



DEVELOPING AN AI-ASSISTED ANIMATED ENGLISH ALPHABET SONG PROTOTYPE FOR INDONESIAN PRIMARY SCHOOL EFL LEARNERS

Ahyadi¹, Hustiana², Asdar³

^{1,2,3}Universitas Sulawesi Barat, Indonesia

E-mail: 1ahyadi@unsulbar.ac.id, 3hustiana@unsulbar.ac.id, 2asdar@unsulbar.ac.id

Accepted :

12 December 2025

Published :

12 December 2025

Corresponding Author:

Ahyadi

Email Corresponding :

ahyadi@unsulbar.ac.id

ABSTRACT

This study aims to design and develop an AI-assisted animated English alphabet song prototype for beginner EFL learners in Indonesian primary schools, focusing on Grades 4–6, applying the Analyze, Design, and Develop stages of the ADDIE model to produce curriculum-aligned multimedia that enhances alphabet recognition and accurate pronunciation. The research employed curriculum review, classroom observations, teacher interviews, and media analysis across five public schools in Majene, Sulawesi Barat, using Suno AI for vocal processing, Gemini for character images, VEO3 for lipsynced animation, and Capcut for video editing. Expert evaluation using a structured evaluation form assessed content validity, language quality, media design, and pedagogical suitability, yielding Aiken's V scores from 0.833 to 1.000, confirming the prototype's validity. The prototype can be applied in primary school English lessons, teacher education, instructional design, and educational technology courses to support early literacy, phonics instruction, and curriculum-aligned multimedia development. This study presents a novel approach by integrating generative AI voice processing, hybrid animated characters, and multimedia learning principles, providing an empirically validated, replicable framework for developing localized, high-quality instructional media for beginner EFL learners.

Keywords: *AI-assisted learning, Animated alphabet song, English language teaching, Primary school EFL learners, multimedia learning, Instructional design.*

1. INTRODUCTION

In response to shifting national goals and international pressures, Indonesia has reinterpreted the role of English in elementary education multiple times. Many public primary schools now only begin formal English instruction in the upper grades, usually from Grade 4 to Grade 6, as a result of the recent Merdeka Belajar reforms, which position English as an optional local content subject rather than a required core subject (Sulistiyo et al., 2020; Widagsa & Khusnia, 2023). Due to this structure, many students encounter English for the first time rather late in their education and over a very

small number of classroom hours. Additionally, research shows that public primary school teachers frequently struggle with time, training, and age-appropriate teaching resources for young EFL students, especially outside major cities (Sulistiyo et al., 2020; Dessler, 2020). When taken as a whole, these factors make it evident that there is a need for interesting and easily accessible materials that may assist in teaching English in ways that are both pleasurable and pedagogically sound (Bhwana, 2023; Ridwan, 2019).

Vol 9, No 1 (2026): ESTEEM

Learning the English alphabet, identifying letters, pronouncing their names, and starting to associate visual symbols with corresponding sounds are among the initial literacy requirements imposed on beginning EFL learners. Alphabet knowledge is a powerful predictor of subsequent reading and spelling outcomes in languages that utilize alphabetic scripts, according to research on early literacy (Piasta & Wagner, 2010; Baleghizadeh & Maryam, 2019). The mismatch between English's more erratic spelling and Bahasa Indonesia's rather constant letter-sound mapping makes this assignment challenging for Indonesian learners. If learners do not receive sufficient, well-structured exposure to English letter names and sounds, they may struggle with the pronunciation of certain letters (such as V, F, G, J, and Z) and carry those inaccuracies into later reading and speaking. It is therefore important to design learning experiences that provide repeated, accurate, and developmentally appropriate practice with the English alphabet (Ditania, 2022; Sunde et al., 2020).

For young learners in EFL contexts, songs are widely regarded as a powerful medium for presenting language input. Musical activities combine rhythm, melody, and repetition in ways that can support attention, memory, and positive attitudes toward learning (Saldiraner & Cinkara, 2021). In Indonesian classrooms, teachers frequently draw on songs to reduce anxiety and make English more approachable, and several classroom-based reports suggest that alphabet songs can help children remember letter sequences and enjoy the learning process (Yeni & Amelia, 2020). However, many of the alphabet songs available on the internet were created for children in English-speaking environments. The speed, accent, and sometimes the cultural references in these songs do not always match the needs of Indonesian primary school beginners, and their integration with the national curriculum

is rarely considered. As a result, teachers may use these materials in an ad hoc way, without a clear instructional design rationale (Akram et al., 2022; Malini et al., 2022).

