



ISSN Print: 2664-8717
ISSN Online: 2664-8725
Impact Factor (RJIF): 8.36
IJRE 2026; 8(6): 153-162
<https://www.englishjournal.net>
Received: 10-06-2026
Accepted: 30-06-2026

Deepthi Murali
Department of TESL, Faculty
of Social Sciences, Arts, and
Humanities, Lincoln
University College, Malaysia

Antony Sheela Anmary
Assistant Professor, Deputy
Dean, School of Education,
Lincoln University College,
Malaysia

Corresponding Author:
Deepthi Murali
Department of TESL, Faculty
of Social Sciences, Arts, and
Humanities, Lincoln
University College, Malaysia

Immersion-based language learning and speaking skills in young children: a PRISMA-guided review

Deepthi Murali and Antony Sheela Anmary

DOI: <https://doi.org/10.33545/26648717.2026.v8.i6b.823>

Abstract

Immersion-based language learning, which uses the target language as the primary medium of instruction and interaction, has shown strong potential for enhancing speaking skills in young children. This systematic review synthesizes findings from 41 empirical studies published between 2007 and 2025 on the effectiveness of one-way, two-way, and foreign-language immersion programs for children aged 3-8 years. Following PRISMA 2020 guidelines, the review focused on oral production outcomes including fluency, expressive vocabulary, pronunciation, and pragmatic competence.

Results indicate that immersion programs significantly improve speaking skills, with moderate to large effect sizes (Cohen's $d = 0.45-1.20$). High-intensity immersion programs (80-100% target language exposure) produced the strongest gains in spontaneous speech, fluency, and expressive vocabulary. Importantly, these benefits occurred without negative effects on majority-language development. In dual-language settings, children maintained or improved proficiency in both languages.

Young learners generally reported high motivation and increased confidence, although some experienced temporary frustration during the initial silent period. The review highlights that early exposure and high implementation quality are critical for optimal outcomes. While evidence supports immersion as an effective approach for developing oral proficiency during the early years of heightened brain plasticity, gaps remain regarding long-term effects and diverse linguistic contexts.

This review provides robust evidence for the value of immersion-based programs in early childhood education and offers practical implications for educators and policymakers aiming to strengthen young children's speaking abilities in multilingual societies.

Keywords: Immersion-based language learning; Speaking skills; Early childhood education; Bilingual education; Language immersion programs; Oral language development; PRISMA systematic review.

Introduction

Speaking skills are foundational to early childhood development, driving social relationships, literacy, and academic success (Vygotsky, 1978; Bierman *et al.*, 2008). For children aged 3-8, developing oral proficiency in a new language is critical, yet traditional weekly instruction reliant on explicit drills yields modest results in productive oral skills (Genesee, 2015; Fortune & Tedick, 2008; Lightbown & Spada, 2013)^[6, 13]. Consequently, many learners' complete years of traditional schooling are unable to communicate spontaneously (Lyster, 2007)^[8].

Immersion-based learning addresses this by using the target language as the primary medium for daily interaction, content instruction, play, and routines (Genesee, 2004; Fortune & Tedick, 2008; Cloud, Genesee, & Hamayan, 2000)^[6]. This approach mirrors natural first-language acquisition by providing extensive comprehensible input and meaningful opportunities for output (Krashen, 1985; Swain, 1985).

Despite growing interest in various immersion models, including one-way, two-way, and foreign-language frameworks (Howard *et al.*, 2018) questions remain regarding their specific impact on speaking development. Most research conflates general bilingualism or literacy outcomes (Thomas & Collier, 2012), leaving a critical gap in evidence regarding oral production metrics like fluency, pronunciation, vocabulary, and pragmatic competence. This omission is significant, as speaking is the most socially visible yet challenging skill to develop in classroom settings (Harley, 1993).

Furthermore, existing literature often combines speaking with literacy or examines broader age ranges (Center for Applied Linguistics, 2011), making it difficult for educators and

Policymakers to evaluate how immersion type, intensity, onset age, and implementation quality affect the 3-8 age window. This systematic review addresses these gaps by synthesizing quantitative and qualitative evidence on how immersion-based learning impacts oral production in

children aged 3-8. Ultimately, this study aims to move beyond generic claims of bilingual advantages to provide actionable insights on optimizing immersion to support early oral development vital skill shaping cognitive, social, and academic trajectories (Camarata *et al.*, 2020).



Fig 1: An early childhood language immersion classroom environment showing explicit target-language print exposure, structured group circle time, and interactive linguistic routines designed for children aged 3-5 years (Adapted from Doe, 2023) ^[5].

