

International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 2, pp. 827-846, February 2026
<https://doi.org/10.26803/ijlter.25.2.37>
Received Nov 28, 2025; Revised Jan 19, 2026; Accepted Feb 20, 2026

Automatic Speech Recognition Enhanced Self-Instructional Learning Material to Improve English Pronunciation for Indonesian Primary School Students

Muh Shofiyuddin*^{ID}, Erna Zumrotun^{ID}, Noor Azizah^{ID}, Muh Muhaimin^{ID}
Universitas Islam Nahdlatul Ulama Jepara
Jepara, Indonesia

Yudie Irawan^{ID}
Universitas Muria Kudus
Kudus, Indonesia

Abstract. English pronunciation of young learners is often limited due to insufficient exposure to the target language, and this affects their self-confidence and communication skills. This study aimed to analyze teacher needs in teaching pronunciation and to develop interactive, systematic, and child-specific self-instructional material based on Automatic Speech Recognition (ASR) to improve pronunciation mastery. The research applied Research and Development with the ADDIE model consisting of analysis, design, development, implementation, and evaluation. However, this study was limited to the analysis, design, and prototype development stages without field testing. Validation was conducted by two English language education experts through a 1–4 scale questionnaire and open-ended comments, assessing aspects of material content, language suitability, visual design, ASR function, illustrations and audio recordings, and repeated practice. The validation results showed that the material was of good to excellent quality, with ASR integration enabling repeated independent practice, real-time feedback, and increased motivation and learning independence. These findings confirm that the material effectively supports pronunciation systematically, although the study was limited to a prototype and a small number of validators. The contribution of this research lies in providing valid, practical, and adaptive ASR-based interactive learning media, offering a solution to the limitations of teachers and media in supporting pronunciation practice in elementary schools. The research can help both teachers and students in the learning process and self-learning with the ASR feature.

*Corresponding author: Muh Shofiyuddin; muh_shofiyuddin@unisnu.ac.id

Keywords: self-instructional material; Automatic Speech Recognition; English Pronunciation; Young Learner; Primary School

1. Introduction

Designing self-instructional materials based on Automatic Speech Recognition (ASR) to improve English pronunciation in elementary school students is crucial due to learners' limited exposure to the target language in their environment. Many students struggle to pronounce certain sounds, affecting their self-confidence and communication skills. Recent research confirms that ASR technology can provide accurate, automated feedback, enabling students to correct their pronunciation errors independently (Ahn et al., 2025; Inceoglu et al., 2024; Liu et al., 2025).

This urgency emphasizes the need to develop self-directed teaching materials that are relevant to the characteristics of elementary school-aged children so that pronunciation practice can be effective even without intensive teacher guidance. This condition aligns with the principles of technology-supported self-directed learning, which emphasizes learner autonomy and continuous practice as essential responses to limited instructional exposure, particularly in foreign-language pronunciation learning at the elementary level.

In response to these challenges and the gaps in existing research, the present study aimed to analyze teachers' needs in teaching pronunciation and to develop interactive, systematic, and child-specific ASR-based self-instructional material. The integration of ASR technology in pronunciation learning has yielded positive results at various levels of education (Guskaroska, 2024). Research trends highlight the need to design materials that are appropriate to children's age and cognitive abilities so that technology use is not merely mechanical but also fosters learning engagement (Y. Liu et al., 2025; Sariyani et al., 2024; Sun, 2023).

Furthermore, ASR-based systems have been shown to encourage student independence, allowing them to adjust their learning pace to suit their individual needs. However, despite the growing body of research on ASR-assisted pronunciation learning, most studies focus on secondary or higher education contexts and emphasize application-based systems rather than structured self-instructional text materials that are tailored to the cognitive and developmental characteristics of elementary school learners. Consequently, empirical evidence on ASR-based self-directed pronunciation materials specifically designed for young learners remains limited. This suggests that designing ASR-based self-instructional text materials can bridge the gap between the demands of intensive practice and the limited access to native speakers.

The novelty of this study lies in the development of an ASR-based self-instructional textbook that was specifically designed for elementary school learners by integrating child-centered design principles. Unlike most existing ASR tools that target older learners, this product features age-appropriate language, step-by-step learning sequences, engaging illustrations, child-friendly audio

models, and repetitive practice that is aligned with children's cognitive characteristics, thereby offering a pedagogically sensitive approach to supporting independent pronunciation learning.

To support this development, the study is grounded in the theoretical framework of technology-based self-directed learning. The aim of this research was to analyze teachers' needs in teaching pronunciation and to develop interactive, systematic, and child-specific ASR-based self-instructional text materials to improve the pronunciation of elementary school students. Research confirms that this technology can accelerate the pronunciation-correction process and increase learning motivation through real-time feedback (Bashori et al., 2024; Spring & Tabuchi, 2021; Wang & Young, 2015).

Through this approach, students not only gain a means of independent practice but also an adaptive learning experience. Thus, the development of ASR-based self-instructional material is an innovative strategy to address the challenges of pronunciation learning in elementary schools. Accordingly, this paper first presents an analysis of teachers' needs relating to pronunciation instruction, followed by the design and development of ASR-based self-instructional text materials and concluding with validation results and pedagogical implications for elementary English learning

The theoretical basis for developing ASR-based self-instructional material rests on the concept of technology-based self-directed learning, which emphasizes student independence in mastering pronunciation. Recent studies have shown that independent practice supported by ASR strengthens phonological awareness and fosters positive attitudes toward foreign language learning (McCrocklin, 2016; Won, 2025; ZIANE & DOUB, 2024). Furthermore, the autocorrect feature in ASR systems allows students to adjust their learning pace to suit their own pace. Based on this foundation, ASR-based self-instructional text materials are positioned as pedagogical tools capable of combining the principles of learning autonomy, pronunciation accuracy, and increased motivation.

