Arabiyatuna
Jurnal Bahasa Arab

Vol. 10. No. 2, November 2026, 531-556
P-ISSN: 2580-5045, E-ISSN: 2580-5053
DOI: http://dx.doi.org/10.29240/jba.v10i2.17325

Embodied Phonological Mapping: Multimodal Scaffolding and Arabic Phoneme Acquisition in Indonesian Early Childhood

Muhammad Ivang Rizaldi[1], Alif Alvian Mawaddi Hamid[2], Achmad Fudhaili[3*], Raswan[4], Asma Wati Wabula[5]
Lembaga Indonesia Endowment Fund for Education Agency, Indonesia[1]
Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia[2,3,4]
Arizona State University, United States[5]

[*]Corresponding Author: fudhaili@uinjkt.ac.id

Cite this article:
Muhammad Ivang Rizaldi, Alif Alvian Mawaddi Hamid, Fudhaili, A., Raswan, & Wati Wabula, A. (2026). Embodied Phonological Mapping: Multimodal Scaffolding and Arabic Phoneme Acquisition in Indonesian Early Childhood. Arabiyatuna: Jurnal Bahasa Arab, 10(2), 531–556. https://doi.org/10.29240/jba.v10i2.17325

Received: 15-06-2026	Revised: 14-09-2026	Accepted: 27-10-2026

Abstract

Arabic phoneme acquisition presents particular challenges for young Indonesian learners, especially when unfamiliar sounds require articulatory movements absent from their first-language repertoires. While previous studies have examined pronunciation errors and phonological interference, limited attention has been given to children's embodied experiences of learning Arabic sounds in non-formal Qur'anic education. This study explores how multimodal scaffolding mediates young children's Arabic phoneme learning at Rumah Qur'an Al-'Izzah, South Jakarta. Employing a qualitative hermeneutic phenomenological approach informed by Van Manen, the study involved three children aged 5–6 years and two Qur'anic instructors. The study gathered the data through classroom observations, semi-structured interviews, instructional documents, and interpreted them through iterative thematic reflection on lived experiences. The findings show that children associated unfamiliar phonemes with familiar bodily sensations, including describing /خ/ as resembling snoring and /ع/ as involving pressure in the throat. Teachers supported these experiences through sensory metaphors, articulatory demonstrations, tactile guidance, and repeated interaction. These findings informed Embodied Phonological Mapping (EPM), a conceptual framework

Copyright © 2026 **Author'(s)**
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

comprising three interconnected dimensions: sensory grounding, motor imitation, and socially guided repetition. Rather than proposing a universal developmental sequence, EPM offers a phenomenological interpretation of how bodily engagement and social mediation contribute to early Arabic phonological learning. The study provides pedagogical insights for developmentally appropriate pronunciation instruction in non-formal Qur'anic learning environments.

Keywords: Arabic phoneme acquisition; multimodal scaffolding; embodied cognition; Rumah Qur'an; early childhood education

Introduction

Language acquisition is a fundamentally embodied, social, and developmental process through which children gradually construct linguistic knowledge by interacting with their physical and sociocultural environments Casey Kennington, 2021). Contemporary research increasingly recognises that language learning cannot be reduced to acquiring abstract linguistic symbols; rather, it emerges from the dynamic integration of perception, action, cognition, and social interaction (Yiyi Huang et al., 2023). Recent studies in developmental linguistics demonstrate that children's language development is inherently multimodal, involving speech, gesture, bodily movement, visual attention, and environmental engagement as interconnected resources for meaning-making (Casey Kennington, 2021). Consequently, phonological acquisition should be understood not merely as the mastery of speech sounds but as a complex process through which children develop sensorimotor representations linking auditory perception, articulatory movement, and communicative intention (Yuxing Wang et al., 2023).

Within this developmental process, phonological acquisition is foundational because phonemes are the smallest linguistic units through which meaning is differentiated and communicated (Marilyn, 2014; Paul Fletcher, 2015). Research in early phonetic acquisition shows that children gradually construct phonological categories through continuous exposure to linguistic input (Thomas Schatz et al., 2021) and repeated interaction with their surrounding environment (Marvin Lavechin et al., 2025). Forming phonological representations is particularly challenging in second-language contexts because learners must perceive and produce sounds that often do not exist in their first-language phonological inventory (Elinor & Hala, 2018). According to the Speech Learning Model, learners often assimilate unfamiliar second-language sounds into existing first-language categories, resulting in systematic phonological substitutions and pronunciation difficulties (James, 1995; Flege et

al., 2021). Such challenges become especially evident in Arabic language learning among Indonesian children, whose native phonological system lacks several Arabic consonants, including interdental, emphatic, and guttural phonemes.

Arabic phonology presents a unique linguistic challenge because many of its consonantal sounds require articulatory gestures unfamiliar to non-native learners. Indonesia's multilingual environment compounds this difficulty. For instance, Fudhaili et al.'s study of Madurese learners found that the phonological distance between Madurese and Arabic contributed to the systematic substitution of Arabic ع and ح with glottal and velar approximations, as well as persistent problems with emphatic consonants (A. Fudhaili et al., 2023). Yet comparable evidence from Javanese- and Sundanese-speaking learners remains scattered, and no study, to our knowledge, has examined how young children in Qur'anic educational settings negotiate these regional phonological repertoires. Research conducted in Indonesia found that learners frequently substitute Arabic guttural and emphatic consonants with more familiar sounds from their first language, resulting in persistent pronunciation errors and phonological simplification (Iris Michael, 2021). Similarly, recent investigations into Arabic phonological awareness emphasise that accurate phoneme perception and articulation constitute essential prerequisites for successful Arabic language acquisition and communicative proficiency (Debora Masoumi et al., 2024). Meanwhile, technological developments have enabled artificial intelligence-based systems to identify pronunciation errors and classify Arabic phonemes with high accuracy (Anca & Anne, 2023). Although these studies have significantly advanced the understanding of Arabic pronunciation difficulties, they primarily focus on phonological outcomes rather than the experiential processes through which phonological knowledge is acquired.

A critical examination of the existing literature reveals that Arabic phonological acquisition research remains largely dominated by phonetic, psycholinguistic, and technological paradigms (Wang et al., 2023). Most studies investigate pronunciation accuracy, phonological interference, speech recognition, or error detection systems (Floriana et al., 2023), while relatively little attention has been devoted to understanding how children construct phonological knowledge through bodily and social experiences (Chia-Hui et al., 2023). This limitation is increasingly problematic in light of recent developments in embodied cognitive science (Casey Kennington, 2021), which argue that linguistic knowledge is grounded in sensorimotor experience rather than abstract symbolic representations alone (Reggin et. al., 2023). Recent psycholinguistic evidence suggests that phonological processing itself may be embodied (Dilay et. al., 2024), as bodily stimulation and motor experience can influence phonological representation and linguistic cognition (Robin Samuelsson, 2025). Likewise, phenomenological research has demonstrated that early language development emerges through embodied interaction between

children and caregivers, highlighting the central role of bodily engagement in meaning construction (Ingrid & Camilla, 2022). Despite these theoretical advances, embodiment remains largely absent from current discussions of Arabic phonological acquisition.

