Elsya : Journal of English Language Studies
ISSN: 2337-6740 (Print) ISSN: 2337-6880 (Electronic)
DOI: https://doi.org/10.31849/r4nc9x58
Vol. 8, No. 1, 2026, pp. 15-27
Journal homepage: https://journal.unilak.ac.id/index.php/elsya

Original Research

Digital Phonological Visuals: Supporting Primary Learners' Early Oral English Competence & Development

Lulu Jola Uktolseja & Melda Agnes Manuhutu
Universitas Victory Sorong, Sorong, Indonesia

Article Info

Article history:
Received 28 November 2025
Revised 21 February 2026
Accepted 24 February 2026

Keywords:
Digital phonological visuals
Adaptive learning
English-speaking development
Primary education
Multimedia learning

Abstract

Early oral English competence is a fundamental foundation for children's academic participation and social communication. However, primary EFL classrooms in Indonesia continue to face challenges such as limited English exposure, memorization-oriented pedagogy, and insufficient phonology-focused learning media aligned with young learners' developmental needs. Although digital media is increasingly used in language education, empirical evidence remains limited on how phonological visual support can be systematically integrated with adaptive learning frameworks to enhance early speaking development at the primary level. To address this gap, this study developed and evaluated a Digital Phonological Visuals (DPV) based adaptive learning framework to support elementary students' English-speaking development. A sequential explanatory mixed methods research and development design was employed, guided by the 4D model. The intervention involved total sampling of 55 sixth-grade students at YPK Klasaman 2 Elementary School, Sorong City. Data were collected through pre- and post-speaking tests assessing pronunciation, fluency, and basic accuracy, complemented by questionnaires and systematic classroom observations. Quantitative results demonstrated a substantial improvement in speaking performance, with mean scores increasing from 44.2 in the pretest to 84.9 in the posttest, yielding a high normalized gain of 0.71. Regression analysis further showed a strong relationship between the DPV-based adaptive learning framework and speaking development ($R = 0.93$; $R^2 = 0.865$). Qualitative findings supported these results, revealing increased learner engagement, enhanced phonological awareness, and greater confidence in oral production. This study contributes a phonology grounded, classroom feasible adaptive digital model with implications for inclusive early EFL speaking in resource constrained primary education contexts.

Corresponding Author: *Uktolseja, lulujola39@gmail.com*

1. Introduction

Children's voices are often the first measure of whether language learning is alive in the classroom, and early English speaking mastery can become a long term foundation for communicative competence and participation in wider learning communities (Alharbi et al., 2024). At the elementary level, speaking is not merely an additional skill, it is a core dimension of English as a foreign language (EFL) learning because it operationalizes vocabulary, grammar, and pronunciation into meaningful interaction (Boonkit, 2010).

Yet, teaching speaking in primary classrooms remains persistently challenging, particularly when teachers are required to translate curricular goals into practical and effective instructional strategies for young learners (Mugo & Bwire, 2024). In Indonesia, these challenges are compounded by limited English exposure beyond school, entrenched memorization-oriented classroom practices, and learning resources that often fail to align with children's developmental needs and preferences (Sukarno & Jinabe, 2024). Although the Merdeka Curriculum promotes student centered learning and deeper oral engagement, its implementation

in everyday classrooms remains uneven due to constrained instructional capacity and limited resources (Irmawati, 2025; Widodo & Khoiriyah, 2025). Cooperative and technology supported practices such as Think Pair Share have shown promise for enhancing speaking skills, yet their effectiveness depends on sustained professional development, context responsive materials, and stronger alignment between family and school learning environments (Mugo & Bwire, 2024; Widodo & Khoiriyah, 2025). This misalignment is even more visible among alpha generation learners who grow up surrounded by smartphones, high speed internet, social media, and AI, shaping expectations for responsive and media rich learning experiences (Blanco & Gutiérrez, 2022).

Evidence from previous classroom research indicates that conventional approaches have not consistently optimized primary students' speaking outcomes, suggesting that instructional routines alone may not adequately support pronunciation, fluency, and willingness to speak (Gaspersz & Uktolseja, 2020; Uktolseja & Gaspersz, 2019). This strand of work implies a practical need for learning media that is contextual, engaging, and responsive to learner needs, particularly media that can scaffold productive and confident oral performance rather than rehearsed output (Derakhshan et al., 2015; Portugal-Toro et al., 2025). However, the pedagogical question is not simply whether media should be used, but what kind of linguistic support the media should foreground to make speaking development more systematic.

A growing body of linguistics informed pedagogy highlights phonological visualization as a promising pathway because it strengthens phonological awareness, a competence that underpins listening and speaking development through sound discrimination, sound to form mapping, and articulatory control (Alasmari, 2024; Huo & Wang, 2017; Stekić et al., 2023). Parallel research on digital and multimedia integration shows that well designed media can raise engagement and improve learning quality, especially when multimodal resources help learners attend to language form while remaining motivated to participate (Boonyopakorn et al., 2024; Sivakami & Gunasekaran, 2025; Staneviciene & Žekienė, 2025; Uktolseja et al., 2025). Building on this direction, Digital Phonological Visuals can be understood as digital media that supports sound comprehension and production by combining phonetic and phonological cues with interactive presentation (Abdoulqadir & Loizides, 2025; AlAli et al., 2025; Meade, 2020), and by integrating visual, textual, and audio inputs that help learners connect sound and meaning more coherently (Krüger & Bodemer, 2022; Pikoli & Lukum, 2021). Still, media effectiveness in real classrooms depends on whether the design can flexibly respond to diverse learner profiles and learning trajectories, which positions adaptive learning as a complementary framework for responsive instructional decisions (Ruan & Lu, 2025).