Moving beyond audio alone, animated multimedia offers additional affordances for early literacy. The cognitive theory of multimedia learning argues that learners can benefit when verbal information is combined with well-chosen visuals, provided that the design avoids unnecessary distractions and supports the integration of words and images in working memory (Mayer, 2021, 2022; Littleton & Mercer, 2013). Principles such as coherence, signalling, and temporal contiguity imply that showing written letters and related imagery in synchrony with spoken letter names can help learners form stronger and more stable associations between what they see and what they hear. For young EFL learners, an animated alphabet song that carefully coordinates visuals, text, and audio has the potential to support both cognitive processing and emotional engagement (Sunde et al., 2020; Boardman et al., 2024)

Recent developments in generative artificial intelligence (AI) have lowered the barrier for producing such multimedia resources. AI-based tools can assist users in generating or transforming audio, music, images, and simple animations, even when users do not have advanced technical or artistic skills. Reviews of generative AI in education show that these tools are increasingly used to support content creation, instructional design, and the production of digital learning materials, although most reported applications to date focus on higher education and adult learners (Yusuf et al., 2024; Zhang et al., 2024). Emerging work in school contexts suggests that AI can function as a creative assistant for designing digital resources and learning tasks, but systematic, classroom-ready examples for young language learners remain relatively rare (Alfarwan, 2025). This gap is particularly

Vol 9, No 1 (2026): ESTEEM

evident in the area of early literacy and alphabet learning.

In Indonesian primary education, existing research on English tends to emphasize policy analysis, teacher perceptions, and traditional teaching strategies (Sulistiyono et al., 2020; Widagda & Khusnia, 2023). While songs and games are often mentioned as recommended techniques for young learners, there is still a shortage of studies that document how multimedia or AI-assisted materials are actually designed, developed, and validated for use with primary EFL learners (Saha et al., 2023). Alphabet songs and animated videos that teachers can find on public platforms such as YouTube are mostly global, ready-made products, and the underlying pedagogical and design considerations are seldom reported or examined in the research literature (Yeni & Amelia, 2020). Consequently, teachers have limited guidance on how to adapt such resources to the Indonesian curriculum or how to create their own materials using emerging AI tools.

In order to address these objectives, the current study takes a design and development approach and uses a few phases of the ADDIE instructional design model Analyze, Design, and Develop as a framework (Branch, 2009). The goal of the project is to develop an animated English alphabet song prototype with AI assistance for Indonesian primary school EFL students in Grades 4-6 who are just starting to study the language. The prototype consists of three primary parts: (1) an English alphabet song with lyrics, tempo, and pronunciation tailored to Indonesian EFL beginners; (2) AI-assisted audio production, where the researcher's recorded singing is converted into a kid-friendly voice with automatically produced background music; and (3) basic AI-generated animations that, in accordance with the principles of multimedia learning, synchronize the audio with the alphabet letters and accompanying images. Classroom

implementation and learning result measurement are outside the purview of this initial study because the main goal is to create and improve a prototype; instead, the evaluation focuses on expert opinion to determine the product's content and medium validity (Firdaus & Firdaus, 2024; Williyani et al., 2024).

Taking into consideration this emphasis, the study addresses two fundamental questions: (1) how can a prototype for an artificial intelligence-assisted animated English alphabet song be systematically designed and developed for Indonesian primary school English as a foreign language (EFL) learners by utilizing the Analyze, Design, and Develop stages of the ADDIE instructional design model? and (2) to what extent do experts consider the prototype to be valid in terms of content accuracy, language quality, and media design, ensuring that it meets the specific needs of beginner EFL learners at the primary school level (Ummah, 2019; Khadafi, 2021). This study seeks to provide a tangible, reproducible example of how generative artificial intelligence (AI) and multimedia design concepts can be coupled to create localized English learning materials for young learners in Indonesia by documenting the creation process and reporting the outcomes of expert-based validation (Nafi'a et al., 2022).

2. METHODS

a. Research Design

This study used a Design and Development (D&D) research methodology, which is often used to generate instructional materials and educational media (Branch, 2009). The goal was to create an AI-assisted animated English alphabet song prototype for beginning EFL learners in Grades 4-6. The study used the ADDIE model's Analyze, Design, and Develop stages, with the Implement and Evaluate stages saved for

Vol 9, No 1 (2026): ESTEEM

future research (Buhaerah, 2024; National Reading Panel, 2022).

b. Context and Target Learners

The study was carried out in five public primary schools in Majene Regency, West Sulawesi, Indonesia, where English is taught from Grades 4 to 6 using the Merdeka Belajar curriculum (Sulistiyo et al., 2020; Widagsa & Khusnia, 2023). The target learners were novice EFL pupils aged 9 to 12 years old who had little prior exposure to English. This study looked at characteristics such as low English vocabulary, limited exposure to English sounds outside of the classroom, and short attention spans, all of which are typical among primary school students (De Wilde et al., 2020).