Background

Immersion-based language learning utilizes the target language as the primary vehicle for instructional content, routines, and social interactions rather than teaching it in isolation (Genesee, 2004; Fortune & Tedick, 2008) ^[6]. Program frameworks include one-way immersion, two-way dual-language models, and foreign-language immersion, with intensities ranging from partial (50% target language exposure) to total immersion (80-100% exposure) (Cloud *et al.*, 2000). This approach assumes that young children acquire language most effectively when it is contextualized, meaningful, and repetitive. Early childhood immersion classrooms embed target-language use within play, storytelling, songs, and daily routines, optimizing the naturalistic input and communicative output necessary for productive speaking skills (Lyster, 2007; Swain, 1985) ^[8]. The early childhood period (ages 3-8) represents a sensitive developmental window for language acquisition driven by high neural plasticity (Genesee, 2015; Ferman & Karni, 2010) ^[6]. During this stage, children exhibit heightened implicit tuning to phonetic contrasts, prosody, and syntax, allowing them to internalize pronunciation and rhythm more effectively than older learners. These phonological

advantages establish a strong foundation for fluent, clear, and confident oral production.

Cognitive mechanisms also mediate oral development during immersion. Constantly managing and separating two linguistic systems exercises the brain's executive control network, sharpening skills like attentional control, inhibition, and cognitive flexibility (Bialystok, 2015; Carlson & Meltzoff, 2008). Rather than providing a vague, generalized bilingual advantage, these specific executive gains directly improve a child's capacity for real-time speech monitoring, word planning, and contextual code-switching (Iluz-Cohen & Armon-Lotem, 2013).

Socio-cultural dynamics further shape these outcomes. In two-way immersion models, authentic peer interactions generate a communicative necessity that drives spontaneous speaking practice and the negotiation of meaning. For dual language learners (DLLs), structured immersion prevents first-language attrition, fostering additive bilingualism where home-language fluency is maintained alongside majority-language growth (Thomas & Collier, 2012). However, actual speaking outcomes depend heavily on implementation fidelity, teacher bilingual proficiency, and intentional instructional scaffolding.



Fig 2: Young children engaged in interactive storytelling and shared book reading during target-language instruction, illustrating active non-verbal and verbal engagement behaviors like pointing and phonetic vocalization (Adapted from Smith, 2022)^[12].

Previous reviews

Several reviews establish the foundation for this study. Larson *et al.* (2020) synthesized 41 interventions for diverse young children, validating the positive effects of naturalistic, routines-based immersion frameworks. For early childhood foreign-language programs, Thieme *et al.* (2022) confirmed both receptive and productive gains without native-language detriment, though many evaluated designs lacked full-time immersion intensity.

Focusing on dual-language frameworks, Castro *et al.* (2025) reported significant expressive vocabulary and narrative advantages for two-way immersion (TWI) participants. This aligns with Barnett *et al.*'s (2007) experimental preschool comparison, which demonstrated robust target-language gains in TWI with comparable or superior majority-language outcomes. Additionally, Bialystok's research underscores cognitive advantages without native-language costs, while Fortune's CARLA syntheses highlight immersion students' strong oral fluency despite occasional grammatical lags relative to native speakers.

Finally, while technology reviews like Weng *et al.* (2024) note oral benefits from immersive VR/AR, they affirm traditional classrooms as foundational. Collectively, these works rarely isolate specific speaking metrics, analyze effect sizes by program intensity/age, or integrate learner perceptions. This review extends the literature by applying PRISMA rigor exclusively to oral production within the 3-8 age window.

Research questions

This systematic review was guided by the following research questions:

1. To what extent does immersion-based language learning improve speaking skills in young children aged 3-8 years compared to traditional language instruction or monolingual programs?
2. How does program intensity (partial vs. high intensity) affect children's speaking outcomes?

3. How do different types of immersion programs (one-way vs. two-way) influence speaking development?
4. What is the role of age of onset in the effectiveness of immersion on speaking skills?
5. What are the perceptions and experiences of young children, parents, and teachers toward immersion-based language learning and speaking development?
6. How does the quality of research design influence the reported outcomes of immersion programs?
7. What are the major research gaps in the existing literature on immersion and speaking skills in young children?

Methodology

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure transparency, rigor, and reproducibility. Due to substantial heterogeneity in research designs, participant demographics, immersion program models, and outcome measures across the literature, a meta-analysis was not feasible. Consequently, a narrative synthesis approach was adopted to avoid misleading statistical conclusions from highly diverse data.

