Govher Kurbanova, Hongyan Wang, Jeroen van de Weijer

Shenzhen University, China

An Experimental Study on the Effects of Audio Assistance in Vocabulary Learning: The Role of Different Word Classes

ABSTRACT

The present study aimed to assess the effectiveness of English vocabulary acquisition with and without audio support among first-year undergraduate students at Shenzhen University. Three core research questions guided the investigation: (1) How effective is vocabulary learning through audio assistance alone? (2) How effective is vocabulary learning without audio assistance (i.e., traditional instruction)? (3) To what extent does the relative effectiveness of audio-aided instruction versus traditional instruction for vocabulary acquisition differ across the four major word classes—nouns, verbs, adjectives, and adverbs—among English as a Foreign Language (EFL) learners? This experimental study recruited over 100 participants, who were divided into an experimental group (receiving audio support) and a control group (without audio support). Results indicated that the control group achieved higher mean scores, 9.09 (phase 1) and 9.92 (phase 2), than the experimental group, 7.52 (phase 1) and 8.90 (phase 1), in both phases, suggesting that audio-only instruction is less effective for vocabulary acquisition than traditional methods.

Keywords: *Audio Aids, Traditional method (of vocabulary teaching), EFL Vocabulary learning.*

Introduction

Competence in vocabulary is crucial for language acquisition, especially in English as a Foreign Language (EFL) settings (Takacs et al., 2015). Language proficiency enables learners to communicate effectively, comprehend texts, and express ideas with fluency. Acquiring vocabulary is frequently reported as difficult due to the extensive range of words and the requirement for sustained recall. Traditional vocabulary acquisition techniques, including rote memorization and repetition, have been widely employed but are frequently criticized for their monotony and ineffectiveness in fostering deep learning.

The incorporation of audiovisual components in language teaching has created considerable interest in recent years (e.g., Belenguer, 2024). Audiovisual aids, including audio recordings, pictures, and videos, are thought to facilitate vocabulary acquisition by activating multiple senses, alleviating cognitive load, and offering contextualized learning experiences. Studies indicate that multimedia technology enhances memory, motivation, and comprehension in learners (Bisson,

2014). The efficacy of these aids relative to conventional approaches, together with their distinct contributions (auditory versus visual), has not been adequately examined. According to Aufderhaar (2004), audio-based methods can improve pronunciation and word stress, resulting in statistically significant gains in recall of audio-sensitive word classes. However, there is still limited comparative data on how including audio aids compare to traditional methods, especially across different word classes like nouns, verbs, adjectives, and adverbs. For instance, we wonder if more concrete words (often nouns or verbs) are remembered more easily than less concrete words, such as adjectives and adverbs.

This research aims to analyze the impact of audio aid on vocabulary acquisition in an EFL context. We will compare the performance of students using audio aids with that of those employing traditional methods, with the aim of finding the most effective techniques for vocabulary acquisition.

Literature Review

Previous Studies on Teaching Vocabulary with Audiovisual aids

Research on the use of audio-visual materials has attracted increased attention in academic circles, as second language (L2) listening instruction has shifted from traditional audio methods to audio-visual techniques due to advancements in science and technology (Nafiah, 2023). Audio-visual aids are becoming increasingly essential instruments for engaging learners and enhancing language teaching, thanks to technological advancements and the growing availability of multimedia resources (Olagbaju & Popoola, 2020). The growth of technology and the extensive accessibility of multimedia resources have transformed language instruction and acquisition (Hermes & King, 2013). Videos, pictures, interactive presentations, animations, and digital resources exemplify audio-visual assistance. These aids offer visual and auditory stimulation to learners, facilitating a multimodal learning experience that engages many senses and improves comprehension and retention of linguistic knowledge (Papanastasiou et al., 2019).

The utilization of audio-visual aids in EFL training presents numerous benefits. More recently, Teng (2022) noticed that audio-visual aids can improve vocabulary recall by engaging auditory senses; however, their investigation did not differentiate the impact of audio alone or contrast it with standard text-only techniques. Lin & Lin (2019) also discovered variable effects of audio aids on vocabulary acquisition across studies, underscoring the necessity to investigate audio's performance in comparison with traditional methods and whether its effectiveness varies by word class (nouns, verbs, adjectives, adverbs).

This study seeks to examine the influence of auditory assistance on vocabulary development in an English as a Foreign Language (ESL) setting. The study aims to compare the performance of students utilizing audio assistance with those employing traditional methods to identify the most effective strategies for vocabulary acquisition.

Studies on Teaching Vocabulary with Audio Aids

Krashen's Input Hypothesis posits that language acquisition advances when learners engage with understandable input that surpasses their current ability in specific areas (Ellis, 2002). Audio aids offer natural language exposure and enhance the acquisition of new vocabulary in authentic contexts. Interacting with audio content allows learners to absorb pronunciation, intonation, and vocabulary application. Previous studies (e.g., Papanastasiou et al., 2019) highlight the relative effectiveness of audio-only compared to audiovisual input.

