

A candidate universal anchor for literacy is offered by the jaw's open-close cycle through its coupling with syllable production—what we term "jaw-drop mapping." Reading science has long treated script-to-sound mapping as a disembodied computation, yet embodied neurocognitive accounts predict that literacy recruits the sensorimotor machinery that originally evolved for speech. Our synthesis provides the first mechanistically explicit account to unify previously disconnected findings from speech-motor, cross-script behavioral, and neural literatures into a set of testable propositions. We propose the Embodied Anchor–Grain-Size Interface Model, comprising five linked propositions (P1–P5). Convergent evidence indicates that the syllable is the maximal phonological unit that is both biomechanically grounded and consciously accessible before literacy. Acquisition cost is determined by the grain size that each writing system demands: phoneme, syllable, or morpheme-syllable. Phonological recoding serves as the shared pivot across scripts; metalinguistic awareness transfers most reliably across languages; and the motor cortex participates functionally in language and print processing, with stronger involvement among second-language users. Like a nautical anchor, the jaw-drop cycle may stabilize early decoding. Critically, no study has directly tested a jaw-anchored curriculum across scripts, and null findings in older remedial readers bound the model's population validity. We therefore advance this model as a mechanistically coherent, peripherally well-supported, but centrally untested research program and specify a five-study agenda for its decisive test.

Keywords: embodied cognition; jaw kinematics; syllable; speech-motor theory; orthographic depth; cross-script literacy; phonological awareness; motor cortex; theoretical integration

Significance Statement

Why do children learn to read faster in some writing systems than in others? Why do oral-language skills predict literacy success even before a child sees printed text? These questions point to a puzzle: reading is typically taught as an abstract, symbol-based skill, yet the body—particularly the speech motor system—may play a foundational role. This article synthesizes evidence across cognitive science, neuroscience, and cross-linguistic research to propose that the jaw's natural opening–closing cycle, which produces the syllable, offers a universal biological anchor for early reading. Syllabaries (e.g., Japanese hiragana) align directly with this anchor; alphabets require additional phoneme-level learning; and Chinese, though morphosyllabic, still relies on syllable-level phonological cues hidden within characters. The model also explains why phonological awareness transfers across languages but script-specific knowledge does not—a critical insight for multilingual education. We do not claim this anchor is ready for classroom use; rather, we specify exactly what studies are needed to test it. The contribution is a reframing of literacy as an embodied process—rooted in the rhythms of speech—with concrete, testable pathways for improving reading instruction across the world's diverse writing systems.

1. Introduction

1.1 From amodal symbol processing to embodied reading

Because it provides a biomechanical anchor for syllabic organization, the jaw's open-close cycle bears directly on reading acquisition. Conventional models of reading acquisition describe a linear pipeline—visual feature extraction, orthographic analysis, phonological recoding, and semantic access—in which the body enters only as input sensor and output effector. This amodal framing struggles to explain two robust facts: why scripts encoding different phonological grain sizes impose systematically different acquisition costs (Seymour et al., 2003; Ellis et al., 2004), and why oral-language foundations predict literacy even before any print is encountered. Embodied cognition offers a principled alternative: cognitive representations are partly constituted by sensorimotor states so that perceiving language re-enacts the motor programs responsible for its production. The educational relevance of this framework has been demonstrated across domains, with embodied approaches showing promise in supporting learning in contexts ranging from STEM to language acquisition (Weisberg & Newcombe, 2017). Our synthesis advances a testable claim: jaw-drop mapping---the coupling between mandibular oscillation and syllable-sized speech units---offers a biomechanically universal anchor. Any writing system can be grafted onto this anchor. Deliberately harnessing it, we argue, should facilitate cross-script literacy acquisition.

Why is a new model needed rather than a straightforward application of existing frameworks? The Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) predicts acquisition cost from orthographic transparency alone, and it correctly anticipates that English learners lag behind learners of shallow orthographies (Seymour et al., 2003). What it does not specify, however, is why the pre-literate skill that transfers most reliably across typologically unrelated scripts is specifically phonological/metalinguistic awareness rather than, for instance, general visual-analytic or executive-function skill—an asymmetry documented across independent meta-analytic lines, including a 156-study transfer-facilitation model (Li, Zhang, & Ke, 2025). Nor does a purely representational grain-size account explain why self-generated articulatory movement outperforms passive or observational analogs during early instruction (Dargue et al., 2019; Boyer & Ehri, 2011). If grain-size matching were a purely computational alignment problem, production and observation would be equally effective routes to the same representational match. A model that grounds the transferable component in a pre-literate, production-based motor unit—the jaw-syllable coupling—supplies the missing mechanistic link between "which unit transfers" and "why producing that unit, not merely perceiving it, is what drives learning." This explanatory gap, rather than a restatement of grain-size theory under a new name, is what the present synthesis is designed to close.

1.2 The jaw-drop mapping: definition and operationalization

The term "jaw-drop mapping" does not appear verbatim in the primary literature; we operationalize it through four convergent constructs drawn from the source pool. Frame/Content theory (MacNeilage, 1998) defines the species-specific organizational property of speech as a continual mouth open–close alternation, proposing that syllabic “frames” derive from mandibular oscillation cycles observable from babbling onset, and that frame and segmental “content” are separately controlled—a dissociation reflected in the syllable-structure constraints that adult speech errors obey. The evolutionary lineage from primate lip-smacks to human syllables matters for reading because it provides a phylogenetic precedent for mandibular oscillation as a vocalization frame. Brown et al. (2021) traced the frame/content architecture along exactly this trajectory. The jaw’s role as a syllable articulator is not merely phylogenetic, however; it is directly measurable. In a detailed EMA study of 18 Southern Swedish speakers, Svensson Lundmark & Erickson (2024) showed that the jaw opens and closes once per syllable, with opening magnitude scaling to syllable strength. Critically, the jaw’s acceleration and deceleration peaks—defining its articulatory steady states—do not align with acoustic segment boundaries (e.g., consonant/vowel onsets); they occur well within the vowel interval. This pattern is consistent with a syllabic frame rather than a segmental content function. The same conclusion emerges from the framing of “the jaw as syllable articulator providing a frame and the other articulators as the segmental articulators providing the content” (Svensson Lundmark & Erickson, 2024). That division of labor—a coordinative frame with superimposed content—was anticipated by the dynamical perspective on speech production (Kelso, Saltzman, & Tuller, 1986). In that view, articulation is a self-organizing system, not unlike a flock of birds that maintains formation without a conductor, and gestures serve as the primitive units. This perspective offers a direct theoretical bridge: if gestures self-organize, the jaw-drop cycle can act as a stable attractor that any writing system might exploit. Tone languages, however, present a challenge: pitch contours must align with segments. Xu’s (2024) syllable-as-synchronization theory addresses this by treating the syllable as the synchronizing mechanism for consonantal, vocalic, and laryngeal gestures. It effectively collapses the temporal degrees-of-freedom problem and explains how suprasegmental events—tone, intonation—fall naturally into place alongside segmental ones, extending the jaw-drop frame to tone languages. On the perception side, the PACT theory (Schwartz et al., 2012) and related dynamical-systems accounts supply a common currency—articulatory gestures—without implying that all perception requires motor simulation. What they do offer is a testable perception-production interface, which, from an embodied standpoint, makes the jaw anchor practically useful for literacy acquisition.

We are explicit at the outset about the epistemic status of this construct. Frame/content theory is influential but contested (e.g., critiques in the open peer commentary accompanying the 1998 target article and subsequent exchanges, including objections that the mandibular cycle is too coarse to carry the full burden of syllable organization), and we present it as a generative hypothesis rather than a settled consensus throughout.