The design of ASR-based self-instructional text materials offers a real opportunity to enhance English pronunciation learning in elementary schools. Various studies have shown that the use of this technology not only helps improve pronunciation accuracy but also encourages students' independence in pronunciation practice (Evers & Chen, 2021, 2022; Xiao et al., 2023). With its age-appropriate, interactive design and automatic feedback, this material complements the limited literature while enabling improvement in pronunciation competency and a more enjoyable and sustainable learning experience.

2. Methodology

This research focused on the development of an ASR-based self-instructional material to improve English pronunciation in elementary school students. The application was developed using Android Studio software with Java programming language. The primary focus was on the design and development of an academically valid prototype without proceeding to the classroom

implementation stage. This strategy was chosen to allow for a comprehensive evaluation of the product's quality across content, language, presentation, and technology integration before widespread testing. This approach aligns with current research trends that emphasize the importance of academic validation in the early stages of learning-media development (Chang & Fisher, 2003; Richey & Klein, 2014). The research concluded at the product design phase.

The research design used a Research and Development method with the ADDIE model framework, which consists of the stages of analysis, design, development, implementation, and evaluation. This research was limited to the needs analysis, design, and development stages, thus focusing on achieving a pedagogically valid prototype. The procedure stages included analyzing elementary school students' needs regarding English pronunciation, designing an ASR-based self-instructional material, developing a prototype of the material and supporting application, and validating the prototype by two experts using questionnaires and open-ended comments. Recent literature confirms that the ADDIE model is adaptive so it can be adjusted to the research context and needs (T. Liu et al., 2023; Moore, 2025). However, this research came to an end after the expert validation.

The research data sources consisted of two expert validators who are lecturers in English language education specializing in pronunciation and teaching methodology. The selection of validators was carried out to ensure that the ASR-based self-instructional material prototype was academically legitimate, pedagogically valid, and relevant to the characteristics of elementary school students. The involvement of experts played a crucial role in assessing the product's suitability in terms of content, language, presentation, and graphical aspects. Expert validation is believed to provide a strong foundation for further revision and development (Plomp, 2013; Xu et al., 2025).

With this procedure, the product design is considered valid and worthy. The use of two expert validators is considered sufficient at the prototype validation stage, as expert judgment in development research prioritizes depth of expertise over quantity. This practice is consistent with instructional design studies that emphasize intensive qualitative and quantitative evaluation by subject-matter experts in early development phases (Plomp, 2013).

2.1 Data Collection and Analysis

The research data were collected through expert judgment using a structured measurement sheet for validation. The instrument consisted of a Likert-based validation questionnaire administered to both validators. The questionnaire assessed the product's suitability in terms of pronunciation content, language appropriateness for elementary school students, visual design, integration of ASR features, illustrations and audio recordings, and repetitive practice. The scale ranged from 1 = Inappropriate, 2 = Less Appropriate, 3 = Appropriate, and 4 = Very Appropriate. The questionnaire also included an open-ended comment section to accommodate detailed qualitative feedback. The combination of quantitative and qualitative data enabled a comprehensive and evidence-based product evaluation (Chang & Fisher, 2003; Richey & Klein, 2014).

Table 1: Expert assessment indicators

Assessment Indicator	Brief Description
Material Systematics	The material is arranged logically and sequentially so students can follow the exercises step by step.
Language Suitability	The language used is appropriate for the developmental level and understanding of elementary school students.
Visual Design	The colors, illustrations, and layout are attractive and age appropriate.
ASR Function	The ASR feature works optimally, providing immediate feedback and supporting independent learning.
Illustrations & Audio Recordings	The illustrations and audio recordings clearly support pronunciation practice and vocabulary comprehension.
Repetitive Practice	Exercises can be repeated as needed for gradual pronunciation mastery.

ASR: Automatic Speech Recognition

The indicators listed in Table 1 were used to assess the book's suitability in relation to various key aspects. The systematic nature of the material would ensure a logical sequence of exercises and make it easy for students to follow each step. Language appropriateness would ensure that the content was appropriate for the developmental level of elementary school students, while the visual design and the illustrations supported engagement and comprehension. The ASR function and repetitive exercises would enable students to practice independently and effectively, with immediate feedback to improve pronunciation. This evaluation ensured that every aspect of the book was thoroughly assessed, providing useful input for revision and prototype development and confirming that the book was ready to be used as an engaging, interactive, and student-centered learning tool.

Data analysis was conducted using quantitative and qualitative descriptive approaches. Questionnaire scores were calculated to obtain an average and categorized according to the product's level of suitability. Open-ended comments from validators were analyzed thematically to identify aspects of the content and design that required revision. The integration of these two analytical approaches allowed for a more valid evaluation because the numerical data were reinforced by contextual meaning. This approach aligns with the views of Creswell and Plano Clark (2023) and Clark and Mayer (2023) who emphasize the importance of combining quantitative and qualitative methods in developmental research for more comprehensive results.

The overall procedures, instruments, and data analysis used in this study ensured that the development of the ASR-based self-instructional textbook had a strong academic foundation, was pedagogically valid, and was relevant to the pronunciation needs of elementary school students. It was, therefore, ready to be used as a basis for further development in the implementation phase.

3. Results and Findings

3.1 Teacher Conditions and Policy Implementation

Based on observations and documentation in six elementary schools across four districts, government support for the provision of specialized English teachers is lacking. In response to the Independent Curriculum policy, which stipulates that English be taught in elementary schools, schools assign classroom teachers to teach the subject even though the teachers are not English language education graduates. Interview and observation data show variations in material mastery and teaching strategies among teachers.