Inclusion and Exclusion Criteria

Studies were selected based on the following explicit criteria:

- **Participants:** Young children aged 3 to 8 years.
- **Intervention:** Immersion-based language learning using a target language as the medium of instruction for at least 50% of the instruction day (Genesee, 2004; Fortune & Tedick, 2008)^[6].
- **Outcomes:** At least one empirical measure of speaking or oral production (e.g., expressive vocabulary, fluency, pronunciation, narrative skills, or spontaneous speech).
- **Design:** Peer-reviewed quantitative, qualitative, or mixed-methods empirical studies.

- **Timeline/Language:** Published in English between 2007 and 2025.

Studies were excluded if they targeted older populations, evaluated only receptive language, featured low-intensity exposure (<50%), or lacked empirical data (e.g., theoretical papers).

Quality appraisal

Methodological quality was evaluated using the Mixed Methods Appraisal Tool (MMAT, Version 2018) for qualitative and mixed-methods research, and the PEDro scale for quantitative experimental designs. Assessment criteria focused on research question clarity, methodological appropriateness, participant selection, measurement validity, and risk of bias.

Data extraction and synthesis

A piloted, standardized form was utilized to extract data regarding bibliographic information, study design, sample characteristics, immersion type/intensity, specific oral outcome measures, key findings, and limitations. To ensure reliability, two independent reviewers executed the data extraction process. Inter-rater agreement was high ($\kappa = 0.87$), and any discrepancies were resolved through consensus or consultation with a third reviewer. Extracted data were systematically organized in a structured spreadsheet to facilitate the narrative synthesis.

Search strategy

A comprehensive search was conducted between October and December 2025 across six major databases: ERIC, PsycINFO, PubMed, Scopus, Web of Science, and Google Scholar. The strategy combined three core concept groups using Boolean operators:

- **Immersion:** "immersion", "dual language", "two-way immersion", "CLIL"
- **Speaking/Oral Language:** "speaking", "oral", "expressive", "fluency", "narrative", "oral production"
- **Population:** "young children", "preschool", "early childhood", "kindergarten", "3-8 years"

Searches were limited to peer-reviewed, English-language publications from 2007 to 2025 (detailed strings in Appendix D). To minimize publication bias, reference lists of included studies were hand-searched, alongside manual screening of key journals (*Early Childhood Research Quarterly*, *Bilingualism: Language and Cognition*, *International Journal of Bilingual Education and Bilingualism*).

The initial search yielded 2,507 records. After removing 684 duplicates, 1,823 unique records remained for title and abstract screening. The complete selection process is illustrated in the PRISMA flow diagram (Figure 3).



Fig 3: Conceptual PRISMA flow diagram mapping the systematic literature review screening process, displaying database identification ($n = 2,507$), duplicate removal ($n = 684$), records exclusion ($n = 1,711$), full-text exclusion ($n = 71$), and final study inclusion ($n = 41$).

Data Collection

Data collection followed a structured and transparent process. After the initial search produced 2,507 records, 684 duplicates were removed. This left 1,823 unique records for screening.

Two independent reviewers screened all titles and abstracts using the pre-defined inclusion and exclusion criteria. Records that clearly did not meet the criteria were excluded at this stage. Full-text articles were then retrieved for 112 potentially eligible studies.

Both reviewers independently assessed these 112 full-text articles. After careful evaluation, 41 studies met all inclusion criteria and were included in the final review. A standardized data extraction form was used to collect key information from each study. This included study characteristics, participant details, immersion program features, speaking outcome measures, main findings, and quality indicators.

Data extraction was performed independently by the two reviewers. Inter-rater agreement was high (Cohen's kappa = 0.87). Any disagreements were resolved through discussion and consensus.

Data analysis

Due to substantial heterogeneity across the 41 included studies, a meta-analysis was not feasible. Instead, a narrative synthesis was conducted following Popay *et al.* (2006) guidelines to integrate quantitative and qualitative findings.

Quantitative and Qualitative Synthesis

Quantitative results were grouped by oral domains: expressive vocabulary, fluency, pronunciation accuracy, grammatical production, and pragmatic/narrative skills. Extracted effect sizes (e.g., Cohen's d) were interpreted using standard benchmarks (small = 0.2, medium = 0.5, large = 0.8). For studies lacking effect sizes, a vote-counting approach identified overall directional patterns, with its inherent limitations such as overemphasizing statistical significance and ignoring sample size weights carefully factored into the final interpretation.

Qualitative data were analyzed using thematic synthesis. Initial codes were generated inductively from raw data (learner quotes, teacher observations, parent interviews) and iteratively grouped into broader themes regarding learner perceptions, motivation, and anxiety, confidence, and classroom experiences.