We retain “jaw-drop mapping” rather than substituting an existing label because it names something intermediate between anatomical fact and psychological theory. Unlike “frame” in Frame/Content theory—a hypothesized control structure inferred from behavioral error patterns—jaw-drop mapping denotes a directly quantifiable kinematic variable: the amplitude and periodicity of mandibular open–close excursions measurable via EMA (Svensson Lundmark & Erickson, 2024). This variable can be logged, manipulated in training protocols, and correlated with print outcomes without committing to any particular account of its neural control. The term, therefore, serves as an operationalization bridge between Frame/Content theory and instructional/experimental application: it specifies what would have to be measured and manipulated to test the frame construct's relevance to literacy—a specification the original theory (formulated for speech production alone) does not itself provide. We use "jaw-drop mapping" strictly in this operational sense throughout, and flag any extrapolation beyond directly measured kinematics.

1.3 The present synthesis: scope, standards, and honest hedging

This is a theory-driven integrative synthesis. Rather than performing formal meta-analytic statistical re-estimation or following the full PRISMA systematic-review protocol, we descriptively aggregate published effect sizes and triangulate convergent primary findings through a rigorous narrative synthesis. It does not involve de novo pooling of raw data, heterogeneity quantification, or publication-bias correction; effect sizes quoted from published meta-analyses are reproduced descriptively, not statistically recombined. Three evidentiary standards were enforced. First, universality claims require two independent methods (e.g., kinematic measurement plus perceptual/developmental evidence), not single-language case studies. Second, instructional claims require controlled designs; effect sizes are quoted from original meta-analytic reports and are reported as descriptive, not re-pooled, estimates. Third, null and contradictory findings are reported with the same prominence as supportive ones, and the certainty of every conclusion is calibrated to the weakest link in its evidence chain. Throughout, perception-side neural evidence on syllabic rhythm processing (the temporal sampling tradition; Goswami, 2011; Giraud & Poeppel, 2012) is kept explicitly distinct from production-side jaw biomechanics evidence, because conflating the two would constitute a causal-category error that this synthesis is designed to avoid (see §3.5 and §4.2). The full search strategy—databases, search strings, and search date windows—is reported in Appendix A for transparency and reproducibility.

2. Method

2.1 Literature base and inclusion criteria

Although we adopted a systematic literature search strategy to ensure coverage, this remains a narrative integrative synthesis and does not constitute a formal systematic review. The synthesis draws on six curated literature tables (two collection waves) totaling 150 records after title/abstract screening, deduplicated by title to 128 unique sources, which were further integrated with backward citation searching from closely related theoretical and empirical reviews within the same six thematic clusters described below. The full search strategy—including database names, search date windows, and representative search strings for each thematic cluster—is reported in Appendix A. Records were included if they (a) reported primary or meta-analytic evidence on speech-motor/syllable theory, cross-script reading acquisition, embodied-instructional effects, or motor–language/print neural coupling, and (b) supplied at least an identifiable author, year, and venue sufficient for bibliographic verification. Records that could not be bibliographically verified through multi-database cross-checking during the final verification pass were excluded from the evidence base and are not cited as supporting any claim in this synthesis (rather than being retained in a provisional, unverified state). The pool clusters into six thematic groups: (i) speech-motor and syllable theory (Frame/Content; EMA descriptive studies; synchronization theory; PACT; dynamical perspectives); (ii) cross-script reading and orthographic depth (Orthographic Depth Hypothesis; Psycholinguistic Grain Size Theory; the 14-country European literacy study; cross-script learning experiments; Finnish letters-versus-syllables processing); (iii) embodied instruction and gesture (multiple embodied-learning meta-analyses; gesture-comprehension meta-analysis; motion-sensing technology meta-analysis; Moved by Reading; Activity/Imagined Activity); (iv) neural evidence on script–sound mapping (VWFA neuronal-recycling series; script-specific N170; PMd-TMS; motor-cortex rTMS and HD-tDCS; Chinese–English phonological neuroanatomy meta-analysis; pre-lexical phonological ERP for Chinese characters; phonetic-radical fMRI and eye-tracking); (v) quantitative cross-language transfer syntheses (Melby-Lervåg & Lervåg, 2011; Chinese–English transfer meta-analysis; 156-study transfer facilitation MASEM; bilingual RAN meta-analysis; cross-modality co-activation meta-analysis); and (vi) meta-analytic methodology (PRISMA 2020; RoB 2; Cochrane risk-of-bias tool; Egger's test; trim-and-fill; heterogeneity quantification; metafor), used for quality calibration rather than substantive evidence.

2.2 Analytic framework

Because re-pooling was infeasible and not intended, we adopted a three-tier narrative-synthesis structure: a theory tier (embodied and speech-motor theories), a behavioral tier (cross-script acquisition and intervention), and a neural tier (motor-system involvement in language and print processing). To ensure methodological rigor in this narrative integration, meta-analytic methodology sources informed the qualitative appraisal of individual studies and meta-analyses cited, rather than being applied as formal statistical procedures within this synthesis. Intervention studies were appraised with reference to the revised Cochrane risk-of-bias tool (RoB 2; Sterne et al., 2019) and its predecessor (Higgins et al., 2011). The descriptive presentation followed the transparency principles of the PRISMA 2020 guideline (Page et al., 2021). We explicitly clarify that all effect sizes quoted from external meta-analyses are treated as descriptive directional indicators derived from their original sources, not as statistically re-estimated or corrected values; this synthesis itself does not perform those diagnostics. We state these scope limitations plainly rather than smoothing them over. Sources lacking sufficient methodological detail for full appraisal (e.g., a record on prior-literacy impacts with a missing abstract) were used only directionally as background context and excluded from the core evidence chain supporting P1–P5. Two peripheral clusters—neural energy-metabolism studies and metaphor-processing neuroimaging—were retained only as background on resource constraints and embodied semantics, respectively. Of the energy-metabolism cluster, only one independently verifiable source (Attwell & Laughlin, 2001) is cited as background, and the cluster is excluded from the core evidence chain

Table 1

Evidence Strength Matrix for the Five Principles of the Embodied Anchor–Grain-Size Interface Model (Organized by the Model's Three-Layer Architecture)

Proposition	Core Supporting Evidence	Evidence Type(s)	Current Strength (Qualitative)	Principal Limitation
Core Mechanism 1 P1: Universal motor entry point (mandibular frame of the syllable)	Barbosa et al., 2026; Brown et al., 2021; Erickson, 2024; Erickson et al., 2012; Kawahara et al., 2015; MacNeilage, 1998; Svensson Lundmark & Erickson, 2024; Xu, 2024	Theory + cross-linguistic EMA kinematics (English, Swedish, Japanese, Mandarin + synchronization theory	Moderate: convergent theory and cross-linguistic kinematics (English, Swedish, Japanese, Mandarin); Erickson (2024) provides a broad synthesis across five languages. However, sample sizes remain small (e.g., N=18 for Swedish, N=9 for Mandarin), and the number of languages with direct kinematic evidence is still limited.	Cross-linguistic kinematic sampling remains thin; larger, balanced samples needed; contested theoretical basis (p3.1)
Core Mechanism 2 P2: Script-determined endpoint (grain-size orthographic-depth/ rhythm-type moderator)	Carioti et al., 2021; Ellis et al., 2004; Kjeldsen, 2019; Seymour et al., 2003; Xu & Ke, 2025; Ziegler & Goswami, 2005	Cross-script behavioral/ experimental; longitudinal classroom intervention (10-year Nordic model)	Strong for the grain-size main effect; moderate-to-strong for the rhythm-type/ early-intervention moderator (now supported by longitudinal quasi-experimental evidence)	Moderator is a live empirical question though longitudinal data support its predictive validity, it has not yet been isolated in a fully controlled experimental design