To facilitate the acquisition of phonological components such as stress patterns, intonation, and connected speech—crucial for lexical retention and pragmatic competence—audio resources, including podcasts and recorded dialogues, provide realistic linguistic models. Empirical studies demonstrate that learner proficiency, age, and contextual factors significantly influence the efficacy of audio-only instruction. Ginther's study on advanced ESL students demonstrated that while audio resources alone were advantageous for improving listening comprehension, they did not influence vocabulary retention unless supplemented with textual or visual reinforcement (Al- Zahrani & Al-Ghamdi, 2025).

Cognitive Load Theory (CLT; Sweller, 1988) states that working memory capacity is limited, and instructional methods that introduce extraneous load (e.g., additional sensory information) can slow learning. This paradigm poses essential enquiries regarding auditory aids: might auditory input impose an unnecessary burden on EFL vocabulary learners, particularly in contrast to text-only techniques?

Although research such as Al-Zahrani & Al-Ghamdi (2025) acknowledges the necessity of textual support for audio, few studies explicitly examine whether audio alone induces cognitive load that hinders vocabulary retention, much less whether this load differs among word classes (nouns, verbs, adjectives, adverbs). This study examines whether there are differences in the effects of audio-assisted and traditional text-based learning, and uses CLT to analyze performance differences.

Methodology

Participants

This study involved around one hundred first-year students from Shenzhen University, evaluating one experimental group and one control group. A sample size exceeding 100 was determined through an a priori power analysis utilizing G*Power (version 3.1) for an independent samples t-test (Faul, Erdfelder, Buchner, & Lang, 2009), which served as the primary statistical method for group comparisons. The analysis showed that 98 participants gave us 80% power to identify a medium effect size (d = 0.5) for differences between the experimental and control groups. We recruited over 100 to account for potential participant loss, which is in line with comparable EFL vocabulary studies (e.g., Lin et al., 2022).

All participants were native Chinese speakers aged 18 to 20. The selection criteria included language proficiency level, previous experience with vocabulary acquisition methods, and willingness to participate in this study. Both groups were equivalent in terms of age, gender, and language competency to ensure that the results were not affected by these variables. The pilot test results demonstrated proficiency in English at the B2, C1, and C2 levels. Ethical considerations were acknowledged, and all participants provided informed consent before the start of the research.

Research Design

The study used a quasi-experimental crossover design to assess the effectiveness of audio aids and traditional methods in vocabulary development. The experiment was divided into two sections, each lasting four weeks (week 1- pre-test; week 2-3 procedure; week 4 – post-test).

Table 1. Research design

Week 1 **Phase 1**	
Group A (experimental group)	
Group B (control group)	With audio aids
	Without audio aids (traditional method)
Week 2 **Phase 2**	
Group A (control group)	Without audio aids
Group B (experimental group)	With audio aids

In Phase 1, Group A (the experimental group) was exposed to vocabulary learning utilizing audio aids, whereas Group B (the control group) utilized traditional methods (see Table 1. Research

design). In Phase 2, the methodologies were altered: Group A employed conventional techniques, while Group B utilized audio resources. This arrangement facilitated a direct comparison of the two procedures within identical groups, mitigating the influence of individual variances and enhancing the dependability of the results. The crossover design guaranteed that both groups encountered both learning circumstances, yielding a more thorough comprehension of the efficacy of each strategy.

Research Instruments

This study employed two principal research instruments to fulfill its objectives: experimental research and a student feedback survey. The initial instrument concentrated on evaluating the efficacy of audio aids vs traditional teaching techniques in vocabulary acquisition. The second measure was employed to gain insights into students' perceptions of the approaches utilized to improve vocabulary acquisition.

The initial instrument was an experimental study design involving two groups of students. The groups were randomly allocated into an experimental group and a control group, each comprising an equal number of individuals.

The audio assistance used pre-recorded MP3 files of each target word, pronounced in British English by a native speaker, accompanied by a brief contextual sentence (e.g., 'Converge: The routes converge near the park'). Each audio clip lasted 5 to 10 seconds, and participants accessed them through a university learning website, without the ability to modify playback speed. The control group utilized a traditional vocabulary instruction method without audio assistance. Instruction was delivered utilizing printed materials, including worksheets. Traditional techniques, such as word lists, direct explanations, and example sentences, were used to instruct vocabulary.

The second instrument involved gathering students' opinions to comprehend their viewpoints on the employed teaching strategies. This instrument aimed to enhance the quantitative data from the vocabulary exam with qualitative insights into the students' experiences and preferences.

Vocabulary List

This study utilized a vocabulary list of thirty carefully chosen terms to guarantee their relevance and suitability for the research aims. The vocabulary was sourced from two main sources: students' textbooks and the Oxford Vocabulary List. The selection approach aimed to incorporate new vocabulary for the participants, facilitating a more precise evaluation of the training methods' efficacy. To do this, only the vocabulary at the C1 and C2 levels of the Common European Framework of Reference for Languages (CEFR) was selected, as these levels imply advanced

proficiency and were less likely to be familiar for with the participants. This methodology ensured that the research concentrated on real vocabulary acquisition instead of the reinforcement of previously acquired vocabulary.

The thirty words were split into two sets of fifteen words each, corresponding with the two weeks of the experiment. Each set was designed to include a balanced distribution of word kinds to investigate if the instructional methods exerted differing impacts on certain aspects of speech. Each set of four nouns, four verbs, four adjectives, and three adverbs. This distribution facilitated a detailed examination of the efficacy of audio aids vs traditional approaches in teaching various word kinds. It offered insights into the efficacy of visual aids for teaching nouns, typically denoting tangible objects, versus the advantages of auditory aids for instructing verbs, which generally convey action. The study sought to ascertain both the general efficacy of each approach and its particular strengths and limitations for word kinds by preserving this balance.