Some teachers are not accustomed to providing systematic pronunciation practice. An inventory of learning resources noted very limited textbooks, some of which were electronic and not always accessible to all students. Supplemental media such as audio recordings of native speakers, interactive tools, and additional practice materials were almost non-existent. Classroom observations showed students following written instructions in textbooks and teacher guidance. Documentation noted that teachers used simple methods such as singing English songs and playing quizzes, with limited pronunciation practice.

3.2 Teacher Working Group (Kelompok Kerja Guru) Participation

The Kelompok Kerja Guru (KKG) observations indicate uneven participation of English teachers, with only a few schools actively involved in curriculum development and teaching strategies. The KKG activity inventory recorded the frequency of teacher attendance, the types of activities in which the teachers participated, and the material discussed at each meeting. Some teachers did not attend the KKG at all, thus limiting opportunities for strategy sharing and competency development. Meeting documentation noted that attending teachers discussed simple teaching methods, media management, and pronunciation materials, but the coverage was not uniform across schools. All participation data were obtained from attendance records, minutes, and direct observation of KKG activities.

3.3 Availability of Media and Pronunciation Practice

The classroom media inventory revealed the availability of books, illustrations, audio, and simple activities to support pronunciation such as singing with English lyrics or playing quizzes. Documentation of the practice units indicated repetition of sounds, words, and sentences, some using electronic media. Observations of student activities noted that they followed teacher instructions and books, but independent practice was limited. Unit and media inventory data indicated the presence of illustrations, audio, and repetitive practice in some units, although the availability of facilities varied across schools and student access was uneven. All findings were recorded based on observations, interviews, and documentation inventory.

3.4 Development of ASR-Based Self-Instructional Books

Based on needs analyses and field findings, self-instructional books are developed with systematically structured materials. Regarding the current material, each unit includes basic sound recognition, word exercises, phrases, and complete sentences. All units are equipped with interactive illustrations and audio

recordings of native speakers. The ASR features are available in all units, allowing students to practice pronunciation independently, receive automatic feedback, and repeat exercises as needed. Development documentation demonstrates the consistent integration of audio, visual, and ASR media across each unit.

A unit inventory records all the repetitive exercises, interactive media, and audio recordings that are available at each stage of the material. The material structure is systematically structured to allow students to follow gradual progression. All units provide repetitive exercises, interactive media, and ASR features. These findings document the book features, unit structure, types of exercises, and the media used to support pronunciation practice at each stage of the material.

3.5 Material Structure and Features

Each unit has clear learning objectives, includes words and sentences appropriate to the student's developmental level, and provides audio, visual, and vocabulary recognition exercises. The ASR feature allows students to check their pronunciation, display pronunciation scores, and provide correction guidance. Interactive illustrations are included in each unit to link sounds to visual contexts. The book provides repeatable exercises that students can complete independently. The unit inventory demonstrates the integration of audio, visual, and ASR throughout the material. The unit structure is systematic, from basic sound recognition to complete sentences, with exercises that allow for repetition and comprehension of pronunciation. All units have consistent supporting media, including audio and interactive illustrations. Unit documentation indicates the ASR feature is active in each exercise, and exercises can be repeated to support continuous pronunciation practice.

3.6 Appropriateness of Material to Student Competencies

The analysis of the material showed that the book is structured in stages, from simple sounds to more complex combinations. Each unit includes repetition exercises that allow students to practice pronouncing sounds consistently. The material is structured according to the students' ability levels, ensuring that the exercises are developmentally relevant. Illustrations, audio, and interactive exercises are available in each unit to support the introduction of new vocabulary and the application of pronunciation in relevant contexts.

Observations conducted during this study indicate that students engage in repeated practice independently. The unit and exercise inventory demonstrated the systematic integration of sound, word, and sentence repetition. Audio support features and interactive exercises were available in all units. Documentation showed that each unit supported pronunciation practice, provided repetitive exercises, and consistent interactive media. These findings demonstrate the systematic organization of the book's materials and the appropriateness of students' abilities at each stage of pronunciation practice.

3.7 Validation of Book Content and Suitability

The book was validated by two English language education experts specializing in pronunciation using a Likert-scale questionnaire and open-ended comments. Validation results showed high scores for content, language appropriateness for students' ages, visual design, and integration of ASR features. Open-ended

comments noted that the material was systematically organized. The illustrations and audio supported pronunciation practice. The ASR feature enabled pronunciation checking and correction guidance. All aspects of the book were rated suitable for use as teaching materials. Validation documentation showed that ASR features were active throughout the units; illustrations and audio were available; and exercises were systematically organized. Questionnaire data demonstrated consistency in assessments between the two experts, while open-ended comments provided technical details about illustration placement, clarity of exercise instructions, and ASR functionality in each unit.

Table 2: Expert assessment results

Aspects Assessed	Validator 1	Validator 2	Average/Description
Material Systematics	Very Good	Good	The material is organized neatly and logically, making it easy for students to follow each step.
Visual Design	Good	Good	The colors, illustrations, and layout are attractive and appropriate for the age characteristics of elementary school students.
ASR Function	Very Good	Very Good	The ASR feature works optimally, providing immediate feedback that encourages independent learning.
Illustrations and Audio Recordings	Good	Good	The illustrations and audio recordings support pronunciation practice clearly and enjoyably.
Repeatable Practice	Good	Good	The exercises can be repeated as needed by students, helping them to master pronunciation gradually.

ASR: Automatic Speech Recognition

Overall, expert validation indicated that the ASR-based self-instructional material demonstrated good to very good quality across all assessed aspects. The material was systematically organized, visually appropriate for elementary learners, and supported by effective illustrations and audio recordings. The ASR function received the highest ratings from both validators, highlighting its optimal performance in providing immediate feedback for independent pronunciation practice. These results confirm that the developed prototype met academic and pedagogical validity criteria and is thus suitable for further refinement and implementation.