The niche of the present study lies in addressing a specific integration gap: although research has examined digital media for speaking, far fewer studies have combined digital phonological visuals with an adaptive learning framework at the elementary level. Existing work often emphasizes particular techniques or methods (Uktolseja & Manuhutu, 2018), gamification designs (Putu Wulantari & Uktolseja, 2023), or locally based media initiatives (Pambudi & Uktolseja, 2024), yet adaptive phonological visual media remains under examined, particularly in how it can systematically support pronunciation and speaking development for young EFL learners.

Accordingly, this study is significant because it positions phonological support as a central linguistic mechanism for speaking growth while operationalizing adaptivity as a classroom responsive design principle rather than a purely technological feature. The study aims to develop digital phonological visuals based on an adaptive learning framework, validate the feasibility of the developed learning media, and analyze the contribution of the media to the development of elementary school students' speaking ability.

The paper closes by consolidating what the findings mean for ELT, particularly for early speaking instruction in contexts with limited exposure and uneven learning resources. By foregrounding digital phonological visuals within an adaptive learning framework, the study offers a practical direction for teacher led media integration that is contextual, student centered, and oriented toward communicative competence. More broadly, the study supports an ELT agenda that treats pronunciation and oral confidence as teachable, designable outcomes in primary education, with implications for scalable speaking interventions in developing country settings.

2. Literature Review

Digital phonological visuals within an adaptive learning framework refer to digital media that integrate visual and phonological elements to support elementary learners' English language development. By linking visual components such as phonetic symbols, articulatory cues, and contextual images with digital sound

representations, learners are better able to understand relationships between sounds, words, and meaning. This approach aligns with dual coding theory and cognitive theory of multimedia learning, which posit that learning is enhanced when verbal and visual channels are activated simultaneously. In addition, adaptive learning systems allow instructional content to be adjusted to learners' engagement levels and abilities, supporting more personalized learning trajectories. As a result, pronunciation accuracy, speaking confidence, and oral participation can improve, positioning digital phonology as a visual, adaptive, and efficient framework for early English learning.

2.1 Phonological Awareness and Speaking Development in Primary EFL Learners

Phonological awareness refers to the ability to recognize, identify, and process the sound structure of spoken language, and it is widely acknowledged as a foundational component of early literacy development (Barakah et al., 2015; Ouazene et al., 2025a). In primary EFL contexts, strengthening phonological competence is also crucial for oral language development, as structured phonological awareness instruction tends to produce stronger learning outcomes than conventional approaches that do not explicitly foreground phonology (Hentasmaka et al., 2022). Although phonology oriented instruction is often introduced through reading activities, research shows that it can effectively transfer to pronunciation and speaking when connected to meaningful oral practice (Huo & Wang, 2017). Empirical studies further indicate that phonological awareness training sharpens phoneme articulation, improves fluency, and helps prevent early pronunciation errors from becoming fossilized (Alfaifi & Saleem, 2025). It also contributes to increased speaking motivation and confidence among young learners (Erudita et al., 2022).

However, much of the existing literature continues to position phonological awareness primarily as a predictor of reading development rather than as a direct driver of speaking ability (Mendes & Barrera, 2017; Ouazene et al., 2025). Moreover, visual and multimodal representations remain underutilized, despite strong evidence that primary learners respond positively to combined auditory and visual input. This gap highlights the need for research on communicative, context responsive, and visually supported phonological approaches that directly target speaking development in elementary EFL classrooms.

2.2 Digital Phonological Visuals and Adaptive Learning in Primary English Instruction

Phonology plays a central role in language learning as it underpins pronunciation accuracy, vocabulary development, and comprehension (Azzahra Iin, n.d.; Kart, 2023). In parallel, digital transformation in education emphasizes the integration of technology and multimedia resources to enhance learning engagement and effectiveness (Alenezi et al., 2023). Multimedia supported instruction, particularly in blended learning contexts, has been linked to improved academic outcomes and learner satisfaction (Meštrić et al., 2024). The Cognitive Theory of Multimedia Learning and dual coding theory explain how coordinated verbal and visual input supports deeper understanding and retention, making multimedia particularly relevant for phonology based instruction (Kanellopoulou et al., 2019; Staneviciene & Žekienė, 2025). Empirical evidence shows that technology enhanced speaking instruction increases engagement and confidence when visual aids help learners connect sounds, words, and meaning (Çeken & Taşkın, 2022). Building on this evidence, the present study conceptualizes Digital Phonological Visuals as an adaptive learning framework that integrates visual and interactive media to support elementary learners' phonological understanding and speaking development, contributing to improved oral language outcomes.

3. Method

This study adopted a mixed-methods sequential explanatory design, in which quantitative data collection and analysis preceded qualitative inquiry. This design was selected to evaluate the effectiveness of the Digital Phonological Visuals (DPV) based adaptive learning framework on students' English speaking development and to subsequently explain the quantitative outcomes through classroom-based evidence. Such an approach is appropriate for early EFL research, where performance gains require contextual interpretation related to learner engagement, phonological awareness, and confidence.

The quantitative phase employed a one-group pretest–posttest design, chosen due to practical and ethical constraints within the school context that necessitated equal instructional access for all students. While this design allows for examination of learning gains following the intervention, it does not fully control for threats to internal validity such as maturation or testing effects. These limitations were addressed through methodological triangulation and cautious interpretation of causal claims.

3.1 Participants and Context

The participants were 55 sixth-grade students (aged 11–12) from YPK Klasaman 2 Elementary School, Sorong City, Indonesia, selected through total sampling. English functioned as a foreign language with limited exposure beyond the classroom. Based on teacher reports and baseline assessment, students' proficiency ranged from beginner to lower-intermediate, with particular challenges in pronunciation accuracy, fluency, and oral confidence. Although students were familiar with basic digital devices, they had minimal prior experience with structured digital or phonology-focused learning tools.