Data Analysis

The quantitative data from pre-tests, weekly evaluations, and post-tests were analyzed using standard statistical techniques. Paired t-tests were employed to evaluate the performance of each group before and after the method shift, whereas independent t-tests were used to analyze any variations between the two groups. Next, the written input was classified and examined. Recurring themes, including the most preferred and least favored aspects of each learning approach, were noted to enhance comprehension of students' viewpoints. Conclusions about the relative efficacy of audio and traditional learning approaches were derived from the extensive data analysis and will be presented in section four.

Research Results

Phase 1 Results

The descriptive statistics indicate a notable disparity between the two groups. The mean vocabulary test score for the experimental group utilizing audio-assisted learning was 7.51 (SD=3.78), whereas the control group employing traditional learning methods without audio assistance had a higher mean score of 9.09 (SD=3.20) (see Table 2).

Table 2. Descriptive analysis of learning vocabulary with audio Aids

Group Statistics

Group	N	Mean	Std. Deviation	Std.Error Mean
Experiment	54	7.5185	3.77550	.51378
Control	54	9.0926	3.19946	.43539

The data indicate that, contrary to the belief that audio aids increase vocabulary acquisition, the group that learned without audio help outperformed the others in the vocabulary assessment.

Table 3. Independent Samples t-test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Score	Equal variances assumed	3.177	.078	-2.337	106	.021	-1.57407	.67345	-2.90926	-.23889
	Equal variances not assumed			-2.337	103.222	.021	-1.57407	.67345	-2.90967	-.23848

The independent samples t-test demonstrated a statistically significant difference ($t(106) = -2.337$, $p = 0.021$, Cohen's $d = 0.45$), suggesting a small-to-moderate effect favoring the control

group (see Table 3). This signifies that the difference between the two groups is statistically significant, indicating that the variation in scores is unlikely to result from random chance.

An analytical interpretation of these findings indicates that the inclusion of audio help may have affected the learning process in ways that were not wholly advantageous for vocabulary retention.

One potential explanation is that participants in the experimental group may have depended more on auditory inputs instead of fully interacting with the written form of the words, perhaps leading to lesser retention. On the other hand, the control group, which utilized traditional techniques without audio assistance, may have exhibited a greater concentration on the textual material, resulting in better recall and retention of vocabulary items.

Analysis by Word Classes with and without Audio Aids

This study also tested the effectiveness of vocabulary learning with and without audio aids, concentrating on distinct word types: **nouns, verbs, adjectives,** and **adverbs**. The experimental group acquired vocabulary with audio support, whereas the control group utilized traditional approaches without any audio input. Both groups were presented with the same collection of words, comprising an equal quantity of nouns, verbs, adjectives, and adverbs. Following the intervention, the participants' test results were evaluated according to word types to ascertain if audio-assisted learning affected vocabulary retention variably across different categories of words. The findings indicated a notable difference in word retention between the two groups, implying that the efficacy of learning strategies differed based on the type of vocabulary being learned (see Figure 1).

The findings indicated a distinct superiority of the control group across all categories of words. The overall scores for each word category were markedly greater in the control group compared to the experimental group. The control group attained the following scores: nouns – 159 (30%), verbs – 138 (26%), adjectives – 140 (27%), and adverbs – 89 (17%). Conversely, the experimental group achieved much lower scores: nouns – 32 (11%), verbs – 122 (40%), adjectives – 88 (29%), and adverbs – 60 (20%) (see Figures 1, 2, and 3). The data demonstrate that traditional learning without audio aids provided superior vocabulary retention compared to learning with audio assistance across all word categories.

