PESHUM : Jurnal Pendidikan, Sosial dan Humaniora
Vol.5, No.5, Agustus 2026

YouTube-Based Pronunciation Instruction through Pronunciation with Emma: Enhancing Students' Mastery of English Diphthong Sounds

Whana Sri Ayu Ningtias[1], Maemuna Muhayyang[2], Haryanto[3]
English Education Study Program, Faculty of Languages and Literature, Universitas Negeri Makassar, Indonesia
Corresponding author: maemarasyid@unm.ac.id

Article History:
Received: 12 Juni 2026
Revised: 29 Juli 2026
Accepted: 07 Agustus 2026

Keywords: *diphthong sounds; EFL learners; pronunciation accuracy; pronunciation instruction; YouTube-based learning*

Abstract: *The study investigated whether the Pronunciation with Emma YouTube channel could have positive effects on junior secondary students' pronunciation of English diphthongs. A pre-experimental one-group pre-test and post-test design was used. The participants were 9th grade students of Class IX.K at SMP Negeri 3 Sungguminasa. The data were obtained from a pronunciation reading test made up of 40 words using eight diphthongs in English. The pre-test was prepared and used to assess students' pronunciation achievement (AO1) at first, then a post-test was conducted after 4 treatment sessions using selected videos from Pronunciation with Emma channel. Quantitative data were also interpreted with* ***descriptive*** *statistics by the paired-sample t-test. It is established that students' mean score increased as from 50.33 in the pretest to 73.93 in the post-test. The paired-sample t-test result was significant, p < .001, which suggests this improvement was significant in the outcomes following the intervention. Improvements were also observed for all eight diphthong sounds. Since pronunciation with Emma can give a useful audio-visual model, articulation support, and repetition effect to EFL learners as additional media in practice, it is a pedagogical view to aiding students in their development of English diphthong pronunciation performance.*

Introduction

Pronunciation is a crucial aspect of learning English as a Foreign Language (EFL) in terms of students' ability to understand spoken English communication. It also affects students' intelligibility, comprehensibility, self-confidence, performance, and willingness to participate in spoken activities. However, pronunciation in classroom language learning is often considered less important than grammar, vocabulary, reading, and writing. This emphasis might hinder students' ability to develop enough awareness to understand how English sounds are produced, formed,

heard, and practiced while listening and speaking. Data from recent years shows that pronunciation difficulties can severely limit a learner's intelligibility and comprehension of spoken English (Sardegna, 2021). Pronunciation is not just about producing native sounds; it serves as a communication tool that enables learners to construct speech that others can understand. Therefore, pronunciation instruction should be central to EFL teaching, especially for students learning to speak for the first time.

The differences in sound patterns between English and Indonesian make pronunciation more challenging for Indonesian EFL learners, particularly when speaking English. Indonesian spelling is inherently more transparent and consistent than English pronunciation, where learners often have to learn complex sound-spelling relationships. As a result, learners often pronounce English words with Indonesian orthographic habits. This is increasingly common as students encounter English sounds that do not translate to Indonesian, including English diphthong sounds. Diphthongs help transition between two vowels in a single syllable [e.g., face /eɪ/, time /aɪ/, boy /ɔɪ/, house /aʊ/, home /oʊ/]. In Indonesia, learners are accustomed to fixed vowel sounds, so they typically split diphthongs into monophthongs or pronounce them according to Indonesian spelling. This can lead to less accurate pronunciation and make spoken English sound less natural and less intelligible.

Early on at SMP Negeri 3 Sungguminasa, it was found that first-year 9th graders were unable to pronounce English words with diphthongs. For example, words like face, name, time, light, house, home, near, sure, etc., are usually pronounced without the proper action of vowel sounds. Students read the words as they appear on the page (rather than how they sound) and adjust them to Indonesian pronunciation patterns. This indicates that explanations are not enough; students need many opportunities to accurately represent and imitate words, along with extensive practice using visual or auditory resources. In this context, teaching pronunciation involves explicit sound recognition and articulation practice. Mora (2026) states that L2 pronunciation is still difficult to learn because language is taught far less in the classroom than other grammar and pronunciation-related areas. This underscores the value of educational media that allows pronunciation practices to be taught in public places and not just during explicit classroom explanations.