Another significant limitation concerns the pedagogical dimension of phonological learning (Sara and Christina, 2024). Existing studies generally examine pronunciation instruction as an isolated linguistic activity, emphasising repetition, articulation drills, or corrective feedback (Zhengyuan Liu et al., 2024). However, research in early childhood education increasingly shows that multimodal scaffolding mediates language learning, including gestures, visual representations, bodily actions (Yusuf et. al., 2026), tactile experiences, and verbal interaction (Bashar & Laura 2024). Teachers' gestures, for instance, significantly support children's vocabulary acquisition and linguistic comprehension by providing nonverbal cues that reinforce meaning (Paige et al., 2024). Similarly, multimodal instructional environments facilitate children's engagement and learning by integrating multiple semiotic resources into the learning process (Yakubu & Obafemi, 2023). Recent scholarship further suggests that scaffolding practices are most effective when they incorporate diverse communicative modes that enable learners to connect abstract concepts with concrete experiences (T. Fanoush et al., 2025). Nevertheless, the role of multimodal scaffolding in Arabic phoneme acquisition remains substantially underexplored, particularly among early childhood learners in Qur'anic educational settings.

Recent early childhood studies indicate that scaffolding is most effective when it combines linguistic support with embodied, multimodal interaction (Paige et al., 2024). Ellis, Scherer, and Lien reported that phonologically focused scaffolding significantly improves children's language development, while gesture-supported instructional approaches have been shown to enhance vocabulary acquisition, language comprehension, and communicative competence among preschool learners (Katie et al., 2024). Furthermore, drama-based and embodied storytelling activities strengthen children's narrative abilities and language retention by linking verbal input to bodily action (Tamara et al., 2018).

At the methodological level, another gap appears. Most current research on Arabic phonological acquisition uses technology-driven, experimental, or quantitative methods to measure pronunciation accuracy. Although useful, these methods overlook learners' subjective and embodied experiences—how children encounter, interpret, and inhabit unfamiliar Arabic phonemes. The current phenomenological investigation is justified by this gap.

Although Embodied Phonological Mapping (EPM) shares with existing accounts of embodied phonology the assumption that phonological knowledge

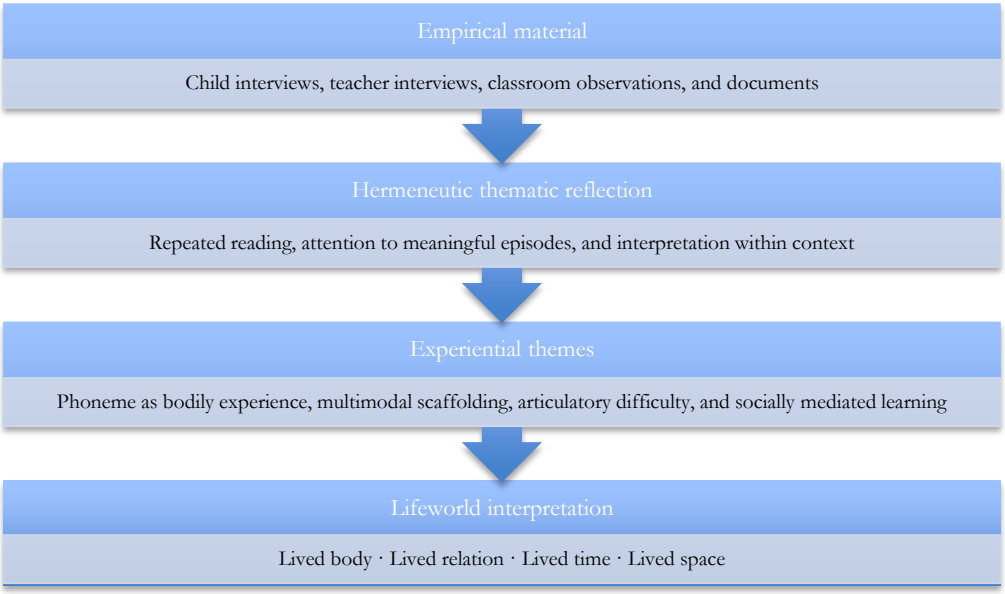
relates to sensorimotor experience, it offers a phenomenological account of how young learners make unfamiliar Arabic phonemes meaningful through socially mediated bodily engagement (Anisa Dwi, 2026). Rather than introducing embodiment as a new principle of phonological representation, EPM foregrounds the interrelationship between sensory grounding, motor imitation, and teacher-guided repetition as experienced in early childhood pronunciation learning. In this respect, EPM offers a contextually grounded interpretive framework that complements existing embodied accounts of phonology by highlighting the pedagogical and lived-experience dimensions of phoneme learning.

In the Indonesian educational context, Rumah Qur'an is a community-based, non-formal learning environment where children learn Arabic sounds through Qur'anic recitation, *hijaiyah* recognition, and basic tajwid instruction. Unlike formal classroom settings, pronunciation learning in Rumah Qur'an is embedded in religious learning routines and close teacher–child interactions. These characteristics provide a relevant context for examining how multimodal scaffolding supports children's encounters with unfamiliar Arabic phonemes. By investigating learning experiences within Rumah Qur'an Al-'Izzah, the present study contributes to understanding Arabic phonological learning in a non-formal educational setting that remains relatively underrepresented in research on early language education.

This study addresses these gaps through an integrated theoretical framework in which each theory serves a distinct analytical function. Two theories explain why Arabic phonemes are difficult. The Speech Learning Model attributes difficulty to first-language assimilation, while Natural Phonology attributes it to articulatory simplification. Two further theories explain how learners overcome these difficulties. Embodied Cognition Theory emphasises sensorimotor grounding, while Sociocultural Theory emphasises socially mediated scaffolding (Farah Hadban et. al., 2026). Finally, Hermeneutic phenomenology provides the methodological orientation for interpreting these embodied and social processes as they are lived by children. This integration shifts the analytical focus from phonological outcomes to the experiential mechanisms through which children make unfamiliar Arabic phonemes meaningful and producible. Accordingly, the study is guided by three research questions: (1) How do early childhood learners experience the acquisition of unfamiliar Arabic phonemes through multimodal scaffolding? (2) What mechanisms facilitate this acquisition? (3) What challenges do learners encounter, and how are these negotiated?.