3.2 Intervention

The intervention consisted of a Digital Phonological Visuals (DPV)–based adaptive learning framework, developed through a research and development process using the 4D model (Okpatrioka, 2023). The DPV materials integrated phonological visualization, modeled audio input, guided articulation cues, and short speaking tasks. Adaptivity was operationalized as instructional responsiveness, enabling flexible pacing, repetition, and differentiated phonological targets based on observed learner performance. The intervention was implemented over six instructional sessions, preceded by a pretest and followed by a posttest.

3.3 Instruments and Data Collection

Data were collected using four instruments. First, an English-speaking test was administered as a pretest and posttest, assessing pronunciation, fluency, and basic accuracy through an analytic rubric adapted for primary EFL learners. Second, a student questionnaire using a five-point Likert scale measured perceptions of the DPV-based adaptive framework and perceived speaking development; the instrument demonstrated high internal consistency (Cronbach's alpha = 0.953). Third, structured classroom observations documented engagement, phonological awareness behaviors, and participation. Fourth, semi-structured interviews with five students provided qualitative insights into learning experiences.

3.4 Data Analysis

Quantitative data were analyzed using descriptive statistics, normalized gain (N-gain) analysis to determine instructional effectiveness, and simple linear regression to examine the relationship between DPV-based adaptive learning and speaking development. Qualitative data from observations and interviews were analyzed thematically to explain learning processes and triangulate quantitative findings.

3.5 Ethical Considerations

The study followed school based ethical procedures suitable for research with minors. Permissions were obtained from the school, and informed consent was secured from parents or guardians with student assent prior to data collection. Participation was voluntary, and confidentiality was protected through anonymized reporting of student data, secure storage of test and questionnaire records, and removal of identifying information from transcripts.

4. Results

The findings indicate that the use of digital phonological visuals significantly improved students' English-speaking ability. A comparison of pretest and posttest results reveals clear progress in pronunciation, fluency, and basic grammatical accuracy, with posttest scores consistently surpassing those obtained in the pretest. These improvements suggest that the instructional intervention had a measurable and positive impact on students' speaking performance.

In addition to the test results, the questionnaire data demonstrate strong student approval of the learning media. Most learners reported that the digital phonological visuals helped them understand pronunciation more clearly, increased their confidence when speaking, and enhanced their overall motivation to learn English. These positive perceptions highlight the role of supportive visual tools in facilitating language acquisition.

Classroom observations further reinforced these findings. Throughout the meetings, students showed higher levels of engagement, participated more actively in speaking activities, and interacted more frequently with the teacher. This increased participation indicates that the use of digital phonological visuals not only improved speaking skills but also created a more dynamic and interactive learning environment.

4.1 N-Gain Test Result

Learning effectiveness was examined by comparing students' pre-test and post-test speaking scores using normalized gain (N-Gain) analysis. This analysis provides an interpretable measure of instructional impact by accounting for students' initial performance levels and the maximum attainable scores.

Table 4.1 Pre-test, Post-test, and N-Gain Results for Speaking Ability

Statistic	Pre-test	Post-test	Gain (Post–Pre)	Ideal Score	N-Gain	N-Gain (%)
Mean	44.2	84.9	40.7	55.8	0.71	70.9
Highest score	65	90	60	70	0.86	85.7
Lowest score	30	70	5	35	0.14	14.3
Number of students	55	55	55	55	55	55

The data in Table 4.1 reveal a pronounced improvement in students' English speaking performance following the implementation of the Digital Phonological Visual-Based Adaptive Learning Framework. The mean score increased from 44.2 in the pre-test to 84.9 in the post-test, representing an average gain of 40.7 points. This magnitude of improvement indicates not only statistical progression but also a substantial pedagogical shift in learners' speaking competence.

The average N-Gain value of 0.71 places overall learning effectiveness in the high category, demonstrating that students achieved more than seventy percent of the possible improvement relative to their initial ability. Notably, improvement occurred across the entire score range. Students with the lowest initial scores showed meaningful progress, while higher-performing students continued to advance, suggesting that the framework supported both remediation and enrichment. This pattern indicates that the combination of phonological visualization and adaptive pacing enabled learners to internalize sound patterns efficiently and apply them in oral production, reducing disparities linked to initial proficiency.

To further clarify the distribution of learning gains, N-Gain values were classified into categorical levels as presented in Table 4.2.

Table 4.2. N-Gain Classification Criteria

Category	N-Gain Range
High	≥ 0.70
Medium	0.30–0.69
Low	< 0.30

The classification framework in Table 4.2 helps contextualize the effectiveness of the intervention across learners. The predominance of high-category N-Gain scores indicates that the majority of students benefited strongly from the instructional framework. Importantly, no students remained in the low-gain category after the intervention, signaling that the learning design successfully prevented stagnation among lower-performing learners.

This distribution suggests that the adaptive features of the framework played a critical role in sustaining engagement and ensuring continuous progress. By allowing repeated exposure to phonological visuals, flexible pacing, and graduated speaking tasks, learners were able to consolidate pronunciation skills before advancing, resulting in consistent learning gains across the classroom. Overall, the N-Gain results demonstrate that the Digital Phonological Visual-Based Adaptive Learning Framework functioned as a high-impact and inclusive approach, promoting collective advancement in English speaking ability rather than isolated improvement among a limited subgroup.

4.2 Questionnaire Results

Students' perceptions of the Digital Phonological Visual-Based Adaptive Learning Framework and its contribution to speaking development were measured through a post-intervention questionnaire assessing two constructs: perceptions of the DPV-based adaptive framework and perceived English-speaking development.

Table 4.3. Descriptive Summary of Questionnaire Responses

Construct	Item Focus	Minimum	Maximum
DPV-based adaptive learning	Pronunciation support, visual clarity, learning pace, responsiveness, motivation	3.00	5.00
English speaking development	Pronunciation clarity, fluency, confidence, participation, sound discrimination	3.00	5.00

Responses ranged from 3.00 to 5.00 across all indicators, demonstrating consistently positive evaluations of both instructional design and speaking outcomes. The restricted upper-range distribution indicates that the framework was broadly perceived as pedagogically supportive rather than differentially effective. This pattern aligns with the inclusive gain distribution identified in the N-Gain analysis.

