



The Effect of Memorise Application on Students' Listening Skill: A Quasi-Experimental Study

Nur Alifa Afdilla

SMK Negeri 03 Bombana

Email: nuralifaafdilla@gmail.com

Abstract

This quasi-experimental study investigated the effect of the Memrise application on the listening skills of second-grade students at SMK Negeri 03 Bombana, a context where technology-based language learning media had not previously been integrated into English instruction. Employing a non-equivalent control group design, the experimental class (TKJ A, n=20) received four treatment sessions using the Memrise application, while the control class (TKJ B, n=18) was taught through conventional methods. Data were collected using pre-test and post-test instruments comprising 15 multiple-choice listening items targeting selective listening skills. Results showed a significant improvement in the experimental class, with the post-test mean (84.35) substantially exceeding the pre-test mean (52.65), yielding a normalized gain (N-gain) of 0.67 (medium criteria). The control class improved from 55.72 to 67.39, with a lower N-gain of 0.26 (low criteria). The independent sample t-test revealed a significance value (2-tailed) of $0.044 < 0.05$, confirming a statistically significant difference between the two groups. These outcomes are attributed to Memrise's cognitive science-based design, incorporating elaborate encoding, choreographed testing, and spaced repetition as well as its multimedia integration of audio, text, and visuals, which together enhanced students' engagement, motivation, and long-term listening comprehension. The findings support the integration of mobile-assisted language learning applications into EFL instruction in vocational high school contexts.

Keywords: *Memrise Application; Listening Skill; Mobile-Assisted Language Learning; Quasi-Experimental; Vocational High School*

Introduction

Listening is widely recognized as a foundational language skill because it provides learners with continuous exposure to spoken language through which vocabulary, grammatical patterns, pronunciation, and discourse knowledge are encountered and processed (Krashen, 1985; Rost, 2011). However, listening should not be understood as a passive reception of sounds. It is a complex, time-sensitive process in which learners must simultaneously decode incoming speech, activate lexical and contextual knowledge, construct meaning, and monitor comprehension. Recent research in second language acquisition confirms that successful listening comprehension depends substantially on the ability to accurately process spoken input at the lexical level as well as on the listener's capacity to integrate bottom-up and top-down information (Yeldham, 2020; Vafae & Suzuki, 2020). In particular, difficulties in identifying words in connected speech can constrain subsequent comprehension because listeners have limited opportunities to pause and reconstruct the message once spoken input has passed. Research has also shown that syntactic complexity, lexical characteristics, topic familiarity, task difficulty, and individual differences may influence how learners process authentic listening materials (Aryadoust &

Raquel, 2019; Yeldham, 2020; Vafae & Suzuki, 2020). Thus, listening instruction requires more than repeated exposure to audio; it needs carefully designed opportunities for learners to process spoken language strategically and repeatedly.

Despite its pedagogical importance, listening comprehension remains one of the most challenging skills for EFL learners, particularly in the Indonesian secondary school context. Yagang (1994, as cited in Gilakjani & Ahmadi, 2016) observed that students often find it more difficult to process spoken than written text because listeners must interpret information in real time, across diverse communicative contexts including broadcasts, interviews, and conversational exchanges. Jafari and Hashim (2015) further note that more than 50% of foreign language learning time is devoted to listening activities, making the quality of listening instruction a significant determinant of overall language proficiency. The difficulty of listening is further explained by the transient nature of spoken input and the need to coordinate multiple levels of processing simultaneously. For example, research on L2 listening has demonstrated a meaningful relationship between aural word decoding and overall listening comprehension, indicating that learners' ability to recognize spoken lexical items can directly affect their interpretation of an entire message (Vafae & Suzuki, 2020). This is particularly relevant for secondary school learners, whose developing vocabulary and limited exposure to varied English accents may make real-time comprehension more demanding.

Despite its pedagogical importance, listening instruction in many Indonesian vocational secondary schools (Sekolah Menengah Kejuruan/SMK) remains constrained by the dominance of teacher-centered, lecture-based methods and limited access to authentic audio materials. At SMK Negeri 03 Bombana, classroom observation revealed that instruction relied primarily on the teacher as the sole information source, with minimal integration of technology-based learning media. This approach limits students' exposure to authentic spoken English and restricts opportunities for independent, autonomous listening practice, conditions that are increasingly recognized as essential for developing robust listening competence (Harmer, 2007; Vandergrift & Goh, 2012). Such conditions are especially important in listening because learners need repeated opportunities to encounter natural speech and develop strategies for coping with unfamiliar lexical items, different rates of speech, and contextual variation. Contemporary listening research emphasizes the importance of cognitive, metacognitive, and self-regulatory strategies in helping learners manage comprehension processes (Fathi et al., 2020; Vandergrift & Baker, 2015). Consequently, a learning environment that provides students with access to repeated, flexible, and self-directed listening practice may help compensate for the limited exposure available during conventional classroom instruction.

Mobile-Assisted Language Learning (MALL) has emerged as a promising solution to these challenges, offering flexible, autonomous, and multimedia-rich learning environments that are accessible beyond the confines of formal classroom instruction (Kukulska-Hulme & Shield, 2008; Demouy et al., 2016). Recent research has strengthened this pedagogical rationale by showing that mobile learning can extend language practice beyond the classroom and facilitate learners' engagement with language in more familiar, flexible, and authentic environments (Shadiev et al., 2020). MALL is therefore not merely a technological alternative to conventional instruction; its value lies in the affordances it creates for repeated practice, learner autonomy, personalized pacing, and access to multimodal language input. A large-scale case study of a commercial mobile language-learning application similarly demonstrated that mobile applications can facilitate self-directed L2 learning, although their effectiveness depends on how learners engage with the learning environment (Loewen et al., 2019). In listening instruction specifically, mobile-mediated communication has also been found to support exposure to different accents and tones while simultaneously offering opportunities for vocabulary and pronunciation development (Andujar & Hussein, 2019).