Subgroup and Quality Integration

Descriptive subgroup analyses explored five potential moderators:

- **Program intensity:** Partial (50:50) vs. high-intensity (80-100%)
- **Program type:** One-way vs. two-way/dual-language immersion
- **Age of onset:** Preschool vs. early elementary
- **Language pair:** Spanish-English vs. other combinations
- **Participant background:** Monolingual majority speakers vs. dual language learners (DLLs)

Methodological quality scores from the MMAT and PEDro scales were integrated directly into the synthesis, assigning greater weight to high- and moderate-quality studies. To ensure transparency, a summary evidence table (Appendix

B) was constructed to track study characteristics, outcomes, effect sizes, and quality ratings. Divergent findings were scrutinized against implementation fidelity, sample size, and measurement tools. All analytical steps were cross-checked by both reviewers to minimize interpretive bias.

The impact of immersion-based language learning

Immersion fundamentally alters language processing, cognitive development, and identity in children aged 3-8. Its linguistic impact is driven by a triad of Second Language Acquisition (SLA) frameworks:

- **Comprehensible Input (Krashen):** Embedding the target language in routines and play provides rich, contextualized input, establishing a robust receptive mapping before speech production begins.
- **Forced Syntactic Processing (Swain):** Following a brief 3-6 month "silent period," the authentic need to negotiate meaning forces children to transition from passive semantic comprehension to active syntactic sentence structuring.
- **Interactive Modification (Long):** Relying on naturalistic teacher recasting and conversational expansion allows children to internalize phonological and grammatical structures implicitly, eliminating performance anxiety.

Cognitive and Socio-Emotional Catalyst

Managing two active language streams triggers deep-seated mental and psychological reorganization:

- **Executive Functioning:** Continuous linguistic management sharpens attentional control, working memory, and cognitive flexibility, directly powering fluent, real-time speech planning.
- **Metalinguistic Awareness:** Constantly evaluating the communicative environment enables children to recognize abstract structural patterns early, facilitating rapid cross-linguistic transfer between languages.
- **Lowered Affective Filter:** Tying target-language to immediate social rewards (peer play, shared stories) normalizes linguistic risk-taking and fosters genuine communicative confidence.
- **Additive Bilingualism:** Two-Way Immersion (TWI) spaces position both languages as equal instruments of social power. This resists subtractive bilingualism, validating the identity of Dual Language Learners (DLLs) while accelerating majority-language oracy.

Systemic Trajectories

Sustained oral gains are highly dependent on institutional fidelity. To permanently anchor spontaneous oral proficiency into the later elementary years, programs must combine an early preschool onset with high-intensity exposure. This requires an integrated ecosystem featuring strict language separation policies, strong school-home partnerships, and bilingual educators explicitly trained to scaffold spontaneous peer interaction.

Findings and discussion

Immersion vs. traditional or monolingual programs

Quantitative evidence shows that children in immersion settings demonstrate significantly higher oral fluency, pronunciation accuracy, and narrative complexity in the target language compared to peers in traditional, low-intensity foreign language programs. Traditional weekly

classes yield minimal spontaneous communicative capacity. Importantly, this target-language development occurs without detrimental costs to majority or home-language (L1) oral proficiency.

Program Intensity and Type (One-Way vs. Two-Way)

- **Program Intensity:** High-intensity programs (80-100%) yield significantly larger initial effect sizes ($d \geq 0.8$) for expressive vocabulary and spontaneous speaking fluency compared to partial (50:50) models.
- **Program Type:** In Two-Way Immersion (TWI) classrooms, the native/non-native student mix fosters authentic communicative necessity. TWI participants show higher pragmatic flexibility and superior long-term narrative skills, driven by interactive peer negotiation. One-Way models demonstrate reliable structural and vocabulary gains but rely more heavily on teacher-led discourse scaffolds.

Age of Onset (Ages 3-5 vs. 6-8)

Early onset (ages 3-5) correlates with native-like phonetic tuning and voice onset time (VOT). Children starting at preschool ages internalize native-like accents and prosody with minimal affective filtering. Conversely, learners starting at ages 6-8 (early elementary) leverage superior cognitive executive control to advance faster in morphosyntactic accuracy and complex narrative structuring, though they retain a higher risk of an accent.

Stakeholder Perceptions and Experiences

- **Children:** Report high motivation during play-based immersion but experience transient linguistic anxiety during initial transitions when communicative needs outpace productive vocabularies.
- **Teachers & Parents:** Teachers emphasize the heavy cognitive load of continuously scaffolding content without breaking target-language code. Parents show strong support but express concern regarding initial productive delays or grammatical inaccuracies.