These characteristics were taken into account to inform the design of instructional media that would be engaging, age-appropriate, and aligned with the cognitive and linguistic needs of the learners (Olowookere et al., 2020).

c. Development Model: ADDIE

The AI-assisted animated English alphabet song prototype was developed using the ADDIE instructional design methodology (Branch, 2009). The focus of this study was on prototype creation and expert validation, hence only the Analyze, Design, and Develop stages were implemented, while the Implement and Evaluate stages were skipped. During the Analyze stage, the researcher evaluated curricular requirements, studied learner characteristics, observed and interviewed teachers from five public primary schools in Majene Regency, and researched current English alphabet instructional resources (Sidik & Fahmi, 2021).

The Design stage involved planning learning objectives, lyrics, melody, tempo, and storyboard animation, applying multimedia learning principles to optimize cognitive processing and learner engagement

(Mayer, 2021, 2022). The Develop stage included producing the AI-assisted audio and animation prototype, followed by expert evaluation to assess content validity and overall quality.

d. Development Procedures

The Analyze stage included three major components: curriculum review, learner needs analysis, and resource review. The curriculum review entailed reviewing official Indonesian primary school English curriculum papers to determine target competencies for Grades 4-6, with a focus on alphabet knowledge and early phonetic abilities (Sulistiyo et al., 2020; Widagsa & Khusnia, 2023). Classroom observations and interviews with instructors were used to analyze learner requirements at five public primary schools in Majene Regency: SDN 26 Pakkola, SDN 58 Pangale, SDN 57 Tangnga-Tangnga, SDN 2 Kampung Baru Majene, and SDN 6 Kampung Baru. This procedure systematically collected information on students' English proficiency, alphabet familiarity, learning difficulties, and teachers' perceptions of instructional needs, providing a qualitative foundation for designing the prototype rather than generating reportable outcome data.

The resource review examined existing instructional media, including English alphabet songs, videos, and other audiovisual materials used in the schools. This review aimed to identify gaps in tempo, pronunciation, and cultural relevance in order to inform the development of the AI-assisted animated prototype (Rahmi & Erlinda, 2014).

In the Design stage, learning objectives were formulated to enable students to recognize letters, associate letters with sounds, and memorize alphabet sequences. Song lyrics and melody were segmented to support repetitive practice, taking into account primary learners' cognitive

Vol 9, No 1 (2026): ESTEEM

processing abilities and attention span (Royani, 2023). Storyboard animation was developed to visually represent letters, incorporate character gestures, and synchronize with the audio track. Multimedia learning principles, including coherence, signaling, contiguity, and segmenting, were applied to ensure effective alignment of audio and visual components (Mayer, 2021, 2022).

The Develop step entailed recording the audio, developing the animations, and incorporating them into a final prototype. The audio production technique involved recording the song with the researcher's own voice in a controlled environment and processing it with an AI-based platform to produce a child-friendly female vocal with musical accompaniment. AI-generated animations were used to depict letters and blend them with the audio track, using multimedia learning principles. Finally, the audio and animation components were blended into a single video file, which was then published to YouTube for documentation and dissemination only (Yousefpoori-Naeim, 2023).

e. Expert Validators

Three experts examined the prototype with the same standardized evaluation form. The first expert specialized in primary EFL education and had previously taught Grades 4-6 in Indonesian public schools. The second specialist specialized in multimedia and music, with prior experience creating instructive audiovisual content for youngsters. The third expert specialized in early childhood and primary education, having knowledge of child development and curriculum design. Using the same assessment form for all three experts guaranteed that the prototype was assessed consistently, comparably, and reliably (Ada et al., 2021).

f. Instrument: Expert Evaluation Form

The expert review form had four major components: content validity, language quality, media design, and pedagogical fit. Content validity especially addressed the accuracy and correctness of the letters provided in the song, the precision of pronunciation, and the material's instructional value for beginning EFL learners (Al-Hajaya & Al-Khresheh, 2012). Language quality assessed the prototype's language clarity, appropriateness, and comprehensibility, ensuring that it corresponded to the target learners' cognitive level and linguistic background (Buyukkal, 2025). Media design evaluated the prototype's visual effectiveness, animation timing, audio synchronization, and overall ability to engage and maintain learners' attention, whereas pedagogical suitability assessed how well the prototype aligned with students' cognitive abilities, attention span, and age-appropriate learning needs (Ummah, 2019).

All items were rated on a five-point Likert scale, ranging from 1 (very inappropriate) to 5 (very appropriate), allowing the experts to provide nuanced judgments on each aspect of the prototype. Content validity was calculated using Aiken's V, with scores of 0.75 or higher considered valid, ensuring that each component of the prototype met established standards for educational media design (Aiken, 1980; Polit, Beck, & Owen, 2007).