This study employed a qualitative hermeneutic phenomenological design, drawing on Van Manen (2016), to examine how young children experience acquiring unfamiliar Arabic phonemes. The research was conducted at Rumah Qur'an Al-'Izzah in Jagakarsa, South Jakarta—an institution

established in 2010 serving 60 students aged 4–12 (20 boys, 40 girls) through the Iqra' method and tajwid instruction. Participants were purposively selected based on three criteria: attendance in tajwid classes for at least six months, a documented history of difficulty pronouncing at least two Arabic phonemes absent from Indonesian (ع، ح، خ، ث), and parental consent. The final sample comprised three children (C1–C3) aged 5–6, all Javanese- or Sundanese-speaking at home, able to recognise *hijaiyah* letters, and having studied Arabic for 6–14 months; two instructors (T1, T2) with eight and three years of Qur'anic teaching experience, both holding *sanad* certification, also participated. Ethical approval was obtained under No. 76/YBA/RQA/05/2026. Written parental consent was secured, verbal child assent was sought before each session, participants were assigned pseudonyms, and data were stored on encrypted devices scheduled for deletion three years after publication. The data were collected over five weeks (October–November 2025) through four participant observations of 60–90 minutes using an observation checklist recording target phoneme, teacher scaffolding moves, child responses, and contextual notes; semi-structured interviews with each child (once, 10–15 minutes) and with one teacher (30–40 minutes), conducted in Bahasa Indonesia with code-switching to Arabic for phoneme names; and document analysis of Iqra' volumes 1–6, tajwid lesson plans, and teacher-written progress notes. Data were analysed using Van Manen's hermeneutic approach through iterative thematic reflection on lived-experience descriptions, identification of lifeworld existentials (lived body, lived space, lived time, lived relation), and interpretive phenomenological writing; coding was manual in Indonesian, with key excerpts translated into English and reviewed by the second author until consensus. Trustworthiness was addressed by comparing children's interview accounts, instructors' explanations, classroom observations, and instructional documents. For example, children's descriptions of /خ/ as resembling snoring were examined alongside teachers' reported use of sensory metaphors and observations of pronunciation activities. This comparison supported interpreting sensory metaphors as a form of embodied pedagogical mediation. We documented interpretive decisions through field notes and analytical memos to maintain a traceable connection between the empirical material and the emerging phenomenological themes (Van Manen, 2016).



Findings and Discussion

Early Childhood Learners' Experience of Arabic Phoneme Acquisition through Multimodal and Embodied Learning?

Before presenting the experiential accounts, we must acknowledge a methodological caveat. Young children do not describe their experiences in the same way that adults do. Their narratives are succinct, sensory, fragmented, and often conveyed through the body rather than through in-depth verbal reflection.

The findings indicate that young learners acquire Arabic phonemes through embodied, multimodal meaning-making rather than auditory perception alone. Across all participants, unfamiliar Arabic phonemes were interpreted by reference to bodily experiences already familiar to the children. For example, when asked about the phoneme /khāʾ/ (خ), one participant described it simply as *"kayak orang ngorok"* ("like someone snoring"), while another repeated *"khā'… khā'… kayak orang ngorok"*. A third participant similarly recalled that *"khā' itu kayak orang ngorok"* because this was how the teacher had introduced the sound.

At first glance, these responses may seem like simple mnemonic associations. However, a hermeneutic phenomenological reading reveals a deeper embodied, interpretive process. The children were not merely memorising an analogy; rather, they were transforming an unfamiliar phonetic category into a meaningful bodily experience. In other words, the sound /khāʾ/ became intelligible because it was anchored to an embodied experience already stored within the child's sensorimotor repertoire. This process closely corresponds to what Gibbs describes as *embodied mapping*, in which abstract

linguistic representations are grounded in bodily experiences rather than processed as disembodied symbolic units (Gibbs Jr., 2005). From a Van Manen perspective, this is not merely cognitive mapping but a disclosure of lived body: the phoneme is experienced as a bodily event before it is understood as a linguistic category.

More importantly, these findings extend existing research on embodied cognition by showing that embodied mapping can operate not only at the semantic level but also at the phonological level. Previous studies have predominantly examined embodiment in vocabulary acquisition and conceptual development (Gibbs Jr., 2005). The snoring metaphor provided learners with an experiential bridge linking an unknown Arabic phoneme to a known bodily sensation. Thus, phonological learning appears to transform acoustic input into embodied perceptual experience. In this case, the phoneme is lived in the breath and throat before it is represented as a sound, supporting Van Manen's (2016) assertion that lived experience is not cognitive until it is physical.

This interpretation is reinforced by participants' descriptions of the pharyngeal phoneme /ʿain/ (ع). One learner explained, *"Iya bener… kayak orang lagi neken leher"* ("Yes, it's like someone pressing the throat"), while another stated, *"Iya… gitu disuruh megang leher"* ("Yes, we were told to hold the throat"). These statements demonstrate that the children's understanding of /ʿain/ was mediated through tactile and kinesthetic experiences rather than through auditory modelling alone. This illustrates how lived body and lived relation are intertwined: the child did not find /'ain/ on his own; rather, the teacher's hand guided his own. Unlike visible articulations such as bilabial or dental sounds, pharyngeal consonants involve articulatory movements that are not easily observed. Consequently, children relied on bodily sensation as an alternative pathway for constructing articulatory awareness.

Recent studies on multimodal language acquisition further support this interpretation. Reggin et al. argue that language learning develops through situated interactions among perception, action, and social engagement rather than through abstract cognitive processing alone (D. Reggin et al., 2023). Similarly, Karadöller, Sümer, and Özyürek demonstrate that children's language acquisition is inherently multimodal, integrating speech, gesture, bodily movement, and environmental interaction (Z. Karadöller et. al., 2025). This also aligns with Van Manen's (2016) concept of pedagogical tact, in which the teacher's bodily guidance becomes an embodied scaffold for the child's emerging articulation.

From a phenomenological perspective, the data further reveal that children do not experience phonemes as isolated linguistic units. Instead, phonemes are experienced as bodily events. The sound /khāʾ/ is experienced as "snoring," while /ʿain/ is experienced as "pressing the throat." This finding

contributes to an emerging body of scholarship suggesting that linguistic knowledge is fundamentally grounded in lived bodily experience rather than exclusively in abstract mental representations (Z. Mahon, 2016). In Van Manen's terms, these accounts unpack lifeworld existentials, specifically the lived body and lived relation and suggest the need to interpret phoneme acquisition as an existential achievement rather than merely a phonetic one.

Table 1. The Essence of Arabic Phoneme Acquisition among Early Childhood Learners

Theme	Participants' Lived Experience (Codes; Frequency)	Observed Manifestation	Essence: What	Essence: How	Phenomenological interpretation
Phoneme as Bodily Event	C1: *"kayak orang ngorok"* (like someone snoring) — /خ/ (C1, C2, C3; n = 9)	All three children spontaneously described /خ/ through the same bodily metaphor; none referred to tongue position or airflow	The phoneme was experienced as a sensation in the throat and breath, not as an acoustic unit to be imitated	Through sensory metaphor introduced by T1 and re-used spontaneously by children across sessions	The unfamiliar sound is interpreted through a familiar bodily sensation.
	C2: *"Iya bener... kayak orang lagi neken leher"* (yes, like someone pressing the throat) — /ع/ (C1, C2; n = 5)	Children located /ع/ through tactile pressure in the throat, an articulatory space they could not see	The phoneme was experienced as pressure and vibration in an invisible bodily region	Through tactile guidance by T2, who placed the child's hand on the throat during production	The unfamiliar phoneme /ع/ as a sensation of pressure in the throat, reflecting an emerging bodily awareness of articulation (lived body).
	C3: *"Iya... gitu disuruh megang leher"* (yes, we were told to hold the throat) (C2, C3; n = 4)	Child recalled the teacher's instruction as the origin of his understanding of /ع/	The phoneme was experienced as a shared bodily event rather than a	Through teacher-guided touch as an embodied scaffold	Articulatory awareness emerges through teacher-guided bodily engagement.