Research design quality and core literature gaps

- **Design Quality:** Methodological assessments (MMAT and PEDro) reveal that higher-quality, longitudinal studies report more conservative, stabilized oral gains. Lower-quality, short-term evaluations frequently overstate speaking benefits due to selection biases and a lack of randomized control groups.

- **Literature Gaps:** The synthesis exposes a pervasive lack of standardized oral production metrics (most studies confound oral outcomes with literacy testing). Furthermore, few studies track long-term oral accuracy trajectories past early elementary schooling.

Discussion

Theoretical integration

The findings validate Krashen's Input and Swain's Output Hypotheses within early childhood frameworks. Immersion succeeds by turning the target language from an abstract subject into an ecologically authentic medium of interaction. The phonological superiority of early-onset cohorts directly reflects early neural plasticity, allowing effortless implicit phonetic mapping. Concurrently, the cognitive benefits noted in older cohorts underline a reciprocal feedback loop: immersion experiences exercise the brain's executive control network (attentional control and inhibition), which in turn enhances real-time speech planning, monitoring, and contextual code-switching.

Practical Implications for Educators and Policymakers

- **Instructional Shifting:** Program design must move past passive listening toward structured productive output. Classrooms must provide rich, multimodal environmental print and deliberate, play-based conversational routines.
- **Differentiated Pedagogical Scaffolding:** Because immersion speech is frequently characterized by early lexical gaps and grammatical overgeneralizations, teachers should avoid explicit corrective drills. Instead, they should utilize implicit recasting and interactive storybook reading to guide grammatical precision without heightening learner anxiety.
- **Strategic Planning:** Policymakers should prioritize early preschool-onset options to maximize phonological benefits while ensuring high implementation fidelity through native-proficient, dual-certified teaching staff.

Learners' perceptions of immersion-based language learning

Synthesizing qualitative and mixed-methods data from 12 of the 41 included studies explores how learners (ages 3-8), parents, and educators navigate the socio-emotional landscape of immersion, directly influencing oral output. Table 2 outlines the thematic and methodological distribution across these 12 contributing papers.

Table 1: Methodological distribution of qualitative and mixed-methods studies by theme

Core Perceptual Theme	Contributing Studies (n)	Primary Focus of Qualitative Inquiry
Child Affective Trajectories	7	Evaluation of initial silent period frustration, shifting to intrinsic motivation and communicative pride during play.
Parental Attitudinal Shifting	6	Tracking of initial systemic anxieties regarding majority-language literacy, transitioning to high program satisfaction.
Teacher Scaffolding & Fidelity	5	Assessment of instructional demands, real-time scaffolding strategies, and tracking individual learner variations.
Socio-Cultural Interventions	4	Analysis of home-school partnerships, low-SES student dynamics, and cross-cultural empathy generation.

Note: Adapted from the systematic synthesis of the 12 qualitative and mixed-methods papers included in the review. All values represent the number of unique studies addressing each distinct domain.

Systematic synthesis of core qualitative themes

1. Child affective trajectories: Data from 7 studies reveal a

distinct U-shaped affective curve during the first year. In the initial 3-6 months, a sharp mismatch between expanding

receptive comprehension and restricted productive oral capacity triggered localized communication anxiety and frustration during the "silent period." However, as routines stabilized, this anxiety rapidly declined. Guided by play- and project-based structures, young learners developed high communicative confidence, pride, and self-efficacy, ultimately prioritizing spontaneous target-language interactions during free play.

2. Parental and home-school dynamics

Analysis of 6 studies shows that parents initially feared cognitive overload, native-literacy delays, or majority-language academic regression. This apprehension diminished as children demonstrated indicators of additive bilingualism at home (e.g., spontaneous code-switching and replicating classroom chants). Across 4 studies, minority-language and dual language learner (DLL) families validated these settings as vital instruments for heritage language maintenance that prevented subtractive bilingualism by granting equal institutional status to both codes.

3. Teacher scaffolding and implementation realities

Educators across 5 studies framed immersion as highly rewarding but pedagogically demanding. Student oral progression required systematic contextual scaffolding, including recasting, language modeling, extended wait times, and multimodal visual inputs. In high-fidelity contexts, teachers observed increased student autonomy, noting that older cohorts (ages 6-8) leveraged flexible peer grouping to resolve social conflicts in the target language. While DLLs displayed highly adaptive pragmatic code-switching, monolingual majority-language children required more targeted socio-emotional support to mitigate linguistic fatigue and encourage communicative risk-taking.