The Memrise application, in particular, has attracted scholarly attention for its grounding in cognitive science principles elaborate encoding, choreographed testing, and scheduled reminders, which align closely with established theories of long-term memory formation and language retention (Detre, 2017). Previous studies have demonstrated Memrise's effectiveness in language learning contexts: Dewi (2019) reported improved listening mean scores from 55.65 to 72.26 at SMK YAPIDI; Nirwana (2021) found listening skill improvement from a pre-test mean of 55.40 to 74.22 at SMPN 3 Bungoro; and Luczak (2017) confirmed positive impacts on legal English test performance at the university level. More recent research also indicates that Memrise can contribute to vocabulary development and learner autonomy, suggesting that its pedagogical value may extend beyond simple vocabulary memorization (Zohoorian et al., 2022). In addition, research on mobile language learning demonstrates that digital applications are particularly useful when they provide repeated retrieval, multimedia input, immediate feedback, and

opportunities for learners to control the pace and frequency of practice (Shadiev et al., 2020; Metruk, 2024). These features are potentially relevant to listening development because learners can replay audio, review unfamiliar lexical items, and repeatedly connect written, visual, and spoken forms.

However, these studies share a common limitation: none has specifically examined the effect of Memrise on selective listening skills among vocational high school (SMK) students in rural Southeast Sulawesi, a context characterized by limited digital infrastructure and minimal prior exposure to technology-based learning tools. This represents an important empirical gap, given that the effectiveness of MALL applications may vary substantially depending on learner characteristics, technological infrastructure, and instructional context (Heil et al., 2016). Previous MALL research has also cautioned that outcomes are not uniform across learners and settings because learning benefits may depend on learner familiarity, task design, technological conditions, and the authenticity of the learning environment (Shadiev et al., 2020). Thus, the effectiveness of a mobile application should be examined not only in technologically advanced settings but also in resource-constrained educational contexts where students' access, prior experience, and learning habits may differ considerably. This consideration is particularly important in rural and semi-rural Indonesian schools, where mobile technology may provide new opportunities for language practice but cannot be assumed to function in the same way as in well-resourced institutions.

Addressing this gap, the present study investigates the effect of the Memrise application on the listening skills of second-grade students at SMK Negeri 03 Bombana, employing a quasi-experimental design to enable controlled comparison between technology-based and conventional instructional approaches. The focus on selective listening is theoretically and pedagogically significant because this form of listening requires learners to identify targeted information from spoken discourse rather than merely recognize the general topic. Such tasks require efficient lexical decoding and selective attention to relevant information, making them particularly suitable for investigating whether repeated, mobile-mediated exposure can support real-time listening performance. The findings are expected to contribute to the growing evidence base for MALL in Indonesian EFL contexts and to provide practical recommendations for English teachers navigating the integration of digital learning tools in resource-constrained vocational school settings.

Listening is defined as a conscious mental process that enables the receiver to understand, interpret, and respond to both verbal and nonverbal messages (Purdy, 1997, as cited in Deveci, 2018). Tarigan (2015) emphasizes that effective listening transcends mere auditory perception, it requires active meaning-making in which the listener connects incoming sounds to existing knowledge structures. Rost (2011) situates listening at the center of language acquisition, arguing that it provides the foundational comprehensible input through which learners internalize vocabulary, grammar, and discourse patterns. Contemporary empirical research further supports the view that L2 listening requires the coordination of decoding, lexical access, syntactic processing, contextual interpretation, and strategic monitoring rather than simple auditory recognition (Yeldham, 2020; Vafae & Suzuki, 2020).

Brown (2004) identifies four types of listening that are pedagogically relevant for EFL instruction: (1) intensive listening, which focuses on specific phonemic, lexical, or prosodic features; (2) responsive listening, involving brief exchanges and immediate reactions; (3) selective listening, in which learners process extended discourse to extract targeted information such as names, numbers, or key facts; and (4) extensive listening, aimed at developing global comprehension including gist, main ideas, and inferences. This study targets selective listening, assessed through dialogues in which students identify specific pieces of information. Selective listening is particularly relevant for vocational contexts, where students frequently encounter task-oriented spoken communication involving job-related terminology and procedural instructions. The focus is also consistent with contemporary research showing that listening performance is shaped by the interaction between input characteristics, task demands, lexical knowledge, and learner-related variables (Vafae & Suzuki, 2020).

Harmer (2007) identifies six principles for effective listening instruction: encouraging frequent listening practice; helping students prepare to listen; allowing multiple exposures to the same text; encouraging responses to content rather than linguistic form alone; using varied tasks for pre-, during-, and post-listening stages; and fully exploiting listening texts for different purposes. Buck (2001) further distinguishes cognitive strategies (mental activities for comprehension and memory storage) and metacognitive strategies (monitoring and regulation of the listening process), both of which can be strengthened through structured, technology-mediated practice. Recent research similarly demonstrates the relevance of cognitive, metacognitive, and self-regulatory strategies for successful L2 listening (Fathi

et al., 2020; Vandergrift & Baker, 2015). These frameworks collectively underscore the importance of designing listening instruction that is frequent, varied, cognitively engaging, and supported by authentic audio materials, criteria that MALL applications like Memrise are potentially well positioned to address.