High ratings on pronunciation support and visual clarity suggest that phonological visualization functioned as a cognitive scaffold. By externalizing articulatory features and sound distinctions, the framework appears to have reduced the abstractness of phonology and strengthened phonological awareness, enabling learners to monitor and adjust oral production more deliberately. This mechanism is consistent with dual-channel processing principles, in which coordinated visual and auditory input enhances encoding and retrieval of linguistic forms.

Perceived speaking development mirrored these instructional evaluations. Students reported improvements in pronunciation clarity, fluency, confidence, and participation. The relatively elevated minimum scores across fluency and sound discrimination items indicate distributed perceived gains, reinforcing the interpretation that adaptive responsiveness mitigated proficiency gaps. Rather than privileging faster learners, flexible pacing and repeated visual feedback appear to have sustained engagement across varied readiness levels.

The instrument demonstrated very high internal consistency (Cronbach's Alpha = 0.953, 20 items), confirming construct coherence and strengthening inferential credibility.

Table 4.4 Model Summary for Regression Analysis

Statistic	Value
R	0.930
R Square	0.865
Adjusted R Square	0.863
Standard Error	1.65

Regression results indicate a strong predictive relationship between perceptions of the DPV-based adaptive framework and perceived speaking development (R = 0.930; R^2 = 0.865). The model explains 86.5% of the variance in perceived speaking growth, suggesting that learners attributed improvement to the structural features of the framework rather than incidental practice effects.

Theoretically, this relationship supports a dual-mechanism interpretation. First, phonological visualization strengthened articulatory awareness and sound discrimination, contributing to cognitive restructuring of pronunciation knowledge. Second, adaptive sequencing reduced performance anxiety and enhanced willingness to communicate, activating affective pathways that sustain oral participation. The convergence of these mechanisms indicates that speaking development emerged from the interaction of phonological precision and adaptive pedagogical design.

Overall, the questionnaire findings corroborate the quantitative gain results and classroom observations, providing theoretically grounded evidence that multimodal phonological scaffolding integrated with structured adaptivity can systematically support inclusive early EFL speaking development.

4.3 Observation and Interview Results

Classroom observations were conducted systematically during the pretest, six instructional meetings implementing the Digital Phonological Visual-Based Adaptive Learning Framework, and the posttest. These observations were complemented by semi-structured interviews with five selected students to deepen the interpretation of behavioral and developmental patterns. The integrated analysis reveals how phonological awareness, engagement, confidence, motivation, and adaptive responsiveness evolved across the instructional cycle.

Table 4.6 Thematic Summary of Classroom Observation Results

No.	Theme	Observed Indicators	Students (n = 55)	Observation Phase
1	Initial learning difficulties	Difficulty operating digital tools and interpreting phonological visuals	9 students (16.4%)	Pretest and Session 1
2	Phonological awareness improvement	Ability to identify English sounds, observe articulation, and pronounce words accurately	46 students (83.6%)	Sessions 3 to 6 and Posttest
3	Increased learning engagement	Active participation, spontaneous repetition, and involvement in pronunciation tasks	48 students (87.3%)	Sessions 2 to 6
4	Improved speaking confidence	Greater willingness to speak, ask questions, and speak audibly in English	42 students (76.4%)	Sessions 4 to 6 and Posttest
5	Motivation to learn English	Sustained attention, positive attitudes, and enthusiasm toward learning activities	44 students (80.0%)	Sessions 2 to 6
6	Support for individual learning pace	Continued engagement despite different learning speeds through repeated visual feedback	50 students (90.9%)	All sessions

The observation data indicate a clear developmental progression from initial adjustment to sustained and confident oral participation. Although 16.4% of students experienced early difficulty navigating the digital interface and interpreting phonological visuals, these challenges were temporary and largely resolved after the first instructional meeting. The difficulty reflected cognitive adaptation to a new multimodal representational system rather than linguistic limitation. This transitional phase is captured in the following interview reflections:

Excerpt S1–IS1–T00:6:34
"Awalnya saya bingung karena gambar mulut dan suaranya belum cocok di kepala saya."
"At first I was confused because the mouth picture and the sound did not match in my head."

Excerpt S2–IS1–T00:18:09
"Saya takut salah, jadi saya tekan suaranya berkali-kali."
"I was afraid of being wrong, so I pressed the sound many times."

These excerpts suggest that learners were recalibrating their attention between auditory input and visual articulatory cues. Once this coordination stabilized, attention shifted toward phonological processing itself. The diminishing of early hesitation demonstrates that structured visual scaffolding can be internalized rapidly by primary learners.

A more substantial transformation emerged from Sessions 3 to 6, where 83.6% of students demonstrated improved phonological awareness. Observational evidence showed students adjusting tongue and lip placement after consulting the visual display and distinguishing previously confusable sounds. Pronunciation moved from imitative repetition to conscious monitoring.

Excerpt S3–IS3–T00:12:41
"Kalau saya lihat posisi lidahnya, saya bisa perbaiki bunyinya."
"When I see the tongue position, I can correct the sound."

Excerpt S4–IS4–T00:5:02
"Sekarang saya bisa dengar beda antara ship dan sheep."
"Now I can hear the difference between ship and sheep."

These reflections reveal the emergence of articulatory awareness. Phonological visualization externalized abstract sound structures, enabling learners to observe and regulate their own pronunciation. The shift from guessing to deliberate correction marks a critical cognitive milestone in early speaking development.

Engagement also intensified across Sessions 2 to 6, with 87.3% of students actively participating in repetition cycles and peer-supported pronunciation tasks. Notably, repetition was not mechanical but reflective. Students used visual feedback as a reference point for collaborative verification.

Excerpt S5–IS2–T00:4:55
"Rasanya seperti latihan, bukan ujian."
"It feels like practice, not a test."