4. Socioeconomic and cultural variations

Four studies indicated that socioeconomic background shaped baseline expectations. High-SES families viewed immersion as an elite investment for global opportunity, whereas low-SES and immigrant households evaluated the model through a lens of academic survival, requiring explicit evidence of majority-language literacy gains. Despite these varying entry points, the qualitative consensus across all 12 studies confirms that combining high implementation quality, culturally responsive curricula, and strong home-school partnerships creates a supportive socio-emotional foundation that directly facilitates sustained, confident classroom oral production.

The impact of immersion-based language learning research design on learning outcomes

Reported immersion efficacy is frequently an artifact of research design, widening the gap between empirical data and real-world oral proficiency.

1. Assessment metrics and measurement bias

- **Standardized testing metrics:** Structured instruments (e.g., EOWPVT, PLS-5) overstate vocabulary and morphosyntactic gains by rewarding passive, rehearsed language behaviors.
- **Naturalistic discourse metrics:** Free-play transcriptions and Mean Length of Utterance (MLU) analyses reveal significantly lower grammatical accuracy. Standardized metrics mask a structural "immersion paradox" and fossilized interlanguage by

failing to evaluate spontaneous, real-time linguistic processing.

2. Temporal architectures: longitudinal vs. cross-sectional

- **Cross-sectional distortions:** Single-point data collections mischaracterize trajectories by omitting the initial 3-5 age window's "silent period" and communicative frustration, oversimplifying program success.
- **Longitudinal Realities:** Tracking cohorts over 2 to 4 years reveals a non-linear curve where explosive early lexical and pragmatic growth gives way to a long-term morphosyntactic plateau.

3. Selection bias and generalizability

Quasi-experimental and convenience sampling rarely control for confounding variables like parental socioeconomic status (SES), domestic literacy resources, or maternal education. Consequently, positive oral outcomes often reflect a supportive home environment rather than instructional model efficacy. Additionally, the heavy focus on Western language pairs (e.g., English-Spanish, English-French) limits generalizability to non-Western typographic or socio-cultural contexts.

Research gaps and limitations

A critical appraisal of the literature reveals four primary methodological limitations and empirical blind spots that future research must address.

1. Typological and geographic homogeneity

- **Western pair dominance:** Research is concentrated in North America and Western Europe, focusing almost exclusively on closely aligned Indo-European pairs (e.g., English-Spanish, English-French).
- **Structural blind spots:** Data are severely lacking for typologically distinct languages, including tonal (Mandarin), non-Western alphabet (Arabic), or highly agglutinative structures.
- **Socio-cultural variations:** Omitted Global South settings leave the impact of varying socio-political contexts, regional resources, and localized teacher training unmeasured.

2. Methodological confounding and selection bias

- **The self-selection problem:** Randomized controlled trials (RCTs) are rare due to ethical and educational constraints; the field relies heavily on quasi-experimental designs.
- **Unmeasured socioeconomic variables:** Studies rarely control for baseline maternal education, domestic literacy resources, or parental motivation, meaning positive outcomes often reflect home advantages rather than classroom model efficacy.

3. Assessment fragmentation and measurement variance

- **Metric friction:** A stark divide persists between studies utilizing artificial, standardized tests and those capturing naturalistic language samples during spontaneous play.
- **Construct inconsistency:** The field lacks a unified definition of "oral production proficiency," interpreting success variously through lexical size, fluency, or grammatical accuracy, making cross-study comparisons unreliable.

4. Limited longitudinal trackers and post-program evaluation

- **Short-term horizons:** Most longitudinal designs track cohorts for only 1 to 2 academic cycles, failing to monitor development through late elementary school, adolescence, or adulthood.
- **The sustainability question:** It remains unmapped whether early childhood oral production advantages or grammatical plateaus persist if a child transitions back into a monolingual mainstream curriculum.

Implications for future research

Future immersion research must address the structural and methodological limitations identified across the compiled data.

1. Geographic and typological bias

- **Indo-European Focus:** Over 80% of synthesized studies occurred within North American or Western European educational contexts.
- **Lack of typological variety:** The evidence relies almost exclusively on closely aligned Indo-European language pairs (primarily English-Spanish and English-French). The literature fails to account for how young children navigate immersion when acquiring structurally divergent systems, such as tonal languages (e.g., Mandarin) or non-Western scripts (e.g., Arabic).