Proposition	Core Supporting Evidence	Evidence Type(s)	Current Strength (Qualitative)	Principal Limitation
Behavioral Consequence 1 P3: Phonological recoding as pivot	Li, Li, & Wang, 2020; Xu & Ke, 2025	Self-teaching experiments + meta-analysis	Strong: convergent experimental and meta-analytic evidence in Chinese; broadly consistent with alphabetic self-teaching literature	Direct tests concentrated in Chinese; less dense evidence in other morphosyllabic/ alphasyllabic scripts
Behavioral Consequence 2 P4: Transfer asymmetry (metalinguistic transfers; orthographic knowledge does not)	Melby-Lerv g & Lerv g, 2011; Li, Zhang, & Ke, 2025; Cui et al., 2025; Zhang et al., 2025 (Hedges g = 0.45)	Multiple independent meta-analyses	Strong: convergent moderate-to-large effects across independent meta-analytic lines	Effect sizes not directly comparable in magnitude across studies (different populations/designs); descriptive convergence only
Neural Basis P5: Motor-cortex participation in language/ print processing, amplified in L2	Vukovic et al., 2017; Johari et al., 2021; Pattamadilok et al., 2016; Tian et al., 2023; Bujok, 2025; Garvin et al., 2025; Momsen et al., 2026; Hoetjes & van Maastricht, 2020; Lin, 2024; Iizuka et al., 2020	Causal neurostimulation (TMS/tDCS) + fMRI + behavioural gesture-speech coupling (EMA, L2 recall, phonological training)	Moderate: causal evidence for functional (not epiphenomenal) motor participation, now supplemented by convergent behavioural and longitudinal evidence on gesture-scaffolding and working-memory compensation	Small-N causal studies; L1 vs. L2 comparison drawn from a limited set of neuroimaging studies; though behavioural/ ecological validity has been strengthened by recent gesture-training and EMA studies

Note. Strength ratings are qualitative syntheses of evidence type, convergence, and sample size, intended as a transparency device rather than a formal GRADE-equivalent certification; no quantitative certainty scoring algorithm was applied.

3. Results

3.1 Theory tier: the jaw–syllable coupling as a candidate universal—arguments and caveats

Three convergent reasons—evolutionary, kinematic, and control-theoretic—qualify the syllable as a biomechanically grounded unit, each illuminating a different facet of this claim. From EMA recordings of 18 Southern Swedish speakers, a systematic pattern emerges: one jaw opening per syllable, amplitude scaled to syllable strength, and timing independent of acoustic segment boundaries (Svensson Lundmark & Erickson, 2024). This dissociation is precisely what one would predict if the jaw provides a frame and the lips and tongue supply the content. The jaw holds the rhythm, much as a stage holds a performance; the actors—the lips and tongue—fill in the details. Frame/content theory (MacNeilage, 1998; Brown et al., 2021) supplies the deeper rationale: syllabic frames appear to descend from ingestive mandibular oscillations. They are already observable in canonical babbling. They persist as a control skeleton that organizes segmental content. Adult speech errors that obey syllable structure are, from this perspective, expected signatures of that skeleton, not anomalies. Synchronization theory (Xu, 2024) takes the argument further, explaining why evolution would settle on such a unit. Bundling consonantal, vocalic, and laryngeal gestures into a single synchronized package collapses the degrees-of-freedom problem. In tone languages, the same synchrony aligns pitch with segments. The jaw-drop frame therefore, extends to Chinese, rather than being confined to alphabet-centric scripts. These three lines of evidence do not merely converge; they force a rethinking of the syllable as biomechanically real, measurable, phylogenetically traceable, and harnessable for literacy.

Cross-linguistic evidence for the jaw–syllable coupling is now emerging from multiple published sources. Erickson (2024) synthesized articulatory prosody data from French, Japanese, Mandarin, Brazilian Portuguese, and English, showing that mandible lowering patterns correspond to syllable prominence values across all five languages, with focus/contrastive emphasis increasing mandible lowering on the emphasized syllable cross-linguistically. Erickson et al. (2012) demonstrated that jaw opening patterns in English correspond to metrically generated syllable stress, with jaw displacement magnitude scaling to stress level. Extending this to a typologically distinct tone language, Barbosa, Silveira, and Erickson (2026) found that Mandarin speakers of English show a universal timing relation between jaw velocity minima and acoustic vowel onset, alongside language-specific transfer of L1 foot structure. Complementary evidence from Japanese further confirms that each syllable in a phrase corresponds to a systematic jaw displacement pattern, with edge-prominence effects modulating overall displacement magnitude (Kawahara, Erickson, & Suemitsu, 2015). These findings collectively suggest that the jaw–syllable coupling reflects a general biophysiological principle of speech production—one mandibular oscillation per syllable—with language-specific modulation (e.g., prominence patterns, L1 prosodic transfer) superimposed on a universal frame.

Three caveats accompany this tier, and a fourth boundary condition—orthographic depth and language rhythm type as a moderator, elaborated in P2 below—is introduced here as an integral part of the model rather than as a late qualification (see §4.1). First, frame/content theory is contested within the field; we therefore frame the jaw–syllable coupling as a strong candidate universal requiring cross-method confirmation, not as an established fact. Second, the EMA evidence derives from 18 speakers of one language; cross-linguistic kinematic sampling remains thin, and small-N L2 kinematic studies suggesting language-independent timing of jaw-opening peak velocity to vowel onset are cited here only as consistent with, not demonstrative of, universality—small samples cannot bear universal claims. Cross-linguistic replication is beginning to emerge: Barbosa et al. (2026) examined jaw kinematics in Mandarin speakers of English, finding that the timing of jaw-opening peak velocity to acoustic vowel onset follows a language-independent pattern, while the distribution of jaw-lowering maxima reflects L1-specific prosodic transfer. This pattern suggests that the jaw–syllable coupling reflects a general speech-production principle modulated by L1 prosodic structure, although further systematic cross-linguistic kinematic sampling is needed. Third, we deliberately separate production-side kinematic evidence from perception-side neural evidence: findings that auditory cortex tracks the speech envelope at syllabic rates (the temporal sampling tradition associated with Goswami, 2011, and Giraud & Poeppel, 2012) are evidence about how listeners parse speech rhythm, not direct evidence that the jaw constitutes the production-side frame; conflating the two is a category error that this synthesis is designed to avoid. Fourth, and this is the boundary condition most consequential for stress-timed languages such as English the syllable's status as a clearly bounded, maximal, universally accessible unit, is weaker in languages where syllable boundaries are acoustically and articulatorily fuzzy (e.g., ambisyllabic consonants, vowel reduction, and consonant clusters that blur onset/coda boundaries). We treat orthographic depth and language rhythm type explicitly as a moderator of how directly the jaw-syllable anchor can be exploited, rather than assuming uniform accessibility across all languages (elaborated as part of P2 in §4.1).

This jaw-syllable coupling has a direct neural correlate in the human motor cortex, and this link is not merely theoretical or kinematic. Using fMRI, Brown et al. (2021) demonstrated somatotopic overlap between larynx and jaw representations in the primary motor cortex, with the dorsal larynx motor cortex (dLMC) situated adjacent to the jaw-depressor area—a configuration not found in non-human primates. This proximity supports the evolutionary hypothesis that the human-specific dLMC evolved to couple voluntary vocalization with pre-existing mandibular oscillatory circuits, thereby enabling the syllabic frame. Cross-linguistic kinematic evidence further corroborates this universality claim: Barbosa et al. (2026) found that the timing of jaw-opening peak velocity to acoustic vowel onset follows a language-independent pattern across Mandarin speakers of English, while the distribution of jaw-lowering maxima reflects L1-specific prosodic transfer. In this view, the universality claim of P1 rests on a phylogenetically conserved cortical architecture that integrates phonation and jaw oscillation, providing a neurobiological substrate for the syllable as the maximal motor unit accessible before literacy.