Figure 1. Analysis by Word Classes with Audio Aids

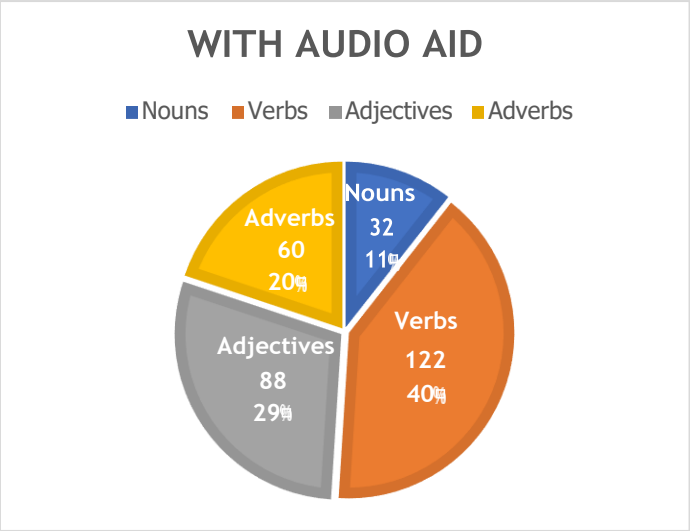

Figure 1 Analysis by word classes with audio aids

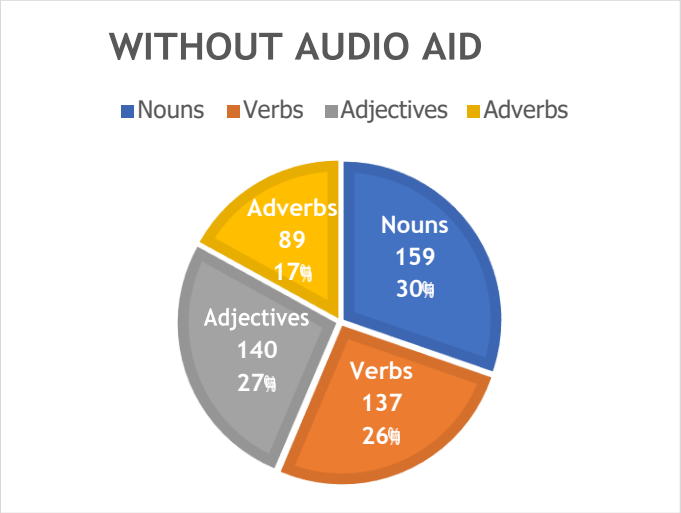

Figure 2 Analysis by word classes without audio aids

Below is an in-depth analysis of word types that provides important details on how students processed and memorized distinct word categories:

Nouns

The experimental group achieved a total score of 32 for nouns, compared to 159 for the control group. This striking difference implies that when audio aids were utilized, learners in the experimental group had a far harder time learning and remembering nouns. The result can be clarified by the intrinsic visual nature of nouns; learners often associate nouns with images or

concrete representations, which are more effectively reinforced through written or visual modalities than through auditory ones. In contrast, noun retention was superior in the control group that employed the traditional method. This outcome aligns with prior studies indicating that traditional association, rather than auditory association, is the most beneficial approach for acquiring concrete nouns.

Verbs

In comparison to nouns, verbs exhibited a smaller retention disparity between the two groups. The control group attained a score of 138, whereas the experimental group achieved a score of 122. The disparity was less pronounced than with nouns; yet, the control group still outperformed the experimental group. This finding implies that verbs may gain more from auditory support than nouns. Hearing verbs in context may help with comprehension and memory because they frequently entail actions. However, the traditional method still produced slightly greater recall, maybe as a result of the control group's students' increased engagement with the words' written form, which strengthened their memorization.

Adjectives

The experimental group scored 88, whereas the control group scored 140, indicating a significant difference in adjective retention as well. This implies that adjectives were better learned using traditional methods rather than audio aids, much like nouns. Adjectives frequently express features or attributes that learners may link more successfully with written context or visual aids than with only auditory input. It is possible that traditional teaching approaches improved recall by encouraging students to concentrate on spelling, sentence structure, and visual reinforcement.

Adverbs

Although the difference was not as great, adverbs showed a similar pattern. The control group received 89 points, whereas the experimental group received 60. This implies that while audio aids were not as successful as traditional strategies, they did offer some advantages in the learning of adverbs. Since they frequently occur in spoken language, adverbs—which modify verbs, adjectives, or other adverbs—may have been a little simpler to learn through aural exposure. However, the findings show that students in the control group were able to recall adverbs better, maybe as a result of their interaction with written materials and contextual signals.

Phase 2 Results

The independent samples t-test for Phase 2 indicated no statistically significant difference in vocabulary scores between the control group (traditional method) and the experimental group (audio aid) ($t(102) = 1.392$, $p = 0.167$, Cohen's $d = 0.27$). The minimal effect size ($d = 0.27$) suggests that although the control group attained a marginally superior mean score ($M = 9.92$, $SD = 3.92$) compared to the experimental group ($M = 8.91$, $SD = 3.50$), the disparity was insufficient to ignore random variation, making it not practically significant for EFL vocabulary instruction (see Table 4).

Table 4 Descriptive statistics of learning vocabulary with audio aids

Group Statistics

Group Mean	N	Mean	Std. Deviation	Std. Error
Experiment	51	9.9216	3.49767	.48977
Control	53	8.9057	3.92362	.53895

The *p*-value of 0.167 (see Table 5) indicates that the disparity between the two groups is not statistically significant. This indicates that although the control group had a higher mean score, the observed disparity is insufficient to assert that the traditional method is, in fact, more effective than the audio-assisted method. However, the findings suggest a trend that warrants further investigation with methodological refinements.

Table 5 Independent Samples t-test

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances Assumed Score									

	.544	.463	1.392	102	.167	1.01591	.72987	-.43178	2.46360
Equal variances Not assumed			1.395	101.417	.166	1.01591	.72825	-.42867	2.46048

The results suggest that audio-assisted vocabulary acquisition could be enhanced by increasing exposure duration, incorporating visual and textual elements, and offering greater flexibility in review options. Although audio learning may not serve as the most effective independent approach, it can be a great asset when integrated with other instructional strategies.

Analysis by Word Classes with and without Audio Assistant

The findings indicated subtle variations in recall rates for each word type between the two groups, offering significant insights into the relationship among teaching methods, word attributes, and cognitive processing. The experimental group employing audio assistance achieved scores of 132 for nouns, 99 for verbs, 145 for adjectives, and 96 for adverbs (see *Figure 3*). Conversely, the control group, which utilized conventional techniques without audio assistance, achieved scores of 147 for nouns, 107 for verbs, 148 for adjectives, and 102 for adverbs (see Figure 4).