The advent of online-based instruction systems offers new possibilities for teaching pronunciation. Technology-enhanced pronunciation learning enables students to practice with pronunciation models, work at their own pace, and receive multiple types of input. Recent research on computer-assisted pronunciation training (CAPT) has shown that technology can be beneficial for learners in pronunciation learning if they receive focused practice on individual segmental features, such as vowels and consonants (Amrate & Tsai, 2025). Additionally, meta-analytical research on automatic speech recognition (ASR) has found that this technology-assisted pronunciation practice can enhance the pronunciation performance of ESL/EFL speakers, especially in the segmental pronunciation domain (Ngo et al., 2024). Although the present study did not make use of ASR-based feedback, the conclusions are interesting, since it shows that technology can help children become more advanced in pronunciation in the field of repeated, structured, and model-mediated practice.

YouTube is one of the most easily obtainable online learning resources for EFL students. YouTube provides an abundance of audiovisual content and can be used as supplementary material on pronunciation instruction. Printed readings don’t, but videos allow learners to listen to pronunciation models, watch mouth movements, repeat target sounds, and rewatch the same things a number of times. The key here is that pronunciation learning includes auditory discrimination and articulation practice. Additionally, a recent meta-analysis of video-based input

in second language learning also reported video-based inputs as effective in improving several points of interest in L2 learning (notably pronunciation, speaking, listening, vocabulary, and grammar), with educational- or language-oriented videos being perceived to be pertinent to the learning process (Sutton & Webb, 2026). Thus YouTube-based pronunciation instruction is pedagogically relevant if it chooses videos that are relevant, accurate, systematic, and appropriate according to the proficiency of the learners.

One YouTube channel that is relevant to pronunciation learning, that is Pronunciation with Emma. This is also the channel that teaches pronunciation: via simple explanations, audio examples, repetition, plus visual demos of sound production. Features like this are relevant to learning diphthongs in English because students can listen to how diphthongs are pronounced and also learn how the mouth moves from one vowel position to another. They will start to notice different pronunciation differences through visualising, then practice sound, the way sounds are spoken, even though they cannot articulate it in writing. Moreover, the channel is quite user friendly and suitable for junior secondary level students that need simplistic, tangible models of pronunciation. Hence, Pronunciation with Emma could serve as some supplementary medium which could support students to reach the complete understanding of the pronunciation of English diphthongs.

A number of research have focused on the impacts of YouTube and electronic media upon learning of English as a foreign language—speaking as well as pronunciation problems, either speaking or pronunciation. However, a substantial amount in the current research pertains to YouTube-based language learning in its true sense, or considers learners' perceptions, speaking ability, vocabulary enhancement, or general pronunciation. More recent research exploring technology-facilitated pronunciation learning has also been dominated by adult learners, ASR-based systems, and general CAPT contexts (Amrate & Tsai, 2025; Ma et al., 2024; Ngo et al., 2024). Little focus has been placed on junior secondary learners' acquisition of individual English diphthong sounds, as expressed in a YouTube channel targeted to pronunciation. This results in a pedagogic and empirical discrepancy. Pedagogical support Teachers need more available and engaging media to support EFL learners to practice high-difficulty English sounds. As a part of a developing research question, the empirical evidence that a given YouTube pronunciation channel can markedly boost learners' pronunciation of English diphthongs is still distant.

It is the background for the present study and we also consider Pronunciation with Emma as a possible additional media that may help us improve students' pronunciation of English diphthongs. This study specifically investigates whether pronunciation skill before and after instruction on the channel affects students differently with a significant difference. This study employs one group pretest and post-test design and thus is hoping to add to the body of knowledge on the effectiveness of delivering pronunciation instruction using YouTube in a junior secondary EFL class. Its findings will be useful in pronunciation teaching in that their findings will show how the type of audiovisual material an instructor is equipped for the learning of pronunciation can affect the pronunciation accuracy among students, their willingness to learn, and their overall confidence when they must produce diphthong English.

Literature Review

Pronunciation in EFL Learning

Pronunciation is a key aspect of English language learning, because it determines learners' intelligibility, speaking confidence, and ability to communicate effectively. In EFL environments, pronunciation involves more than making certain sounds (that is, putting them together as they

are spoken) for their listeners to understand. As Kelly (2000) says, pronunciation encompasses the production of sounds: sounds, stress, rhythm, intonation, etc. These elements work together to help speakers convey meaning clearly and naturally. Thus, pronunciation must be understood as not an isolated language capability but rather as a form of oral communication. Pronunciation is also difficult for EFL students due to the unique sound system that occurs in English and is not the same as their first language. They could transfer phonological patterns from their mother tongue to English, and subsequently produce inaccurate pronunciation. But Sardegna (2022) has mentioned that this may be true as pronunciation should not only mimic native-sounding speech, and it is also used to train learners to develop strategies for intelligibility as opposed to mimicking native-like speech. Effective pronunciation instruction, then, must address students' ability to detect sound contrast, practice difficult sounds, and produce speech that can be understood in English communication.