Excerpt S3–IS5–T00:4:10
"Kami lihat lagi gambarnya bersama."
"We looked at the picture again together."

These accounts indicate that the learning environment reframed pronunciation practice as iterative refinement rather than evaluative performance. Visual mediation redistributed authority from teacher to learner, fostering collaborative sound exploration. Engagement thus reflected cognitive investment and shared problem-solving rather than simple compliance.

Confidence gains became particularly visible during Sessions 4 to 6, where 76.4% of students demonstrated greater audibility and willingness to speak publicly. The adaptive structure allowed rehearsal before performance, reducing fear of immediate correction.

Excerpt S2–IS4–T00:3:20
"Sekarang saya bisa cek dulu, jadi saya berani bicara."
"Now I can check first, so I am brave to speak."

Excerpt S4–IS6–T00:8:03
"Saya bisa lihat dan coba lagi."
"I can see it and try again."

These statements suggest that confidence developed through perceived control. When pronunciation errors became adjustable rather than permanent, anxiety diminished. The visual scaffolding therefore functioned not only cognitively but affectively, strengthening willingness to communicate.

Finally, adaptive pacing emerged as a central inclusive mechanism. Across all sessions, 90.9% of students remained engaged despite differing learning speeds. Slower learners repeated sound sequences without stigma, while faster learners extended their practice independently.

Excerpt S1–IS5–T00:37:44
"Saya tidak merasa tertinggal karena bisa ulang."
"I do not feel left behind because I can repeat it."

Excerpt S5–IS6–T00:51:12
"Saya coba kata yang lebih sulit supaya tidak bosan."
"I try more difficult words so I do not feel bored."

This balance between repetition and progression preserved classroom cohesion while accommodating individual variation. Adaptivity operated not as technological complexity but as pedagogical responsiveness embedded in structured visual feedback.

The integrated observation and interview findings reveal a coherent developmental arc: learners moved from representational confusion to articulatory awareness, from passive imitation to collaborative refinement, and from hesitation to confident oral participation. The Digital Phonological Visual-Based Adaptive Learning Framework did more than increase speaking practice; it reconfigured the epistemology of pronunciation learning. By making sound visible and errors revisable, it transformed phonology from an abstract auditory demand into a tangible and negotiable learning object.

This transformation explains the widespread improvements observed across engagement, confidence, and phonological precision. Speaking development occurred not merely because students practiced more, but because they understood how sounds were formed, why differences mattered, and how correction could be achieved. In this way, visualization and adaptive pacing functioned synergistically to create a cognitively grounded and affectively secure environment for early EFL oral development.

5. Discussion

This study demonstrates a substantial improvement in primary EFL learners' speaking development after the implementation of the Digital Phonological Visuals based Adaptive Learning Framework. Quantitatively, students' mean speaking score increased from 44.2 in the pretest to 84.9 in the posttest, with a normalized gain of 0.71 classified as high. In parallel, the regression results indicate a strong relationship between the DPV based adaptive learning framework and speaking development, with R squared of 0.865, suggesting that the framework explains a large proportion of variance in students' speaking outcomes. Qualitatively, classroom observations corroborate these outcomes by showing increased phonological awareness, engagement, motivation, confidence, and support for varied learning pace across most learners. Taken together, the convergence of test gains, learner perceptions, and observation themes strengthens the internal coherence of the findings and indicates that DPV supported adaptivity can function as an effective route to early speaking growth in contexts with limited English exposure (Boonkit, 2010; Huo & Wang, 2017).

From a language learning perspective, the most plausible mechanism underlying these gains is the strengthening of phonological awareness as a gateway to more accurate and confident oral production. Prior work positions phonological awareness as foundational for sound discrimination, articulation, and error prevention, particularly when instruction is explicit and structured (Alasmari, 2024; Hentasmaka et al., 2022; Alfaifi & Saleem, 2025). This study extends that logic by showing that phonological knowledge can be activated through digital visual representations that make segmental features and articulation more tangible for children. In other words, DPV likely reduces the abstractness of phonology for young learners by externalizing sound structure into perceivable visual cues, which can accelerate mapping between phonemes, words, and meaning, and in turn support fluency and pronunciation accuracy (Meade, 2020; Kart, 2023). The observation pattern, where early difficulties with digital tools diminished while phonological awareness and confidence rose, is consistent with a progression from initial cognitive load toward more automated decoding and production as learners become familiar with the representational system and practice routines (Çeken & Taşkın, 2022).

The multimedia nature of DPV also aligns strongly with dual coding and cognitive theory of multimedia learning principles. When instruction coordinates auditory input with visual representations and guided attention, learners can distribute processing across verbal and visual channels, potentially improving retention and reducing confusion in pronunciation learning (Kanellopoulou et al., 2019; Krüger & Bodemer, 2022; Staneviciene & Žekienė, 2025). For primary learners who are often more responsive to multimodal cues, DPV can help stabilize sound perception and production by offering repeated, consistent visual

feedback that complements teacher modeling. This may also explain the reported increases in motivation and engagement, since interactive multimodal materials tend to support sustained attention and reduce anxiety in speaking tasks by providing scaffolding that makes success more attainable (Derakhshan et al., 2015; Boonyopakorn et al., 2024). Importantly, the participants are situated in a digital era where learners' everyday experiences are saturated with screens, icons, and media forms, so DPV can also be interpreted as instructionally congruent with children's multimodal literacy practices (Blanco & Gutiérrez, 2022).