Table 1. Indicators of Expert Evaluation Form

Component	Indicator Description	Rating Scale
Content Validity	Accuracy of letters, correctness of pronunciation, and instructional relevance	1-5
Language Quality	Clarity, appropriateness of language for beginner EFL learners	1-5
Media Design	Effectiveness of visual presentation, animation timing, and engagement	1-5

potential		
Pedagogical Suitability	Alignment with learners' cognitive level, attention span, and age-appropriate design	1–5

Note: The same form was used by all three experts independently to ensure uniformity in scoring and reliability of content validity analysis.

g. Data Collection and Analysis

The previously disclosed expert evaluation form was used to collect expert judgments. Each expert separately completed the same structured assessment form, evaluating all items on a Likert scale and providing written comments and suggestions for improvement. The quantitative analysis concentrated on computing Aiken's V for each item to determine content validity. Following established norms, items with Aiken's V values of 0.75 or higher were considered to have adequate content validity (Aiken, 1980; Polit, Beck, & Owen, 2007). Items with values less than 0.75 were deemed insufficiently valid and scheduled for modification.

Qualitative analysis involved reviewing the experts' open-ended comments to identify recurring themes and specific recommendations related to content, language, media design, and pedagogical suitability. These qualitative insights were used to refine the prototype and, where necessary, to revise items on the evaluation form before finalizing the product.

h. Scope and Limitations

This study centered solely on the creation and expert validation of an AI-assisted animated English alphabet song prototype. Only the Analyze, Design, and Develop stages of the ADDIE paradigm were implemented; the Implement and Evaluate stages, which entail experiments with learners, were purposefully left out of this phase. As a result, the prototype's usefulness

at improving alphabet recognition and pronunciation among primary school students has yet to be empirically verified.

The findings therefore reflect expert-based judgments rather than classroom performance data. Future research should incorporate the Implement and Evaluate stages by conducting classroom trials, gathering learner outcome data, and obtaining teacher and student feedback to assess the instructional impact, usability, and scalability of the prototype in diverse educational contexts.

3. RESULTS

a. Analyze

The Analyze stage offered a comprehensive picture of English learning conditions and learner characteristics in the five public primary schools in Majene Regency, West Sulawesi. According to the Merdeka Belajar curriculum, English instruction in Grades 4-6 emphasizes hearing, speaking, and letter recognition as a basis for early foreign language literacy (Sulistiyo et al., 2020; Widagsa & Khusnia, 2023). Observations and interviews showed that English was typically taught only 1–2 hours per week, and learners' exposure to the alphabet was largely limited to textbooks with minimal oral modeling.

The requirements analysis found several reoccurring issues with letter pronunciation. Because of inconsistencies between the English and Indonesian vowel systems, learners frequently mispronounced vowel letters such as A, I, and E. Letters O and U were also difficult to speak, and semi-consonant letters like W, Y, and H were frequently mispronounced or articulated using Indonesian phonetic patterns. In addition, pupils struggled to discern between letters with similar sounds, particularly F, V, and P. These data revealed that alphabet training must clearly address letters whose sound is notably different from Indonesian.

Vol 9, No 1 (2026): ESTEEM

The analysis also highlighted limitations in existing instructional media. Most alphabet songs used in the schools were global products with fast tempo, non-Indonesian accents, and limited alignment with the national curriculum. Visual aids or animations that clearly linked written letters to their sounds were scarce. Several schools lacked functioning projectors or audio equipment, making it difficult to deliver engaging audiovisual lessons. Students tended to become passive when presented with static materials, and teachers expressed a strong need for song- and animation-based media that emphasized correct pronunciation, sustained attention, and reflected the local context of Majene Regency. Overall, the Analyze stage underscored the importance of interactive audio-visual media to support accurate phonetic development for specific letters (A, I, E, O, U, W, H, Y, F, V, P) and to make alphabet learning more engaging and accessible.

b. Design

The findings from the Analyze step directly influenced the prototype's Design stage. The learning objectives were designed to help students recognize alphabet letters, associate each letter with its associated sound, and memorize the alphabet's sequence. These objectives were developed in accordance with the Merdeka Belajar competences for Grades 4-6 and suited to the cognitive characteristics of beginning EFL learners, such as low working memory, developing phonological awareness, and short attention span.

The English alphabet song was created with segmented lyrics and music to allow for repeated repetition without overwhelming students. Special emphasis was placed on letters that commonly present pronunciation challenges for Indonesian learners, particularly A, I, E, O, U, W, H, Y, F, V, and P. The tempo and rhythmic pattern were selected to be slow and clear enough for

beginners to follow while remaining sufficiently lively to maintain motivation.

