



THE EFFECT OF STRUCTURED YOUTUBE-BASED PRACTICE ON VIETNAMESE EFL UNDERGRADUATES' SPEAKING PERFORMANCE

Nguyen Tran Nhu Hao¹,

Kim Thanh Tuyen²ⁱ

¹English Language Education,
Can Tho University,
Vietnam

²Department of Social Science,
Can Tho University of Technology,
Vietnam

Abstract:

This study examined the effect of structured YouTube-based speaking practice on the speaking performance of second-year non-English-major undergraduates at Tay Do University, Vietnam. A randomized pre-test/post-test control-group design was employed. Seventy students from one General English Speaking 2 class were individually and randomly assigned to an experimental group (n = 35) or a control group (n = 35). Over 12 weeks, the experimental group completed a sequence of video viewing, shadowing, self-recording, and guided spoken response, while the control group undertook time-matched conventional speaking practice. Two curriculum-aligned speaking-test forms assessed grammar and vocabulary, pronunciation, fluency, and interactive communication. Independent- and paired-samples t-tests were used, with non-parametric robustness checks. No statistically significant between-group differences were found at pre-test. At post-test, the experimental group achieved a higher total speaking score (M = 13.69, SD = 1.41) than the control group (M = 12.09, SD = 1.44), $t(68) = 4.69$, $p < .001$, $d = 1.12$. Significant advantages were also found for grammar and vocabulary, pronunciation, and fluency, but not for interactive communication. Both groups improved over time. The findings support the use of the integrated YouTube-based protocol as supplementary speaking practice, particularly for linguistic resources and speech delivery. They do not establish the independent effect of video viewing or the long-term maintenance of gains. Additional reciprocal interaction is recommended when interactive communication is an instructional priority.

Keywords: YouTube; EFL speaking performance; shadowing; non-English-major undergraduates; randomized experiment

ⁱ Correspondence: email kttuyen@ctu.edu.vn, kimthanhtuyen.chntmk@soctrang.edu.vn

1. Introduction

Speaking is a central concern in English language education because it requires learners to use linguistic knowledge in real time. For non-English-major undergraduates, however, opportunities to develop oral performance may be restricted by limited classroom practice and a lack of regular English-speaking environments outside lessons. This study examines a structured supplementary approach in which pre-existing YouTube videos are combined with shadowing and guided speaking tasks. Rather than treating access to online video as an intervention in itself, the study evaluates a defined instructional sequence against conventional speaking practice.

1.1 Background of the Study

Communicative competence involves more than knowledge of grammatical rules. Learners must select appropriate language, organize ideas, pronounce words intelligibly, and respond to other speakers while communication is taking place (Canale & Swain, 1980; Bygate, 1987). Unlike writing, speaking usually offers little opportunity for extended planning or revision. Its real-time demands therefore make sustained practice important for the development of fluency and effective oral production (Thornbury, 2005).

The literature reviewed for the research identifies linguistic, affective, and pedagogical constraints on speaking development. Learners may have difficulty retrieving vocabulary, applying grammar spontaneously, and maintaining intelligible pronunciation. Speaking anxiety and fear of negative evaluation can further restrict participation (Horwitz *et al.*, 1986). In East Asian instructional settings, communicative teaching can also be constrained by established classroom practices and curriculum demands (Littlewood, 2007). These considerations make supplementary practice relevant, especially where increased classroom speaking time is not feasible.

The local rationale was supported by a preliminary needs assessment conducted before the intervention. Among 525 accessible second-year students at Tay Do University, 246 (46.9%) identified speaking as their most difficult English skill, 278 (53.0%) identified it as the skill they most wanted to improve, and 409 (77.9%) reported a need for additional opportunities to practice speaking outside regular class time. This preliminary survey was distinct from the 70-student intervention sample. It provided contextual justification for a speaking-practice program rather than evidence of the program's effectiveness.