A further contribution of this study lies in its adaptive learning framing. Here, adaptivity is conceptualized as responsiveness to learners' needs and pace through instructional design rather than reliance on AI automation alone. This design centric perspective is theoretically meaningful because personalization can be achieved through structured sequencing, scaffolding, and targeted supports that accommodate learner variability even in contexts with limited technological infrastructure (Akavova et al., 2023; Kasztelnik, 2024). While the literature recognizes AI as a powerful enabler of adaptive pathways capable of tailoring content, pacing, and feedback, AI should be viewed as a complement rather than a prerequisite for adaptivity (Aggarwal et al., 2023; Akavova et al., 2023; Kasztelnik, 2024). When available, generative AI tools can further enhance personalization through features such as real time summarization and interactive tutoring (Jayavardhini, 2024). Overall, this framing supports a pragmatic approach to personalization that sustains engagement among slower learners while appropriately challenging faster learners, even in resource constrained classroom settings. Adaptive learning research emphasizes the value of adjusting pathways and feedback based on learner response patterns and progress. Even when the adaptivity is teacher guided, a structured framework can operationalize personalization in practical ways (Ruan & Lu, 2025; Sharma et al., 2025). The observation theme showing that most students remained engaged despite differing learning speeds indicates that the DPV framework likely offered repeated opportunities and differentiated pacing that reduced participation gaps, which is essential in inclusive primary EFL classrooms.

Despite the strong outcomes, several gaps limit the inferences that can be drawn and should be acknowledged explicitly. First, the design appears to rely on a single group pretest posttest structure within one school context and one grade level, which weakens causal attribution because maturation, teacher effects, novelty effects, or test familiarity could contribute to score changes. Second, the sample is context specific and relatively small, so generalizability across regions, school types, and learner profiles remains uncertain. Third, although questionnaire and observation data enrich interpretation, self-reported perceptions may be influenced by social desirability and teacher presence, while observational coding could benefit from inter rater reliability procedures. Fourth, the study focuses primarily on short term gains. It remains unclear whether improvements persist over time and transfer to spontaneous speaking outside structured tasks, or to related skills such as listening discrimination and early literacy.

The manuscript itself positions a broader gap in prior research, namely limited studies that integrate digital phonological visuals with an adaptive learning framework at the elementary level. This study addresses that gap, yet it also creates a new set of questions about which component drives the strongest effect. For example, the relative contribution of phonological visualization versus adaptivity is not separated analytically. Similarly, while speaking was assessed in terms of pronunciation, fluency, and basic accuracy, the study does not yet clarify whether gains are driven more by segmental improvements, rhythm and stress alignment, increased willingness to communicate, or task familiarity. These unanswered questions define the next boundary of empirical work needed to strengthen theoretical claims.

The novelty of this study lies in its integration of digital phonological visuals with an adaptive learning framework to target early speaking development, not merely vocabulary recognition or reading oriented phonological tasks. The framework operationalizes phonology as a multimodal, learnable system that can be taught responsively to learners' pace, making phonological awareness instruction directly consequential for oral performance, confidence, and engagement. This is a meaningful contribution because many phonological awareness interventions in EFL are still positioned as literacy support rather than as a direct lever for speaking development in young learners (Huo & Wang, 2017; Erudita et al., 2022).

Moreover, the strong quantitative effect paired with convergent qualitative evidence positions DPV based adaptivity as a viable classroom level innovation that does not depend on high end AI infrastructure, which is particularly relevant for developing country contexts and resource limited schools.

Practically, the findings imply that teachers can strengthen speaking development by embedding explicit phonological visualization into routine speaking tasks, using repeated guided practice and paced progression so that learners with different readiness levels can participate meaningfully. Instructional designers and curriculum teams can treat DPV as a design principle for early speaking materials, ensuring that sound symbol articulation links are visible and systematically recycled across lessons. For teacher education, the results suggest that training should not only cover communicative activities but also develop teachers' capacity to teach pronunciation through multimodal scaffolding, including how to diagnose phonological difficulties and select visuals that match common learner errors (Liu, 2021). At the policy level, the study supports broader calls for earlier and more sustainable English instruction that is developmentally appropriate and aligned with learners' digital realities (Alharbi et al., 2024).

Future studies should test DPV based adaptive learning using more rigorous comparative designs, such as quasi experimental control groups or randomized classroom trials, to strengthen causal inference and isolate the contribution of visualization versus adaptivity. Longitudinal follow ups are needed to evaluate retention, transfer to spontaneous interaction, and downstream effects on listening discrimination and early literacy. Researchers should also investigate contextual moderators such as learners' initial phonological awareness, home language phonology, teacher expertise, and technology access, and should expand to diverse school settings and grade levels. Methodologically, future work can combine rubric based speaking assessment with acoustic or articulatory measures to capture micro level pronunciation changes, and can implement stronger qualitative trustworthiness procedures such as inter rater coding, member checking, and triangulation across teacher interviews, learner artifacts, and classroom video data.

6. Conclusion

This study concludes that the integration of Digital Phonological Visuals within an adaptive learning framework constitutes a robust and pedagogically meaningful approach to enhancing early EFL speaking development in primary classrooms. The key findings demonstrate substantial improvement in learners' pronunciation accuracy, fluency, and oral confidence, supported by strong learning gains, positive learner perceptions, and convergent classroom evidence of increased engagement, motivation, and participation. Collectively, these results confirm that phonology-centered visual scaffolding, when combined with responsive instructional pacing, can function as a powerful mechanism for reducing early speaking barriers and supporting inclusive learning progress across diverse proficiency levels.

The central novelty of this study lies in repositioning phonological visuals not as a supplementary aid for reading or listening, but as a direct pedagogical lever for oral language development, thereby extending phonological awareness from a literacy-oriented construct to a core driver of communicative competence in early ELT. From a research perspective, the study contributes empirical evidence to mixed-methods ELT scholarship by linking measurable speaking gains with observable classroom processes, while from a practical standpoint it offers a classroom-feasible model that can be implemented without reliance on advanced technological infrastructure, making it particularly relevant for resource-constrained contexts. The findings imply that ELT practitioners and curriculum designers should embed explicit, multimodal phonological instruction into routine speaking activities and adopt adaptive strategies that accommodate learners' varied developmental trajectories. Future research should pursue controlled or comparative designs to isolate the relative contributions of phonological visualization and adaptive sequencing, examine longitudinal effects on retention and transfer to spontaneous interaction, and explore theoretically grounded extensions such as the interaction between phonological visuals and embodied cognition, working memory load, or willingness to communicate across different age groups, proficiency levels, and sociolinguistic contexts.