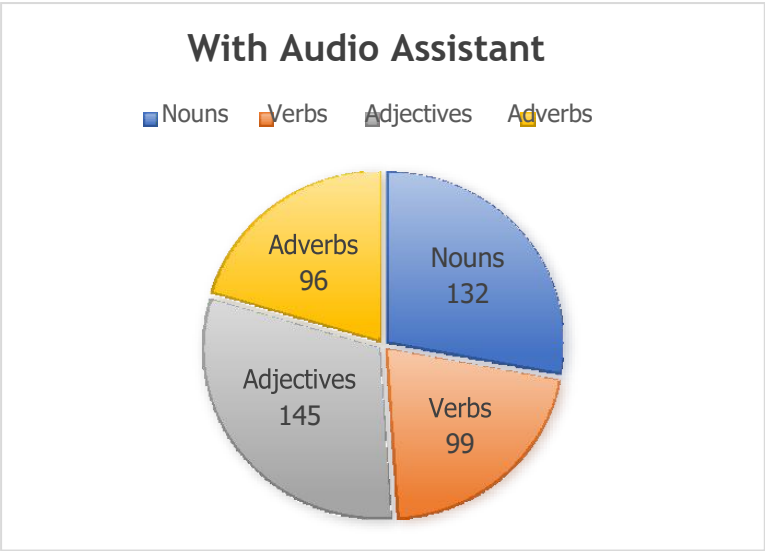

Figure 3 Analysis by word classes with audio aids

Figure 4 Analysis by word classes without audio aids

Although the control group outperformed the experimental group in three of the four word categories, the variations were insignificant, indicating that the efficacy of audio aids could vary based on the specific type of word being acquired.

Nouns

Nouns, often concrete and easily visualizable, demonstrated a slight advantage in the control group, scoring 147 against 132 in the experimental group. This research indicates that traditional methods may be slightly more effective for teaching nouns than those that use audio aids. The concrete quality of nouns enables learners to create vivid mental representations, which can be enhanced through repetition and contextual illustrations in conventional pedagogical approaches. Audio aids, although they may improve engagement and pronunciation accuracy, cannot offer substantial advantages for the retention of nouns. The brain's capacity to visualize tangible objects and concepts diminishes dependence on auditory reinforcement. Nevertheless, the minor disparity in scores between the two groups suggests that audio aids may not hinder noun acquisition; instead, they may provide a less significant benefit compared to conventional approaches.

Verbs

Verbs, representing acts and processes, demonstrated lower memory rates in both groups relative to nouns and adjectives, with the experimental group achieving a score of 99 and the control group attaining a score of 107. The marginally superior score in the control group indicates that traditional techniques may be more efficacious for instructing verbs than approaches that integrate audio aids.

Verbs are organically more complex than nouns, as they frequently necessitate comprehension of tense, aspect, and syntactic frameworks. Traditional techniques, often characterized by repeating exercises, contextual illustrations, and direct instruction, may offer a more systematic strategy for understanding these intricacies. Conversely, audio aids, although beneficial for enhancing pronunciation and auditory recognition, may insufficiently address the semantic and grammatical subtleties of verbs. The abstract nature of several verbs may render them less receptive to aural reinforcement, as they lack the tangible imagery linked to nouns. This discovery underscores the necessity of customizing pedagogical approaches to address the distinct obstacles presented by various word categories, especially those that are more abstract and reliant on context.

Adjectives

Adjectives, which describe features and qualities, had essentially comparable recall rates in both groups, with the experimental group achieving a score of 145 and the control group attaining a score of 148. The small change indicates that the efficacy of teaching methods for adjectives is predominantly unaffected by the incorporation of audio aids. Adjectives, like nouns, are frequently concrete and visualizable, facilitating their encoding and retrieval. The descriptive quality of adjectives enables learners to connect them to specific nouns or circumstances, enhancing retention regardless of the instructional approach used. Audio aids may improve learning through auditory reinforcement; however, they do not appear to provide a substantial benefit compared to traditional approaches for acquiring adjectives. This discovery highlights the necessity of accounting for the intrinsic properties of word types in the development of vocabulary instruction, since certain categories may be less affected by the selected teaching approach than others.

Adverbs

Adverbs exhibited approximately comparable recall rates in both groups, with the experimental group achieving a score of 96 and the control group attaining a score of 102. The significantly superior score in the control group suggests that traditional techniques may be slightly more efficacious for instructing adverbs compared to approaches that utilize audio assistance. Traditional techniques, which prioritize repetition and contextual application, may offer a more systematic manner for acquiring the various meanings and functions of adverbs. Conversely, audio aids, although possibly beneficial for enhancing pronunciation and intonation, may insufficiently tackle the semantic and syntactic complexities of adverbs.

Results of Students' Feedback on Vocabulary Teaching Methods

The secondary research instrument (student feedback surveys) was designed to capture learners'

subjective perceptions of the two vocabulary acquisition methods indicated nuanced individual differences in participants' engagement with audio- assisted compared to traditional text-based learning. One student, Mia emphasized a distinct advantage of audio aids, stating: "Utilizing audio aids facilitated my acquisition of new terminology. Listening to their pronunciation and contextual usage facilitated quicker recollection and enhanced my confidence in speaking." We can say that audio can help with confidence in pronunciation and comprehending the context, even though Mia's pleasant experience is different from the experimental group's overall lower test scores.