Pronunciation is an essential component of English language learning because it influences learners' intelligibility, speaking confidence, and ability to communicate effectively. In EFL contexts, pronunciation is not only concerned with producing individual sounds correctly but also with making spoken language understandable to listeners. Kelly (2000) explains that pronunciation includes the production of sounds, stress, rhythm, and intonation. These elements work together to help speakers convey meaning clearly and naturally. Therefore, pronunciation should not be viewed as an isolated language skill, but as an integral part of oral communication.
For EFL learners, pronunciation often becomes challenging because the sound system of English differs from that of their first language. Learners may transfer the phonological patterns of their mother tongue into English, which can result in inaccurate pronunciation. Sardegna (2022) emphasizes that pronunciation instruction should help learners develop strategies for improving intelligibility rather than merely imitating native-like speech. This means that effective pronunciation learning should focus on helping students notice sound differences, practice difficult sounds, and produce speech that can be understood in communication.

English Diphthong Sounds

Vowel production, specifically diphthong sounds, is one of the key components in the pronunciation of English. Diphthongs, as noted by Roach (2009), are vowel sounds that glide or otherwise move between a vowel position on a single syllable. In contrast to monophthongs, which are produced with a relatively stable vowel quality, diphthongs have to show a well-evolved movement of two different vowel sounds. English diphthongs consist of /eɪ/ in face, /aɪ/ in time, /ɔɪ/ in boy, /aʊ/ in house, /oʊ/ in home, /ɪə/ in near, /eə/ in care, and /ʊə/ in sure. So diphthongs are difficult for Indonesian EFL learners who speak English because Indonesian vowels are more stable and are generally pronounced according to writing. This can encourage students to simplify the diphthongs into one vowel sound in English or produce English words based on Indonesian speech patterns. For instance, face may be pronounced as /fes/ rather than /feɪs/. These pronunciation errors can diminish accuracy and intelligibility. For this reason, English diphthongs need explicit teaching, adequate modeling, and long practice to ensure that students can learn and render the vowel movement.

One important aspect of English pronunciation is vowel production, particularly diphthong sounds. According to Roach (2009), diphthongs are vowel sounds that involve a glide or movement from one vowel position to another within a single syllable. Unlike monophthongs, which are produced with a relatively stable vowel quality, diphthongs require a smooth transition between two vowel sounds. English diphthongs include sounds such as /eɪ/ in *face*, /aɪ/ in *time*, /ɔɪ/ in *boy*, /aʊ/ in *house*, /oʊ/ in *home*, /ɪə/ in *near*, /eə/ in *care*, and /ʊə/ in *sure*.

Diphthongs are often difficult for Indonesian EFL learners because Indonesian vowels tend to be more stable and are commonly pronounced according to written forms. As a result, students may simplify English diphthongs into single vowel sounds or pronounce English words based on Indonesian spelling patterns. For example, the word *face* may be pronounced as /fes/ instead of /feɪs/. Such pronunciation problems can reduce accuracy and intelligibility. Therefore, English diphthongs require explicit teaching, sufficient modelling, and repeated practice so that students can recognize and produce the vowel movement accurately.

YouTube as a Medium for Pronunciation Learning

Digital technology has opened new avenues for language instruction, specifically speaking pronunciation. YouTube is one popular digital platform with educational videos that may be helpful during teaching language. YouTube is a platform to view videos of relevant subjects anytime and anywhere, which enables students to pick up, or download, material at any time while learning with audio and visuals. These features offer particular value to students in pronunciation learning, as people need to listen to accurate pronunciation models and watch the pronunciation process. Audiovisual learning supports pronunciation as it links listening input (the sounds we hear and see) with visual articulation stimuli. Using videos, students can watch mouth movement, lip positions, and sound transitions as they listen to pronunciation examples. Sutton and Webb (2026) posit that audiovisual input can aid in second language acquisition by giving learners multiple sources of information.

In pronunciation education, including such sound features which would never be clear to students through written explanations with these tools. Literature Review in recent years has also demonstrated that technology-facilitated pronunciation can aid learners' pronunciation learning through digital tools. As reported by Amrate and Tsai (2025) there is potential effectiveness with computer-assisted pronunciation training by providing targeted practice in the target speech sound and repeated exposures to it. Similarly, Ngo et al. (2024) found that technology and training for pronunciation can enhance pronunciation performance of ESL/EFL learners. YouTube is not always available for automatic feedback in this case, however, although it can still provide valuable supplemental medium if teachers use YouTube to add appropriate video and apply it to classroom activities when selected and included in lessons (incorporating these videos in context).