The following figure illustrates the ASR application that accompanies the book and can be used by students for independent study.



Figure 1.
Application cover

Figure 2. Menu in
the application

Figure 3. Activities
topic menu

Figure 4. Room
topic menu



Figure

Figure

Figure

Figure

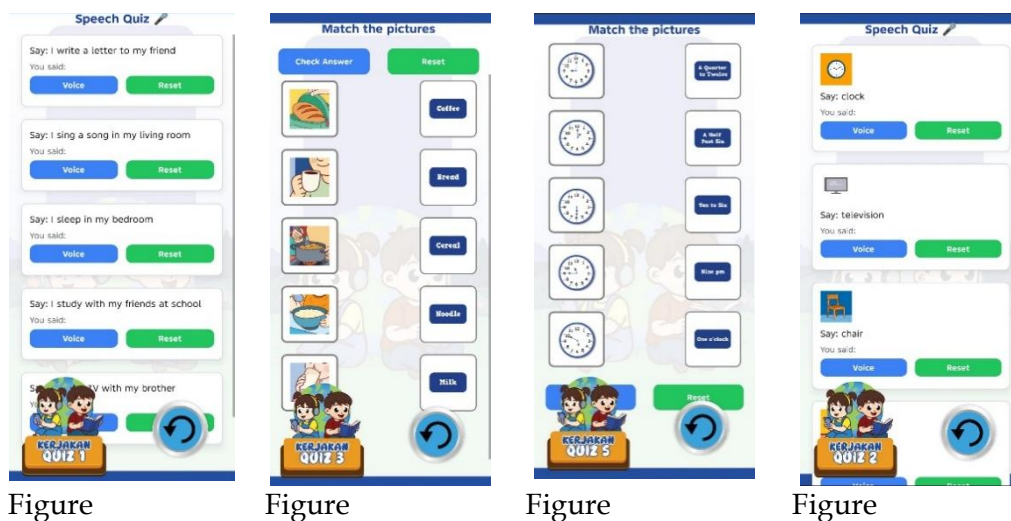


Figure 1: Illustration of the ASR app

Figure 1 displays screenshots of the ASR-based English learning app that was developed for elementary school students. Each image displays a different stage of the app interface, beginning with the cover page, followed by the main menu, and then the specific topic menus such as Activities and Room. This visual presentation aims to provide a clear overview of the navigation structure and available features, including interactive buttons that allow students to practice pronunciation independently. By displaying each menu, users can understand the flow of the app from start to finish. This is important to ensure that teachers and students can use all the features optimally and navigate the application without confusion.

Furthermore, Figure 1 also demonstrates the consistency of the visual design and the child-friendly interface with icons, character illustrations, and attractive colors. Each menu has a specific function; for example, the 'activities' menu provides interactive exercises, while the 'room' menu presents vocabulary contexts that are relevant to students' everyday environments. The sequential arrangement of the screenshots makes it easier for readers to understand how the app supports pronunciation practice step by step. This visual approach makes the learning process more intuitive and enjoyable, while emphasizing the integration of the self-instructional textbook and the ASR application as complementary learning media.

This planned structure further supports the alignment of the material with the student's competency level, allowing pronunciation practice to occur gradually and consistently. Moreover, Automatic Speech Recognition-based self-instructional books have a gradual structure and are tailored to the needs of the elementary school curriculum.

4. Discussion

Teacher preparedness is a crucial factor in the effectiveness of English pronunciation instruction in elementary schools, especially when classroom teachers are not English language education graduates. Differences in teacher

competency limit teaching strategies and the availability of systematic pronunciation practice, leading students to rely more on teacher instruction than on independent practice. Limited learning media, including textbooks, native speaker recordings, and interactive tools, as well as simple methods such as quizzes or singing, hinders consistent, repetitive practice. The integration of technology, particularly ASR, allows students to practice independently more effectively and to improve their pronunciation accuracy. Research shows that ASR can improve pronunciation, while teacher training in its use strengthens learning outcomes (X. Liu et al., 2019; McCrocklin, 2016; Spring & Tabuchi, 2021). With adequate technological support and teacher competency, pronunciation learning can become optimal and sustainable.

In addition to teacher competency, the availability of learning media is also a significant challenge in supporting pronunciation practice. Limited media and resources are major obstacles to teaching English pronunciation in elementary schools. The lack of exercise books, native speaker audio recordings, and interactive tools makes it difficult for students to practice independently and consistently. Without adequate media support, students simply follow teachers' instructions, limiting opportunities to correct repeated pronunciation errors. The integration of ASR technology offers a solution to increase student independence in pronunciation practice.

Research shows that ASR can recognize variations in children's voices with high accuracy, support adaptive pronunciation practice (Bhardwaj et al., 2022; Shivakumar et al., 2014; Sun, 2023), and provide real-time feedback. Caiza et al. (2025) emphasized that combining ASR with multimedia materials can increase student learning motivation, while García et al. (2020) demonstrated the effectiveness of ASR compared to traditional methods. With media support and teacher training, the use of ASR can create repetitive, systematic, and effective pronunciation practice.

However, media limitations cannot be overcome solely with technology; they also require competency development through active teacher participation in professional forums. Teacher participation in KKG is a crucial factor in improving pronunciation teaching competency; however, participation levels remain uneven. Some teachers rarely attend, limiting opportunities to share strategies and experiences, which risks creating disparities in teaching quality between classes. Active teachers more often implement varied methods and experiment with media, while passive teachers rely solely on textbooks. Studies show that collaborative professional development can improve instructional quality and encourage the adoption of innovative practices (Amemasor et al., 2025).