References

Abdoulqadir, C., & Loizides, F. (2025). Interaction, artificial intelligence, and motivation in children's speech learning and rehabilitation through digital games: A systematic literature review. *Information*, *16*(7), Article 599. https://doi.org/10.3390/info16070599

Aggarwal, D., Sharma, D., & Saxena, A. (2023). Exploring the role of artificial intelligence for augmentation of adaptable sustainable education. *Asian Journal of Advanced Research and Reports*, *17*(11), 179–184. https://doi.org/10.9734/ajarr/2023/v17i11563

Akavova, A., Temirkhanova, Z., & Lorsanova, Z. (2023). Adaptive learning and artificial intelligence in the educational space. *E3S Web of Conferences*, *451*, Article 06011. https://doi.org/10.1051/e3sconf/202345106011

AlAli, R., Al-Barakat, A., Bataineh, R. F., & Alqatawna, M. (2025). From pixels to prose: Teachers' views on the power of digital imagery in early language development. *Forum for Linguistic Studies*, *7*(2), 160–173. https://doi.org/10.30564/fls.v7i2.8186

Alasmari, A. (2024). Do linguistic factors impact phonological awareness in L2? A study of phonological awareness of English by adult Saudi learners. *Arab World English Journal*, *15*(2), 337–361. https://doi.org/10.24093/awej/vol15no2.21

Alenezi, M. (2023). Digital learning and digital institution in higher education. *Education Sciences, 13*(1), Article 88. https://doi.org/10.3390/educsci13010088

Alfaifi, A. A., & Saleem, M. (2025). Raising phonological awareness to rectify the misarticulation of plosive consonants among adult EFL learners. *World Journal of English Language*, *15*(8), 84–95. https://doi.org/10.5430/wjel.v15n8p84

Alharbi, A., Almazroei, N., Alfaleh, R., Alsubaie, S., Almutairi, M., & Alruwibkh, Y. (2024). Fostering early English education in Saudi Arabia: A pathway to achieving the United Nations' Sustainable Development Goals in education. *Forum for Linguistic Studies*, *6*(5), 550–563. https://doi.org/10.30564/fls.v6i5.7107

Azzahra Iin, F. (2025). Navigating the landscape of phonology research: A systematic review of current trends and interdisciplinary applications. *ELite Journal: International Journal of Education, Language, and Literature*, *5*(2), 21-33. https://doi.org/10.26740/elitejournal.v5n2.p21-33

Barakah, M. A. S., Elshobary, A. M., El-Assal, N. N., Habil, I. S., Eldin, S. T. K., & El-Refaie, D. A. (2015). Phonological awareness deficits in Arabic-speaking children with learning disabilities. *Egyptian Journal of Otolaryngology*, *31*(2), 140–142. https://doi.org/10.4103/1012-5574.156103

Blanco, E. F., & Gutiérrez, M. R. (2022). Kid influencers as content creators: Legal implications. *VISUAL Review: International Visual Culture Review*, *9*, Article 3553. https://doi.org/10.37467/revvisual.v9.3553

Boonkit, K. (2010). Enhancing the development of speaking skills for non-native speakers of English. *Procedia – Social and Behavioral Sciences*, *2*(2), 1305–1309. https://doi.org/10.1016/j.sbspro.2010.03.191

Boonyopakorn, J., Tasatanattakool, P., Nilsook, P., & Wannapiroon, P. (2024). Mobile language learning: A digital approach to improving English communication. *International Journal of Interactive Mobile Technologies*, *18*(22), 159–173. https://doi.org/10.3991/ijim.v18i22.50001

Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: A systematic review. *Smart Learning Environments*, *9*(1), Article 20. https://doi.org/10.1186/s40561-022-00200-2

Derakhshan, A., Tahery, F., & Mirarab, N. (2015). Helping adult and young learners to communicate in speaking classes with confidence. *Mediterranean Journal of Social Sciences*, *6*(2), 520–525. https://doi.org/10.5901/mjss.2015.v6n2p520

Erudita, K., Astuti, P. S., & Sukanadi, N. L. (2022). Examining the effect of phonological awareness instruction on EFL learners' pronunciation and motivation. *Erudita: Journal of English Language Teaching*, *2*(2), 129–147. https://doi.org/10.28918/erudita.v2i2.796

Gaspersz, S., & Uktolseja, L. J. (2020). The use of serial picture media to improve English speaking ability. *Jurnal Pendidikan Bahasa*, *7*(2).

Hentasmaka, D., Cahyono, B. Y., Basthomi, Y., & Puspitasari, Y. (2022). The effect of embedding phonological awareness training on adult EFL learners' phonological awareness skill. *European Journal of Educational Research*, *11*(4), 2023–2030. https://doi.org/10.12973/eu-jer.11.4.2023

Huo, S., & Wang, S. (2017). The effectiveness of phonological-based instruction in English as a foreign language students at primary school level: A research synthesis. *Frontiers in Education*, *2*, Article 15. https://doi.org/10.3389/feduc.2017.00015

Irmawati, M. (2025). Think pair share and Google Classroom in teaching English for young learners: A literature review. *Language Circle: Journal of Language and Literature*, *19*(2S), 294–310. https://doi.org/10.15294/lc.v19i2s.31214

Jayavardhini, P. (2024). AI study partner: Development of an LLM and GenAI-enhanced study assistant tool. *International Journal of Scientific Research in Engineering and Management*, *8*(3), 1–5. https://doi.org/10.55041/ijsrem29626

Kanellopoulou, C., Kermanidis, K. L., & Giannakoulopoulos, A. (2019). The dual-coding and multimedia learning theories: Film subtitles as a vocabulary teaching tool. *Education Sciences*, *9*(3), Article 210. https://doi.org/10.3390/educsci9030210