3.2 Behavioral tier I: script grain size and orthographic depth determine acquisition cost.

An unusually regular pattern

retention of spatio-motor information, allowing learners to allocate more resources to phonological and semantic processing. These findings provide convergent behavioral evidence that the jaw-syllable anchor, as a production-based motor unit, may similarly reduce cognitive load during the critical mapping between speech and print—a function hypothesized to transfer across scripts precisely because it is grounded in a universal motor frame, not in script-specific visual forms.

Connection to the thesis: the components that transfer most strongly—phonological awareness (including syllable awareness) and decoding—are precisely the levels at which the jaw-drop–syllable anchor operates; what never transfers automatically is script-specific orthographic knowledge, which grain-size studies show adjusts item-by-item. This asymmetry licenses the syllable-level embodied anchor as a transfer-ready entry point while denying any shortcut around script-specific teaching.

3.5 Neural tier: motor-system involvement in language comprehension and print processing

Neural evidence separates into two strands, and we keep their evidential roles distinct. A single question cuts across both strands: is the motor system's involvement functional or merely epiphenomenal? The rTMS evidence provides a compelling answer. When Vukovic and colleagues (2017) delivered online pulses over left primary motor cortex within 200 ms of word onset, they did not simply observe activation---they altered semantic decision latencies in a left-lateralized, meaning-specific manner. These findings demonstrate direct interference with cognitive function, not mere correlation. Behavioral evidence shows that co-speech gestures influence both speech perception and production. Bujok (2025) demonstrated that beat gestures---simple rhythmic hand movements---can bias lexical stress perception in real time, affecting eye movements during spoken-word recognition. At the neural level, Momsen et al. (2026) found that congruent versus incongruent gesture–speech pairings modulate the cortical tracking of the speech envelope, with incongruent gestures reducing the fidelity of the neural representation of the amplitude envelope. Garvin et al. (2025) used EMA to show that speaking with a co-speech gesture increases tongue and jaw displacement magnitude and reduces temporal variability in articulatory coordination---an effect they term "coupling enhancement." The efficacy of such gesture–speech coupling depends critically on the cognitive load imposed by the target linguistic unit itself. Hoetjes & van Maastricht (2020) showed that observing complex iconic gestures aided the production of a familiar L2 phoneme (/u/) but hindered the acquisition of a novel, complex phoneme (/θ/). This double dissociation suggests that the motor system's scaffolding role is bounded by task difficulty and stimulus complexity, not universally beneficial. This boundary condition aligns with P5: motor leverage appears strongest when the learner maps speech to an already accessible phonological unit (e.g., the syllable), but may be less effective or counterproductive when the learning goal demands fine-grained, unfamiliar segmental discriminations without adequate prior phonological grounding. These findings collectively suggest that the motor system is not an epiphenomenal bystander in language processing. Gestural–speech coupling actively stabilizes and amplifies articulatory output, providing a direct behavioral link to the motor cortex causality established by TMS (Vukovic et al., 2017).

This pattern is consistent with the prediction that understanding language re-enacts the motor programs for producing it, so disrupting the motor cortex should affect comprehension. HD-tDCS work extends this claim. Stimulating the motor cortex facilitates the processing of literal, metaphoric, and idiomatic action sentences compared with visual controls (Johari et al., 2021). The motor system's contribution is not restricted to concrete action verbs; it extends to figurative language as well.

Tian et al. (2023) reported that graded motor recruitment during action-language comprehension is stronger in L2 than L1. This finding suggests a compensatory reliance on sensorimotor resources precisely in the bilingual, cross-script populations of central interest. When lexical-semantic access is less automatic, the system may fall back on richer modal simulations, using the body as a scaffold. Related neuroimaging adds anatomical plausibility: a laryngeal motor cortex specialized for vocal pitch control has been identified (Simonyan & Horwitz, 2011), and mouth versus hand effectors are differentially recruited for abstract, emotional, and concrete concept processing (Mazzuca, Lugli, Benassi, Nicoletti, & Borghi, 2018), with flexible sensorimotor grounding even extending to abstract words (Mazzuca et al., 2021). These results fit within broader accounts of grounded cognition (Barsalou, 2008, 2020), where conceptual processing is partly constituted by modal sensorimotor states.

On the print-processing side, scaffolding gives way to recycling. The neuronal-recycling proposal (Dehaene & Cohen, 2007) holds that literacy does not invent new brain circuitry; it colonizes older visual maps. This claim has been simulated, tracked, and measured. Convolutional-network simulations show written-word codes emerging when a visual network originally tuned to natural images is recycled for reading (Hannagan et al., 2021). Longitudinal fMRI traces the visual word form area emerging over the first school year on the cortex that was only weakly specialized before (Dehaene-Lambertz et al., 2018). Additional computational work confirms that letter perception can self-organize through unsupervised deep learning that recycles natural-image features (Testolin, Stoianov, & Zorzi, 2017). Recycling does not bulldoze prior functions: learning to read reshapes visual cortical networks while preserving their capacity for object recognition (Hervais-Adelman et al., 2019). This process resembles converting a general-purpose hall into a specialized library without tearing down the walls—the structure remains, but the furniture is rearranged.

Within this recycled circuitry, the visual word form area occupies a central niche (Dehaene & Cohen, 2011), and dual-route neurobiological models map it onto parallel pathways for phonological recoding and direct orthographic access (Perfetti & Tan, 2013; Coltheart et al., 1993). Comparative evidence points to a cortical precursor for orthographic processing in the primate inferior temporal cortex (Rajalingham et al., 2020), reminding us that the neural raw material predates writing itself.

The boundary between universal and script-specific processing is illuminated by a coordinate-based meta-analysis of Chinese and English reading, which located a common network augmented by writing-system-specific variation (Tan, Laird, Li, & Fox, 2005). The N170 component brings this tension into sharp relief: reading-skill associations are lateralized left for German children (r = -0.531, 17.1% variance) and right for Chinese children (r = -0.423, 11.2% variance)---a double dissociation that supports fine-tuning rather than destructive competition (Ma, Li, Cao, & Sommer, 2025). A Chinese–English phonological neuroanatomy meta-analysis similarly documents shared and script-distinct circuits (Tan, Laird, Li, & Fox, 2005). More broadly, machine-learning approaches to neuroimaging data have shown promise in predicting literacy outcomes, yet the field faces substantial variability in algorithms and brain circuits across studies, alongside a relative scarcity of longitudinal predictions in young children before formal schooling (Nakai, Tirou, & Prado, 2024). The reading network is universally grounded, with script-specific variation superimposed on that ground.

Print processing should not be reduced to passive visual recycling, however. Motor transcription leaves a measurable footprint. Disrupting left dorsal premotor cortex with TMS slows pseudoword lexical decisions, directly implicating writing motor representations in early visual word recognition (Pattamadilok et al., 2016). Handwriting skill and the reading network are neuroanatomically linked in the developing brain (Palmis, Danna, Velay, & Longcamp, 2017), and intrinsic and extrinsic factors of handwriting jointly shape reading development (Xu & Liu, 2023). Typing versus handwriting in Chinese language learning yields distinct cognitive footprints (Lyu, Lai, Lin, & Gong, 2021)—a reminder that how a child produces script matters. A systematic review and meta-analysis of embodied action-language processing corroborates consistent sensorimotor recruitment across paradigms (Giacobbe, Raimo, Cropano, & Santangelo, 2022), closing the loop between the motor-comprehension and the print-recycling strands.