Mobile-assisted learning and ASR have also been shown to be effective in strengthening pronunciation and supporting student learning independence (Metruk, 2024; Sun, 2023). Low engagement limits access to ongoing training and technology integration, so external support is needed to bridge the gap in teaching quality. Uneven teacher involvement further complicates the need for innovative media to minimize the gap in pronunciation learning quality. The availability of

learning media is a crucial factor in mastering English pronunciation; many schools provide only simple textbooks and activities, depriving students of access to independent, repetitive practice, which is essential for developing pronunciation. Observations indicate limitations in interactive media, native-speaker recordings, and supplementary practice tools, resulting in pronunciation practice remaining dependent on teacher guidance rather than being fully independent. Research demonstrates that automated feedback-based systems and peer assessment can enhance independent practice opportunities and student pronunciation outcomes (Liu et al., 2025).

Another study confirmed that interactive audio applications can act as pronunciation coaches, encouraging consistent student engagement (Alharthi, 2024). A recent review also showed that the use of interactive digital media with listening and repetition tasks accompanied by immediate feedback significantly contributed to pronunciation improvement (Amrate & Tsai, 2024). This situation emphasizes the need for learning media innovation to focus on providing independent practice tools with continuous feedback support.

Automatic Speech Recognition-based self-instructional textbooks were developed to address the limitations of teachers and classroom media. Each unit includes an introduction to basic sounds, word exercises, phrases, and complete sentences, complete with interactive illustrations and audio recordings of native speakers. The ASR feature allows students to practice pronunciation independently, receive automatic feedback, and repeat exercises as needed. The material is structured in stages to ensure students systematically progress in their competencies. The integration of audio, visual, and ASR technology supports repeated and consistent pronunciation practice. Research shows that the use of ASR in a flipped learning context can improve students' pronunciation progress (Leis, 2025).

Another study comparing traditional instruction and ASR-based pronunciation found that ASR was more effective in correcting specific concrete sounds in the short term (García et al., 2020). Similar findings were also demonstrated by research confirming the role of ASR corrective feedback in improving ESL/EFL pronunciation accuracy (Ngo et al., 2024). This confirms that ASR-based self-instructional textbooks serve as a practical solution to support students' independent pronunciation practice.

The following is a flowchart of the ASR Direct Feedback Training.

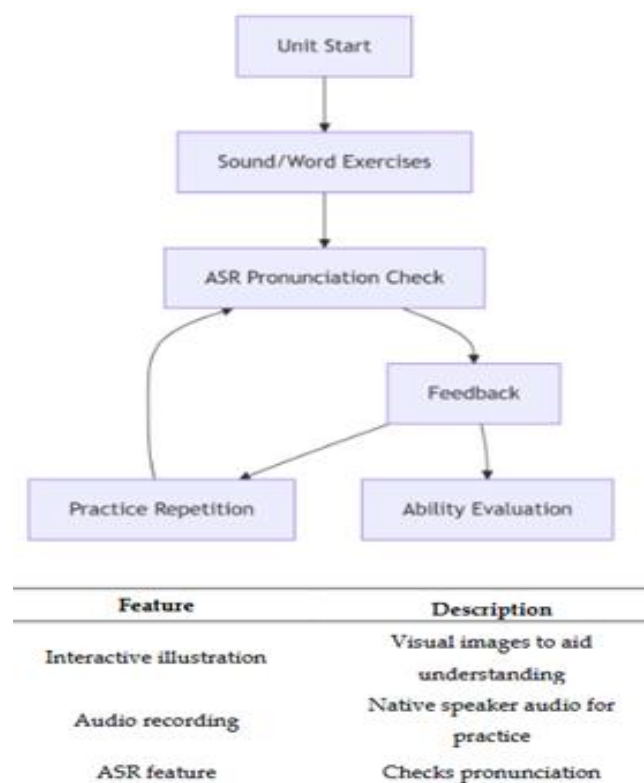


Figure 2: Practice flow with direct ASR feedback

Figure 2 shows the practice flow in an ASR-based self-instructional book designed to support English pronunciation mastery in elementary school students. The process begins with understanding the lesson unit, followed by practicing specific sounds or words. Each exercise is reviewed using the ASR feature, which provides immediate feedback and allows students to identify pronunciation errors. After receiving feedback, students repeat the exercise to reinforce sound and word mastery. Finally, students' skills are evaluated to assess pronunciation competence. The features of the book and the app, including interactive illustrations, audio recordings, and the ASR system, support each stage of this flow in an engaging and accessible way, enabling students to learn independently, repeatedly, and effectively.

This application, designed to complement English textbooks, is equipped with materials and exercises. The exercises are effective when students can practice independently while receiving automatic feedback through ASR technology, allowing errors to be identified and corrected immediately (Dai & Wu, 2023). Each unit features interactive illustrations, audio recordings, and a speech-processing system that supports pronunciation accuracy (Bogach et al., 2021). Consistent visual and audio media enable students to master English vocabulary and sounds gradually while fostering independent learning (Lai & Cheong, 2022). This combination of features makes this book an effective tool for systematically and sustainably improving pronunciation.

The book's material is structured in stages, starting with simple sounds and progressing to more complex combinations to align with the student's competency development. Repetition exercises combined with illustrations and interactive audio help students master vocabulary and pronunciation more systematically, which is in line with findings that digital multimodal scaffolding effectively supports language comprehension (Vena & Yuliana, 2023). The gradual structure allows students to practice sounds and words at their level of ability and reduces errors through technology-based adaptations that are relevant to the elementary school context (Aravantinos et al., 2024). Consistent audio-visual support has also been shown to improve children's mastery of language structures, leading to increased confidence in pronunciation (Gowenlock et al., 2024). This systematic arrangement of materials with interactive features shows strong potential for continuously increasing the effectiveness of pronunciation learning.