Storyboard animations were created to accompany each lyric segment. Visuals included the hybrid main character and supporting elements that represented letters and sounds in a concrete and appealing way. Character gestures and letter animations were synchronized with the audio track to reinforce the mapping between what learners heard and what they saw. Throughout the design process, multimedia learning principles, coherence (removing irrelevant details), signaling (highlighting target letters), contiguity (presenting spoken and written forms together), and segmenting (breaking content into manageable chunks), were applied to support efficient cognitive processing and sustained engagement (Mayer, 2021, 2022). The output of the Design stage was a detailed plan for the song structure, visuals, and timing that guided subsequent development.

c. Develop and Expert Validation

During the Develop stage, a prototype was created based on design criteria. The audio production process began with the researcher recording the song in a controlled atmosphere with his own voice. The recording was then processed using Suno, an AI-powered platform, to produce a child-friendly female vocal in a "kids song" style with a cheery tone and a moderate tempo appropriate for primary students. The software also generated musical accompaniment, and multiple iterations were carried out to provide a crisp and natural-sounding audio track.

For the visual component, the main animated character was created using Gemini by combining features of a cat, bat, and rabbit into a single hybrid figure designed to be visually attractive and memorable for children. Images for each segment of the song were produced to illustrate the letters,

Vol 9, No 1 (2026): ESTEEM

the character, and relevant visual cues in line with the storyboard. These images were converted into animated clips using VEO3, which provided lipsync functionality so that the character's mouth movements matched the AI-generated vocals. Final video editing and synchronization of audio, character movement, and letter animations were completed in CapCut, resulting in a unified AI-assisted animated English alphabet song prototype. The video was uploaded to YouTube for documentation and dissemination purposes only, not for data collection.

After completion, the prototype was validated by experts. Three experts one in primary EFL education, one in multimedia and educational music, and one in early childhood and primary education evaluated the product using a structured evaluation form that included four components: content validity, language quality, media design, and pedagogical suitability. Each item was scored on a five-point Likert scale, and Aiken's V was calculated using the expert ratings.

Table 2 presents the experts' ratings and the resulting Aiken's V indices for the four items. As shown in the table, Aiken's V values ranged from 0.833 to 1.000, all exceeding the 0.75 cut-off for acceptable content validity. These results indicate that the prototype demonstrates strong validity in terms of letter accuracy, pronunciation modeling, language clarity, visual and animation quality, and alignment with learners' cognitive characteristics. The validated prototype is therefore ready for classroom implementation and learner-based evaluation in subsequent research phases.

Table 2. Expert Validation Results Based on Aiken's V

Ite m	Valid ator 1	Valid ator 2	Valid ator 3	Su m	n(c -1)	Aike n's V
1	4	4	5	10	12	0.83

						3
2	4	5	4	10	12	0.83
						3
3	5	5	4	11	12	0.91
						7
4	5	5	5	12	12	1.00
						0

Note. Aiken's V values ≥ 0.75 indicate acceptable content validity (Aiken, 1980; Polit, Beck, & Owen, 2007).

4. DISCUSSION

a. Development of the Prototype

The first research question focused on how an AI-assisted animated English alphabet song prototype could be systematically built and developed for novice EFL learners in Indonesian primary schools utilizing the ADDIE model's Analyze, Design, and Develop stages. The findings show that the ADDIE-based strategy provided a structured and iterative framework for the researcher to integrate learning objectives, song content, and visual animations with the cognitive and linguistic demands of students in Grades 4-6.

The Analyze stage highlighted the inadequate exposure to English, the phonetic issues for letters such as A, I, E, O, U, W, H, Y, F, V, and P, and the need for compelling, age-appropriate audiovisual content. The Design stage applied multimedia learning principles, coherence, signaling, contiguity, and segmenting, to optimize information processing and maintain attention. Finally, the Develop stage successfully produced a prototype integrating AI-processed child-friendly vocals, a hybrid animated cat-bat-rabbit character, and synchronized on-screen letters.

This approach extends prior research on early EFL literacy by combining AI-assisted audio processing with animated media, whereas previous studies largely focused on conventional alphabet songs,

Vol 9, No 1 (2026): ESTEEM

teacher-led instruction, or non-AI digital resources (Sari, 2024; Booton et al., 2021; Trisyani, 2022). The hybrid character animation provides a novel strategy to support letter-sound associations while maintaining engagement, which aligns with findings from multimedia learning theory (Mayer, 2021, 2022) and research on music-based language acquisition for young learners (Saldiraner & Cinkara, 2021).

b. Validation of the Prototype

The second research question looked at how expert evaluators perceived the prototype's validity. Three specialists in primary EFL teaching, multimedia and music, and early childhood education independently evaluated the prototype on four criteria: content validity, language quality, media design, and pedagogical fit. The resulting Aiken's V scores ranged between 0.833 and 1.000, suggesting that all questions met or above the recognized validity criterion. These results show that the prototype meets curriculum-aligned teaching criteria, accurately represents target letters, and follows age-appropriate pedagogical concepts.