The two strands converge on a single principle: the brain that reads is the brain that acts—recycled, fine-tuned, and scaffolded by the motor system.

Two discipline notes. (1) Perception-side envelope-tracking findings (Goswami, 2011; Giraud & Poeppel, 2012) are used here only to establish that syllabic rhythm is neurally privileged for listeners; they are not cited as jaw-biomechanics evidence. (2) Of the neural energy-metabolism cluster, only Attwell and Laughlin (2001) are independently verifiable; their energy-budget analysis—showing that signaling consumes the large majority of cortical energy expenditure—is invoked solely as a biological-plausibility background for why recycling pre-existing articulatory–auditory circuitry is metabolically favored over building new networks. That study did not measure reading or speech-motor tasks, and we do not treat it as reading evidence.

The reading network is not purely visual; handwriting gestures are neurally coupled to word recognition. Perfetti & Tan (2013) reviewed evidence that the left premotor Exner’s area, involved in handwriting, is activated during the reading of handwritten words in both alphabetic and logographic scripts. This finding points to a universal sensorimotor component in reading that parallels the motor involvement in speech comprehension. It reinforces the idea that literacy co-opts motor programs originally evolved for action—consistent with the neuronal recycling hypothesis—and provides a direct link between motor learning (e.g., writing) and reading acquisition. This link, the present model argues, is anchored at the syllabic, jaw-driven level.

4. Discussion

4.1 The Embodied Anchor–Grain-Size Interface Model

We integrate the three tiers into a single model with five linked propositions, which are schematized in Figure 1.

Figure 1. The Embodied Anchor–Grain-Size Interface Model: convergence of three evidence tiers onto five linked propositions (P1–P5).

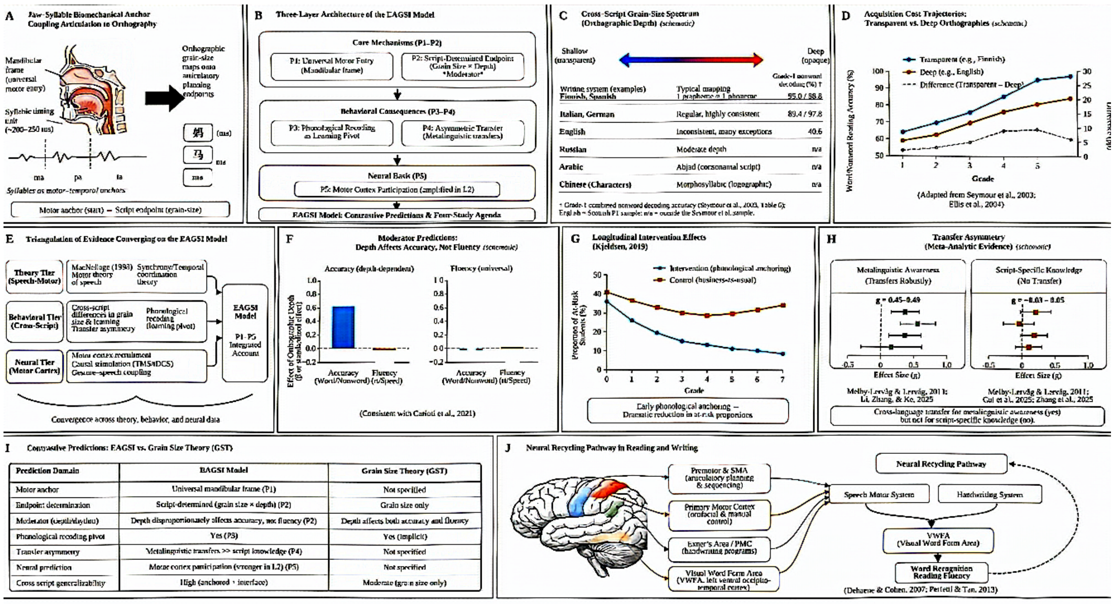

Figure 1. *Comprehensive multi-panel illustration of the Embodied Anchor–Grain-Size Interface (EAGSI) Model.*

Note. (A) Schematic of the jaw–syllable biomechanical anchor coupling articulation to orthography. (B) Three-layer architecture of the model: Core Mechanisms (P1–P2), Behavioral Consequences (P3–P4), and Neural Basis (P5). (C) Cross-script grain-size spectrum showing orthographic depth across writing systems. (D) Acquisition cost trajectories comparing transparent versus deep orthographies (adapted from Seymour et al., 2003; Ellis et al., 2004). (E) Triangulation of evidence from theoretical, behavioral, and neural tiers converging on the EAGSI model. (F) Moderator predictions: orthographic depth disproportionately affects accuracy (depth-dependent) but not fluency (universal), consistent with Carioti et al. (2021). (G) Longitudinal intervention effects showing reduced at-risk proportions following early phonological anchoring (Kjeldsen, 2019). (H) Transfer asymmetry: robust cross-language transfer for metalinguistic awareness ($g \approx 0.45$–0.49) versus null transfer for script-specific knowledge (Melby-Lervåg & Lervåg, 2011; Li, Zhang, & Ke, 2025). (I) Contrastive predictions distinguishing EAGSI from Grain Size Theory. (J) Neural recycling pathway from motor cortex and handwriting areas to the visual word form area (VWFA) (Dehaene & Cohen, 2007; Perfetti & Tan, 2013). The figure integrates theoretical, empirical, and predictive elements of the model; it is a conceptual synthesis, not a statistical model fit.
Panels C, D, F, and H are schematic visualizations rather than plots of source statistics; axis values and bar heights are illustrative only.

(P1) Universal entry point—with a boundary condition built in. Mandibular open–close oscillation provides the production-side frame of the syllable (MacNeilage, 1998; Svensson Lundmark & Erickson, 2024), and syllabic synchronization bundles consonantal, vocalic, and laryngeal gestures into one controllable unit (Xu, 2024)—making the syllable the maximal speech unit accessible to any human before literacy, subject to the boundary condition specified immediately below. This proposition would be falsified by the observation that syllable production in any spoken language systematically fails to exhibit a one-to-one correspondence with mandibular oscillatory cycles, or that a population of preliterate learners reliably accesses sub-syllabic units without prior explicit instruction.

(P2) Script-determined endpoint, with orthographic-depth/rhythm-type as an explicit boundary condition. Literacy 2004). Shallow alphabets require phoneme-level descent at a cost set by consistency (Ziegler & Goswami, 2005; Seymour et al., 2003). Chinese binds morphemes–syllables via phonetic-radical sub-character cues, with phonological recoding as the self-teaching pivot (Li et al., 2018; Xu & Ke, 2025). We build the moderator explicitly into P2 itself, rather than treating it as an afterthought: in stress-timed languages with syllable-boundary ambiguity (most notably English, where consonant clusters, ambisyllabicity, and vowel reduction blur the syllable's edges), the syllable-level anchor is expected to be less cleanly exploitable. Instruction in such languages may need to combine syllable-level and sub-syllabic (phoneme-level) framing rather than relying on the syllable anchor alone. Indirect support for this moderating role comes from Barbosa (2026), who found that prose reading—with its more flexible rhythmic structure—elicited larger and more variable jaw movements than poetry reading, which is constrained by line boundaries. This finding suggests that text structure, analogous to the grain size imposed by a writing system, systematically alters jaw-syllable dynamics. The efficiency of the jaw anchor thus appears to depend on prosodic and orthographic context, as the model predicts. The P2 boundary condition receives empirical support from a meta-analysis on developmental dyslexia (Carioti et al., 2021). That analysis showed that reading accuracy deficits in dyslexia are significantly moderated by an interaction between orthographic depth and age, whereas reading fluency deficits are requires binding symbols to speech at the grain size the target script encodes. Syllabaries bind the syllable directly—the fastest acquisition route (Ellis et al.,
universal across orthographies and persist into adulthood. This pattern aligns precisely with the model's prediction: the syllable-level motor anchor underpins fluency as a universal rhythmic process, while accuracy—which depends on mapping script to speech at finer grain sizes—is more susceptible to orthographic transparency. The dyslexia literature therefore, provides strong external validation of the model's core distinction. This proposition would be falsified by the observation that acquisition speed across scripts is unrelated to the grain size they encode (e.g., that a deep alphabet is acquired as rapidly as a syllabary), or that instruction targeting a smaller-than-encoded unit consistently outperforms grain-size-matched instruction at the entry stage.