			private discovery		
Multimodal Integration	T1: *"Kalau hanya dijelaskan, mereka lupa. Kalau dipegang lehernya, mereka ingat."* (If merely explained, they forget; if they hold their throat, they remember.) (T1; n = 3)	Teacher distinguished verbal-only from tactile-verbal instruction based on observed outcomes	Phonological learning was experienced as more durable when multiple sensory channels converged	Through simultaneous speech, gesture, touch, and metaphor in a single instructional moment	Tactile engagement transforms verbal instruction into a memorable bodily experience, highlighting the role of embodied interaction in phonological learning (lived body and lived relation).
	C1: *"Huruf kha itu kayak orang ngorok"* (the letter kha is like someone snoring) (C1, C2; n = 4)	Child integrated letter recognition, sound, and bodily metaphor in one utterance	The phoneme was experienced as a unified experiential object across modalities	Through repeated multimodal rehearsal across observations 2–4	The unfamiliar phoneme /خ/ with the familiar bodily experience of snoring, revealing how sensory metaphors shape phonological understanding (lived body)
	T2: *"Yang sering pegang leher lebih cepat bisa* /ع/*."* (Those who often hold their throat master /ع/ faster.) (T2; n = 2)	Teacher noted differential outcomes between children who received tactile scaffolding and those who did not	The phoneme was experienced as more accessible through touch than through hearing alone	Through tactile-kinesthetic scaffolding layered onto verbal explanation	The unfamiliar phoneme /ع/, highlighting the connection between sensory engagement and socially mediated articulation (lived body and lived relation).

These two themes articulate the study's theoretical contribution. First, they extend Gibbs (2005) by showing that embodied mapping operates not only at the semantic level but also at the phonological level: an unfamiliar Arabic phoneme becomes intelligible through a familiar bodily event. Second, they refine Van Manen's (2016) pedagogical tact, demonstrating that the teacher does not explain the phoneme but shows it with the body, thereby intertwining lived body and lived relation. Third, they challenge auditory-only models of phonological acquisition and refine SLM-r (Flege & Wayland, 2021), suggesting that sustained exposure operates through embodied re-enactment rather than mere acoustic input.

That Arabic phoneme acquisition is fundamentally an embodied and experiential process rather than a purely auditory activity. The recurring use of sensory analogies, such as associating /*khaʾ* / with snoring and /ʿ *ain*/ with pressing the throat, indicates that children construct phonological understanding through familiar bodily experiences. These findings suggest that learners do not perceive phonemes as abstract linguistic units but as meaningful sensory events grounded in their lived experiences. From a phenomenological perspective, phonological awareness emerges through the interaction between perception, bodily engagement, and social mediation. This pattern supports Embodied Cognition Theory by illustrating that phonological representations are developed through sensorimotor experiences rather than through cognitive processing alone. Consequently, the findings challenge traditional views of pronunciation learning that prioritise auditory imitation, highlighting instead the importance of embodied meaning-making in early Arabic phonological development.

Overall, these various experiences indicate that the acquisition of Arabic phonemes is not a fragmented cognitive operation but rather involves the simultaneous engagement of existential dimensions of the lifeworld: the lived body, lived relations, lived space, and lived time. This convergence gives rise to a concept termed in this study as "Embodied Phonological Mapping" a process wherein unfamiliar Arabic phonemes become meaningful and pronounceable through sensory, motor, relational, and temporal experiences. These phonemes are not learned merely as sounds but are experienced as events involving the body, the social dimension, and the dimension of time.

Multimodal Scaffolding Facilitate Arabic Phoneme Acquisition

The findings reveal that the acquisition of Arabic phonemes among young learners is not facilitated by verbal explanation alone but by a complex system of multimodal scaffolding that includes bodily demonstration, sensory metaphors, tactile guidance, repetitive practice, and social mediation. The data indicate that teachers continuously transformed abstract phonetic concepts into concrete, observable, and physically experienced actions, enabling children to

access articulatory knowledge that would otherwise remain beyond their cognitive and linguistic capacities.

One of the most prominent forms of scaffolding identified in the study was the use of sensory metaphors. When introducing the phoneme /khāʾ/ (خ), teachers repeatedly associated the sound with *"orang ngorok"* (snoring). This metaphor was subsequently internalised by the learners, who consistently recalled the phoneme through this bodily experience. One participant explained, *"Huruf kha itu kayak orang ngorok,"* while another stated, *"Kha… kha… kayak suara orang ngorok."* These responses suggest that the metaphor functioned not merely as a memory aid but as a cognitive scaffold that enabled learners to connect an unfamiliar phonetic category with a familiar sensory experience. Through this process, the abstract articulatory characteristics of /khāʾ/ became experientially meaningful and therefore easier to reproduce.

From a sociocultural perspective, such instructional practices exemplify what Lev Vygotsky conceptualised as scaffolding within the Zone of Proximal Development (ZPD). Rather than expecting learners to discover articulatory mechanisms independently, teachers provided temporary support that enabled children to perform beyond their current developmental level (S. Vygotsky, 1978). The metaphor of snoring effectively reduced cognitive complexity by translating an unfamiliar Arabic phoneme into a familiar embodied experience. Consequently, phonological learning occurred not through direct transmission of linguistic knowledge but through mediated participation in socially organised learning activities (P. Lantolf, et. al., 2018).

The data further demonstrate that teachers employed tactile and kinesthetic scaffolding when teaching pharyngeal sounds, particularly /ʿain/ (ع). Several participants reported that teachers instructed them to hold or press their throats while producing the sound. One learner explained, *"Iya, disuruh pegang leher,"* while another described the articulation as *"kayak orang lagi neken leher."* These statements indicate that bodily sensation became an instructional resource through which learners developed awareness of otherwise invisible articulatory movements.

This finding is particularly significant because pharyngeal consonants are among the most difficult categories for non-native Arabic learners. The tactile guidance observed in this study, therefore, functioned as an embodied scaffold that compensated for the articulation's perceptual invisibility. Such findings align with contemporary research on embodied language acquisition, which argues that language learning emerges from the coordination of perceptual, motor, and sensory experiences rather than from auditory processing alone (D. Reggin et al., 2023).

Beyond tactile guidance, the study identified articulatory demonstration as another critical scaffolding mechanism. Teachers frequently modelled tongue

placement and mouth positioning when introducing interdental consonants such as /t̲āʾ/ (ﺙ). Participants reported that teachers instructed them to place their tongues between their teeth and repeatedly demonstrated the movement until learners could imitate it successfully. These demonstrations transformed abstract phonetic descriptions into observable motor actions, allowing children to construct articulatory knowledge through visual and motor imitation.

The effectiveness of this strategy can be explained through multimodal learning theory. According to multimodal scholars, learning becomes more effective when information is distributed across multiple semiotic channels rather than presented through a single modality (Gunther Kress, 2010). In the present study, phonological information was communicated simultaneously through speech, gestures, facial expressions, bodily movements, tactile sensations, and visual demonstrations. Such multimodal integration reduced cognitive load and increased learners' opportunities to construct meaningful phonological representations (Kahriman & Samuelsson, 2026). Similar findings have been reported by Masoumi, Bourbour, and Lindqvist, who found that multimodal interactions enable young children to negotiate meaning more effectively by drawing upon multiple communicative resources simultaneously (Pesco and Gagné, 2015).