Pronunciation with Emma as a Supplementary Learning Medium

The video Pronunciation with Emma is a specific channel on YouTube for pronunciation. This channel specializes in English pronunciation. The channel gives lessons covering English sounds like vowels, consonants, stress, rhythm, intonation, as well as some frequently encountered pronunciation issues. Most of its explanations are fairly straightforward, clear and evidence based, with examples of pronunciation. This renders the channel useful for EFL learners who need the simplified but structured pronunciation input. With regards to English diphthong learning, Pronunciation with Emma helps students in a number of ways. First as the channel provides auditory models, we can listen to the pronunciation of English diphthongs. Second, it provides visual cues as it demonstrates mouth movement for sound production. Third, students can repeat some of the pronunciation examples multiple times which is essential for acquiring pronunciation accuracy. Fourth, the channel adds a possibility for independent learning as students can watch the videos outside the classroom for additional practice. Nonetheless, YouTube channels in pronunciation learning should not replace the role of the teacher in the pronunciation learning process.

Teachers must choose suitable videos, work with students to steer them to match up clips for specific pronunciation exercises, plan pronunciation practice, and provide corrective feedback, etc. Students who engage in this practice after they learn pronunciation without the teacher's guidance will passively pick up the videos without a full understanding of pronunciation. Thus Pronunciation with Emma is most useful as an optional, supportive aid to classroom-based pronunciation teaching. From the evidence, from the literature reviewed, pronunciation learning requires specific models, practice, repetition of exposures and targeted applications. Particularly when students are learning increasingly hard sounds like English diphthongs. Pronunciation videos on YouTube and Pronunciation with Emma are both potential methods used for delivering auditory input and for teaching pronunciation and Pronunciation with Emma provides a set of particular models that are useful for teaching pronunciation (English pronunciation). It's for this reason that the current study aimed to examine Pronunciation with Emma as an added medium to improve junior secondary students' pronunciation of the diphthong in English.

Method

Quantitative method was used in this research with pre-experimental design. The design used in this study was a one-group pre-test and post-test design. In this study, students underwent a pre-test before, and a post-test after treatment. Using this design the effect on students' pronunciation ability of using videos from the Pronunciation with Emma YouTube channel was measured. The research was carried out at SMP Negeri 3 Sungguminasa between February 2nd and March 2nd, 2026. The study population was composed of ninth-grade students, totaling 363. The sample was selected using purposive sampling. For example, one class, namely class IX.K, consisting of 30 students, was selected as the research sample because the students consistently struggled when trying to pronounce English diphthongs correctly. The test consisted of 40 words containing diphthong sounds. The test was conducted in two stages: pre-test and post-test. The treatment took place in four meetings. In treatment sessions, students viewed the videos presented through the Pronunciation with Emma channel, listened to some pronunciation examples they could practice, and repeated the words together while mimicking the native speaker in the videos. The students were evaluated on their clarity, accuracy, and the number of correctly pronounced words. Scoring criteria were listed as a 1-5 scale with 5 as excellent pronunciation and 1 as very poor pronunciation in English (Brown, 2004). Students' scores were also categorized into several categories — very good (86–100), good (71–85), fair (56–70), poor (41–55), and very poor (<40) according to criteria set by Depdiknas (2006).

The data in this study were analyzed quantitatively. Students' pronunciation scores from the pre-test and post-test were tabulated and compared to identify their pronunciation development after the treatment. Descriptive statistics were used to calculate the mean score, standard deviation, standard error, frequency, and percentage. In addition, students' scores were classified into five achievement categories: very good, good, fair, poor, and very poor. To examine the improvement of each English diphthong sound, the total scores of the eight diphthongs in the pre-test and post-test were compared. Finally, a paired-sample t-test was conducted to determine whether the difference between the pre-test and post-test scores was statistically significant. The following table illustrate the ways of data analysis.