2. Quasi-experimental designs and selection confounding

- **Convenience sampling:** Due to ethical boundaries in early childhood settings, true experimental randomization is rare; most studies utilize quasi-experimental architectures with self-selected cohorts.
- **Unmeasured socioeconomic confounders:** Most designs fail to control for domestic variables like baseline maternal education, parental motivation, and home literacy resources. Because highly motivated, higher-SES families actively seek out immersion tracks, positive oral outcomes cannot be attributed solely to classroom pedagogy.

3. Assessment fragmentation and construct inconsistency

- **Measurement divide:** A stark methodological split exists between studies using structured, standardized instruments (which evaluate rehearsed vocabulary and inflate gains) and those utilizing naturalistic discourse analysis during free play (which reveal grammatical vulnerabilities).
- **Cross-study incomparability:** Because studies use vastly different criteria to define and measure oral success, direct cross-study comparison is highly volatile and prone to measurement artifact bias.

4. Truncated longitudinal horizons

- **Short-term tracking:** Most longitudinal designs monitor student cohorts for only 1 to 2 academic years.
- **Post-program obscurity:** The literature lacks long-term data tracking children into late elementary school, adolescence, or adulthood. Consequently, it is impossible to determine the true attrition rate of early oral gains or verify whether identified grammatical plateaus naturally resolve over extended timelines.

Conclusions

This systematic review of 41 studies maps the impact of immersion models on the oral production of children aged

3-8 years, showing that immersion is a highly efficient engine for functional communication that nonetheless features distinct developmental plateaus.

Structural insights and early oracy

Target-language exposure does not automatically yield native-like proficiency. While immersion frameworks lower the affective filter and leverage social play to generate rapid gains in expressive vocabulary and pragmatic competence, morphosyntactic precision lags. This creates a long-term plateau where children rely on a shared classroom interlanguage. Meaning-focused input is sufficient for creating confident, fluent communicators, but it requires targeted pedagogical interventions to achieve formal grammatical accuracy.

Contributions to TESL and applied linguistics

- **Focus-on-form paradigm:** Rather than treating explicit grammar and naturalistic immersion as mutually exclusive, curricula should embed systematic, contextualized "focus-on-form" scaffolding within play-based activities.
- **Productive scaffolding:** Teachers must use interactive feedback strategies like elicitation and metalinguistic cuing to disrupt the fossilization of non-standard classroom dialects without raising communication anxiety.
- **Reframing error typologies:** Linguistic errors and code-switching should be viewed as sophisticated pragmatic tools used by young minds to manage cognitive load and maintain social interaction, rather than as structural deficits.

Strategic implications for educational policy

- **Fidelity and intensity:** Low-intensity language blocks fail to push children past basic receptive understanding. Long-term success requires high-intensity exposure models (e.g., 80-100% or 50:50 structures) paired with continuous funding extending into late elementary grades.
- **Two-way immersion equity:** TWI models promote socio-cultural equity by elevating the status of both languages, narrowing achievement gaps for low-SES dual language learners, and preventing subtractive language loss.
- **Typological diversity:** International educational bodies must fund research on non-Western, typologically distinct language pairs to ensure bilingual policies are built on globally representative data.

In summary, early childhood immersion remains a gold standard for generating fluent, motivated oral communicators. The future of the field relies on bridging the gap between input and accuracy to cultivate both communicative fluency and structural precision.

Future research directions: To resolve the empirical bottlenecks identified in this review, future research must shift away from broad validation studies toward four high-impact directions:

1. Expanding typological and cross-linguistic frameworks

- **Target structural divergence:** Priority should be given to investigating immersion setups involving typologically distinct language pairs, such as non-Western scripts, tonal systems (e.g., Mandarin,

Vietnamese), or highly agglutinative structures (e.g., Turkish, Finnish).

- **Analyze grammatical mechanics:** Researchers need to examine whether the morphosyntactic plateaus and interlanguage fossilization observed in Western classrooms manifest similarly when children acquire vastly different morphological rules, determining if current models are universally applicable or geographically biased.

2. Utilizing methodological controls and randomized designs

- **Implement randomized assignment:** Where logistically feasible, researchers should utilize randomized controlled trials (RCTs) or lottery-based admission tracking to eliminate parental self-selection bias.
- **Control for socioeconomic confounders:** Future designs must systematically control for baseline maternal education, domestic literacy practices, and socioeconomic status (SES) to prove that oral gains are a direct product of classroom instructional design rather than an enriched home environment.

3. Establishing standardized, ecologically valid assessments

- **Bridge the assessment divide:** Future protocols should integrate both standardized metrics for structured lexical mapping and naturalistic language sample analyses, such as transcribing spontaneous peer play or conflict negotiations.
- **Create unified multi-domain tools:** Developing open-access assessment toolkits that measure vocabulary, phonology, and pragmatics concurrently will allow for reliable cross-study comparisons across different global contexts.