The ASR-based self-instructional book was validated by English language education experts. The validation results indicated that the material is systematically structured, the visual design is attractive, and the ASR feature functions optimally. Illustrations and audio recordings support pronunciation practice, while repetition exercises provide students with opportunities to repeat as needed. This validation aligns with findings that the success of learning media is determined by the integration of content, design, and technology functions that support learning interactions (Wiśniewska, 2016). Furthermore, the use of ASR features that enable immediate feedback has been shown to increase student engagement and the effectiveness of pronunciation training (Caiza et al., 2025). Therefore, the results of the content and feasibility validations indicate that the book is suitable for use as a learning medium that promotes independent and effective pronunciation learning.

Based on the validation results of two English language education experts, the ASR-based self-instructional textbook demonstrates good to excellent quality across all assessed aspects. The material is structured logically and systematically, making it easy for students to follow the pronunciation-practice steps, step by step. The structure of the ASR-based self-instructional textbook demonstrates a logical structure from basic sound recognition to complete sentences, enabling students to follow pronunciation practice step by step. This structure facilitates the gradual development of children's phonological skills and prevents confusion when encountering more complex sounds. The validation results showed a score of 87.5%, which is consistent with research emphasizing the importance of a tiered material sequence to improve comprehension and retention (Morrison et al., 2019). This structure allows students to gain a strong pronunciation foundation before facing more complex material while providing a focused and enjoyable learning experience. Therefore, the material structure is a crucial component in supporting the effectiveness of independent pronunciation practice.

The book's visual design features attractive illustrations, colors, and layouts that suit the characteristics of elementary school children, thus supporting student

engagement and motivation. Appropriate visuals facilitate focus and comprehension of the material and promote active student interaction with the book. This aspect scored 80%, which is consistent with literature showing that engaging visual elements enhance cognitive engagement and information retention in language learning (Clark & Mayer, 2023). The child-friendly design ensures each unit of material is easy to follow, minimizes confusion, and fosters independent learning. Therefore, effective visual integration is crucial for building a coherent, enjoyable, and motivating learning experience for students practicing pronunciation.

The ASR feature works optimally, providing real-time feedback that allows students to identify pronunciation errors and correct them independently. Immediate feedback is a crucial element in accelerating the correction process and improving pronunciation accuracy while also encouraging independent learning. Validation results showed a perfect score of 100%, which is in line with research confirming the effectiveness of ASR in improving children's pronunciation and facilitating consistent independent practice (Elimat & AbuSeileek, 2014; Ngo et al., 2024). This technology integration helps students to adjust their learning pace to their individual abilities, making the practice process more adaptive and enjoyable. Thus, optimal ASR functionality serves as the core of an interactive learning experience that supports continuous pronunciation mastery.

Interactive materials and audio recordings support the material, helping students to understand vocabulary and practice pronunciation through multisensory means. The presence of audio-visual media strengthens the connection between sound and meaning, enhancing phonological skills and learning motivation. This aspect achieved an average score of 80%, which is consistent with findings that the combination of text, images, and audio effectively supports children's language comprehension (Mohammed Cherif, 2024). The consistent presentation of illustrations and audio recordings throughout each unit allows students to practice repeatedly with clear references, facilitating the acquisition of new vocabulary and language sounds. Thus, these media are a crucial component in reinforcing pronunciation practice and encouraging active and sustained learning engagement.

Repeated practice is provided throughout each unit to allow students to strengthen their pronunciation skills at their own pace. Systematic repetition facilitates long-term retention and strengthens children's phonological awareness. The validation results showed an average score of 80%, which is consistent with literature emphasizing the importance of structured repetition in language learning (Aravantinos et al., 2024). Through repeated practice, students can independently correct errors and adjust the intensity of practice to their individual needs. This approach supports adaptive learning and ensures pronunciation mastery.

The visual design is engaging and age-appropriate, increasing motivation and engagement. The ASR feature works optimally, providing automatic feedback that allows students to improve their pronunciation independently, thus

supporting independent learning. Illustrations and audio recordings effectively reinforce pronunciation practice, while repetitive exercises provide opportunities for students to repeat as needed. These findings align with research showing that interactive learning media integrated with adaptive technology can enhance learning effectiveness and student engagement (Gavriushenko et al., 2015; Sabri et al., 2024). Hence, this validation confirms that the ASR-based self-instructional textbook is ready to be used as an interactive, enjoyable, and effective learning medium to improve the English pronunciation competence of elementary students.

5. Conclusion

This research addressed the challenge of supporting the English pronunciation mastery of elementary school students through the development of a self-instructional textbook integrated with ASR. The findings demonstrate that the ASR-based textbook effectively facilitates independent, repetitive, and systematic pronunciation practice by providing learners with immediate and real-time feedback. Expert validation results indicate that the instructional content, visual design, and ASR functionality are structured from good to excellent quality, contributing positively to students' engagement, motivation, and learning autonomy.

Although the study is limited to the prototype development and academic validation stages and involves a relatively small number of expert validators, the results underscore the potential of systematic and adaptive learning media design. The integration of step-by-step instructional materials, repetitive practice activities, interactive illustrations, audio models, and ASR features creates a coherent learning environment that enables learners to monitor their pronunciation progress gradually. These findings suggest that ASR-based instructional materials are effective pedagogical tools for consistent pronunciation practice and learner independence, while highlighting the importance of aligning technological innovation with sound pedagogical principles in elementary English language learning.