Table 1. Data Analysis Procedures

No.	Data Analyzed	Source of	Statistical	Purpose of	Interpretation

		Data	Technique	Analysis	
1	Students' pronunciation scores before treatment	Pre-test	Mean, standard deviation, and standard error	To identify students' initial pronunciation ability before using *Pronunciation with Emma*	The pre-test score represents students' baseline ability in pronouncing English diphthongs
2	Students' pronunciation scores after treatment	Post-test	Mean, standard deviation, and standard error	To identify students' pronunciation ability after four treatment sessions	The post-test score represents students' pronunciation achievement after the intervention
3	Students' achievement levels	Pre-test and post-test scores	Frequency and percentage	To classify students' pronunciation ability into achievement categories	The classification shows the distribution of students in very good, good, fair, poor, and very poor categories
4	Improvement of each diphthong sound	Scores of eight English diphthongs	Score comparison and gain analysis	To determine which diphthong sounds improved after treatment	The gain score indicates the degree of improvement for each diphthong sound
5	Difference between pre-test and post-test scores	Paired pre-test and post-test scores	Paired-sample t-test	To examine whether the improvement was statistically significant	The result is significant if the p-value is lower than .05

FINDINGS AND DISCUSSION

Findings

This research's results following the reading test. In the next part we show the result of pupils' score. The increase in students' pronunciation scores before and after the treatment are summarized in Table 2.

Table 2. students' mean score of pre-test and post-test

Pair 1	Mean	N	Std. deviation	Std. Error Mean
Pre-test	50.33	30	9.327	1.703
Post-test	73.93	30	9.157	1.672

The mean score on the pre-test was 50.33 and the mean score on the post-test was 73.93. This means that students' pronunciation performance was enhanced following the Pronunciation with

Emma YouTube channel. The standard deviation from the pre-test was 9.327 and from the post-test was 9.157. That means the post-test scores were slightly more consistent than the pre-test scores. Also, the standard error decreased from 1.703 in the pre-test to 1.672 in the post-test, which is indicative of more stable post-test scores.

Table 3. Student Grade Percentage

Classification	Score	Pre-test		Post-test	
		Frequency	Percentage	Frequency	Percentage
Very Good	86-100	0	0%	1	3,33%
Good	71-85	0	0%	20	66,67%
Fair	56-70	7	23,33%	7	23,33%
Poor	41-55	20	66,67%	2	6,67%
Very Poor	<40	3	10,00%	0	0%

The table above indicated that when the pre-test results were compared, the large majority of the students were classified as poor. There were 20 students (66.67%) with scores between 41 and 55; seven students (23.33%) in the fair category with scores between 56 and 70; and three students (10.00%) in the very poor category. Following treatment, posttest performance of students in pronunciation improved. The vast majority shifted to the good category, 20 students (66.67%) scoring between 71 and 85. Seven students (23.33%) fell into the fair category; only two students (6.67%) stayed in the poor category. Moreover, one student (3.33%) came from the very good category, while none of the children had reached this level during the pre-test. The use of the Pronunciation with Emma channel enhanced students' pronunciation ability as revealed from these findings.

Table 4. Paired Sample T-Test

		Paired Differences							Significance	
					95% confidence interval of the difference					
		Mean	Std. deviation	Std. error mean	Lower	Upper	t	df	One side p	Two-sided p
Pair	pre- test post-test	-23.600	4.952	.904	-25.449	-21.751	-26.102	29	<.001	<.001

Based on the results of the t-test, and Table 4 shows, average difference from pre-test to post-test is = -23.600. The resulting t-value is -26.102, df = 29. Meanwhile, Sig., 2-tailed, significant value < 0.001, which is lower than the set significance level of 0.05. Hence the null hypothesis (H0) is rejected and alternative hypothesis (H1) is accepted. Moreover, according to 95% confidence interval, the difference has a range of -25.449 and -21.751, which means that the difference between pre-test and post-test is consistently significant and does not cross the zero

value. This further indicates that the two test results showed a considerable difference.

Table 5. Improvement Analysis Table

No	Diphthongs	Examples	Pre-test	Post-test	Increase
1	/eɪ/	Day, name, face, rain, game	433	597	+164
2	/aɪ/	My, time, light, night, five	503	632	+129
3	/ɔɪ/	Boy, toy, enjoy, noise, coin	547	647	+100
4	/aʊ/	Now, out, house, down, about	323	515	**+192**
5	/oʊ/	Go, home, no, show, boat	368	498	+130
6	/ɪə/	Ear, here, near, clear, idea	296	484	**+188**
7	/eə/	Hair, care, chair, fair, where	310	447	+137
8	/ʊə/	tour, pure, sure, cure secure	309	515	**+206**