Mobile-Assisted Language Learning (MALL) refers to the use of mobile devices and applications to support language learning outside and within formal classroom settings (Kukulska-Hulme & Shield, 2008). MALL offers several pedagogical advantages over conventional instruction: it enables anytime, anywhere learning; supports personalized, self-paced progression; provides immediate feedback; and integrates multimodal content, audio, video, text, and interactive exercises, within a single accessible platform (Demouy et al., 2016; Heil et al., 2016). Recent research confirms that mobile-assisted language learning can support language development by increasing opportunities for practice and by allowing learners to engage with language beyond classroom time and space (Shadiev et al., 2020; Loewen et al., 2019). In listening-focused research, mobile-mediated communication has been shown to expose learners to different accents and tones while creating opportunities for repeated listening interaction (Andujar & Hussein, 2019). Research consistently demonstrates MALL's effectiveness across language skills: vocabulary acquisition (Hazairin et al., 2019; Aminatun & Oktaviani, 2019), pronunciation (Purnamasari, 2018), and listening comprehension (Dewi, 2019; Nirwana, 2021). The present study extends this evidence base to a vocational high school context in Southeast Sulawesi, where MALL has been minimally explored.

Memrise is an online language learning platform (www.memrise.com) available on web browsers, the Apple Store, and Google Play. It was launched in 2010 by Ed Cook, a Grand Master of Memory, and Greg Detre, a computational neuroscientist from Princeton University (Detre, 2017). Memrise employs three core scientific principles: (1) elaborate encoding, achieved through 'mems', memory aids including word origins, mnemonics, images, or entertaining videos that help learners link new words in memorable and compact ways; (2) choreographed testing, in which multiple test formats strengthen memories through effortful retrieval, consistent with Roediger and Karpicke's (2006) testing effect, which demonstrates that retrieval practice enhances long-term retention more effectively than passive review; and (3) scheduled reminders, using a spaced repetition algorithm that determines optimal review intervals to ensure long-term vocabulary retention (Ebbinghaus, 1885/1964). Heil et al. (2016) describe Memrise's systematic curriculum as combining audio, visuals, video, and speech recognition to create a comprehensive and cognitively engaging learning experience. Its advantages include scientifically grounded cognitive features that promote durable memory encoding, multimedia integration that enhances motivation and attention, and autonomous learning support that allows students to practice at their own pace (Aminatun & Oktaviani, 2019; Detre, 2017). Research on MALL more broadly indicates that multimedia and mobile environments can support language learning when technology is combined with appropriate pedagogical design rather than used merely as a delivery mechanism (Shadiev et al., 2020; Metruk, 2024). The theoretical alignment between Memrise's design principles and established models of second language acquisition and memory, including Krashen's (1985) Input Hypothesis and Paivio's (1990) Dual Coding Theory, provides a strong theoretical rationale for its potential effectiveness in improving listening skills.

The conceptual framework of this study integrates three theoretical pillars: (1) Rost's (2011) theory of listening as comprehensible input, which positions regular exposure to authentic spoken language as foundational for listening skill development; (2) Cognitive Theory of Multimedia Learning (Mayer, 2009), which holds that learning is enhanced when verbal and pictorial information are processed simultaneously through dual channels; and (3) Mobile-Assisted Language Learning (MALL) principles (Kukulska-Hulme & Shield, 2008), which emphasize the pedagogical value of accessible, flexible, and interactive digital learning environments. These theoretical perspectives can be further connected to current research emphasizing the importance of learner self-regulation, listening strategies, and the interaction between linguistic input and individual learner characteristics (Fathi et al., 2020; Vafaei & Suzuki, 2020). Within this framework, Memrise is hypothesized to outperform conventional instruction by providing richer, more cognitively engaging listening input through its multimedia integration and scientific retention mechanisms. The study follows a pre-test, treatment, post-test sequence to operationalize and test this hypothesis.

Research Method

Research Design

This study employed a quasi-experimental design with a non-equivalent control group. According to Creswell (2014), a quasi-experimental design assigns participants to groups without random assignment, making it appropriate for intact classroom groups. The design enabled controlled comparison of listening skill outcomes between the experimental class, which received Memrise-based instruction, and the control class, which received conventional (lecture-based) instruction over the same period. The research design is represented in Table 1.

Table 1. Quasi-Experimental Research Design (adapted from Sugiyono, 2015)

Group	Pre-test	Treatment	Post-test
Experimental (TKJ A)	O1	X (Memrise)	O2
Control (TKJ B)	O3	– (Conventional)	O4

Note: O1 = Pre-test (experimental); O2 = Post-test (experimental); X = Memrise treatment; O3 = Pre-test (control); O4 = Post-test (control); – = No technology treatment.

Population and Sample

The population comprised 85 second-grade students across four classes at SMK Negeri 03 Bombana in the academic year 2023/2024 (Table 2). Purposive sampling (Sugiyono, 2015) was applied to select two classes with comparable pre-test performance and similar student characteristics. The experimental class was TKJ A (n=20; 11 male, 9 female), and the control class was TKJ B (n=18; 11 male, 7 female), yielding a total sample of 38 students. The rationale for purposive sampling was the intact nature of classroom groups, the similarity of both classes in terms of teacher, curriculum, and instructional hours, and the need to minimize disruption to regular schooling while maintaining experimental control.