Another important form of scaffolding identified in the data was structured repetition. Both learners and teachers emphasised the importance of repeatedly practising difficult phonemes during Qur'anic recitation sessions. One participant explained that difficult sounds gradually became easier because they were encountered repeatedly during daily recitation. Teachers similarly observed that learners at higher instructional levels exhibited greater articulatory accuracy due to prolonged exposure and repeated practice.

Although repetition is commonly associated with behaviourist approaches to language learning, the present findings suggest a more nuanced interpretation. Repetition did not function merely as mechanical drilling; rather, it operated as guided repetition embedded within meaningful social interaction. Teachers continuously monitored learners' pronunciation, provided corrective feedback, and adjusted scaffolding according to individual needs. Thus, repetition became a socially mediated process of phonological refinement rather than a simple habit-formation exercise. This interpretation aligns with the revised Speech Learning Model (SLM-r), which emphasises that the formation of new phonetic categories requires sustained exposure, perceptual attention, and repeated interaction with target-language sounds (E. Flege and Wayland, 2021).

Perhaps the most significant finding concerns the synergistic interaction among multiple forms of scaffolding. The data indicate that no single instructional strategy operated independently. Instead, sensory metaphors,

bodily demonstrations, tactile guidance, repetition, and teacher feedback functioned as an integrated pedagogical system. For example, when teaching /ʿain/, teachers simultaneously explained the sound verbally, demonstrated articulation physically, instructed learners to touch their throats, corrected production attempts, and encouraged repeated practice. This integrated process created a rich multimodal learning environment in which children could access phonological information through multiple pathways simultaneously.

This finding extends existing research on scaffolding by suggesting that phonological acquisition is best understood as a process of multimodal mediation. Previous studies have often conceptualised scaffolding primarily as verbal support or instructional guidance (Zulfatun Anisah, 2016). The present study demonstrates that scaffolding in Arabic phoneme acquisition involves the orchestration of diverse sensory and semiotic resources that collectively transform inaccessible phonetic knowledge into meaningful embodied experience. Consequently, the effectiveness of scaffolding lies not in any individual technique but in the dynamic interaction among multiple modalities that support learners' phonological development.

Table 2. Multimodal Scaffolding Mechanisms Supporting Arabic Phoneme Acquisition

Scaffolding experience (T1/T2)	**Child's response**	**Embodied meaning**	**Interpretive insight**	**Context**
T1 associated the sound with *"orang ngorok"* (snoring) and demonstrated the breath through an exaggerated exhale	C1, C2, and C3 immediately repeated *"kayak orang ngorok"* and produced /خ/ with the correct breathy friction within the same session	/خ/	High — mastered in 1 session	Whole-group halaqah; Iqra' volume 3
T2 placed the child's hand on her own throat, then on the child's throat, while producing /ع/ slowly	C2 located the sensation after 2–3 trials but reverted to a glottal stop when reciting quickly; C3 required 4 sessions before consistent production	/ع/	Medium — 2–3 trials for recognition, 4 sessions for production	Individual scaffolding within group halaqah
T1 modelled tongue-between-teeth position for /ث/ using a	C1 and C2 imitated correctly in isolation but substituted /s/	/ث/	Medium — accurate in isolation, unstable in	Whole-group halaqah; tajwid drill

mirror and her own mouth	during connected recitation; C3 required repeated mirror checks		recitation	
T1 used a hand gesture sweeping from the throat outward to represent the pharyngeal constriction of /ح/	C1 reproduced the gesture spontaneously while reciting; C2 and C3 followed after 2 sessions	/ح/	High — gesture internalised and used independently	Whole-group halaqah; tajwid drill
T2 combined verbal explanation, tactile guidance, and gesture simultaneously during a single correction episode for /ع/	C2 produced /ع/ correctly immediately after the combined scaffold, though the effect faded in the next session	/ع/	Medium — immediate success, weak retention	Individual correction during group recitation
T1 used guided repetition across four consecutive sessions, returning to /خ/ and /ع/ at the start of each halaqah	All three children produced /خ/ accurately from session 2 onward; /ع/ remained unstable for C2 and C3 until session 4	/ع/ ,/خ/	High for /خ/; Medium for /ع/	Routine opening drill across 4 observations

Successful Arabic phoneme acquisition is facilitated by integrating multiple scaffolding mechanisms rather than relying on a single instructional strategy. From a Van Manenian perspective, successful Arabic phoneme acquisition is not the accumulation of instructional techniques but the orchestration of embodied and relational mediations that allow the child to inhabit unfamiliar sounds. Sensory metaphors, articulatory demonstrations, tactile guidance, repetition, and corrective feedback collectively function as interconnected mediational tools that help learners bridge the gap between unfamiliar phonetic categories and concrete articulatory experiences. The findings indicate that teachers transform abstract phonological knowledge into observable and physically experienced actions, thereby reducing learners' cognitive and articulatory burdens. This supports Sociocultural Theory, particularly the concept of the Zone of Proximal Development, where learning occurs through guided participation and expert mediation. However, this study expands upon that sociocultural interpretation by demonstrating that mediation is not merely verbal or social in nature, but is fundamentally embodied and affective, what Van Manen (2016) terms "pedagogical tact." More importantly,

the data suggest that multimodal scaffolding should be viewed not as supplementary support but as a central mechanism of phonological development. The effectiveness of instruction appears to derive from the simultaneous orchestration of verbal, visual, tactile, and motor resources, creating a rich multimodal learning ecology that facilitates phonological internalisation. From a phenomenological perspective, this ecology reveals the intertwining of the lived body, lived relations, and lived time as the child perceives the sound, shares the experience of it with the teacher, and gradually inhabits that sound through repeated encounters.

Challenges Children Encounter in Acquiring Arabic Phonemes

The findings reveal that Arabic phoneme acquisition among early childhood learners is characterised by a complex interplay of linguistic, articulatory, cognitive, and sociocultural challenges. While learners demonstrated gradual improvement through multimodal scaffolding and repeated exposure, the acquisition process characterized by a persistent existential tension between the child's familiar bodily repertoire and the demands of Arabic articulation, which are alien to them. Instead, children encountered persistent difficulties arising from the phonological distance between Arabic and their first language, the physiological complexity of Arabic articulation, prior learning experiences, and the influence of local linguistic environments. These challenges not only affected pronunciation accuracy but also shaped how children perceived, interpreted, and internalised unfamiliar phonological categories.

One of the most significant challenges identified in the data concerns the acquisition of Arabic phonemes that have no equivalent in the Indonesian phonological inventory. Teachers consistently reported that learners struggled to differentiate and produce interdental consonants such as /t̲āʾ/ (ث), /d̲āl/ (ذ), and /ẓāʾ/ (ظ), as well as pharyngeal and uvular sounds such as /ʿain/ (ع), /ḥāʾ/ (ح), and /khāʾ/ (خ). One teacher explained that children often perceived these sounds as similar to more familiar Indonesian consonants and therefore replaced them with approximations that felt easier to articulate. This tendency was reflected in recurring substitutions, such as /t̲āʾ/ pronounced as /s/, /d̲āl/ as /z/, and /ʿain/ as a simple vowel onset. Such patterns indicate that learners initially relied on existing phonological categories when encountering unfamiliar Arabic sounds.