The novelty of this validation lies in the targeted focus on letters that typically present pronunciation challenges for Indonesian beginner learners and the integration of AI-assisted audio with interactive animated visuals. While prior studies have validated alphabet learning media for children (Sari, 2024; Trisyani, 2022), none combined generative AI voice processing with a curriculum-aligned, multimedia learning approach tailored for primary EFL contexts. This provides empirical evidence for the effectiveness of low-cost AI tools in producing high-quality instructional media that are culturally and pedagogically adapted.

c. Implications for Practice

The verified prototype has the potential to be an innovative instructional resource for

Grades 4-6 in Indonesian primary schools. Its uses include early literacy and phonics training, which provides consistent, precise exposure to English letters and pronunciation. Furthermore, the prototype can serve as a model for teacher education programs, instructional design courses, and educational technology workshops, demonstrating how AI-assisted media can improve learning results. By providing a reproducible framework, this study helps to produce EFL learning materials that are sustainable, locally relevant, and technologically upgraded.

5. CONCLUSION

This project successfully built an AI-assisted animated English alphabet song prototype for beginner EFL learners in Indonesian primary schools, specifically Grades 4-6, using the ADDIE instructional design model's Analyze, Design, and Develop stages. The prototype had AI-processed child-friendly vocals, a hybrid animated cat-bat-rabbit mascot, and synced on-screen letters to address phonetic issues typically encountered by Indonesian learners. Expert evaluation of content validity, language quality, media design, and pedagogical fit yielded Aiken's V ratings ranging from 0.833 to 1.000, demonstrating the prototype's compliance with curriculum standards and instructional quality requirements.

The results indicate that a systematic, design-based approach can successfully guide the creation of interactive and pedagogically appropriate multimedia for young language learners, ensuring both cognitive and affective needs are met. The prototype demonstrates how carefully designed multimedia can reinforce letter recognition, phonetic accuracy, and learner engagement, addressing common challenges in early EFL instruction.

This study introduces an innovative method to early EFL instruction by merging

Vol 9, No 1 (2026): ESTEEM

generative AI voice processing with interactive multimedia design, resulting in an experimentally validated, curriculum-aligned framework for creating localized educational resources. Low-cost AI techniques were successfully used to create high-quality, engaging, and pedagogically sound learning materials. The study emphasizes the potential for AI-assisted animation to promote individualized training, increase student motivation, improve attention and memory retention, and boost confidence in English pronunciation. The incorporation of multimedia learning concepts demonstrates the practical application of cognitive theory to educational materials, bridging the gap between research and classroom practice. The prototype also provides direction for integrating technology developments with curriculum objectives, ensuring that instructional media remains relevant, culturally appropriate, and developmentally appropriate for primary students.

Despite its achievements, the study is limited by the absence of classroom implementation and the relatively small number of expert validators. Future research should examine the prototype's effectiveness in actual classroom settings, measuring alphabet recognition, pronunciation accuracy, and learner engagement through mixed-method trials. Additional research may look into adaptation for additional linguistic contexts, the use of gamification, learner-controlled pacing, and scalability to diverse educational environments in order to establish generalizability and sustainability. These steps will provide evidence-based recommendations for teachers, curriculum creators, and educational technologists who want to include AI-powered resources into early language learning programs. This study also provides a reproducible paradigm for building AI-enhanced instructional media, illustrating how new technology can be used to create interactive, learner-centered,

curriculum-aligned resources that promote early EFL literacy.

6. ACKNOWLEDGEMENT

The researcher would like to acknowledge their own efforts in carrying out this study, such as designing, developing, and validating the AI-assisted animated English alphabet song prototype. This study was conducted out with no external funding or institutional sponsorship. The researcher completed all stages, including curriculum analysis, learner needs assessment, multimedia production, and expert review. The author also thanks the specialists who gave their time and professional judgment to validate the prototype, ensuring its instructional, linguistic, and pedagogical integrity.

7. REFERENCES

- Ada, K., Tapan Broutin, M. S., Kaleli Yılmaz, G., & Bayram, G. M. (2021). Investigation of documentation processes of students with low and high level of mathematical literacy: A case study. *Journal of Pedagogical Research*, 5(4), 189–213. <https://doi.org/10.33902/JPR.2021474615>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- Akram, H., Abdelrady, A. H., Al-Adwan, A. S., & Ramzan, M. (2022). Teachers' perceptions of technology integration in teaching-learning practices: A systematic review. *Frontiers in Psychology*, 13(June), 1–9. <https://doi.org/10.3389/fpsyg.2022.920317>
- Alfarwan, A. (2025). Generative AI use in K-12 education: A systematic review.