Online video offers a possible source of supplementary spoken input. Learners can encounter vocabulary and pronunciation in context and revisit the same model at a personally manageable pace. This extends the opportunities described in research on extra-mural English learning (Sundqvist & Sylvén, 2016). Nevertheless, exposure does not necessarily involve speaking. The present intervention therefore linked video viewing to repeated oral production and guided responses, allowing its effect on performance to be examined directly.

2. Research Problem Statements

Previous studies have reported favorable outcomes for YouTube- and video-supported speaking practice, but their evidence differs in design and focus. Some primarily examine learners' perceptions of improvement, while others compare measured speaking performance between instructional conditions. Purnamasari (2018), for example, focused on learners' reported pronunciation development, whereas Saed *et al.* (2021) investigated YouTube use in developing Jordanian university students' speaking skills. The distinction matters because perceived improvement is not equivalent to improvement on a speaking assessment.

In Vietnam, video-related research has involved technical university students, English majors, and secondary learners. Nhat (2018) identified a need for additional out-of-class speaking opportunities among technical students; Oanh and Van (2020) examined vlog-based practice; and Dinh and Phung (2023) compared speaking performance following video-supported lessons among high-school learners in Kien Giang. The research identified a need for further performance-based evidence concerning non-English-major undergraduates at Tay Do University and a replicable sequence using existing YouTube material. The present article addresses that performance-focused gap.

2.1 Research Objective

The objective was to examine whether learners receiving a 12-week structured YouTube-based speaking-practice program achieved better speaking performance than learners receiving time-matched conventional practice. Overall speaking performance was the principal outcome, with grammar and vocabulary, pronunciation, fluency, and interactive communication examined as its four rubric components.

2.2 Research Question

- To what extent does the use of YouTube videos affect EFL learners' speaking performance?

In this question, "the use of YouTube videos" refers to the integrated viewing-shadowing-guided response protocol, not to passive or unrestricted viewing.

2.3 Hypotheses

H1: The experimental group will achieve significantly higher speaking scores than the control group after the intervention.

H0: There is no significant difference in speaking scores between the experimental group and the control group after the intervention.

3. Literature Review

3.1 Speaking Performance and Its Assessment

Speaking combines production and interaction. Bygate (1987) distinguishes the resources used to produce utterances from the skills needed to manage communication with an interlocutor. This distinction is important because improvement in individual delivery does not necessarily imply equivalent development in spontaneous interaction. An analytic assessment can make such differences visible by rating distinct dimensions rather than reducing performance to a single global judgment (Brown, 2004).

Speaking performance was operationalized through four criteria. Grammar and Vocabulary concerned the range, accuracy, and appropriateness of linguistic resources. Pronunciation concerned intelligibility and control of stress, rhythm, and intonation. Fluency concerned the continuity and smoothness of speech, including hesitation and pausing. Interactive Communication concerned initiating, responding to, sustaining, and repairing an exchange. Each criterion was scored from 1 to 5, giving a total score between 4 and 20 (Nguyen, 2026).

The rubric was a researcher-adapted instrument, not an unchanged published test. Its linguistic and delivery dimensions drew on analytic assessment principles discussed by Brown (2004). The research also grounded the interactive dimension in the interactional competence work of Galaczi (2014) and Galaczi and Taylor (2018). Retaining a separate interactive criterion allowed the study to examine whether a largely model-based practice routine transferred equally to reciprocal communication.

3.2 Input, Output, and Scaffolded Speaking Practice

Three complementary theoretical perspectives informed the intervention. Krashen's (1982) Input Hypothesis highlights the importance of comprehensible target-language input. In this study, researcher-selected videos presented spoken English in meaningful contexts and at a level intended to be manageable for the learners. Repeated viewing, visual information, and optional captions supported access to the language of the model. The theoretical expectation was that repeated exposure would provide resources that learners could subsequently use in their own speech.

Swain's (1985) Output Hypothesis complements this account by emphasizing productive language use. Watching a video does not require learners to retrieve words or articulate their own utterances. Shadowing, by contrast, requires them to reproduce the model's pronunciation, rhythm, and phrasing. Guided responses require a further step: learners must draw on their own linguistic resources to formulate a meaningful answer. These activities provide opportunities to notice differences between intended and actual production rather than remaining at the level of receptive understanding.