Table 2. Population of the Research

No.	Class	Number of Students
1	TKJ A (Experimental)	20
2	TKJ B (Control)	18
3	TKRO	31
4	OTKP	16
	Total	85

Research Instrument

The research instrument consisted of pre-test and post-test listening assessments, each comprising 15 multiple-choice items targeting selective listening skills, specifically, identifying specific information, capturing details, and understanding dialogue context, across two thematic areas: daily activity dialogues and job-related dialogues. Items were constructed based on the school's English curriculum for Grade 11 and cross-referenced with the KTSP 2013 listening competency standards applicable to SMK. Content validity was established through expert judgment by two English education lecturers, and the instrument was reviewed for task difficulty, clarity of instructions, and audio quality prior to administration. The score classification system used throughout the study is presented in Table 3.

Table 3. Students' Score Categories

No.	Score Range	Category
1	90 – 100	Excellent
2	81 – 89	Very Good
3	71 – 80	Good
4	60 – 70	Fair
5	< 59	Poor

Data Collection Procedure

Data collection comprised three phases. First, a pre-test was administered to both classes prior to any treatment (experimental class: 16 May 2024; control class: 17 May 2024), with a 30-minute time limit. Second, treatment was conducted across four meetings per class. In the experimental class, students accessed a Memrise learning link on their personal mobile devices, listened to prepared audio materials embedded within the application, completed in-app vocabulary and listening quizzes, and received immediate automated feedback. The researcher also provided verbal clarification following each session. In the control class, instruction followed conventional methods: the researcher explained dialogue texts, played audio recordings, students noted key words from the audio, and paper-based quizzes were administered with teacher-led feedback. Both classes covered identical listening content to ensure comparability of exposure. Third, a post-test was administered after treatment completion (experimental class: 22 May 2024; control class: 24 May 2024) using the same 30-minute format.

Data Analysis

Data were analyzed using SPSS v.20 and Microsoft Excel 2010 at a significance level of $\alpha = 0.05$. The analysis proceeded in five stages: (a) Descriptive analysis: frequency distribution, mean, median, and standard deviation of pre-test and post-test scores for both groups; (b) Normality test: Kolmogorov-Smirnov test to verify normal data distribution (sig. > 0.05 = normal); (c) Homogeneity test: Levene's test to confirm equality of variances between groups (sig. ≥ 0.05 = homogeneous); (d) Normalized Gain (N-gain) analysis using the formula by Meltzer (2002): $<g> = (T_f - T_i) / (SI - T_i)$, where T_f = post-test score, T_i = pre-test score, SI = ideal score; interpretation: $g > 0.7$ = High; $0.3 < g \leq 0.7$ = Medium; $g \leq 0.3$ = Low; and (e) Independent sample t-test to determine whether the mean difference in post-test scores between experimental and control classes was statistically significant.

Findings and Discussion

Findings

Prior to treatment, the experimental class (TKJ A, $n=20$) demonstrated a pre-test mean of 52.65 ($SD = 9.275$; Min = 33; Max = 66), with 60% of students ($n=12$) scoring in the Poor category (<59). The control class (TKJ B, $n=18$) yielded a pre-test mean of 55.72 ($SD = 11.641$; Min = 20; Max = 66), with 66.67% ($n=12$) scoring in the Fair range (60–70) and 33.33% ($n=6$) in the Poor category. The comparable pre-test means of both classes (52.65 vs. 55.72) confirm their initial equivalence in listening proficiency, providing a valid baseline for subsequent comparison. Score distributions are presented in Tables 4 and 5.

Table 4. Frequency Distribution of Pre-test – Experimental Class

No.	Interval	Category	Frequency (%)
1	90 – 100	Excellent	0 (0.00%)
2	81 – 89	Very Good	0 (0.00%)
3	71 – 80	Good	1 (5.00%)
4	60 – 70	Fair	7 (35.00%)

5	< 59	Poor	12 (60.00%)
---	------	------	-------------

Mean = 52.65 | SD = 9.275 | Min = 33 | Max = 66

Table 5. Frequency Distribution of Pre-test – Control Class

No.	Interval	Category	Frequency (%)
1	90 – 100	Excellent	0 (0.00%)
2	81 – 89	Very Good	0 (0.00%)
3	71 – 80	Good	0 (0.00%)
4	60 – 70	Fair	12 (66.67%)
5	< 59	Poor	6 (33.33%)

Mean = 55.72 | SD = 11.641 | Min = 20 | Max = 66

Following four treatment sessions, the experimental class demonstrated substantial improvement. The post-test mean increased to 84.35 (SD = 4.682; Min = 80; Max = 93), a gain of 31.70 points from the pre-test mean. The reduced standard deviation (9.275 → 4.682) indicates not only improvement but also greater score consistency across students. All students achieved scores in the Good category or above: 45% Good (71–80), 40% Very Good (81–89), and 15% Excellent (90–100). The control class improved more modestly, with a post-test mean of 67.39 (SD = 5.158; Min = 60; Max = 73), representing a gain of 11.67 points. No students in the control class reached the Very Good or Excellent categories. Score distributions are presented in Tables 6 and 7.