- Frontiers in Education*, 10, 1647573. <https://doi.org/10.3389/feduc.2025.1647573>
- Al-Hajaya, N., & Al-Khresheh, T. (2012). The effect of cognitive learning style-based reading program on the achievement of Jordanian freshmen english majors. *International Education Studies*, 5(3), 235–246. <https://doi.org/10.5539/ies.v5n3p235>
- Baleghizadeh, S., & Maryam, B. (2019). The Effect of Summary Writing on Reading Comprehension: the Role of Mediation in Efl Classroom. *New England Reading Association Journal*, 47(1), 44–48. <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=88399940&site=ehost-live>
- Bhwana, P. G. (2023). *Ministry claims Merdeka curriculum improves quality of education in Indonesia*. <https://en.tempo.co/read/1685279/ministry-claims-merdeka-curriculum-improves-quality-of-education-in-indonesia>
- Boardman, A. G., Polman, J. L., Scornavacco, K., Potvin, A. S., Garcia, A., Dalton, B., Stamatis, K., Guggenheim, A., & Alzen, J. L. (2024). Examining enactments of project-based learning in secondary english language arts. *AERA Open*, 10(1), 1–17. <https://doi.org/10.1177/23328584241269829>
- Booton, S. A., Hodgkiss, A., & Murphy, V. A. (2021). The impact of mobile application features on children's language and literacy learning: a systematic review. *Computer Assisted Language Learning*, 36(3), 400–429. <https://doi.org/10.1080/09588221.2021.1930057>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Buhaerah, N. (2024). Implikasi peran pendidikan dan pelatihan. *Bongaya Journal of Research in Management*, 7, 85–94. <https://doi.org/https://doi.org/10.37888/bjrm.v7i1.583>
- Buyukkal, B. (2025). *Politeness Strategies*. Büyükkal, Burcu. “Politeness Strategies.” Socio-Cultural Concepts of (Im) Politeness: Interpretation, Linguistics, and Discourse Analysis: Interpretation, Linguistics, and Discourse Analysis.
- Dessler, G. (2020). *Human resource management* (Fifth Edit). Pearson Education.
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism*, 23(1), 171–185. <https://doi.org/10.1017/S1366728918001062>
- Ditania, S. (2022). *An anlysys on students' problem in speaking skill faced by the eleventh grade of social science of sma negeri 3 pekanbaru in academic Year 2019/2020* [Universitas Islam Riau]. <https://repository.uir.ac.id/11477/>
- Firdaus, R., & Firdaus, R. (2024). Implementation of the ADDIE model for developing multimedia learning media and gamification on space objects material on elementary school. *JUTIF: Jurnal Teknik Informatika*, 5(4), 365–373. <https://doi.org/10.52436/1.jutif.2024.5.4.2307>
- Khadafi, B. I. (2021). Feminist and pessimist existentialism in kate chopin's “the story of an hour”: a systemic functional grammar analysis. *Journal of English Language Studies*, 6(2), 138.

- Laksana, S. D., Setyosari, P., Praherdhiono, H., Kuswandi, D., Priatna, Y., Abidin, N., ... Hile, A. (2025). Enhancing reading literacy through augmented reality among Indonesian migrant children in Malaysia. *Journal of Integrated Elementary Education*, 5(2), 469–487. <https://doi.org/10.21580/jieed.v5i2.26498>
- Littleton, K., & Mercer, N. (2013). Educational dialogues. *International Handbook of Research on Children's Literacy, Learning, and Culture*, 291–303.
- Malini, N. L. N. S., Sukarini, N. W., Yadnya, I. B. P., & Maharani, S. A. I. (2022). Exploring needs analysis of English language training: An evidence from small hotel and restaurant employees in Nusa Lembongan, Bali, Indonesia. *Indonesian Journal of Applied Linguistics*, 12(1), 212–223. <https://doi.org/10.17509/ijal.v12i1.46437>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Nafi'a, M. Z. I., Kuswandi, D., & Wedi, A. (2022). Pengembangan desain pembelajaran multiliterasi berbasis tringo dalam pembelajaran menulis teks eksposisi siswa SMA. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 5(4), 349. <https://doi.org/10.17977/um038v5i42022p349>
- National Reading Panel. (2022). *Report of the National Reading Panel: Teaching Children to Read*.
- Olowookere, E. I., Omonijo, D. O., Odukoya, J. A., & Anyaegbunam, M. C. (2020). Exploring the effect of gender and personality characteristics on educational performance. *Journal of Educational and Social Research*, 10(5), 99–107. <https://doi.org/10.36941/JESR-2020-0091>
- Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, 45(1), 8–38. <https://doi.org/10.1598/RRQ.45.1.2>
- Rahmi, M. M., & Erlinda, R. (2014). Teachers' role in building students' character through English language teaching classroom. *Proceeding of the Third International Seminar on Languages and Arts*, 20, 320–327.
- Ridwan, M. (2019). GALLERY WALK; An alternative learning strategy in increasing students' active learning. *Nady Al-Adab: Jurnal Bahasa Arab*, 16(1), 49. <https://doi.org/10.20956/jna.v16i1.6662>
- Royani, I. (2023). Pronunciation learning strategies used by EFL learners in university context. *Proceeding of Conference on English Language Teaching*, 3, 421–432.
- Saha, G. C., Akber, S. M., & Roy, A. (2023). Impact of outcome-based education (obe) on learners' performance in business courses. *International Journal of Professional Business Review*, 8(8), e02394. <https://doi.org/10.26668/businessreview/2023.v8i8.2394>
- Saldiraner, G., & Cinkara, E. (2021). Using songs in teaching pronunciation to young EFL learners. *PASAA*, 62(1), 119–141. <https://doi.org/10.58837/chula.pasaa.62.1.5>