Vygotsky's (1978) sociocultural perspective provides an account of the support embedded in the sequence. Model utterances, captions, repeated playback, worksheets, and teacher-led follow-up activities can mediate performance that learners might initially find difficult to sustain independently. The sequence moves from exposure to supported

reproduction and then toward more independent response. These theoretical perspectives explain the instructional design; they were not measured as separate mediating variables in the statistical analysis.

Theories of practice and interaction also help explain why outcomes may differ across speaking criteria. Repetition can support more efficient retrieval and delivery (DeKeyser, 2007), but responding to an unpredictable interlocutor involves demands that a recorded model cannot fully reproduce. Long's (1996) emphasis on interaction and the negotiation of meaning therefore suggests a reason to examine interactive communication separately from pronunciation and fluency.

3.3 YouTube Videos, Shadowing, and Guided Responses

YouTube material can combine auditory, visual, and textual information. Mayer's (2009) multimedia-learning perspective offers a rationale for considering these channels together, while the meta-analysis by Montero-Perez *et al.* (2013) reports benefits of captioned video for listening and vocabulary learning. Such findings support the use of video as input, but they should not be treated as direct evidence that viewing alone improves speaking. Productive tasks remain necessary to connect audiovisual exposure with the outcomes assessed here.

The intervention used pre-existing conversational and language-learning videos rather than teacher-produced recordings. Their value lay in the availability of clear spoken models and contextualized language. Authenticity and task suitability nevertheless require careful consideration: materials need to be relevant and appropriately demanding rather than merely unedited or entertaining (Gilmore, 2007; Guariento & Morley, 2001). The intervention therefore used explicit selection criteria and a common weekly procedure.

Shadowing was defined as repeating the model during or immediately after listening, with attention to pronunciation, timing, rhythm, and intonation. Short segments and replay allowed the activity to be adjusted to the learner's needs. Self-recording then provided a means of listening back to one's own production. Guided comprehension and opinion questions extended practice beyond imitation by requiring learners to formulate responses connected to the content of the video. The systematic review by Whitworth and Rose (2025) supports the relevance of shadowing for pronunciation-related instruction while also showing that outcomes should be differentiated rather than assumed to improve uniformly.

3.4 Previous Empirical Studies and the Research Gap

The international literature provides a basis for investigating the proposed sequence. Purnamasari (2018) reported positive learner perceptions of YouTube-supported pronunciation practice in a blended environment. Saed *et al.* (2021) found that YouTube-supported instruction contributed to the speaking development of Jordanian EFL university students. Muslem *et al.* (2022) similarly reported improved speaking following the use of authentic English YouTube videos in a rural Indonesian school. These studies

suggest pedagogical potential, but their settings, participant profiles, and measures differ from those of the present study.

Vietnamese studies also support further investigation rather than a universal claim of effectiveness. Oanh and Van (2020) examined spoken production through vlog-based practice, while Dinh and Phung (2023) found a video-related advantage on a speaking assessment among secondary learners. Extending this work to a regional university required attention to the characteristics of non-English-major students, the regular course syllabus, and the nature of the comparison condition. A protocol based on existing videos, explicit output tasks, and an analytic speaking measure offered a way to address these concerns.

3.5 Conceptual Research Framework

The independent variable was the instructional condition: structured YouTube-based speaking practice or conventional speaking practice. The dependent variable was speaking performance, represented by the total rubric score and its four criteria. Both conditions shared the regular syllabus and intended practice time. The conceptual proposition was that the integrated input, shadowing, and guided-response sequence would produce a post-test advantage, particularly in dimensions directly rehearsed through repeated listening and production. Because the components were delivered together, the framework concerned the package as a whole rather than the isolated contribution of YouTube, shadowing, or teacher guidance.