From the perspective of the Speech Learning Model-Revised (SLM-r), these substitutions are not random errors but predictable outcomes of cross-linguistic phonological transfer. According to Flege and Wayland, second-language learners tend to map unfamiliar sounds onto the closest available categories in their first-language phonological system when perceptual distinctions are not yet well established (Flege & Wayland, 2021). Because

Indonesian lacks interdental and pharyngeal consonants, learners often perceive Arabic phonemes through the filter of their native phonological repertoire. Consequently, phonological acquisition involves not merely learning new articulatory movements but restructuring existing perceptual categories. These findings therefore support contemporary research showing that perception and production difficulties are fundamentally interconnected in second-language phonological development (Zheng & Gor, 2024).

The findings further suggest that articulatory complexity is a substantial barrier, particularly in participants' accounts of learning /'ain/. In contrast to bilabial and alveolar consonants, /'ain/ is articulated deep in the throat and cannot be directly observed or learned through visual imitation. Several children described it using bodily metaphors, including "*kayak orang lagi neken leher*" ("like someone pressing the throat"), or recalled being asked to touch their throats while producing it. These accounts suggest difficulty locating and controlling the articulatory space required for pharyngeal production. From a phenomenological perspective, this is more than a technical articulatory challenge: it disrupts the child's experience of the lived body. The sound cannot yet be situated within the child's familiar bodily map; without a corresponding sensorimotor schema, the phoneme remains difficult to inhabit (Berent & Platt, 2022). This interpretation is consistent with research on embodied phonology, which emphasises the integration of auditory perception, motor control, and bodily awareness in phonological acquisition. Berent and Platt argue that phonological knowledge is grounded in sensorimotor systems rather than represented solely through abstract cognitive structures (Corrine Occhino, 2017). The present study extends this view by showing how young learners' difficulties with Arabic phonemes are closely associated with limited access to embodied articulatory experience. The challenge, therefore, is not only to hear a sound accurately, but also to experience, locate, and reproduce the bodily actions involved in its production. This account echoes Van Manen's (2016) description of the pre-reflective dimension of the lived body: children must first experience the sound bodily before they can produce it intentionally.

Another challenge emerging from the data concerns phonological simplification. Teachers reported that learners frequently replaced difficult Arabic phonemes with easier-to-articulate, more familiar sounds. For example, pharyngeal and emphatic consonants were often reduced to their non-emphatic counterparts, while complex articulations were simplified into sounds already present in Indonesian. Such substitutions reflect a broader tendency toward articulatory economy, whereby learners reduce phonological complexity to match their current developmental capabilities.

This phenomenon can be interpreted through the lens of Natural Phonology Theory, which posits that speakers naturally favour simpler articulatory patterns and gradually suppress these simplification processes as

linguistic competence develops (David Stampe, 1973). Although Natural Phonology was originally formulated to explain first-language acquisition, the present findings suggest that similar processes operate during the acquisition of Arabic as an additional language. The children's substitutions, therefore, represent developmental strategies rather than mere pronunciation failures. Phenomenologically, these substitutions can be understood as a form of resistance by the lived body against the demands of unfamiliar articulation a physical self-preservation strategy that favors the familiar over the foreign. Viewed from this perspective, phonological errors serve as valuable indicators of how learners navigate an unfamiliar sound system and gradually construct new, embodied representations.

The findings also highlight the role of pedagogical interference arising from previous learning experiences. Teachers observed that some learners had received instruction in Arabic pronunciation before enrolling at the institution. However, earlier instruction often emphasised rote repetition rather than articulatory awareness. As a result, children sometimes arrived with entrenched pronunciation habits that were difficult to modify. From a Van Manenian perspective, this illustrates the temporal dimension of lived body: past bodily habits sediment into the child's present articulation, requiring not merely correction but a re-education of the senses. One teacher noted that learners frequently found it difficult to break these habits because the old sounds had become part of their bodily repertoire; they insisted on previously learned pronunciations even after corrective feedback. This observation suggests that inaccurate phonological representations can become fossilised when early instruction fails to establish appropriate articulatory foundations.

From a sociocultural perspective, this finding underscores the importance of instructional mediation in phonological development. Learning is not merely the accumulation of linguistic input but the gradual internalisation of socially mediated practices (S. Vygotsky, 1978). Consequently, the quality of early pedagogical experiences plays a decisive role in shaping subsequent phonological development. The present findings, therefore, support recent scholarship emphasising that pronunciation learning is influenced not only by linguistic factors but also by educational histories and instructional ecologies (Mihai & Classen, 2023).

In addition to pedagogical factors, the study identified the influence of local linguistic environments. Teachers reported that regional accents and speech habits associated with Indonesian and local languages, particularly Sundanese and Javanese, affected learners' pronunciation patterns. Certain phonemes were consistently produced with articulatory features characteristic of local speech varieties, resulting in deviations from standard Arabic pronunciation. This observation aligns with contemporary research showing that first-language phonology continues to influence second-language speech

production even after substantial exposure to the target language (Elinor & Hala, 2018).

However, the most theoretically significant finding concerns the interaction among these challenges. The data suggest that phonological difficulties cannot be attributed to a single factor. Rather, learners experienced challenges because linguistic distance, articulatory complexity, perceptual limitations, previous instructional experiences, and sociolinguistic influences operated simultaneously. For example, the difficulty of producing /ʿain/ was not merely an articulatory issue. It was compounded by the absence of a comparable Indonesian phoneme, limited bodily awareness of pharyngeal articulation, insufficient prior exposure, and the influence of existing pronunciation habits. This multidimensional interaction indicates that Arabic phoneme acquisition is best understood as an ecological process in which multiple developmental systems intersect.

The findings therefore challenge reductionist explanations that attribute pronunciation difficulties solely to articulatory mechanics or first-language interference. Instead, they suggest that Arabic phonological acquisition among young learners is a dynamic process shaped by reciprocal interactions among cognitive, bodily, linguistic, pedagogical, and sociocultural factors. In this sense, pronunciation errors should not be viewed merely as deficits but as developmental manifestations of learners' ongoing attempts to reconcile unfamiliar phonological categories with existing perceptual and experiential frameworks.

Table 3. Challenges in Arabic Phoneme Acquisition as Lived Struggles

Child's Lived Struggle	Observed Manifestation	Negative Case / Notes
C2 (n = 11): *"Aku selalu bilang /s/ padahal /ث/, Bu. Lidahku susah."* (I always say /s/ instead of /ثائب/, Miss. My tongue finds it difficult.)	Substituted /s/ for /ث/ in 9 of 12 recitation attempts during observation 3; correct only in isolation when T1 held a mirror	C3 did not improve despite three sessions of mirror-guided tongue placement; T1: *"Kalau cepat, dia balik ke /s/."* (When reading fast, he reverts to /s/.)
C3 (n = 8): *"Kalau /ع/ rasanya di tengah, tapi aku nggak bisa nahan."* (For /ع/العلا/ it feels like it's in the middle, but I can't hold it.)	Produced /ع/ correctly in isolation after tactile guidance but reverted to a glottal stop in connected recitation	C2 succeeded only after the fourth session; C3 remained unstable — tactile guidance alone is insufficient without stabilised articulatory awareness

C1 (n = 5): *"Kha' itu kayak orang ngorok, tapi kalau baca jadi kayak /h/ الحاكم."* (Kha' is like someone snoring, but when I recite it becomes like /h/.)	/خ/ produced correctly in drill but weakened to /h/ during Qur'anic recitation	T2: *"Kalau sudah masuk ayat, fokusnya pindah."* (Once they enter the verse, their focus shifts.)
C2 (n = 4): *"Aku takut salah, jadi aku diam."* (I'm afraid of being wrong, so I stay silent.)	Withdrew from recitation after repeated correction; produced fewer attempts than C1 and C3 in observations 2–3	Affective dimension not anticipated in the initial framework; T2: *"Terlalu banyak koreksi bikin dia menarik diri."* (Too much correction makes him withdraw.)
T1 (n = 3): *"Yang paling susah itu /ع/, karena nggak kelihatan."* (The hardest is /ع/, because it's invisible.)	Teacher identified the invisibility of pharyngeal articulation as the core pedagogical obstacle	T2 corroborated: *"Anak butuh rasa, bukan cuma lihat."* (Children need to feel it, not just see it.)