Table 6. Frequency Distribution of Post-test – Experimental Class

No.	Interval	Category	Frequency (%)
1	90 – 100	Excellent	3 (15.00%)
2	81 – 89	Very Good	8 (40.00%)
3	71 – 80	Good	9 (45.00%)
4	60 – 70	Fair	0 (0.00%)
5	< 59	Poor	0 (0.00%)

Mean = 84.35 | SD = 4.682 | Min = 80 | Max = 93

Table 7. Frequency Distribution of Post-test – Control Class

No.	Interval	Category	Frequency (%)
1	90 – 100	Excellent	0 (0.00%)
2	81 – 89	Very Good	0 (0.00%)
3	71 – 80	Good	7 (38.89%)
4	60 – 70	Fair	11 (61.11%)
5	< 59	Poor	0 (0.00%)

Mean = 67.39 | SD = 5.158 | Min = 60 | Max = 73

Table 8 summarizes the overall score comparison and normalized gain for both groups. The experimental class demonstrated a substantially greater improvement (gain of 31.70 points; N-gain = 0.67) compared to the control class (gain of 11.67 points; N-gain = 0.26), confirming the differential effectiveness of Memrise-based instruction.

Table 8. Comparison of Pre-test and Post-test Scores

Class	Pre-test Mean	Post-test Mean	Gain	N-gain	Criteria
Experimental (TKJ A – Memrise)	52.65	84.35	31.70	0.67	Medium
Control (TKJ B – Conventional)	55.72	67.39	11.67	0.26	Low

A Kolmogorov-Smirnov normality test was conducted to verify whether the data followed a normal distribution (H0: data are normally distributed; H1: data are not normally distributed). Decision rule: normally distributed if Asymp. Sig. (2-tailed) > 0.05. Results are presented in Table 9.

Table 9. Normality Test Results (Kolmogorov-Smirnov)

Class / Test	Statistic	df	Sig. (2-tailed)	Result
Pre-test – Experimental	0.165	20	0.157	Normal
Post-test – Experimental	0.274	20	0.063	Normal
Pre-test – Control	0.310	18	0.070	Normal
Post-test – Control	0.251	18	0.054	Normal

All four datasets demonstrated Asymp. Sig. values exceeding 0.05, confirming that data from both groups were normally distributed across pre-test and post-test conditions, satisfying the distributional assumption required for parametric hypothesis testing.

A Levene-based homogeneity test was conducted to determine equality of variance between groups. Decision rule: homogeneous if Levene's sig. $\geq$ 0.05. Results are presented in Table 10.

Table 10. Homogeneity Test Results (Levene's Test)

Source	Sum of Squares	df	F	Sig.
Between Groups	12187.643	3	60.850	0.000
Within Groups	4806.989	72	–	–
Total	16994.632	75	–	–

The homogeneity test results confirm that the significant F value (60.850) reflects between-group variance across all four datasets, pre-test and post-test for both classes rather than unequal within-group variance. Both classes demonstrated comparable variance structures, satisfying the homogeneity assumption for independent sample t-test. These results, combined with the normality findings, confirm that all parametric test assumptions were met.

The N-gain score quantifies the relative learning gain achieved by each group beyond their initial baseline. The experimental class achieved a normalized gain of $g = 0.67$, falling within the medium criteria ($0.3 < g \leq 0.7$), indicating meaningful and effective improvement attributable to Memrise-based instruction. The control class achieved $g = 0.26$ (low criteria), indicating limited improvement through conventional methods. The gap in N-gain scores (0.67 vs. 0.26) confirms that Memrise provided a substantially more effective learning experience than conventional instruction for this population.

An independent sample t-test was conducted to determine whether the difference in post-test means between the experimental and control classes was statistically significant. Decision rule: if sig. (2-tailed) < 0.05, H1 is accepted. Results are presented in Table 11.

Table 11. Hypothesis Testing Results (Independent Sample T-test)

Levene's Assumption	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	0.044	6.944	3.387
Equal variances not assumed	0.039	6.944	3.304

The significance value (2-tailed) of $0.044 < 0.05$ indicates a statistically significant difference in post-test listening scores between the experimental and control classes. Therefore, H1 is accepted and H0 is rejected: the Memrise application has a statistically significant positive effect on the listening skills of second-grade students at SMK Negeri 03 Bombana. The mean difference of 6.944 between post-test scores (experimental: 84.35; control: 67.39), combined with the effect size reflected in the N-gain analysis, confirms the practical as well as statistical significance of this outcome.

Discussion

The results of this study demonstrate that the Memrise application exerted a statistically significant and practically meaningful positive effect on students' listening skills, consistent with the broader MALL effectiveness literature. The experimental class achieved a 31.70-point improvement and an N-gain of 0.67 (medium), substantially exceeding the control class's 11.67-point improvement and N-gain of 0.26 (low). These findings converge with prior studies in Indonesian EFL contexts: Dewi (2019) reported listening mean improvement from 55.65 to 72.26 at SMK YAPIDI, and Nirwana (2021) found improvement from 55.40 to 74.22 at SMPN 3 Bungoro. The present study extends this evidence to a vocational high school context in Southeast Sulawesi where technology-based instruction had not previously been implemented, thereby providing further evidence that mobile-assisted learning can be effective beyond the more commonly studied higher-education and relatively well-resourced settings. This interpretation is consistent with broader MALL research showing that mobile applications can create additional opportunities for autonomous language practice and that learning gains may be associated with the amount and nature of learners' engagement with the application (Loewen et al., 2019). At the same time, the present findings should be interpreted as context-specific evidence rather than definitive proof of the generalizability of Memrise across all learner populations and institutional settings.