4. Material and Methods

4.1 Research Design

A randomized pre-test/post-test control-group design was employed. Seventy participants from one class were individually and randomly assigned to the experimental group or the control group. Both groups completed a speaking pre-test before the intervention and a parallel post-test after 12 weeks. The pre-test and post-test used different, curriculum-aligned forms with the same task structure. The experimental group received the structured video-based protocol, while the control group received conventional speaking practice. The comparison was therefore between two forms of instruction, not between instruction and an inactive condition.

4.2 Research Site, Population, and Sample

The study took place at Tay Do University in Can Tho City, Vietnam. Participants were second-year Finance and Banking undergraduates enrolled in General English Speaking 2. They were non-English majors who had completed foundational General English coursework. The accessible sample comprised all 70 eligible students enrolled in one General English Speaking 2 class taught by the same lecturer. This accessible institutional sample was not a random sample of the wider population of Vietnamese EFL learners.

A computerized random-number procedure was used to allocate individual students to the experimental group ($n = 35$) and control group ($n = 35$). The experimental group included 19 female and 16 male students; the control group included 18 female and 17 male students. Students continued to attend their original timetabled sections. During the weekly follow-up slot, condition-specific breakout groups were used so that the assigned practice could be implemented without changing the overall timetable. All 70 participants completed both speaking tests, and there were no missing cases in the paired comparisons.

4.3 Speaking Tests and Scoring

Form A was used at pre-test and Form B at post-test. Each form included a short warm-up interview of approximately one to two minutes, a topic-selection and preparation stage, and a two-minute individual long turn. The warm-up elicited responses to two or three questions; learners then prepared for approximately one minute before speaking on the agreed topic. An optional closing question could provide further evidence of interactive communication. The tests were audio-recorded for scoring.

The forms were mapped to the communicative content of Four Corners 3, the coursebook used in General English Speaking 2. Different but related syllabus topics were used to reduce direct rehearsal of an identical test prompt. A panel of three experienced English-language-teaching colleagues reviewed the forms for clarity, appropriateness, topical familiarity, and comparable task demands. This review followed the instrument-development emphasis on checking suitability before main data collection (Dörnyei, 2007). The equivalence claim rested on content mapping and expert judgment, not on a separately reported statistical equating study.

The four-criterion analytic rubric described in Section 2.1 was applied at both testing occasions. Ratings ranged from 1, indicating very limited performance, to 5, indicating excellent performance. The total score was the sum of the four criterion scores. Trained raters used common descriptors and discussed scoring procedures before the main assessment. An absolute-agreement intraclass correlation of .84 was obtained from a randomly selected 20% subsample of recordings.

4.4 Materials and Video Selection

The intervention combined the regular coursebook, researcher-selected YouTube videos, viewing-shadowing-response worksheets, and teacher-led follow-up tasks. Videos came from Speak English With Vanessa, Learn English with Bob the Canadian, and the BBC 6 Minute English series. Selection considered topic alignment, difficulty relative to the learners' course, clear audio and visual quality, an identifiable accent within each video, and content suitable for university learners. The materials included conversational lessons and topic-based language-learning videos.

Video running times ranged approximately from 6 to 27 minutes. For longer videos, learners shadowed two or three selected segments of approximately three to five minutes rather than the entire recording. The materials were selected in advance by the

researcher and reviewed against the course syllabus. Learners controlled when to practice, the use of replay and captions, and the segments used within the prescribed procedure; they did not independently determine the whole weekly curriculum.

4.5 Implementation of the Experimental Condition

At the beginning of each week, the teacher posted the assigned video link and worksheet through the class communication platforms. Learners first watched the video for general understanding and then watched it again while identifying new vocabulary and a useful expression. They subsequently shadowed short segments, repeating aloud and pausing or replaying when needed. At least one self-recording was made with a phone. The worksheet included comprehension questions, opinion questions, and a brief self-reflection checklist.