Kart, A. N. (2023). Visual phonics: Relevant for all early readers, especially struggling readers. *Human Research in Rehabilitation*, *13*(1), 141–147. https://doi.org/10.21554/hrr.042316

Kasztelnik, K. (2024). Artificial intelligence-assisted curriculum development: Innovations in designing educational content for the 21st century learner. *Journal of Higher Education Theory and Practice*, *24*(11). https://doi.org/10.33423/jhetp.v24i11.7367

Krüger, J. M., & Bodemer, D. (2022). Application and investigation of multimedia design principles in augmented reality learning environments. *Information*, *13*(2), Article 74. https://doi.org/10.3390/info13020074

Liu, X. (2021). A study of English phonetic teaching strategies from the perspective of embodied cognition. *Theory and Practice in Language Studies*, *11*(5), 556–560. https://doi.org/10.17507/tpls.1105.14

Meade, G. (2020). The role of phonology during visual word learning in adults: An integrative review. *Psychonomic Bulletin & Review*, *27*(1), 15–23. https://doi.org/10.3758/s13423-019-01647-0

Mendes, G. G., & Barrera, S. D. (2017). Phonological processing and reading and writing skills in literacy. *Paidéia (Ribeirão Preto), 27*, 298–305. https://doi.org/10.1590/1982-43272766201711

Meštrić, K. B., Maravić, J., & Hribar, M. (2024). Multimedia use in higher education: e-Universities project.

Mugo, T. M., & Bwire, A. M. (2024). Strategies used in teaching English language oral skills and effects on primary school learner participation in Embu County, Kenya. *International Journal of Multidisciplinary Research and Growth Evaluation, 5*(2), 588–593.

Okpatrioka. (2023). Research and development (R&D) penelitian yang inovatif dalam pendidikan. *Dharma Acariya Nusantara: Jurnal Pendidikan, Bahasa dan Budaya*, *1*(1).

Ouazene, Z., Karroum, A., Amraoui, J., & Gougil, R. (2025). Phonological awareness and early Arabic literacy skills: Predictive insights from a Moroccan case study. *Statistics, Optimization and Information Computing*, *15*(1), 835–847. https://doi.org/10.19139/soic-2310-5070-2886

Pambudi, N., & Uktolseja, L. J. (2024). Bridging cultures enriching English language learning through local heritage. *Journal of English Research*, *5*(3). https://doi.org/10.47827/jer.v5i3.410

Pikoli, M., & Lukum, A. (2021). Development of audio-visual learning media integrating character education in chemistry learning. *Journal of Physics: Conference Series*, *1968*(1), Article 012007. https://doi.org/10.1088/1742-6596/1968/1/012007

Portugal-Toro, A., García-Peñalvo, F. J., Balderas Ruiz, L. A., & Vences Esparza, A. (2025). Cognitive and affective processes in second language oral communication: A mixed methods research. *Frontiers in Education*, *10*, Article 1571099. https://doi.org/10.3389/feduc.2025.1571099

Putu Wulantari, N., & Uktolseja, L. J. (2023). The role of gamification in English language teaching: A literature review. *Journal on Education*, *6*(1), 2847–2856. https://doi.org/10.31004/joe.v6i1.3328

Ruan, S., & Lu, K. (2025). Adaptive deep reinforcement learning for personalized learning pathways. *Computers and Education: Artificial Intelligence*, *9*, Article 100463. https://doi.org/10.1016/j.caeai.2025.100463

Sharma, A., Kaushik, N., Srivastav, A., Verma, S. K., Khurana, J., & Bhowmik, M. (2025). Adaptive learning systems for multimedia design education. *ShodhKosh: Journal of Visual and Performing Arts*, *6*(2s), 23–33. https://doi.org/10.29121/shodhkosh.v6.i2s.2025.6701

Sivakami, N., & Gunasekaran, S. (2025). Innovative instructional strategies for the digital age: Multimedia-enhanced writing pedagogy in English language classrooms. *International Journal of Basic and Applied Sciences*, *14*(2), 721–728. https://doi.org/10.14419/m63na081

Staneviciene, E., & Žekienė, G. (2025). The use of multimedia in the teaching and learning process of higher education: A systematic review. *Sustainability*, *17*(19), Article 8859. https://doi.org/10.3390/su17198859

Stekić, K., Ilić, O., Ković, V., & Savić, A. M. (2023). ERP indicators of phonological awareness development in children: A systematic review. *Brain Sciences*, *13*(2), Article 290. https://doi.org/10.3390/brainsci13020290

Sukarno, S., Riyadini, M. V., Jinabe, M., & Lestari, A. E. (2024). Unveiling pre-service English teachers' perceptions of differentiated instruction. *Diksi, 32*(1), 47–60. https://doi.org/10.21831/diksi.v32i1.64077

Uktolseja, L. J., & Gaspersz, S. (2019). Implementation of chain story techniques to improve English speaking ability. *Journal of Education and Learning*, *13*(4), 490–497. https://doi.org/10.11591/edulearn.v13i4.13327

Uktolseja, L. J., & Manuhutu, M. A. (2018). The use of whispering team to enhance students' listening skill. *Getsempena English Education Journal*, *5*(2).

Uktolseja, L. J., Manuhutu, M. A., & Wonmaly, W. (2025). A hybrid ADDIE-RAD approach to developing a multimedia learning application. *International Journal of Engineering Trends and Technology*, *73*(9), 79–90. https://doi.org/10.14445/22315381/IJETT-V73I9P108

Widodo, E., & Khoiriyah, K. (2025). Teaching English to young learners: Professional development trainings for primary school teachers in implementing Kurikulum Merdeka. *Abdi Jurnal Pengabdian dan Pemberdayaan Masyarakat*, *7*(2), 635–642. https://doi.org/10.24036/abdi.v7i2.1466