The student Dragon, on the other hand, talked about the benefits of studying simply from text. He said, "Without audio help, I was able to focus better on understanding the definitions and practicing the vocabulary." Dragon's comment is in line with the Cognitive Load Theory framework that this study has referred to. It suggests that taking away auditory input made it easier to focus on vocabulary in a more semantic (definition-based) and practical (application-based) way. This is consistent with the control group's stronger objective performance in both phases. These two points of view show that while audio aids may help with certain learning goals (like pronunciation and speaking confidence), traditional text-based methods work better for focused comprehension and retention, which are both important for vocabulary acquisition in this EFL context. This supports the study's main finding that how well a method works depends on both objective learning outcomes and the learner's own priorities.

Comparison between Phase 1 and Phase 2

Traditional learning methods led to higher recall in Phase 1 across all categories— nouns, verbs, adjectives, and adverbs—when comparing vocabulary acquisition with and without audio aids for various word categories. Nouns were the area where learners in the experimental group struggled the most, indicating the biggest difference. Overall, the control group fared better than the experimental group in every word category, but verbs demonstrated some benefit from audio help.

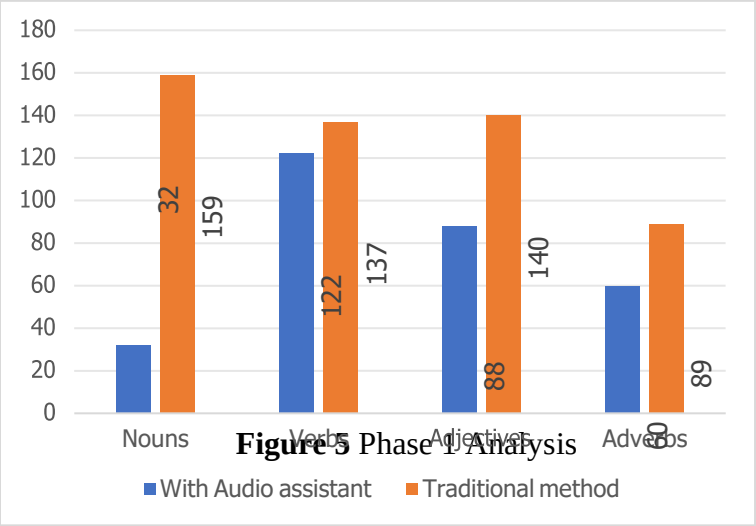

Figure 5 Phase 1 Analysis

These results emphasize how crucial it is to take word kinds and cognitive processing into account when creating vocabulary learning plans.

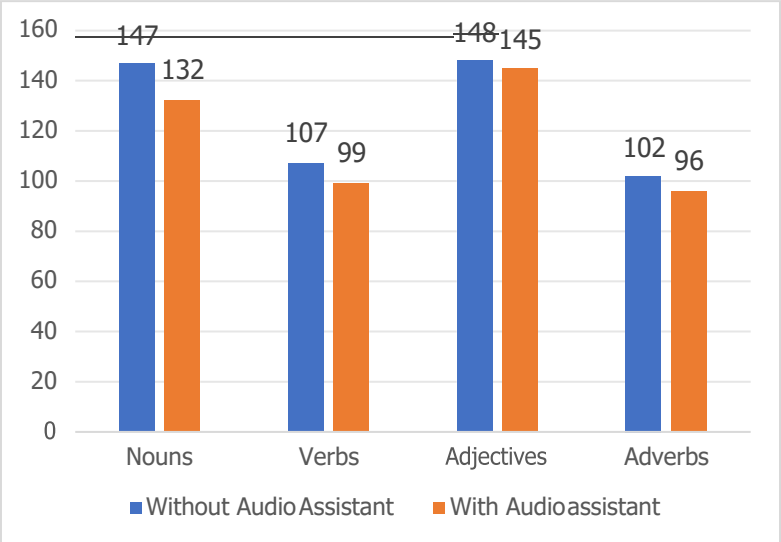

Figure 6 Phase 2 Analysis

A possible explanation for the disparities in memory rates between the two groups is the notion of cognitive load, which denotes the mental work necessary for processing and retaining information. Audio aids, although stimulating, may elevate cognitive load by adding extra sensory information that learners must concurrently absorb. For concrete word categories such as nouns and adjectives, this supplementary input may not considerably influence learning efficiency, as these terms are more readily visualized and linked to preexisting information. Nevertheless, for more abstract word categories such as verbs and adverbs, the augmented cognitive burden linked to audio aids may impede recall. Traditional techniques, emphasizing repetition and contextual practice, may alleviate cognitive load by enabling learners to focus on the semantic and syntactic dimensions of words without the interference of auditory inputs. This explanation corresponds with the marginally elevated scores noted in the control group for verbs and adverbs, indicating that conventional approaches may be more effective for instructing these categories of words.