The independent component was planned to take approximately 45–60 minutes per week. Learners submitted their worksheet and a recording or short recording extract before the weekly deadline. The teacher checked completion and listened to a sample of recordings. A further 15–20-minute teacher-led follow-up slot was embedded in the regular speaking lesson. During this slot, learners responded to the prepared questions and clarified relevant vocabulary or comprehension points. The protocol therefore involved both independent preparation and guided classroom use of the prepared language.

4.6 Implementation of the Control Condition

The control group followed the same general syllabus, weekly topics, instructor, and assessment procedures. Its supplementary practice used conventional speaking tasks from the regular course materials without the structured YouTube viewing and shadowing sequence. The intended amount of practice time was matched to the experimental condition. Scheduled contact time was not increased for the experimental group; its in-class follow-up took place within the common timetable. The principal planned contrast concerned the type of instructional materials and speaking-practice procedure.

4.7 Data Analysis

Means and standard deviations were used to summarize the four criterion scores and the total score. Independent-samples t-tests compared groups at pre-test and post-test, and paired-samples t-tests examined pre-test-to-post-test change within each group. Tests were two-tailed and interpreted at $\alpha = .05$. Cohen's d was reported alongside significance tests to describe the magnitude of the reported differences (Cohen, 1988; Plonsky & Oswald, 2014).

The study tested post-test total-score normality and reported significant departures from normality in both groups. The primary t-tests were retained on the basis of the equal group sizes, and Mann–Whitney U and Wilcoxon signed-rank tests were used as robustness checks. Their significance patterns were compared with those of the

corresponding parametric analyses (Field, 2018). The pre-test comparison was used to describe the initial scores rather than to establish that all measured and unmeasured characteristics were identical.

4.8 Ethical Considerations

Participation was voluntary, and informed consent was obtained before data collection. Learners were informed that withdrawal would not result in a grade penalty. Recordings and responses were stored securely, and reporting used aggregate results or anonymized codes. The control group continued to receive its regular curriculum. The research also recognized the need to reduce perceived pressure arising from the researcher's instructional role at the institution (Nguyen, 2026).

5. Results and Discussions

5.1 Preliminary Checks and Pre-test Performance

All 70 learners were included in the pre-test and post-test analyses. The Shapiro–Wilk tests for the post-test total score indicated departures from normality in the experimental group ($p = .018$) and control group ($p = .043$). The non-parametric analyses reproduced the substantive significance pattern of the primary analyses, supporting the stability of the conclusions across the two types of procedure.

Table 1: Pre-test speaking performance of the experimental and control groups

Criterion	Experimental M (SD)	Control M (SD)	t(68)	p	Cohen's d
Grammar and vocabulary	2.31 (0.72)	2.60 (0.60)	-1.80	.076	-0.43
Pronunciation	2.43 (0.70)	2.57 (0.74)	-0.83	.409	-0.20
Fluency	2.20 (0.87)	2.43 (0.70)	-1.21	.229	-0.29
Interactive communication	2.46 (0.74)	2.46 (0.78)	0.00	1.000	0.00
Total score	9.40 (1.61)	10.06 (1.30)	-1.87	.065	-0.45

Note: $n = 35$ per group. Each criterion is scored from 1 to 5; total scores range from 4 to 20.

As shown in Table 1, none of the five pre-test comparisons reached the .05 significance level. The groups had the same reported mean for interactive communication, while the control group had higher means for the other criteria and the total score. The initial total-score difference was not statistically significant, $t(68) = -1.87$, $p = .065$, but its direction and magnitude should still be recognized: the control group began with a 0.66-point advantage in the rounded means and a reported standardized difference of $d = -0.45$. Thus, the results show no detected pre-test difference at the selected threshold, not proof of exact equivalence.