Beyond validating the technical feasibility of ASR integration, this study highlights the pedagogical value of embedding technology within a self-directed learning framework tailored to young learners. The structured sequencing of activities supports gradual phonological development, while repetitive practice combined with automated feedback enables learners to internalize accurate pronunciation patterns at their own pace. Importantly, the age-appropriate visual and audio components reduce cognitive load and sustain attention, which are critical factors in elementary language learning. By emphasizing autonomy without eliminating instructional scaffolding, the developed textbook demonstrates how ASR-based materials can balance technological affordances with pedagogical sensitivity, thereby offering a scalable and sustainable approach to pronunciation instruction in resource-limited educational contexts.

6. Acknowledgments

The authors gratefully acknowledge the financial support provided by the Directorate of Research and Community Service (DPPM) through the Program Penelitian Fundamental-Reguler (PFR). The authors also extend their sincere appreciation to Universitas Islam Nahdlatul Ulama Jepara for the institutional support that enabled the successful completion of this research and particularly to the Institute for Research and Community Service (LPPM) for its guidance, facilitation, and administrative assistance throughout the research process. Gratitude is also extended to everyone who played a role in this research. The authors acknowledge the use of artificial intelligence-based tools for language editing and clarity improvement. The authors confirm that all scientific content, interpretations, and conclusions remain their sole responsibility.

7. References

- Ahn, T., Hong, Y., Im, Y., Kim, D. H., Kang, D., Jeong, J. W., Kim, J. W., Kim, M. J., Cho, A.-R., & Nam, H. (2025). Automatic speech recognition (ASR) for the diagnosis of pronunciation of speech sound disorders in Korean children. *Clinical Linguistics & Phonetics*, 39(10), 913–926. <https://doi.org/10.1080/02699206.2024.2387609>
- Alharthi, S. M. (2024). Siri as an interactive pronunciation coach: its impact on EFL learners. *Cogent Education*, 11(1), 2304245. <https://doi.org/10.1080/2331186X.2024.2304245>
- Amemasor, S. K., Oppong, S. O., Ghansah, B., Benuwa, B.-B., & Essel, D. D. (2025). A systematic review on the impact of teacher professional development on digital instructional integration and teaching practices. *Frontiers in Education*, 10, 1541031. <https://doi.org/10.3389/educ.2025.1541031>
- Amrate, M., & Tsai, P. (2024). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 1–21. <https://doi.org/10.1017/S0958344024000181>
- Aravantinos, S., Lavidas, K., Voulgari, I., Papadakis, S., Karalis, T., & Komis, V. (2024). Educational approaches with AI in primary school settings: A systematic review of the literature available in Scopus. *Education Sciences*, 14(7), 744. <https://doi.org/10.3390/educsci14070744>
- Bashori, M., van Hout, R., Strik, H., & Cucchiari, C. (2024). I Can Speak: improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, 18(5), 443–461. <https://doi.org/10.1080/17501229.2024.2315101>
- Bhardwaj, V., Ben Othman, M. T., Kukreja, V., Belkhier, Y., Bajaj, M., Goud, B. S., Rehman, A. U., Shafiq, M., & Hamam, H. (2022). Automatic speech recognition (ASR) systems for children: A systematic literature review. *Applied Sciences*, 12(9), 4419. <https://doi.org/10.3390/app12094419>
- Bogach, N., Boitsova, E., Chernonog, S., Lamtev, A., Lesnichaya, M., Lezhenin, I., Novopashenny, A., Svechnikov, R., Tsikach, D., & Vasiliev, K. (2021). Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching. *Electronics*, 10(3), 235. <https://doi.org/10.3390/electronics10030235>
- Caiza, G., Villafuerte, C., & Guanuche, A. (2025). Interactive Application with Virtual Reality and Artificial Intelligence for Improving Pronunciation in English Learning. *Applied Sciences*, 15(17), 9270. <https://doi.org/10.3390/app15179270>
- Chang, V., & Fisher, D. (2003). The validation and application of a new learning environment instrument for online learning in higher education. In *Technology-rich learning environments: A future perspective* (pp. 1–20). World Scientific.

- https://doi.org/10.1142/9789812564412_0001
- Clark, R. C., & Mayer, R. E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.
- Creswell, J. W., & Plano Clark, V. L. (2023). Revisiting mixed methods research designs twenty years later. *Handbook of Mixed Methods Research Designs*, 1(1), 21–36.
- Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, 36(5–6), 861–884. <https://doi.org/10.1080/09588221.2021.1952272>
- Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The Jalt Call Journal*, 10(1), 21–47. https://doi.org/https://files.eric.ed.gov/fulltext/EJ1107929.pdf?utm_source
- Evers, K., & Chen, S. (2021). Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*, 59(4), 669–685. <https://doi.org/10.1177/0735633120972011>
- Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, 35(8), 1869–1889. <https://doi.org/10.1080/09588221.2020.1839504>
- García, C., Nickolai, D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction. *Calico Journal*, 37(3), 213–232. <https://doi.org/10.1558/cj.40379>
- Gavriushenko, M., Karilainen, L., & Kankaanranta, M. (2015). Adaptive systems as enablers of feedback in English language learning game-based environments. *2015 IEEE Frontiers in Education Conference (FIE)*, 1–8. <https://doi.org/10.1109/FIE.2015.7344107>
- Gowenlock, A. E., Norbury, C., & Rodd, J. M. (2024). Exposure to language in video and its impact on linguistic development in children aged 3–11: A scoping review. *Journal of Cognition*, 7(1), 57. <https://doi.org/10.5334/joc.385>
- Guskaroska, A. (2024). *Exploring technology acceptance of ASR for pronunciation learning*. Iowa State University.
- Inceoglu, S., Chen, W.-H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *System*, 124, 103387. <https://doi.org/10.1016/j.system.2024.103387>
- Lai, J. W., & Cheong, K. H. (2022). Educational opportunities and challenges in augmented reality: Featuring implementations in physics education. *Ieee Access*, 10, 43143–43158. <https://doi.org/10.1109/ACCESS.2022.3166478>
- Leis, A. (2025). How speech-to-text technology affects pronunciation gains and self-confidence in EFL learners. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2025.2534498>
- Liu, Hwang, G.-J., Yu, P., Tu, Y.-F., & Wang, Y. (2025). Effects of an automated corrective feedback-based peer assessment approach on students' learning achievement, motivation, and self-regulated learning conceptions in foreign language pronunciation. *Educational Technology Research and Development*, 1–22. <https://doi.org/10.1007/s11423-025-10484-z>
- Liu, T., Zhang, Z., & Gao, X. (2023). Pedagogical design in technology-enhanced language education research: A scoping review. *Sustainability*, 15(7), 6069. <https://doi.org/10.3390/su15076069>
- Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., Chen, X., & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, 25(2), 126–140. <https://doi.org/10.1504/IJIL.2019.097674>
- Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature

- review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, 12(1), 2466288. <https://doi.org/10.1080/2331186X.2025.2466288>
- McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57(1), 25–42. <https://doi.org/10.1016/j.system.2015.12.013>
- Metruk, R. (2024). Mobile-assisted language learning and pronunciation instruction: A systematic literature review. *Education and Information Technologies*, 29(13), 16255–16282. <https://doi.org/10.1007/s10639-024-12345-6>
- Mohammed Cherif, D. (2024). *The Use of the Multi-Sensory Teaching Approach in Enhancing Vocabulary Learning for EFL Young Learners*. Université KASDI MERBAH-OUARGLA. <https://dspace.univ-ouargla.dz/jspui/handle/123456789/36913>
- Moore, P. D. (2025). Smarter Learning: Integrating AI into Instructional Design for 21st-Century Education. *SCIENTIA MORALITAS-International Journal of Multidisciplinary Research*, 10(1), 86–108. <https://doi.org/10.5281/zenodo.16335328>
- Morrison, G. R., Ross, S. J., Morrison, J. R., & Kalman, H. K. (2019). *Designing effective instruction*. John Wiley & Sons.
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/S0958344023000113>
- Plomp, T. (2013). Educational design research: An introduction. *Educational Design Research*, 1, 11–50. https://www.fi.uu.nl/publicaties/literatuur/educational-design-research-part-a.pdf?utm_source
- Richey, R. C., & Klein, J. D. (2014). *Design and development research: Methods, strategies, and issues*. Routledge. <https://doi.org/10.4324/9780203826034>
- Sabri, S. M., Ismail, I., Annuar, N., Rahman, N. R. A., Abd Hamid, N. Z., & Abd Mutalib, H. (2024). A conceptual analysis of technology integration in classroom instruction towards enhancing student engagement and learning outcomes. *Integration*, 9(55), 750–769. <https://doi.org/10.35631/IJEPC.955051>
- Sariani, S., Khairat, M. El, Haslina, W., & Baetty, B. (2024). ASR-based system for promoting pronunciation: Promoting collaborative approach for higher education ELF learners. *AILA Review*. <https://doi.org/10.1075/aila.23021.sar>
- Shivakumar, P. G., Potamianos, A., Lee, S., & Narayanan, S. (2014). *Improving speech recognition for children using acoustic adaptation and pronunciation modeling*. [https://sail.usc.edu/publications/files/Improving Speech Recognition for Children using Acoustic Adaptation and Pronunciation Modeling.pdf](https://sail.usc.edu/publications/files/Improving_Speech_Recognition_for_Children_using_Acoustic_Adaptation_and_Pronunciation_Modeling.pdf)
- Spring, R., & Tabuchi, R. (2021). Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online. *Journal of English Teaching through Movies and Media*, 22(2), 93–104. <https://doi.org/10.16875/stem.2021.22.2.93>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, 14, 1210187. <https://doi.org/10.3389/fpsyg.2023.1210187>
- Vena, A., & Yuliana, Y. G. S. (2023). The Role of Multisensory Approach in Fostering Student Well-Being in English Language Teaching for Young Learners: English. *Journal of English Development*, 3(01), 229–242. <https://doi.org/https://doi.org/10.25217/jed.v3i01.4784>
- Wang, Y., & Young, S. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, 31(6), 493–504. <https://doi.org/10.1111/jcal.12079>

- Wiśniewska, H. (2016). Learner autonomy: The role of educational materials in fostering self-evaluation. In *Autonomy in second language learning: Managing the resources* (pp. 85–98). Springer. https://link.springer.com/chapter/10.1007/978-3-319-07764-2_6
- Won, Y. (2025). Assessing the efficacy of word error rate as a proxy for pronunciation quality: A comparative study of ASR systems and human evaluations among young EFL learners. *Journal of Second Language Pronunciation*.
- Xiao, H., Ou, K., Wang, H., & van de Weijer, J. (2023). The effect of ASR apps on monophthong pronunciation improvement and generalization to new words in English. *Science and Information Conference*, 1410–1433. https://doi.org/10.1007/978-3-031-37717-4_93
- Xu, G., Yu, A., & Liu, L. (2025). A meta-analysis examining AI-assisted L2 learning. *International Review of Applied Linguistics in Language Teaching*, 0. <https://doi.org/10.1515/iral-2024-0213>
- ZIANE, S., & DOUB, A. (2024). Using ASR Technology to Assess EFL Learners' Pronunciation. *Proc. IAC Prague 2024*, 95. https://www.researchgate.net/publication/384528886_Using_ASR_Technology_to_Assess_EFL_Learners%27_Pronunciation