As for the total scores produced for each of the diphthong sounds within the pre-test and post-test, it is evident that all diphthong sounds are improving after the treatment with videos from the Pronunciation with Emma channel. In the pre-test, that sound /ɔɪ/ received the most overall score, at 547, and the one with the lowest score – /ɪə/, at 296. This suggests the following statement; at baseline, students sound level performance on some diphthong sounds was low and it had not reached consistency across the different vowel sounds. The treatment showed a strong improvement in all diphthong sounds. On the post-test, though the highest score remained in /ɔɪ/, with 647, the maximum sound change occurred: +206 for /ʊə/ and +192 for /aʊ/ and +188 for /ɪə/, respectively. This means that the students had taken significant strides toward mastering sounds formerly deemed at a higher level of difficulty, especially in relation to sounds seldom used in casual conversation. The /ɔɪ/ sound also gained the least at +100, but represented improvements in comprehension. This could be due to the fact that students had improved quite less than the others in early sound, so it was not as remarkable as at earlier stages of sound development, where they remained at a very low level.

Discussion

The result of this study shows that the use of the Pronunciation with Emma YouTube channel helps significantly to improve the pronunciation of English diphthongs by the students. This change is indicated by an increase in the mean score from 50.33 in the pre-test to 73.93 in the post-test of the students. The paired sample t-test also revealed that the difference between the two tests was significant, p-value less than .001. Results indicated that students' pronunciation performance increased following four treatment sessions with support from the Pronunciation with Emma channel. Note (note: although it was pre-experimental, there was no control group and so the study was not dependent on a standardised test) that statistically, students showed a significant improvement post-intervention.

The improvement can be explained by exposure to clear audio-visual pronunciation models in small doses. In English diphthongs, an individual must make smooth movement from vowel position to another within the same syllable. But for Indonesian EFL learners, the result is to give them a tough time: Indonesian vowels are usually less changeable, with many vowel sounds spelling this pattern more consistently. Before this treatment, a lot of students still treated diphthongs like one single sound in its whole or even pronounced diphthongs according to

Indonesia's script. After treatment, a student is said to be more aware that English diphthongs need sound movement, not just one fixed sound that is produced by one vowel. This realization allowed them to say words like face, time, house, home, near, and sure more faithfully.

Pronunciation in the case of Emma is apparently utilized in helping students' pronunciation learning via two channels: auditory input and visual articulation support. Learners were invited to listen to repeated regular pronunciation models supported by auditory input, the visual input allowed students to pay attention not only to mouth movement or lips, but also to the transition between sounds. It matters because it's not enough just to get aural sounds, but also to be able to understand how the sounds are formed. Sutton and Webb (2026) note that audio-visual input might be important for second language learning on account of the fact it gives language and visual cues to those who learn to speak by themselves in a second language. In the current instance, this combination proved to be especially effective as students learned how to relate the sounds they heard to the articulation moves in the videos.

The findings also corroborate with a recent survey examining technology-mediated pronunciation learning. While this analysis did not utilize automatic or computer-aided speech recognition or corrective feedback, this study does support a broader suggestion that digital technologies can be beneficial in pronunciation learning when provided with the opportunity for systemized, repeated, and model-based practice. Amrate and Tsai (2025), in their systematic review on computer aided Pronunciation Training, point out that technology-based pronunciation instruction is effective if it addresses the specific pronunciation target and if learners have opportunities for repeated practice throughout the learning process. The target for this task of study focused on English diphthong sounds, so that students can concentrate on a single pronunciation feature rather than trying to change pronunciation at all.

This is additionally supported by the interpretation derived in this diphthong-related approach. The improvement of all eight diphthong sounds followed treatment. The top increase recorded in /ʊə/ was 206 points, and /aʊ/ has gone through 192 points with an increase, whereas /ɪə/ has a net of 188 points. According to their lower pre-test scores, these sounds were some of the most challenging items at beginning of study. This notable increase indicates that students most benefited from direct modeling and practice of certain sounds that were both difficult and new. Conversely /ɔɪ/ experienced the lowest increase, though it did improve. Perhaps students had a slightly better baseline on this diphthong, so there wasn't as much space for dramatic change. In that sense, this pattern of improvement indicates that YouTube-based pronunciation instruction proved particularly beneficial for diphthongs students more struggled with historically. That shift in students' score categorization also reflects a significant pedagogical shift. On the pre-test, the majority of students were poor and none showed up to the good or very good categories. Most students transitioned into the good category after the treatment and one student entered the very good category. This not only suggests statistical improvement, but also practical betterment in students' pronunciation performance. The improvement indicates that the intervention allowed students to transition from limited pronunciation control to more accurate English diphthong production. In classroom terms, this is key since pronunciation gets significant when students can produce sounds more clearly and with increased confidence during oral practice.