5.2 Post-test Performance of the Two Groups

Table 2: Post-test speaking performance of the experimental and control groups

Criterion	Experimental M (SD)	Control M (SD)	t(68)	p	Cohen's d
Grammar and vocabulary	3.49 (0.70)	2.94 (0.76)	3.09	.003	0.74
Pronunciation	3.63 (0.65)	3.11 (0.72)	3.15	.002	0.75
Fluency	3.40 (0.69)	2.91 (0.85)	2.61	.011	0.62
Interactive communication	3.17 (0.75)	3.11 (0.63)	0.35	.731	0.08
Total score	13.69 (1.41)	12.09 (1.44)	4.69	< .001	1.12

Note: n = 35 per group. Positive d values favor the experimental group. Mann-Whitney tests reproduced the same significance pattern.

At post-test, the experimental group obtained a higher overall speaking score than the control group, $t(68) = 4.69$, $p < .001$, $d = 1.12$ (Table 2). Its total mean was 13.69, compared with 12.09 for the control group. This result supports the alternative hypothesis for overall speaking performance. The reported effect size indicates a substantial standardized between-group difference in this sample, but it concerns the combined instructional protocol rather than the platform alone.

The criterion-level results were differentiated. The experimental group scored higher in grammar and vocabulary, $t(68) = 3.09$, $p = .003$, $d = 0.74$; pronunciation, $t(68) = 3.15$, $p = .002$, $d = 0.75$; and fluency, $t(68) = 2.61$, $p = .011$, $d = 0.62$. Pronunciation, grammar and vocabulary showed the largest of these criterion-level standardized differences. In contrast, interactive communication means were very similar: 3.17 in the experimental group and 3.11 in the control group, $t(68) = 0.35$, $p = .731$, $d = 0.08$. The intervention therefore did not produce a statistically significant advantage on every aspect of speaking.

5.3 Within-group Changes from Pre-test to Post-test

Table 3: Pre-test-to-post-test changes within each group

Group/criterion	Pre-test M	Post-test M	Gain	SD of gain	t(34)	p	d
Experimental: grammar and vocabulary	2.31	3.49	1.17	1.10	6.31	< .001	1.07
Experimental: pronunciation	2.43	3.63	1.20	0.99	7.14	< .001	1.21
Experimental: fluency	2.20	3.40	1.20	1.05	6.75	< .001	1.14
Experimental: interactive communication	2.46	3.17	0.71	1.07	3.94	< .001	0.67
Experimental: total score	9.40	13.69	4.29	2.14	11.87	< .001	2.01
Control: grammar and vocabulary	2.60	2.94	0.34	0.68	2.97	.005	0.50
Control: pronunciation	2.57	3.11	0.54	0.95	3.38	.002	0.57
Control: fluency	2.43	2.91	0.49	1.07	2.69	.011	0.46
Control: interactive communication	2.46	3.11	0.66	1.06	3.68	< .001	0.62
Control: total score	10.06	12.09	2.03	1.62	7.42	< .001	1.25

Notes: n = 35 per group. Gain = post-test minus pre-test. Paired-samples effect sizes represent within-group changes, not between-group effects. Displayed means and gains were rounded independently. Wilcoxon signed-rank tests indicated significant changes for all comparisons.

Both groups improved significantly on every criterion and on the total score (Table 3). The experimental group's total gain was 4.29 points, compared with 2.03 points in the control group. The experimental group's gains in grammar and vocabulary, pronunciation, and fluency were also descriptively larger. Interactive communication showed a different pattern: gains were 0.71 and 0.66 points, respectively. This near-parallel improvement is consistent with the absence of a significant post-test difference on the interactive criterion.

The control-group gains are important for interpreting the comparison. These learners continued to receive speaking instruction and supplementary practice, so improvement in both groups was plausible. The intervention's advantage is supported by the between-group post-test analyses, not simply by the fact that the experimental group improved. Likewise, the within-group effect sizes describe change within each condition and should not be substituted for the between-group treatment comparison. Mean gains were calculated in the analysis before rounding, which explains small differences between reported gains and the subtraction of displayed means.