(P3) Recoding as pivot. Phonological recoding drives orthographic learning. Self-teaching studies show that reading aloud, not silent inspection, locks print to speech, especially when phonetic radicals are transparent (Li, Li, & Wang, 2020; Xu & Ke, 2025). Recoding, in our account, is the engine that converts the jaw-drop anchor into orthographic knowledge across diverse scripts. This proposition would be falsified by the observation that silent reading produces equivalent or superior orthographic learning gains compared to reading aloud in self-teaching paradigms, or that orthographic learning proceeds normally in populations with impaired phonological recoding ability.

(P4) Transfer asymmetry. Metalinguistic awareness transfers across

languages; script-specific knowledge does not. The evidence is consistent (Melby-Lervåg & Lervåg, 2011; Li, Zhang, & Ke, 2025; Cui et al., 2025; Zhang et al., 2025). The syllable anchor is biomechanical—it carries phonology, not font. Universal currency travels; local bills must be minted locally. Script-specific overlays require local teaching. This proposition would be falsified by the observation that script-specific orthographic knowledge (e.g., grapheme–phoneme correspondence rules) transfers automatically across typologically unrelated scripts to the same degree as phonological awareness, or that metalinguistic awareness fails to show reliable cross-language transfer in bilingual populations.

(P5) Motor leverage: Motor cortex is functional, not epiphenomenal. Causal evidence for motor-cortex involvement in language comprehension comes from an online repetitive TMS study (Vukovic et al., 2017). When a 20-Hz rTMS burst was applied over the left primary motor cortex (hand area) within 200 ms of word onset, participants showed a task- and meaning-specific modulation of reaction times: semantic judgments of hand-action words were slowed, whereas abstract words were facilitated. No such effect was observed during lexical decision, and no effect occurred with right-hemisphere stimulation. This double dissociation demonstrates that M1 contributes causally and selectively to the retrieval of motor-semantic features—a pattern that aligns with the simulation theory of semantics and provides the strongest evidence to date that motor areas are functionally necessary for full meaning access (Vukovic et al., 2017; Johari et al., 2021; Pattamadilok et al., 2016), with effects stronger in L2 (Tian et al., 2023). P5 is further corroborated by studies showing that gesture–speech synchrony enhances articulatory magnitude and temporal stability (Garvin et al., 2025), and that beat-gesture alignment influences online lexical stress perception (Bujok, 2025) and even recalibrates subsequent auditory-only stress perception (Bujok, Peeters, & Ozyurek, 2025). This chain of evidence—from neurostimulation (Vukovic et al., 2017) to behavioral gesture–speech coupling (Garvin et al., 2025; Momsen et al., 2026)—indicates that motor engagement is functional and automatic and can induce lasting perceptual changes. That explains why production-based instruction outperforms observation: self-generated articulatory output is the active ingredient, not a mere bonus. This proposition would be falsified by the observation that disrupting motor cortex (via TMS, tDCS, or lesion) produces no detectable effect on language comprehension or written-word recognition, or that embodiment effects in instruction are entirely accounted for by attention/engagement confounds rather than by motor-system recruitment per se.

Causal chain: universal motor frame entry (bounded by P2) → grain-size alignment → recoding drives learning → metalinguistic transfer → motor leverage. Each link rests on measured evidence. The jaw opens, the syllable anchors, and the brain recycles—and the child reads. Table 1 summarizes the current evidence strength and principal limitation for each of the five propositions. For a direct visual comparison of evidence strength across the five propositions, see Supplementary Figure S1.

We underscore what the model is not: it is not a claim that grain-size theory and the jaw–syllable construct are interchangeable. They operate at different descriptive levels—grain-size theory describes how orthography recruits phonological units; the jaw–syllable construct describes which phonological units are biomechanically and consciously available before orthography arrives—and the interface between them (P2) is a hypothesis about instructional leverage, not a demonstrated identity. The mechanism is specified as the P1→P2 alignment above, together with the testable prediction that instruction targeting the largest jaw-salient unit a script encodes should outperform instruction targeting smaller units at the entry stage—exactly the comparison a syllable-versus-grapheme Hangul study instantiated, finding higher trained- and novel-word accuracy under syllable-focused instruction for English-speaking learners of Korean (Brennan & Heneman-Lee, 2025).

4.2 Limitations and information gaps

Five limitations bound every claim above. (1) The decisive experiment is missing. No study in the pool manipulates jaw-level motor training and measures cross-script literacy outcomes; the link between production-side universality evidence (theory + EMA) and instructional evidence (articulatory gestures, self-teaching) remains inferential, not experimental. (2) Intervention evidence carries a bias risk. Against RoB 2 dimensions: the single-case dyslexia study shows irrelevant-task improvement (testing effects); the Moved by Reading English-DLL implementation shows null transfer to an untrained text (Adams, Restrepo, & Glenberg, 2018), a boundary condition discussed in §3.3 alongside its Spanish-transparent-orthography counterpart (Adams, Glenberg, & Restrepo, 2019); the articulatory-segmentation studies are externally valid mainly for novices. (3) Heterogeneity and publication bias are uncorrected. Source records did not report I^2 or funnel diagnostics for the quoted meta-analyses; all g/SMD values should be read as upper-bound estimates pending second-order correction. (4) Pool composition skew. Chinese and English/European-orthography studies dominate; Arabic and Indic akshara appear only sparsely, limiting the reach of "cross-script" generalizations; one key record exists only as a conference abstract and should be weighted accordingly as preliminary evidence. (5) Live tensions remain unresolved. In Spanish, prereading phonological awareness failed to predict reading acquisition beyond letter knowledge (Zugarramurdi et al., 2022), complicating strong PA-centrism, and Wise et al.'s (1999) null articulation-specific effect in older remedial readers directly bounds the populations for whom articulatory scaffolds should be expected to help. Reading disability itself is multifactorial rather than syllable/PA-specific: the double-deficit hypothesis identifies phonological deficits and naming-speed deficits as separable sources of dyslexia (Wolf & Bowers, 1999), and in Chinese, orthographic–phonological mapping impairments and morphological-awareness deficits characterize developmental dyslexia in ways the jaw framework does not directly address (Shu, McBride-Chang, Wu, & Liu, 2006; Song, Gu, Song, & Quan, 2025). The "phonetic radicals versus coding systems" question is resolved convergently: both Li et al. (2018) and the Zhuyin-null result in Xu and Ke (2025) converge against auxiliary coding systems as self-teaching supports. The model's parameters are therefore population- and stage-conditional, not universal constants, and the orthographic-depth/rhythm-type moderator specified in P2 (§4.1) is itself an open empirical question rather than a settled boundary.