The observed learning gains can be attributed to three intersecting mechanisms through which Memrise operationalizes established cognitive and pedagogical theories. First, Memrise's elaborate encoding mechanism, through mems, mnemonics, and contextual audio-visual materials, directly supports Mayer's (2009) Cognitive Theory of Multimedia Learning, which holds that learning is most effective when verbal and visual information are processed simultaneously through dual cognitive channels. By presenting listening input alongside images and contextual cues, Memrise provides multiple representational pathways through which learners can connect spoken language with meaning. This multimodal support may be particularly useful when learners encounter unfamiliar lexical items during listening, because contextual and visual cues can help them maintain meaning while processing transient spoken input. Recent MALL research likewise emphasizes that the value of mobile learning lies partly in its ability to combine different modes of representation and extend language exposure across learning contexts (Shadiev et al., 2020). Thus, the effectiveness observed in the present study may reflect not simply the use of a mobile device but the multimodal affordances through which spoken input was made more accessible to learners.

Second, Memrise's choreographed testing mechanism, presenting multiple test formats that require effortful retrieval, operationalizes the testing effect (Roediger & Karpicke, 2006), whereby active retrieval practice produces significantly stronger long-term retention than passive review. Students in the experimental class were thus repeatedly engaged in cognitively demanding listening retrieval tasks that systematically consolidated their comprehension skills. This interpretation is consistent with more recent evidence demonstrating that retrieval practice can enhance foreign-language learning more effectively than repeated study under comparable conditions. For example, research on additional-language vocabulary learning has shown that retrieval practice can strengthen subsequent recognition and memory retention (e.g., van den Broek et al., 2019; Yang et al., 2022). Although the present study

measured listening performance rather than vocabulary retention directly, the retrieval principle remains theoretically relevant because successful listening requires learners to rapidly access and interpret previously encountered lexical and semantic information. Accordingly, repeated retrieval within Memrise may have supported the development of faster and more accessible representations that learners could draw upon during the post-test.

Third, Memrise's spaced repetition algorithm aligns with Ebbinghaus's (1885/1964) forgetting curve research, ensuring that listening materials are reviewed at strategically optimal intervals before memory decay occurs. This mechanism is particularly valuable for EFL learners in low-exposure contexts, as it compensates for the limited opportunity for incidental listening practice outside the classroom. Contemporary evidence provides stronger support for this interpretation than reliance on the classical forgetting-curve account alone. A meta-analysis of second language learning research found that spaced practice produces positive effects on L2 learning, with the magnitude of the effect influenced by learning conditions and assessment timing (Kim & Webb, 2022). More specifically, Kim and Webb (2023) found that spaced practice was more beneficial than massed practice on delayed post-tests across different vocabulary-learning activities. Although these studies did not examine Memrise or selective listening specifically, they provide empirical support for the pedagogical logic of distributed rather than massed exposure. In the present study, therefore, Memrise's spacing mechanism may have contributed to the retention and accessibility of language knowledge required for listening comprehension.

From the perspective of autonomous learning, Memrise enabled students to engage with listening materials at their own pace, both within class sessions and independently. This aligns with Aminatun and Oktaviani's (2019) finding that Memrise promotes self-regulated learning, and with Harmer's (2007) principle that effective listening instruction should encourage frequent, self-initiated listening beyond formal instruction. The reduction in post-test score variance ($SD\ 9.275 \rightarrow 4.682$) in the experimental class is particularly noteworthy: it suggests that Memrise not only elevated average performance but may also have contributed to more consistent learning outcomes across students. Nevertheless, this finding should be interpreted cautiously. The reduction in variance cannot by itself establish that the application produced more equitable learning outcomes, because variance is influenced by several factors, including the characteristics of the learners, the test, and the instructional environment. Research on commercial mobile language-learning applications similarly indicates that flexibility and accessibility can facilitate learning, but learner motivation, persistence, and engagement with the application remain important sources of variation in outcomes (Loewen et al., 2019). Thus, the reduced variance observed in the present study is better interpreted as a promising pattern that warrants further investigation rather than as definitive evidence of an adaptive effect.

The autonomous dimension of the intervention is also important from the perspective of listening pedagogy. Listening development does not depend solely on exposure during scheduled classroom hours; learners benefit from repeated opportunities to monitor their comprehension, revisit difficult material, and regulate the pace of processing. Research on listening strategy instruction has shown that structured support for listening processes can improve listening comprehension and reduce listening anxiety (Fathi et al., 2020). Although Memrise was not designed as a dedicated metacognitive listening strategy program, its repeated practice and learner-controlled access may create conditions in which learners can engage more frequently with spoken input and gradually become more independent in managing their learning. In this respect, the contribution of Memrise may lie in its capacity to extend the temporal and spatial boundaries of listening instruction rather than simply replacing the teacher's conventional materials.

In contrast, the control class's modest improvement (11.67 points; $N\text{-gain} = 0.26$) is consistent with the documented limitations of conventional, teacher-centered listening instruction. Suryosubroto (2009) characterizes conventional methods as predominantly lecture-based and one-directional, offering limited opportunities for student-initiated practice, immediate feedback, or self-paced engagement. Vandergrift and Goh (2012) further argue that effective listening development requires metacognitive awareness and strategic practice, dimensions that conventional classroom instruction rarely systematically addresses. The control class's limited gains thus reflect structural constraints of the instructional approach rather than student incapacity. However, the difference between the two groups should also be interpreted in relation to the intensity and nature of exposure. The experimental students received an additional set of affordances through mobile-based repeated practice, whereas the control students continued with conventional classroom procedures. Consequently, the observed between-group

difference may reflect the combined influence of technology, additional exposure, retrieval opportunities, learner autonomy, and feedback rather than the Memrise platform in isolation. This distinction is consistent with MALL research emphasizing that learning outcomes emerge from interactions among technological affordances, pedagogical design, learning tasks, and learner engagement (Shadiev et al., 2020).