5.4 Discussion of the Linguistic and Delivery-related Outcomes

The positive results for grammar and vocabulary, pronunciation, and fluency are consistent with the instructional activities most directly emphasized in the protocol. Video viewing provided repeated exposure to contextualized language, while shadowing required learners to articulate the model rather than merely understand it. Self-recording and guided response offered additional opportunities to attend to delivery and use language in personally formulated answers. This interpretation fits the complementary input and output perspectives of Krashen (1982) and Swain (1985). It remains a theoretical explanation of the observed pattern rather than a tested mediation model.

Pronunciation practice was especially explicit. Learners repeatedly heard a model and attempted to reproduce its stress, rhythm, intonation, and phrasing. Repetition may also have reduced the processing demands of producing familiar sequences, supporting smoother delivery as discussed by DeKeyser (2007) and Thornbury (2005). The findings align with the potential of shadowing identified by Whitworth and Rose (2025), while the rubric-based measurement here does not permit conclusions about each individual sound or suprasegmental feature.

The results also converge with the speaking-related benefits reported by Saed *et al.* (2021) and Muslem *et al.* (2022). In the Vietnamese context, the positive performance comparison is broadly consistent with Dinh and Phung (2023), although their study involved high-school learners rather than undergraduates. The present study contributes evidence from non-English-major Finance and Banking students at a regional university. It should not be treated as demonstrating that the same effect size will occur with other proficiency levels, disciplines, or institutional settings.

5.5 Interactive Communication and the Boundaries of the Effect

Interactive communication was the exception to the experimental group's post-test advantage. This does not mean that learners failed to develop in this dimension: both groups made significant within-group gains. Rather, the structured video-based condition did not outperform conventional practice on the measured ability to initiate, respond to, and sustain an exchange. This distinction prevents the total-score result from being interpreted as a uniform improvement in all speaking capacities.

Much of the experimental routine was completed individually. Shadowing required a response to a recorded model, but the model could not react to the learner, request clarification, or jointly negotiate meaning. Guided answers and the weekly follow-up introduced interaction, yet they did not fully reproduce sustained, unpredictable conversation. Vygotsky's (1978) account of socially mediated development and Long's (1996) interaction-oriented perspective offer a plausible explanation for the more limited differential effect on interactive communication.

The assessment context also matters. Interactive communication was elicited mainly through a short examiner-learner exchange rather than an extended peer discussion. The result therefore concerns the criterion as measured by this test. Overall, the findings suggest that individual model-based production and reciprocal interaction should be treated as related but distinguishable targets when designing supplementary speaking practice.

6. Recommendations

Teachers may use existing YouTube videos as part of a clearly specified speaking routine rather than giving an unrestricted instruction to watch English videos. Selection should consider syllabus relevance, manageable difficulty, clear spoken models, and realistic workload. Viewing should be followed by output: segment-based shadowing, self-recording, and questions that require learners to express meaning in their own words. These elements reflect the integrated protocol evaluated in the study, not independently established effects of each activity.

When interactive communication is a central objective, the protocol should be complemented by pair or small-group exchanges in which learners respond to unpredictable contributions, ask for clarification, and develop ideas collaboratively. Such activities would address demands that recorded models and rehearsed individual answers cannot fully supply. Their added effectiveness would still need to be tested rather than assumed from the present comparison.

For learners, active participation in the complete routine is preferable to passive viewing alone. Short recordings and comparison with the model can make recurring pronunciation and fluency concerns more visible. For course designers, the procedure offers a documented way to connect supplementary practice to an existing syllabus without treating digital material as a replacement for the regular course or the teacher's pedagogical role.

Further research should use larger and more diverse samples, include delayed assessment, and examine ways of minimizing the exchange of intervention material between conditions. Component-based studies could distinguish the contribution of viewing, shadowing, self-recording, and guided response. More extended interaction tasks would also help establish whether additional reciprocal practice changes the interactive-communication outcome. These directions follow from the limits of the original design rather than from analyses newly undertaken for this article.