4.3 Practical implications—stated at the evidentiary confidence level they deserve

Three hypothesis-generating design principles, awaiting RCT validation, follow from the evidence reviewed above; none should be read as validated instructional prescriptions. (1) Syllable-level embodied activities are a plausible, hypothesis-generating universal front-end for cross-script instruction, worth testing because the syllable is the largest unit learners already embody (subject to the orthographic-depth boundary condition in P2) and because syllable/PA skills are among the most transferable components (Melby-Lervåg & Lervåg, 2011; Brennan & Heneman-Lee, 2025). Classroom-level evidence is at least consistent with the leverage of the syllable unit: syllable awareness combined with syllable-controlled text supports the development of finger-point reading in beginners (Mesmer & Lake, 2010). However, no direct trial of jaw-monitoring activities (e.g., chin-drop counting, syllable tapping) exists in the pool, so this remains a mechanism-derived hypothesis to be tested, not a validated prescription for classroom adoption. (2) Articulatory-gesture instruction may have stage-specific value for beginners' reading, as a testable design principle rather than an established recommendation—the Boyer and Ehri (2011) finding that only mouth pictures (not letters-only training) uniquely improved word reading, together with motor-cortex causal evidence, motivates trialing oral-motor imagery when binding speech to print, always with letters attached (recoding as pivot), never as oral-only training, and never as a substitute for systematic decoding instruction (see moderator analyses in Ehri et al., 2001). (3) Grain-size alignment is proposed as a design hypothesis to organize cross-script curricula: prioritize phonetic-radical syllabic cues for Chinese learners rather than leaning on auxiliary symbol layers (Li et al., 2018; Xu & Ke, 2025); teach multi-grain-size flexibility explicitly for deep alphabets; and exploit L1 metalinguistic awareness as an L2 facilitator (Li, Zhang, & Ke, 2025). For reading disability, articulatory-consciousness training shows promise but must be evaluated with transfer measures and irrelevant-task controls to exclude testing artifacts (Thurmann-Moe et al., 2021).

4.4 Future research agenda

Five studies would decisively advance the field. (1) A pre-registered cross-script RCT of a syllable-level oral-motor front-end (identical activities across scripts), spanning at minimum one syllabic and one morphosyllabic script, with transfer tests and irrelevant-task controls---the experiment whose absence is the field's central gap. (2) Longitudinal motor–neural tracking combining EMA jaw kinematics, prereading phonological awareness, and first-year VWFA emergence (longitudinal fMRI; Dehaene-Lambertz et al., 2018), testing the developmental continuity of motor frame → syllable awareness → print mapping. (3) Adjudication studies with pre-registered moderators addressing the Spanish PA-prediction failure (Zugarramurdi et al., 2022) and the boundary conditions of syllable-anchor leverage across scripts, including the orthographic-depth/rhythm-type moderator formalized in P2. (4) A formal second-order meta-analysis---which the present synthesis explicitly does not attempt---applying trim-and-fill/selection-model corrections and heterogeneity quantification to the embodied-instruction effect sizes surveyed descriptively here, separating genuine effects from reporting bias; this is precisely the quantitative undertaking that would convert the present integrative model into a fully quantified meta-analysis. (5) A longitudinal neuroimaging study that combines infant or kindergarten EMA jaw kinematics with resting-state or task-based fMRI (or fNIRS/EEG) and subsequent literacy measures would provide a direct test of the model‘s developmental and universal claims. Following the broader machine-learning framework reviewed by Nakai et al. (2024), such a study could determine whether early individual differences in jaw–syllable coupling predict later reading fluency and whether this prediction is moderated by orthographic depth---offering a decisive test of the model’s core predictions.

5. Conclusions

The EAGSI model (Figure 1) converts a diffuse set of previously disconnected findings—spanning speech-motor theory, cross-script behavioral evidence, and motor–language neuroscience—into a single falsifiable model with a concrete five-study test agenda. This is the contribution the synthesis is intended to make available now, rather than after the decisive experiment exists. Three conclusions are warranted. First, the jaw–syllable coupling is a strong—but contested—candidate universal: one mandibular oscillation per syllable with frame/content dissociation is supported by theory and articulography (MacNeilage, 1998; Svensson Lundmark & Erickson, 2024), while perception-side neural privilege for syllabic rhythm is established (Goswami, 2011; Giraud & Poeppel, 2012) but must not be misread as production-side biomechanical proof; this universal candidacy is bounded by the orthographic-depth/rhythm-type moderator formalized in P2. Second, the critical variable in cross-script literacy acquisition is the interface between this embodied anchor and the grain size each script encodes: syllabaries dock directly (Ellis et al., 2004), alphabets must descend to phonemes at a consistency-dependent cost (Ziegler & Goswami, 2005; Seymour et al., 2003), and Chinese docks via phonetic-radical sub-character cues rather than auxiliary coding systems (Li et al., 2018; Xu & Ke, 2025); Phonological recoding is the shared orthographic-learning pivot (Li, Li, & Wang, 2020), and metalinguistic awareness is the strongest transferable component, with broadly convergent moderate-to-large meta-analytic magnitudes across independent evidence lines (including a 156-study facilitation effect and a corrected d = 0.49 for phonological co-activation in written word recognition; Li, Zhang, & Ke, 2025; Cui et al., 2025). Third, the motor cortex functionally participates in language comprehension and written-word recognition—more strongly in L2 (Tian et al., 2023)—and embodied instruction yields reliable, graded effects (largest for self-produced movement; Dargue et al., 2019) that are stage-contingent for articulatory scaffolds (novices benefit; older remedial readers show null unique effects; Wise et al., 1999).

The responsible verdict is neither an enthusiastic “instructional framework” nor dismissal: the jaw-drop mapping with the syllable is a plausible, mechanistically specified, and empirically tractable universal anchor for cross-script literacy acquisition, awaiting its decisive direct test. By formalizing this embodied anchor within the vocabulary of cognitive mechanisms (universal entry, grain-size flexibility, recoding pivots, transfer asymmetry, and functional motor leverage), our synthesis fundamentally reframes literacy acquisition not as a disembodied symbolic computation, but as a process of anchoring arbitrary visual symbols to an evolutionarily conserved, phylogenetically privileged sensorimotor apparatus. The theoretical contribution of the EAGSI model lies precisely in its insistence that the biomechanical dynamics of the speech act—rather than merely the resultant acoustic signal—actively shape the cognitive architecture supporting reading. This conception carries deep implications for how cognitive science understands the multimodal, body-based nature of representational systems, inviting an integrative agenda that directly ties motor planning to the emergence of abstract linguistic competence.

Acknowledgements: Omitted for blind review.

Appendix A: Search Strategy

Databases searched: Web of Science (Core Collection), Scopus, PubMed/MEDLINE, PsycINFO, and CNKI (for Chinese morphosyllabic-script sources), supplemented by Google Scholar for grey literature and forward-citation chaining on core theoretical sources.

Search date windows: Initial collection wave: [1995-2026]. Second collection wave: [2015-2026], targeting newly indexed 2024–2025 records identified through forward-citation alerts on seminal papers (MacNeilage, 1998; Ziegler & Goswami, 2005; Melby-Lervåg & Lervåg, 2011).

Total initial hits: 3150 records across all databases before deduplication.

Records after title/abstract screening:150.

Total unique records included in the final evidence base:128 (deduplicated by title and bibliographic verification).