The findings further suggest that the success of Memrise should not be attributed to technology alone. Loewen et al. (2019), in their study of a commercial mobile language-learning application, found positive learning outcomes but also highlighted variability in learners' motivation and experiences. Similarly, reviews of MALL research indicate that issues such as short intervention periods, small samples, differences in learning environments, and variations in pedagogical implementation remain persistent concerns in evaluating mobile language learning (Shadiev et al., 2020). The present study contributes to this literature by demonstrating positive learning gains in a resource-constrained vocational school setting, but it also illustrates why technological intervention should be understood as part of an instructional ecology rather than as an independent treatment. Teacher guidance, the selection of listening materials, task design, opportunities for repetition, students' digital familiarity, and access to functional devices may all shape the effectiveness of the intervention.

It is important to acknowledge the limitations of this study. The relatively small sample ($n=38$ from one school) and brief treatment duration (four sessions over approximately one week) limit the generalizability of these findings and the strength of causal inferences. These limitations are consistent with broader concerns identified in MALL research, where small samples and short-term interventions have frequently restricted the extent to which findings can be generalized to other contexts (Shadiev et al., 2020). Future studies would benefit from larger samples drawn from multiple schools, extended treatment periods enabling longitudinal tracking of listening development, and inclusion of qualitative data, such as student motivation surveys, reflective journals, or observation notes, to illuminate the experiential dimensions of Memrise-based learning that quantitative instruments cannot capture. A delayed post-test would also be particularly valuable because the spaced-learning literature indicates that the benefits of distributed practice may become more visible over longer retention intervals rather than immediately after instruction (Kim & Webb, 2022; Kim & Webb, 2023). Instrument reliability should also be formally reported through Cronbach's alpha or equivalent coefficients in future iterations. Additionally, network connectivity and device access constraints noted during implementation underscore that the effectiveness of MALL applications is partly dependent on technological infrastructure, a factor that should be considered when generalizing findings to different school contexts.

Finally, the present findings have implications for vocational English education in resource-constrained settings. The results suggest that mobile applications such as Memrise can provide an additional layer of listening exposure without requiring schools to rely exclusively on sophisticated classroom technology. This is particularly relevant where opportunities for authentic English listening outside the classroom are limited. Nevertheless, implementation should remain pedagogically guided rather than technology-driven. Teachers need to determine which listening objectives can be supported through mobile practice, how activities relate to classroom instruction, and how learners can be encouraged to use the application as a complement to—not a replacement for—interactive listening instruction. Viewed in this way, the contribution of the present study is not simply that Memrise improved test scores, but that carefully mediated mobile-assisted practice may provide a feasible mechanism for increasing the frequency, flexibility, and distribution of listening exposure among vocational EFL learners. Further research comparing different MALL applications, treatment durations, learner proficiency levels, and combinations of mobile practice with explicit listening-strategy instruction would help establish the conditions under which such benefits are most sustainable.

Conclusion

This study confirmed a statistically significant positive effect of the Memrise application on the listening skills of second-grade students at SMK Negeri 03 Bombana. The experimental class, receiving four treatment sessions with Memrise, achieved a post-test mean of 84.35 compared to a pre-test mean of 52.65 ($N\text{-gain} = 0.67$, medium), substantially outperforming the control class which improved from 55.72 to 67.39 through conventional instruction ($N\text{-gain} = 0.26$, low). The independent sample t-test yielded a significance value of $0.044 < 0.05$, confirming the statistical significance of this difference and supporting acceptance of H1. These outcomes are theoretically grounded in Memrise's alignment with Multimedia

Learning Theory, the testing effect, and spaced repetition principles, which together facilitate deeper encoding, durable retention, and more equitable learning outcomes.

These findings carry several practical implications. English teachers in vocational high schools are encouraged to integrate memorise or comparable MALL applications as regular supplementary tools for listening instruction, particularly in contexts with limited access to authentic audio resources. School administrators should prioritize stable internet connectivity and accessible device provision to enable effective MALL implementation. Teacher training programs should include practical guidance on curating and embedding digital learning applications within curriculum-aligned listening instructional sequences. Future research is recommended to extend this inquiry across larger and more diverse samples, employ mixed-method designs to capture both quantitative outcomes and qualitative learning experiences, and explore the effect of Memrise on other English language skills, including speaking and vocabulary, across different educational levels and geographic contexts.