As shown in Table 3, the difficulties young learners face are not isolated technical problems; rather, they are interconnected lived experiences. Three patterns are particularly evident. First, substitution functions as an ingrained bodily habit, not simply as a momentary error. C2 substituted /ث/ with /s/ in nine of twelve attempts, and the substitution reappeared whenever the pace of recitation increased. This indicates that accurate production in isolation does not necessarily transfer to connected recitation. Second, the hidden articulation of /ع/ remains a central obstacle. Although the children could hear the phoneme, they could not locate it within a bodily map they had not yet developed. C3's lack of improvement after three sessions of tactile intervention further suggests that scaffolding needs to be sustained rather than provided in a single episode. Third, affective withdrawal emerged as an unexpected challenge: C2's silence following repeated correction indicates that feedback should be calibrated to each child's emotional tolerance. Taken together, these findings refine SLM-r (Flege & Wayland, 2021) by showing that perceptual restructuring is mediated by bodily and affective factors. They also extend Natural Phonology (Stampe, 1973) by suggesting that simplification may serve as a form of bodily self-preservation when children encounter unfamiliar articulatory demands.

Conclusion

The central claim of this study is the difficulty of Arabic phonemes for young Indonesian learners is not primarily a difficulty of hearing, but a difficulty

of inhabiting. Children did not fail to perceive /ع/ and /ث/; they failed to locate these phonemes within a body that had not yet learned to feel them. This is the take-home message that the data compel: phonological acquisition is an existential achievement, not merely an auditory one.

This claim is formalised in the concept of Embodied Phonological Mapping (EPM), which this study defines as a three-stage process through which an unfamiliar phoneme is transformed from an acoustic event into an inhabitable bodily one. In the first stage, sensory grounding, the phoneme is anchored to a familiar bodily sensation /خ/ to the breath of snoring, /ع/ to the pressure of a hand on the throat. In the second stage, motor imitation, this sensation is converted into an observable and reproducible articulatory gesture through teacher demonstration, touch, and mirroring. In the third stage, socially guided repetition, the gesture is stabilised through repeated, affectively calibrated interaction until the child can produce the phoneme without external support. EPM is not a teaching technique and not a replacement for the Speech Learning Model or Natural Phonology; it is a descriptive-explanatory framework that specifies the experiential mechanism those models presuppose but do not articulate. Where SLM-r explains *why* assimilation occurs, and Natural Phonology explains *why* simplification occurs, EPM explains *how* a child moves from assimilation and simplification toward stable production through the body, through relation, and through time.

Three limitations must be acknowledged honestly. First, the study involved three children in one institution over five weeks. This sample is defensible for hermeneutic phenomenology, which prioritises depth over breadth, but it cannot support claims about prevalence or generalisability. Second, the affective dimension emerged late in the analysis and was not systematically designed for; C2's withdrawal from recitation was documented through observation and teacher interview, but a study designed from the outset to capture affect might have yielded a more robust account. Third, the first author served simultaneously as observer and analyst, which introduces the risk of over-interpretation; peer debriefing with the academic supervisors mitigated but did not eliminate this risk. A fourth limitation is methodological: children's verbal accounts were brief and sensory, and while triangulation with observation and teacher narratives strengthened the interpretation, the study cannot claim to have captured the full interiority of the children's experience.

These limitations open four specific directions for future research. First, longitudinal studies should track whether phonemes scaffolded through EPM remain stable six and twelve months after instruction, particularly for /ع/, which remained unstable for two of three children in this study. Second, comparative studies across regional phonological repertoires such as Javanese, Sundanese, and Madurese should test whether EPM operates identically when the first-

language inventory differs, or whether the sensory metaphors that ground each phoneme must be culturally and linguistically recalibrated. Third, design-based research should translate EPM into a teacher-training protocol for Rumah Qur'an and TPQ instructors, specifying which sensory metaphors work for which phonemes and how corrective feedback can be calibrated to children's affective tolerance. Fourth, quasi-experimental studies should test whether EPM-informed instruction outperforms auditory-only instruction on phoneme retention, thereby establishing the framework's pedagogical efficacy beyond the phenomenological warrant offered here. Until such studies are conducted, EPM should be read as a theoretically grounded, empirically illustrated hypothesis about how young children inhabit unfamiliar phonemes not as a validated instructional model.

Acknowledgment

The authors would like to express their sincere gratitude to Rumah Qur'an Al-Izzah Jagakarsa for granting permission and facilitating the data collection process. We also extend our deepest appreciation to the teachers and early childhood learners who participated in this study, whose invaluable experiences and insights made this research possible. Furthermore, we acknowledge the institutional support provided by Universitas Islam Negeri Syarif Hidayatullah Jakarta. The authors also thank Asma for her assistance in language proofreading and editing.

Ethical Statement

This study was conducted in accordance with research ethics principles applicable to social and educational research. Institutional research permission was granted through official letter No. 76/YBA/RQA/05/2026, dated Mei 20, 2026, for the qualitative case study conducted at Rumah Qur'an Al-'Izzah Jakarta. The research involved no clinical intervention and was designed to minimise potential risks to participants. Ethical procedures included informed consent, voluntary participation, confidentiality, and anonymisation of participants' identifying information.

Author Contributions Statement

MI, AA and AF contributed to the conceptualisation, methodology, data collection, and formal analysis. R provided supervision, conceptual guidance, and critical feedback as academic advisors. MI drafted the original manuscript. All authors contributed to the review, editing, and final approval of the manuscript. Language proofreading and editing were assisted by a professional language editor.

AI Writing Statement

During the preparation of this manuscript, the authors used QuillBot to assist with language editing, grammar checking, and readability enhancement. The authors carefully reviewed and edited the AI-assisted content and take full responsibility for the final manuscript.