Discussion

The main aim of this study was to examine the effectiveness of audio-assistedlearning on L2 vocabulary acquisition, specifically regarding its varying effects on different word classes. The findings, derived from a two-phase experimental design, reveal an interesting as well as contradictory

context. In contrast to the dominant educational tendency favoring multi-modal learning, our findings repeatedly indicated that traditional, text-based learning approaches without audio accompaniment produced superior, or at least comparable, results in vocabulary recall and retention. This part will analyze the important findings, contextualize them within the current academic discourse, and offer theoretical explanations for the observed patterns.

Interpreting the Core Findings: The Superiority of Text-Based Learning

The most notable outcome from Phase 1 was the statistically significant benefit ($p = 0.021$) shown by the control group, which utilized traditional methods of learning. This group outscored the experimental group in all word categories—nouns, verbs, adjectives, and adverbs—with a notably significant difference in noun acquisition. Phase 2, however, did not produce a statistically significant difference ($p = 0.167$), had a consistent pattern of increased mean scores for the traditional learning group across three of the four word categories. The persistent tendency seen in both stages indicates that the incorporation of an audio element did not enhance, and may have even somewhat inconvenienced the vocabulary acquisition process in this particular learning environment.

A reasonable framework for interpreting these data is Cognitive Load Theory (CLT), Sweller (1988). CLT posits that working memory has a limited capacity, and learning is enhanced when instructional strategies avoid overloading it. Within the framework of this investigation, the audio-assisted approach can be construed as imposing an additional cognitive load. Participants in the experimental group were required to simultaneously process the auditory representation of a word and gits definition. This allocation of focus may have reduced the cognitive effort available for the basic task of establishing durable lexical representations in long-term memory. In contrast, the control group, concentrating exclusively on textual and semantic aspects, might focus their whole cognitive resources on establishing and reinforcing associations between the written word and its meaning, resulting in enhanced encoding and recall.

Although CLT discusses the extraneous auditory load, another potential reason is the participants' familiarity with text-based learning; as Chinese EFL learners, they may possess greater experience with traditional vocabulary acquisition methods (e.g., textbook memorization), resulting in superior performance in the control group.

These findings possess immediate implications for EFL application developers: Features such as 'auto-play audio' for flashcards may not improve vocabulary memory and might be discretionary rather than obligatory, permitting learners to select based on their preferences.

Differential effects on word classes: A nuanced perspective

An in-depth examination shows notable differences in the effectiveness of audio assistance among various word classes. In Phase 1, the experimental group exhibited the lowest scores on nouns, a result that seems paradoxical considering that nouns often represent tangible, highly visualizable notions. It might be posited that audio would facilitate the anchoring of these tangible referents. This significant disadvantage for nouns in the auditory condition corresponds with the notion of "desirable difficulties" (Bjork, 1994). The demanding retrieval process necessitated by the text-only approach—where learners must actively produce the phonological representation from the orthographic form—may foster a more profound and enduring memory trace. The auditory group, possessing the given phonological form, was exempt from this arduous procedure, likely resulting in superficial processing and reduced retention, especially for high-frequency, concrete nouns when such effort is most beneficial.

On the other hand, the findings from both phases indicate a less detrimental and occasionally slightly advantageous impact of audio on the acquisition of verbs and adverbs. This discovery can be compared with the study of Kelly et al. (2017), which indicated that sound symbolism and prosodic characteristics contribute to verb acquisition. They contend that the acoustic characteristics of specific words can discreetly indicate their meaning or grammatical category. Although our study did not explicitly examine sound-symbolic pairs, it is conceivable that the prosodic elements in the audio (e.g., stress patterns, intonation) offered supplementary, non-redundant syntactic cues for the processing of verbs and adverbs, which are inherently more abstract and relationally defined than nouns. Consequently, for these word classes, the audio may have transitioned from a source of extraneous load to offering a moderate degree of relevant load, facilitating syntactic categorisation and integration.

However, it is necessary to acknowledge that the traditional strategy proved to be equally or more effective for verbs and adverbs. This suggests that any possible advantage of audio is subtle and likely dependent on various factors, including learner proficiency, the phonological distance between L1 and L2, and the particular learning task, a complexity emphasized in the meta-analysis by Lin et al. (2022) regarding multimedia learning in vocabulary acquisition. Their integration of multiple studies demonstrated significant variability in impact sizes, highlighting that the effectiveness of audio is not universal but influenced by a range of internal and external learner characteristics.

Limitations and methodological reflections

This study has basic limitations. The learning treatment was short, and the assessments evaluated immediate or near-instant memory. The dynamics of long-term retention may vary, and the initial "desirable difficulty" associated with the text-only strategy could potentially result in enhanced lasting vocabulary knowledge. The audio was presented in a static, non-interactive style with a predetermined tempo. The outcomes may vary with self-directed audio or interactive audio activities that necessitate a response from the student. Ultimately, individual variations in learning styles and auditory processing capacities were not assessed, potentially explaining some of the differences among the groups.

Conclusion

This study aimed to carefully evaluate the assumed advantages of audio-assisted vocabulary acquisition. The results clearly indicate that, for adult L2 learners in this environment, conventional text-based methodologies were more effective for adult EFL learners at Shenzhen University (CEFR B2–C1). The statistically significant findings from Phase 1 and the constant performance trend in Phase 2 challenge the uncritical adoption of audio as a universal learning aid.