Representative search strings (by thematic cluster):

(i) Speech-motor and syllable theory: ("frame/content theory" OR "syllable frame" OR "mandibular oscillation" OR "jaw kinematics") AND (speech OR articulation OR syllable)

(ii) Cross-script reading and orthographic depth: ("orthographic depth" OR "grain size" OR "psycholinguistic grain size theory") AND (reading OR literacy) AND (cross-linguistic OR "cross-script" OR orthograph)

(iii) Embodied instruction and gesture: ("embodied learning" OR "embodied cognition" OR gesture OR "motion-sensing") AND (reading OR literacy OR "language learning") AND (meta-analysis OR "systematic review")

(iv) Neural evidence on script-sound mapping: (fMRI OR TMS OR tDCS OR EEG OR ERP) AND ("visual word form area" OR "motor cortex" OR "premotor cortex") AND (reading OR literacy OR "word recognition")

(v) Quantitative cross-language transfer syntheses: ("cross-language transfer" OR "biliteracy" OR "metalinguistic awareness") AND (meta-analysis OR "structural equation model") AND (reading OR phonological)

Inclusion/Exclusion operationalization: Records were retained only if bibliographically verifiable (author/year/venue confirmed via database cross-checking or publisher DOI resolution) and directly relevant to one of the six thematic clusters. Conference abstracts were retained only where explicitly flagged as preliminary evidence in the main text. Records failing bibliographic verification were excluded from the evidence base.

Note: Consistent with the scope of a theoretical narrative synthesis rather than a formal systematic review, no dual-reviewer screening or formal PRISMA flow diagram was produced; the transparency level reported herein meets the reproducibility standard for a SWiM-adjacent narrative integration.

Table A1.
Cross-domain Effect-size Summary (descriptive, not re-pooled)

Construct / Domain	**Effect size (as reported)**	**k studies / N**	**Source**
Embodied learning, general achievement	g = 0.52, with reduced cognitive load	17 studies, N = 1,046	Lyu & Deng, 2024
Embodied learning, general achievement	g = 0.406, 95% CI [0.264, 0.548]	46 studies	Liu, Zuo, Zhao, & Lu, 2025
Embodied effectsinver bprocessing/learning	Significant, moderate (three-level model)	not separately reported here	Huang, Zhai, Jiang, Wang, & Li, 2026
Motion-sensing-technology-assisted language learning	Larger, sensorimotor-channel-specific effect	not separately reported here	Yu, 2024
Gesture effects on comprehension (meta-analysis)	Positive; larger for self-produced gesture	meta-analytic	Dargue, Sweller, & Jones, 2019
Chinese self-teaching: phonetic radicals (orthographic choice)	g = 0.89	13 studies, 26 samples, N = 1,403	Xu & Ke, 2025
Chinese self-teaching: phonetic radicals (character writing)	g = 0.97	13 studies, 26 samples, N = 1,403	Xu & Ke, 2025
Chinese self-teaching: semantic radicals (orthographic choice /	g = 0.18/ g = 0.55	13 studies, 26 samples, N = 1,403	Xu & Ke, 2025

developmental dyslexia: Morphological awareness as a core cognitive construct. *Journal of Educational Psychology, 98* (1), 122–133.

[95]Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist, 17* (2), 197–208.

[96]So, W. C., Sim, C.-H. C., & Low, J. W.-S. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes, 27* (5), 665–681.

[97]Song, Y., Gu, J., Song, S., & Quan, X. (2025). Orthographic-phonological mapping impairments in Chinese children with developmental dyslexia: Insights from an ERP investigation. *Annals of Dyslexia, 75 (*2), 289–310.

[98]Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., 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., McAleenan, A., & Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomized trials. *BMJ, 366*, l4898.

[99]Svensson Lundmark, M., & Erickson, D. (2024). Segmental and syllabic articulations: A descriptive approach. *Journal of Speech, Language, and Hearing Research, 67* (10S), 3974–4001.

[100]Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping, 25* (1), 83–91.

[101]Testolin, A., Stoianov, I., & Zorzi, M. (2017). Letter perception emerges from unsupervised deep learning and recycling of natural image features. *Nature Human Behaviour, 1* (9), 657–664.

[102]Thurmann-Moe, A. C., Melby-Lervåg, M., & Lervåg, A. (2021). The impact of articulatory consciousness training on reading and spelling literacy in students with severe dyslexia: An experimental single case study. *Annals of Dyslexia, 71* (1), 1–22.

[103]Tian, L., Chen, H., Heikkinen, P.-P., Liu, W., & Parviainen, T. (2023). Spatiotemporal dynamics of activation in motor and language areas suggest a compensatory role of the motor cortex in second language processing. *Neurobiology of Language, 4* (1), 178–197.

[104]Tong, X., & McBride, C. (2018). Toward a graded psycholexical space mapping model: Sublexical and lexical representations in Chinese character reading development. *Journal of Learning Disabilities, 51* (5), 482–489.

[105]Tong, X., Tong, X., & McBride, C. (2017). Radical sensitivity is the key to understanding Chinese character acquisition in children. *Reading and Writing, 30* (6), 1251–1265.

[106]Treiman, R. (1994). Use of consonant letter names in beginning spelling. *Developmental Psychology, 30 (*4), 567–580.

[107]Verhoeven, L., Perfetti, C., & Pugh, K. (Eds.). (2019). *Developmental dyslexia across languages and writing systems.* Cambridge University Press.

[108]Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36* (3), 1–48.

[109]Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally

contributes to language comprehension: An online rTMS study. *Neuropsychologia, 96*, 222–229.

[110]Wang, M., Cheng, C., & Chen, S.-W. (2006). Contribution of morphological awareness to Chinese–English biliteracy acquisition. *Journal of Educational Psychology, 98* (3), 542–553.

[111]Wang, M., Park, Y., & Lee, K. R. (2006). Korean–English biliteracy acquisition: Cross-language phonological and orthographic transfer. *Journal of Educational Psychology, 98* (1), 148–158.

[112]Wang, M., Perfetti, C. A., & Liu, Y. (2005). Chinese–English biliteracy acquisition: Cross-language and writing system transfer. *Cognition, 97* (1), 67–88.

[113]Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: Overview of a topical collection in CR:PI. *Cognitive Research: Principles and Implications, 2* (1), 38. https://doi.org/10.1186/s41235-017-0071-6

[114]Wise, B. W., Ring, J., & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology, 72* (4), 271–304.

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

[116]Xu, X., & Ke, S. E. (2025). What is Chinese orthographic learning via self-teaching? A systematic review and meta-analysis. *Reading and Writing.* https://doi.org/10.1007/s11145-025-10739-4

[117]Xu, Y. (2024). Syllable as a synchronization mechanism that makes human speech possible. *Brain Sciences, 15* (1), 33.

[118]Xu, Z., & Liu, D. (2023). *Intrinsic and extrinsic factors of handwriting and their impacts on reading: A view from embodied cognition.* In the Routledge International Handbook of Visual-Motor Skills, Handwriting, and Spelling (pp. 136–148). Routledge.

[119]Yang, M., Cooc, N., & Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education, 2* (1), 15.

[120]Yu, Q. (2024). Effects of motion-sensing technology on language learning: Evidence from a meta-analysis. *Interactive Learning Environments, 32* (10), 7507–7523.

[121]Zhang, J., Li, H., Dong, Q., Xu, J., & Sholar, E. (2016). Implicit use of radicals in learning characters for nonnative learners of Chinese. *Applied Psycholinguistics, 37* (3), 507–527.

[122]Zhang, N., Wu, Z., & Wang, M. (2025). Cross-language phonological activation in bilingual visual word recognition: A meta-analysis. *Psychonomic Bulletin & Review, 32* (5), 2059–2089.

[123]Zhou, L., Fong, M. C. M., Minett, J. W., Peng, G., & Wang, W. S.-Y. (2014). Pre-lexical phonological processing in reading Chinese characters: An ERP study. Journal of *Neurolinguistics, 30*, 14–26.

[124]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.

[125]Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the orthography: Revisiting the contribution of prereading phonological awareness to reading acquisition. *Developmental Psychology, 58* (6), 1003–1016.