References

Anisa Dwi Nurchayati, Eric Kunto Aribowo, and Akhmad Fadjeri, "Enhancing Arabic Phonetic Learning through YOLO-Based Grapheme Recognition and Audio Feedback," An Nabighoh 28, no. 1 (2026): 319–340, https://doi.org/10.32332/an-nabighoh.v28i1.319-340

Anisah, Z. (2024). Scaffolding in early childhood language learning. International Journal of Early Childhood Inclusive Education, 8(2), 466–479. https://doi.org/10.35896/ijecie.v8i2.868

Berent, I., & Platt, M. (2022). The embodied nature of phonological knowledge. Psychonomic Bulletin & Review, 29(4), 1257–1273. https://doi.org/10.3758/s13423-022-02066-0

Bernstein, K. A., et al. (2024). Gesture like a kitten and you won't forget your tale: Drama-based, embodied story time supports preschoolers' narrative skills. Early Childhood Research Quarterly, 66, 178–190. https://doi.org/10.1016/j.ecresq.2023.10.004

Carranza-Pinedo, V., & Diprossimo, L. (2025). Scaffolding and individuality in early childhood development. Topoi, 44(2), 491–503. https://doi.org/10.1007/s11245-024-10155-3

Ciaglia, F., Huang, Y., et al. (2023). Investigating preferential acquisition and attachment in early word learning through cognitive, visual and latent multiplex lexical networks. Physica A: Statistical Mechanics and Its Applications, 612, 128468. https://doi.org/10.1016/j.physa.2023.128468

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Thousand Oaks, CA: Sage Publications.

Ellis, P. K., Scherer, N. J., & Lien, K. M. (2024). Enhanced milieu teaching with phonological emphasis. International Journal of Speech-Language Pathology, 26(2), 179–193. https://doi.org/10.1080/17549507.2023.2199176

Fanoush, T., et al. (2025). Mispronunciation detection and diagnosis for young Arabic learners using transfer learning. IEEE Access, 13, 175047–175068. https://doi.org/10.1109/ACCESS.2025.3616335

Farah Hadban, Bassil Mashaqba, and Nisreen Al-Khawaldeh, "Acquisition of Arabic Guttural Consonants in Jordanian Arabic-English Bilingual Preschoolers," International Journal of Bilingualism (2026), https://doi.org/10.1177/13670069261466514

Farran, B. M., & Morett, L. M. (2024). Multimodal cues in L2 lexical tone acquisition: Current research and future directions. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1410795

Flege, J. E., & Wayland, R. (2021). The revised Speech Learning Model (SLM-r): Recent developments in second language speech acquisition research. Second Language Speech Learning, 7(2), 84–118. https://doi.org/10.1017/9781108886901.003

Fletcher, P., & MacWhinney, B. (Eds.). (2015). The handbook of child language. Oxford: Blackwell Publishing.

Fudhaili, A., et al. (2023). Arabic phoneme learning challenges for Madurese students and the solutions. Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab, 15(2), 417–431. https://doi.org/10.24042/albayan.v15i2.15837

Gibbs, R. W. (2005). Embodiment and cognitive science. Cambridge: Cambridge University Press.

Kahriman, D., & Pramling Samuelsson, I. (2026). Early childhood education for sustainability: Empowering toddlers' ecological wisdom in a polycrisis context. In Confronting the Polycrisis (pp. 245–259). Cham: Springer. https://doi.org/10.1007/978-3-032-22167-4_15

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. First Language, 45(6). https://doi.org/10.1177/01427237241290678

Kennington, C. (2021). Language acquisition is embodied, interactive, emotive: A research proposal. arXiv. https://doi.org/10.48550/arXiv.2105.04633

Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. London: Routledge.

Lantolf, J. P., Poehner, M. E., & Swain, M. (2018). The Routledge handbook of sociocultural theory and second language development. New York: Routledge. DOI:10.1080/13670050.2019.1652559

Lavechin, M., Dupoux, E., Cristia, A., et al. (2025). Simulating Early Phonetic and Word Learning Without Linguistic Categories. Developmental Science, 27(2), e13606. https://doi.org/10.1111/desc.13606

Mahon, B. Z. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. Psychonomic Bulletin & Review, 23, 941–958. https://doi.org/10.3758/s13423-016-1045-2

Masoumi, D., Bourbour, M., & Lindqvist, G. (2024). Scaffolding young children's learning through multimodal interactions in early childhood education. Early Childhood Education Journal, 52(3), 411–425. https://doi.org/10.1007/s10643-023-01510-x

Mihai, A., & Classen, A. (2023). A framework for scaffolding language development through meaningful interactions. Young Exceptional Children, 26(4), 233–244. https://doi.org/10.1177/10962506221149701

Occhino, C. (2017). An introduction to embodied cognitive phonology: Claw-5 hand-shape distribution in ASL and Libras. Complutense Journal of English Studies, 25, 69–103. https://doi.org/10.5209/CJES.57198

Pesco, D., & Gagné, A. (2015). Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. Early Education and Development, 26(7), 1052–1072. https://doi.org/10.1080/10409289.2015.1060800

Pramling Samuelsson, I., & Björklund, C. (2022). The relation of play and learning empirically studied and conceptualised. International Journal of Early Years Education, 31(1), 1–15. https://doi.org/10.1080/09669760.2022.2079075

Reggin, L. D., Franco, L. E. G., Horchak, O. V., Labrecque, D., Lana, N., Rio, L., & Vigliocco, G. (2023). Consensus paper: Situated and embodied language acquisition. Journal of Cognition, 6(1), 63. https://doi.org/10.5334/joc.308

Saiegh-Haddad, E., & Ghawi-Dakwar, H. (2018). Does linguistic distance affect phonological representation? Evidence from Arabic language learning. Journal of Child Language, 45(6), 1377–1399. https://doi.org/10.1017/S0305000918000302

Samuelsson, R. (2025). Creating a web of multimodal resources: Examining meaning-making during a children's book project in a multilingual community. Journal of Early Childhood Literacy, 25, 749–776. https://doi.org/10.1177/14687984231195179

Stampe, D. (1973). A dissertation on natural phonology. Chicago: University of Chicago Press.

van Manen, M. (2016). Researching lived experience: Human science for an action sensitive pedagogy (2nd ed.). New York: Routledge.

Vihman, M. M. (2014). Phonological development: The first two years (2nd ed.). Malden, MA: Wiley-Blackwell.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Wang, C. H., Chen, Y. C., & Tseng, W. T. (2023). Multimodal instruction and young learners' language development: A systematic investigation. SAGE Open, 13(4), 1–15. https://doi.org/10.1177/21582440231200151

Wang, Y., et al. (2023). A sculpting effect of reading on later representational quality of phonology revealed by multi-voxel pattern analysis in young children. Brain and Language, 239, 105252. https://doi.org/10.1016/j.bandl.2023.105252

Yakubu, & Obafemi. (2023). Systematic review of multiliteracies skills, multimodal tools to facilitate learning in early childhood classroom. Jurnal Inovasi Teknologi Pendidikan, 10, 431–440. https://doi.org/10.21831/jitp.v10i4.66343

Yusuf Arisandi, M. Yusuf Harun, and Aufa Hanin Salsabil, "Embodied Arabic Orthography through Rasm 'Uthmānī Copybook Practice: A Mixed-Methods Study in Malaysia," Arabiyatuna: Jurnal Bahasa Arab 10, no. 1 (2026): 317–340, https://doi.org/10.29240/jba.v10i1.16109

Zheng, Q., & Gor, K. (2024). The influence of native phonology, allophony, and phonotactics on nonnative lexical encoding. Language Learning, 74(1), 146–183. https://doi.org/10.1111/lang.12581