We expect that YouTube pronunciation instruction made a more flexible and less intimidating learning environment leading also to increased improvement. Students could hear the pronunciation model repeated multiple times, copy it and practice without relying on the teacher for direct correction. It could lead to reduced anxiety and potential for learners to be more eager to practice spoken English in such a climate. Bashori et al. (2021) demonstrated that

language learning supported by technology can improve language enjoyment in learners, and minimise speaking-related anxiety. While in this study there was no measurement of anxiety or enjoyment, classroom observation during the treatment indicated that students were more prone to be engaged and more interested in replicating the pronunciation patterns. This is suggestive of the potential contribution of accessible digital media for producing a more supportive environment for pronunciation.

The results are also consistent with those of Ngo et al. (2024) found in a meta-analysis that the implementation of technology-supported pronunciation learning promotes the pronunciation performance of ESL/EFL learners. They've only studied automatic speech recognition of speech data and this study utilized YouTube videos. However, both studies share the common pedagogical implication: learning pronunciation gets better if learners are given sufficient input, targeted practice, repeated exposure to models of pronunciation. Pronunciation with Emma represented a readily available pronunciation model that found its way to the classroom without the need for technology or expensive apps. Pronunciation with Emma can also be considered as a supplementary medium that can be leveraged for instruction of English diphthongs in a junior secondary level, based on the above results. The channel is particularly useful since it offers the easy and explicit explanation of pronunciation, practical examples, and the use of repetition to model. But it's not a substitute for the teacher, not really. Teachers still must choose suitable videos, direct a student's attention onto the target sounds, offer the opportunity for corrective feedback, and create follow-up practice. With lack of the teacher's instruction, students can passively watch these videos and do not learn proper pronunciation awareness. As such, the best use of YouTube in pronunciation development is not to be seen as an independent source of entertainment but rather as structured and informative educational resource incorporated into the learning process of teachers.

These results notwithstanding, there are several limitations that need to be noted. The first was that the study was one-group in its pre-test and post-test design and the improvement could not necessarily be ascribed purely to treatment at all. Other variables, including repeated testing, student familiarity with the word list or extra English exposure outside the classroom, may have contributed to the results. Second, the sample of the study was small based on just 30 students from a single class, therefore our findings cannot be generalized to junior secondary EFL learners in general. Third, the treatment took place in just four meetings, thus measuring a rapid improvement instead of sustained retention of pronunciation. Future studies need a control group, more samples, delayed post-tests, and more detailed examination of students' pronunciation mistakes before and after treatment.

The overall results show that the Pronunciation with Emma YouTube channel can provide students assistance in improving their pronunciation of English diphthongs as well. The substantial improvement of post-test scores, the advancement of the students to an upper order performance category and the positive gains of the diphthong sounds demonstrate the advantages of audio-visual pronunciation instruction for the junior higher educated EFL learners. These findings support the concept that an accessible computerized medium for language is valuable, provided the learning modality for teaching is concrete, the teaching medium is designed properly, and the teacher offers guidance in structure and sequence.

Conclusion

The study concludes that the use of the Pronunciation with Emma YouTube channel positively influenced students' English diphthong pronunciation. The mean score increases for student performances in the post-test compared with the pre-test also showed significant

difference as indicated by paired-sample t-test, suggesting that YouTube based pronunciation instruction can improve students' performance in producing English diphthong sounds. The difference in the performance of all eight diphthongs indicates that audio-visual input, native-speaker models, visual articulation, and continuous pronunciation development facilitated students more accurately comprehending and producing difficult English vowel glides.

From a pedagogical viewpoint, the findings suggest that pronunciation instruction should not rely exclusively on teacher explanation or textbook drills. The goal of pronunciation instructions must be concrete, where students get to hear, see, imitate and repeat target sounds in authentic ways. The Pronunciation with Emma channel provides this kind of support because it combines sound models, mouth movement demonstrations and accessible explanations. Hence English teachers could offer YouTube pronunciation videos that target certain pronunciation features difficult for Indonesian students, such as diphthongs, as a supplementary media during classroom instruction.

The results also imply that YouTube pronunciation learning can enhance more flexible and autonomous practice. Such study materials could also be reviewed outside the classroom by the students to be heard by the pronunciation practice, to be repeated in repetition in the pronunciation, with time leading to confidence with the pronunciation speech to develop in the second language. The teacher's role, however, is key. Teachers should choose suitable videos, direct students' focus to certain sounds, create controlled and communicative exercises, and offer correction suggestions so that learners do not just watch video but actively respond to that.