Vol 9, No 1 (2026): ESTEEM

- Sari, L. J. (2024). Application-based assembled edu media to stimulate alphabet recognition ability for early childhood children. *Jurnal Pendidikan Anak Usia Dini Undiksha*, 12(1), 44–53.
<https://doi.org/10.23887/paud.v12i1.75456>
- Sidik, A. S., & Fahmi, F. (2021). Pengembangan media video pembelajaran speaking berbasis Bahasa Ibu (Bahasa Sasak) di SMK Qamamrul Huda. *Jurnal Ilmiah Mandala Education*, 7(4), 2019–2022.
<https://doi.org/10.36312/jime.v7i4.2401>
- Sulistiyono, U., Mukminin, A., Cahyono, B. Y., & Haryanto, E. (2020). The portrait of primary school English in Indonesia: Policy recommendations. *Education 3–13*, 47(6), 751–761.
<https://doi.org/10.1080/03004279.2019.1680721>
- Sunde, K., Furnes, B., & Lundetrae, K. (2020). Does Introducing the Letters Faster Boost the Development of Children's Letter Knowledge, Word Reading and Spelling in the First Year of School? *Scientific Studies of Reading*, 24(2), 141–158.
<https://doi.org/https://doi.org/10.1080/10888438.2019.1615491>
- Trisyani, R. A. (2022). The use of alphabet box media in letter recognition in children aged 4–5 years. *ECEDS (Early Childhood Education Development Studies)*, 3(2).
<https://doi.org/10.35508/eceds.v3i2.12382>
- Ummah, M. S. (2019). Panduan Asistensi Mengajar. *Institut Al Fithrah Surabaya*, 11(1), 1–14.
http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_sistem_pembetulan_terpusat_strategi_melestari
- Widagsa, R., & Khusnia, A. N. (2023). Should English be a primary school compulsory subject in Indonesia's latest "Merdeka Belajar" curriculum? *Journal of English Language and Language Teaching*, 7(1), 1–19.
<https://doi.org/10.36597/jell.v7i1.14408>
- Williyan, A., Fitriati, S. W., Pratama, H., & Sakhiyya, Z. (2024). AI as co-creator: exploring Indonesian EFL teachers' collaboration with AI in content development. *Teaching English with Technology*, 24(2), 5–21.
<https://doi.org/10.56297/vaca6841/LRD X3699/RZOH5366>
- Yeni, M., & Amelia, R. (2020). Teaching alphabet for young learners through song. *Journal of English Language Education*, 5(2), 12–22. Retrieved from
<https://www.jele.or.id/index.php/jele/article/view/69>
- Yousefpoori-Naeim, M. (2023). A Review of Duolingo Stories. *International Journal on Innovations in Online Education*, 7(2), 1–7.
<https://doi.org/10.1615/intjinnovonline>
- Yusuf, A., Pervin, N., Román-González, M., & Md Noor, N. (2024). Generative AI in education and research: A systematic mapping review. *Review of Education*, 12(2), e3489.
<https://doi.org/10.1002/rev3.3489>
- Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers of Digital Education*, 1(3), 223–245.
<https://doi.org/10.1007/s44366-024-0028-5>

Vol 9, No 1 (2026): ESTEEM

Appendix A. Full Lyrics and Prototype Video

Full Lyrics of the AI-Assisted Animated English Alphabet Song

Created by Ahyadi

Verse 1

A B C D E F G
H I J K L M N
O P Q R S T U and V
W X Y and Z

Chorus

Let us sing A B C
Let us read and see
We are happy, you and me
Learning letters A to Z

Verse 2

A B C D E F G
H I J K L M N
O P Q R S T U and V
W X Y and Z

Chorus

Let us sing A B C
Let us read and see
We are happy, you and me
Learning letters A to Z

Prototype Video

The AI-assisted animated English alphabet song prototype video, demonstrating audiovisual synchronization, character animation, and lip-sync, can be accessed at the following link: <https://youtu.be/fIPa2fxujrw?si=4BzYdDHmsCIJSC1V>

Video Details:

- Duration: 1 minutes 25 seconds
- Format: MP4
- Purpose: Demonstrates the final output produced according to the ADDIE-based development procedures described in the Methods section.

Note for Reviewers:

This video is provided exclusively for reviewer access to evaluate the audio quality, animation design, and overall usability of the prototype. It is for review purposes only and should not be reproduced or distributed without permission.