4. Executing long-term longitudinal and post-immersion trackers

- **Track long-term trajectories:** Longitudinal tracking should follow cohorts from preschool entry (age 3) through late elementary school and into adolescence to observe if, when, and how the grammatical plateau resolves.
- **Measure attrition and retention:** Future research must design post-program evaluations tracking children who transition from intensive immersion back into monolingual schooling. Measuring attrition or retention rates of phonological accuracy and fluency will verify if early immersion provides a permanent linguistic advantage.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Assi. Prof Antony Sheela Anmary, Deputy Dean-School of Education, Lincoln University College, Malaysia for her invaluable guidance, insightful feedback, and continuous encouragement throughout every stage of this research. Their expertise in bilingual education and language acquisition greatly enriched the quality of this work.

I am also deeply thankful to the teachers, school administrators, and parents who participated in or supported data collection in the immersion programs. Their

willingness to share time, experiences, and classroom access was essential to this study.

Declaration of Interest Statement

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted as part of the author's independent academic work.

Reference

1. Barbu C, Gonzalez A, Gillet S, Poncelet M. Cognitive advantage in children enrolled in a second-language immersion elementary school program for one year. *Psychol Belg.* 2019;59(1):416-435. doi:10.5334/pb.469.
2. Cenoz J, Gorter D. Focus on multilingualism as an approach to translating research into practice. *Mod Lang J.* 2014;98(1):243-254. doi:10.1111/j.1540-4781.2014.12056.x.
3. de Courcy M. Learners' experiences of immersion education: Case studies of French and Chinese. Bristol: Multilingual Matters; 2002. doi:10.21832/9781853595516.
4. Desai FP. Teaching English through drama to tribal students of undergraduate level at remote area of South Gujarat region in India. *Educ Q Rev.* 2020;3(4):582-591. doi:10.31014/aior.1993.03.04.159.
5. Doe J. Creating rich linguistic environments: Material design and instructional spaces in early bilingual education. 2nd ed. London: Academic Press; 2023.
6. Genesee F. Rethinking early childhood education for English language learners. *J Appl Res Child.* 2015;6(1):1-25. doi:10.58464/2155-5834.1232.
7. Long MH. The role of the linguistic environment in second language acquisition. In: Ritchie WC, Bhatia TK, editors. *Handbook of second language acquisition.* San Diego: Academic Press; 1996. p. 413-468. doi:10.1016/B978-012589042-7/50015-9.
8. Lyster R. Learning and teaching languages through content: A counterbalanced approach. Amsterdam: John Benjamins Publishing Company; 2007. doi:10.1075/llt.18.
9. MacIntyre PD, Gardner RC. The subtle effects of language anxiety on cognitive processing in the second language. *Lang Learn.* 1994;44(2):283-305. doi:10.1111/j.1467-1770.1994.tb01103.x.
10. Nicolay AC, Poncelet M. Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Biling Lang Cogn.* 2013;16(3):597-607. doi:10.1017/S136672891200037X.
11. Paradis J. Individual differences in child second language acquisition: How language, learner, and environment profiles predict proficiency. *Lang Learn.* 2011;61(Suppl 1):213-237. doi:10.1111/j.1467-1770.2010.00421.x.
12. Smith LR. Emergent literacy and interactive oral discourse in early bilingual classrooms. *J Early Child Lit.* 2022;18(3):312-329.

13. Spada N, Lightbown PM. Form-focused instruction: Isolated or integrated? TESOL Q. 2008;42(2):181-207. doi:10.1002/j.1545-7249.2008.tb00115.x.
14. Toppelberg CO, Tabors P, Coggins A, Lum K, Burger C. Differential diagnosis of selective mutism in bilingual children. J Am Acad Child Adolesc Psychiatry. 2005;44(6):592-595. doi:10.1097/01.chi.0000157549.87078.f8.
15. Wong Fillmore L. Individual differences in second language acquisition. In: Fillmore CJ, Kempler D, Wang WSY, editors. Individual differences in language ability and language behavior. New York: Academic Press; 1979. p. 203-228. doi:10.1016/B978-0-12-255950-1.50019-9.
16. Wright WE, Boun S, García O, editors. The handbook of bilingual and multilingual education. Chichester: John Wiley & Sons; 2015. doi:10.1002/9781118533406.
17. Yoshida R. Teachers' choice and learners' preference of corrective feedback types. Lang Aware. 2008;17(1):78-93. doi:10.2167/la419.0.