However, the present study was constrained by the fact it was a pre-experimental study with a small sample size and a shorter experimental period. Future studies should include control groups, larger samples, longer intervention duration, and delayed post-tests in order to explore the long-term outcomes of teaching pronunciation via YouTube. Pronunciation with Emma can therefore be seen as a practical and pedagogical option to facilitate English pronunciation learning in junior secondary EFL classrooms.

References

Aulia, V. (2019). English pronunciation practices: From tongue twisters to YouTube channel. *Script Journal*, *5*(1), 45-54. https://doi.org/10.24903/sj.v5i1.351

Amrate, M., & Tsai, P.-H. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL, 37*(1), 22–42. https://doi.org/10.1017/S0958344024000181

Al-Jarf, R. (2022). YouTube videos as a resource for self-regulated pronunciation practice in EFL distance learning environments. *Journal of English Language Teaching and Applied Linguistics, 4(2*), 44–52. https://doi.org/10.32996/jeltal.2022.4.2.4

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System, 99*, *Article 102496,* 1-16 https://doi.org/10.1016/j.system.2021.102496

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.

Fatimah, A. S., Fitria, I. N., Saputra, Y., Auliya, A. G., & Demir, K. (2024). Autonomous learning of English pronunciation through vlogs: Evidence from two Indonesian EFL students. *Journal of English Educators Society, 9*(2), 104-112. https://doi.org/10.21070/jees.v9i2.1824

Jailani, A. (2022). The use of YouTube videos in English language teaching: A perspective of Indonesian undergraduate students majoring in English language education. *J-SHMIC:*

Journal of English for Academic, 9(2), 13-24 . https://doi.org/10.25299/jshmic.2022.vol9(2).9949

Ma, Q., Mei, F., & Qian, B. (2026). Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. *Computer Assisted Language Learning*, *39*(4), 915–941. https://doi.org/10.1080/09588221.2024.2432965

Mukarrama, A., Kalsum, & Fajriansyah, S. (2025). Video-assisted instruction for pronunciation improvement: *An intervention study with Indonesian EFL students. Issues in Applied Linguistics & Language Teaching, 7(*1*),* 272-282 https://doi.org/10.37253/iallteach.v7i1.10417

Mora, J. C. (2025). *Phonological acquisition in instructed SLA: L2 pronunciation training in and beyond the classroom and individual differences*. *Language Teaching*, 1–16. https://doi.org/10.1017/S026144482510102X

Kabooha, R., & Elyas, T. (2018). The effects of YouTube in multimedia instruction for vocabulary learning: Perceptions of EFL students and teachers. *English Language Teaching*, *11*(2), 72–81. https://doi.org/10.5539/elt.v11n2p72

Kelly, G. (2000). *How to Teach Pronunciation*. Harlow, England: Longman.

Kusuma, S., Mardiana, W., & Iklil Saifulloh, A. (2024). Integrating Digital Media in Language Education: A Study on YouTube's Effectiveness in Pronunciation Instruction. *Fonologi: Jurnal Ilmuan Bahasa Dan Sastra Inggris, 2*(2*), 271–276.* https://doi.org/10.61132/fonologi.v2i2.972

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

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Salsabila, D. A., Salam, U., & Rezeki, Y. S. (2025). YouTube as a tool for English language learning: A meta-analysis in EFL contexts. English Education: Journal of English Teaching and Research, *10*(1).12-29. https://doi.org/10.29407/jetar.v10i1.23641

Simbolon, S. R., & Ismahani, S. (2024). Developing English proficiency through qualitative study of "English With Lucy" YouTube channel. *The Journal of Educational Development, 12*(2).73-83 https://doi.org/10.15294/jed.v12i2.4207

Sutton, D., & Webb, S. (2026). The effects of audiovisual input on second language learning: A meta-analysis. *Studies in Second Language Acquisition*, 1–25. https://doi.org/10.1017/S0272263126101612

Suwastini, N. K. A., Rastari, K. L., & Suarcaya, P. (2023). YouTube as instructional media in promoting EFL Indonesian students' pronunciation. *Polyglot: Jurnal Ilmiah, 19*(1*)*. 1-24. https://doi.org/10.19166/pji.v1i19.6309

Tahmina, T. (2023). Students' perception of the use of YouTube in English language learning. *Journal of Languages and Language Teaching, 11*(1). 151-159 https://doi.org/10.33394/jollt.v11i1.6883

Veronica G. Sardegna. (2021). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching, 55*(3), 1–5. https://doi.org/10.1017/S0261444821000380