References

- Aminatun, D., & Oktaviani, L. (2019). Memrise: Promoting students' autonomous learning skill through language learning application. *Metathesis: Journal of English Language, Literature, and Teaching*, 3(2), 214–223. <https://doi.org/10.31002/metathesis.v3i2.1982>
- Andujar, A., & Hussein, S. A. (2019). Mobile-mediated communication and students' listening skills: A case study. *International Journal of Mobile Learning and Organisation*, 13(3), 309–332. <https://doi.org/10.1504/IJMLO.2019.100432>
- Aryadoust, V., & Raquel, M. (Eds.). (2019). *Quantitative data analysis for language assessment volume II: Advanced methods*. Routledge. <https://doi.org/10.4324/9781315187808>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.
- Demouy, V., Jones, A., Kan, Q., Kukulska-Hulme, A., & Eardley, A. (2016). Why and how do distance learners use mobile devices for language learning? *The EUROCALL Review*, 24(1), 10–24.
- Detre, E. C., & Cook, E. (2017). Memrise. *The Electronic Journal for English as a Second Language*, 21(1), 4–5.
- Deveci, T. (2018). Listening as a lifelong learning skill: What, why, and how. *16th International Conference on Literature, Languages, Humanities and Social Sciences*, 85–89.
- Dewi, R. (2019). *The effectiveness of Memrise towards students' listening skill: An experimental research at the second grade of SMK YAPIDI Jayanti Kab. Tangerang* [Unpublished undergraduate thesis]. Universitas Muhammadiyah Tangerang.
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Dover. (Original work published 1885)
- Fathi, J., Derakhshan, A., & Torabi, S. (2020). The effect of listening strategy instruction on second language listening anxiety and self-efficacy of Iranian EFL learners. *SAGE Open*, 10(2), 1–13. <https://doi.org/10.1177/2158244020933878>

- Gilakjani, A. P., & Ahmadi, S. M. (2016). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching*, 9(6), 123–133. <https://doi.org/10.5539/elt.v9n6p123>
- Harmer, J. (2007). *How to teach English*. Pearson Education Limited.
- Hazairin, P. H., Mujiyanto, J., & Sutopo, D. (2019). Optimizing EIL using Memrise as builder of students' vocabularies. *SELTICS Journal*, 2(2), 190–206.
- Heil, C. R., Wu, J. S., Lee, J. J., & Schmidt, T. (2016). A review of mobile language learning applications: Trends, challenges, and opportunities. *The EUROCALL Review*, 24(2), 23–27.
- Jafari, K., & Hashim, F. (2015). The effects of using advance organizers on improving EFL learners' listening comprehension. *International Journal of Foreign Language Teaching in the Islamic World*, 3(3), 5–11.
- Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, 72(2), 269–319. <https://doi.org/10.1111/lang.12479>
- Kim, S. K., & Webb, S. (2023). Does spaced practice have the same effects on different second language vocabulary learning activities? Fill-in-the-blanks versus flashcards. *The Modern Language Journal*, 107(4), 944–964. <https://doi.org/10.1111/modl.12879>
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile-assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20(3), 271–289. <https://doi.org/10.1017/S0958344008000335>
- Lee, P.-J., Liu, Y.-T., & Tseng, W.-T. (2021). One size fits all? In search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, 25(3), 400–430. <https://doi.org/10.1177/1362168819856451>
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawlings, J. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311. <https://doi.org/10.1017/S0958344019000065>
- Luczak, A. (2017). Using Memrise in legal English teaching. *Studies in Logic, Grammar and Rhetoric*, 49(62), 1–14. <https://doi.org/10.1515/slgr-2017-0002>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics. *American Journal of Physics*, 70(12), 1259–1268. <https://doi.org/10.1119/1.1514215>
- Metruk, R. (2024). Mobile-assisted language learning and pronunciation instruction: A systematic literature review. *Education and Information Technologies*, 29, 16255–16282. <https://doi.org/10.1007/s10639-024-12453-0>
- Nirwana. (2021). *The use of Memrise application to improve students' listening skill: A pre-experimental research at eighth grade of SMPN 3 Bungoro Kab. Pangkep* [Unpublished undergraduate thesis]. Universitas Muhammadiyah Makassar.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

- Purnamasari, A. (2018). What EFL learners say about YouTube use to improve pronunciation in a blended learning environment. *JET (Journal of English Teaching)*, 4(3), 205–215.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Pearson Education Limited.
- Shadiev, R., Liu, T., & Hwang, W.-Y. (2020). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology*, 51(3), 709–720. <https://doi.org/10.1111/bjet.12839>
- Sugiyono. (2015). *Metode penelitian kombinasi (mix methods)*. Alfabeta.
- Suryosubroto, B. (2009). *Proses belajar mengajar di sekolah*. PT Rineka Cipta.
- Tarigan, H. G. (2015). *Mendengarkan sebagai suatu keterampilan berbahasa*. Angkasa.
- Vafaei, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, 42(2), 383–402. <https://doi.org/10.1017/S0272263119000670>
- van den Broek, G. S. E., Segers, E., van Rijn, H., Takashima, A., & Verhoeven, L. (2019). Effects of elaborate feedback during practice tests: Costs and benefits of retrieval prompts. *Journal of Experimental Psychology: Applied*, 25(4), 588–601. <https://doi.org/10.1037/xap0000212>
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390–416. <https://doi.org/10.1111/lang.12105>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Yang, J., Zhang, E.-H., He, X., Zhang, H., Gou, H., Wang, X., Wang, S., & Cao, H.-W. (2022). Retrieval practice enhances learning and memory retention of French words in Chinese-English bilinguals. *Lingua*, 272, 103294. <https://doi.org/10.1016/j.lingua.2022.103294>
- Yeldham, M. (2020). Does the presence of formulaic language help or hinder second language listeners' lower-level processing? *Language Teaching Research*, 24(3), 338–363. <https://doi.org/10.1177/1362168818787828>
- Zohoorian, Z., Noorbakhsh, M., & Zeraatpisheh, M. (2022). EFL learners' vocabulary achievement and autonomy: Using Memrise mobile application. *Indonesian Journal of EFL and Linguistics*, 7(2), 233–249. <https://doi.org/10.21462/ijefl.v7i2.487>