This study's theoretical contribution is its comprehensive use of Cognitive Load Theory to vocabulary acquisition. It indicates that for literate adult learners, audio may serve as a source of insignificant cognitive load, hindering the effective encoding of new vocabulary. The varying effects on word classes enhance our comprehension, suggesting that the grammatical and semantic attributes of target vocabulary interact intricately with the mode of training. Although audio provided little benefit for tangible nouns, its slightly reduced negative impact on verbs and adverbs suggests a possible function in conveying syntactic information, necessitating further research.

From an educational perspective, these findings highlight the importance for material designers and language educators. Although multi-modal techniques are valuable, they are not a universal solution. The automatic incorporation of audio assistance in digital flashcards or educational applications warrants reevaluation. Instructors should support practices that facilitate effortful retrieval and profound textual engagement, particularly during initial vocabulary acquisition. Audio may be more effectively utilized in advanced phases for listening comprehension enhancement, pronunciation improvement, or as part of more intricate, integrated tasks rather than merely serving as a basic aid for first form-meaning association.

Suggestions for Future Research

Future studies are expected to expand upon these findings in several directions. First, longitudinal

studies are required to examine the long-term retention effects of audio-assisted learning compared to text-only learning. Second, studies should include assessments of individual characteristics, such as working memory ability and auditory learning preference, to identify which learners, if any, may authentically benefit from audio help. Third, research might investigate more interactive and dynamic applications of audio, such as sentence-level dictation or variable-speed playback, to determine if they minimize the unnecessary cognitive strain noted in this study. Lastly, conducting this study again with learners of lower proficiency or with languages that have clearer spelling rules could help us find critical limits on how these results can be used in other situations.

In conclusion, this study advocates for a more analytical, evidence-driven methodology in pedagogy. It proposes that, at times, the most efficacious route to learning is through increased cognitive resistance, wherein silent, concentrated interaction with text cultivates a stronger and more flexible vocabulary.

Acknowledgments

I want to thank everyone who helped me complete this paper from the bottom of my heart. This is a revised and enhanced version of the research design from my Master's thesis. I owe a lot to my MA supervisor, Professor Wang Hongyan, for her help at the beginning of this research.

I want to thank Professor Lin Yue for her aid in getting research volunteers to participate. I would like to thank Professor Zheng Xiaobei for her help with the data analysis methods, which made the empirical core of our paper much stronger. I am also very grateful to my PhD supervisor, Professor Jeroen van de Weijer, for his helpful comments and patient advice during the revision process, which were very important in forming the final manuscript.

Declaration of Interest Statement

The authors state that they have no known financial interests or personal relationships that could have influenced the work presented in this paper.

References

Al-Zahrani, A. A., & Al-Ghamdi, M. A. (2025). Challenges That Examiners Encounter When Assessing Speaking Skills at Yanbu English Language Institute and Preparatory Year Program (YELI-PYP). Journal of Language Teaching and Research, 16(2), 505-514. (n.d.).

Aufderhaar, Carolyn. “Learner Views of Using Authentic Audio to Aid Pronunciation: 'You Can Just Grab Some Feelings.'” TESOL Quarterly, vol. 38, no. 4, 1 Dec. 2004, p. 735, https://doi.org/10.2307/3588293. Accessed 21 Apr. 2019.

Belenguer Cortés, Luz. “Audiovisual Translation and Language Teaching.” Journal of Audiovisual Translation, vol. 7, no. 1, 12 Dec. 2024,

https://doi.org/10.47476/jat.v7i1.2024.292. Accessed 4 Jan. 2025.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). The MIT Press.

Faul, F., Erdfelder, E., Buchner, A. *et al.* Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* **41**, 1149–1160 (2009). https://doi.org/10.3758/BRM.41.4.1149

Ginther, A. (2002). Context and content visuals, and performance on listening comprehension stimuli. *Language Testing*, *19*, 133 - 167.

Hermes, Mary & Kendall A. King. 2013. Ojibwe language revitalization, multimedia technology, and family language learning. Language Learning and Technology 17(1). 125–144. https://doi.org/10.64152/10125/24513.

Krashen, S. (1994). The input hypothesis and its rivals. In Ellis N (eds.) Implicit and Explicit Learning of Languages. London: Academic Press, 45-77.

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Nafiah, S. U. (2023, August 31). Audio-Visual Aids in English Foreign Language (EFL) Teaching: A Multimodal Discourse Analysis. Dinamika Bahasa dan Budaya, 18(2).

Olagbaju, O. O., & Popoola, A. G. (2020). Effects of Audio-Visual Social Media Resources on Supported Instruction on Learning Outcomes in Reading. International Journal of Technology in Education, 3(2), 92-104. (n.d.).

Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M., & Papanastasiou, E. (2019). Virtual and augmented reality effects on K-12, higher and tertiary education students’ twenty-first-

century skills. Virtual Reality, 23, 425-436. (n.d.).

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12 (2), 257-285.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks: A Meta-Analysis. Review of Educational Research, 85(4), 698–739. https://doi.org/10.3102/0034654314566989

Teng, M. F. (2022). The effectiveness of multimedia input on vocabulary learning and retention. Innovation in Language Learning and Teaching, 1-17. (n.d.).