7. Conclusions

The study provides context-specific evidence that a 12-week structured YouTube-based speaking-practice program can support the speaking performance of non-English-major EFL undergraduates compared with time-matched conventional practice. The experimental group achieved a higher post-test total score and higher scores in grammar and vocabulary, pronunciation, and fluency. The overall hypothesis was supported, but the criterion-level advantage did not extend to interactive communication. Both groups improved, highlighting the value of interpreting the experimental comparison alongside the continuing contribution of conventional instruction.

Several limitations qualify the conclusion. Participants came from one accessible cohort, one academic discipline, and one university. Individual random assignment strengthened the design's internal comparison but did not make the sample representative of all EFL learners (Shadish *et al.*, 2002). Students remained within the same cohort, so exchange of materials or practice procedures across conditions could not be completely prevented. The intervention combined several instructional elements, making attribution to video viewing alone inappropriate.

The study also covered only one semester and included no delayed post-test. Its evidence therefore concerns immediate post-intervention performance rather than long-term maintenance. The test forms were supported through curriculum mapping and expert review, and the reported analyses compared group means without isolating every possible source of performance variation. These constraints should be considered alongside the observed baseline control-group advantage and the post-test normality findings. Within these boundaries, the main instructional implication is specific: structured access to spoken models becomes useful for speaking when it is linked to purposeful production and teacher-supported practice. The program is best understood as a supplementary instructional package, with stronger evidence for linguistic resources and delivery than for a differential improvement in spontaneous interactive communication.

Acknowledgements

The authors thank the participating students and the teachers and colleagues at Tay Do University who supported the research procedures.

Creative Commons License Statement

This research work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0>. To view the complete legal code, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.en>. Under the terms of this license, members of the community may copy, distribute, and transmit the article, provided that proper, prominent, and unambiguous attribution is given to the authors, and the material is not used for commercial purposes or modified in any way. Reuse is only allowed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Conflict of Interest Statement

The authors declare no conflict of interest.

About the Author

Nguyen Tran Nhu Hao is a student of the Master of Education program in Principles and Methods in English Language Education at Can Tho University, Vietnam.

Kim Thanh Tuyen is a Doctor of Philosophy. She is currently a lecturer in the Social Science Department at Can Tho University of Technology, Vietnam. Her research interests include Autonomy learning, Culture, Process Genre Approach (PGA) in Writing, Speaking, Teaching, Testing and Assessment, and Curriculum Development.

References

- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Longman. Retrieved from https://books.google.ro/books/about/Language_Assessment.html?id=NqglAQAAIAAJ&redir_esc=y
- Bygate, M. (1987). *Speaking*. Oxford University Press. Retrieved from https://books.google.ro/books/about/Speaking.html?id=XozCwaqSJFIC&redir_esc=y
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. <https://doi.org/10.4324/9781315456539>
- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Lawrence Erlbaum Associates. Retrieved from <https://psycnet.apa.org/record/2006-20180-006>

- Dinh, T. H. T., & Phung, V. D. (2023). The effects of using videos in English speaking lessons on high school students' speaking performance, Kien Giang Province, Vietnam. *European Journal of English Language Teaching*, 8(4), 129–157. <https://doi.org/10.46827/ejel.v8i4.5011>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118. <https://doi.org/10.1017/S0261444807004144>
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 553–574. <https://doi.org/10.1093/applin/amt017>
- Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219–236. <https://doi.org/10.1080/15434303.2018.1453816>
- Guariento, W., & Morley, J. (2001). Text and task authenticity in the EFL classroom. *ELT Journal*, 55(4), 347–353. <https://doi.org/10.1093/elt/55.4.347>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. Retrieved from <https://www.ling.ohio-state.edu/research/groups/lacqueys/readings/horowitz.pdf>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press. Retrieved from https://www.sdkrashen.com/content/books/principles_and_practice.pdf
- Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching*, 40(3), 243–249. Retrieved from <https://www.cambridge.org/core/journals/language-teaching/article/abs/communicative-and-taskbased-language-teaching-in-east-asian-classrooms/42FAB897C9A7255F9843C090649E1987>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press. Retrieved from <https://www.library.brawnblog.com/Long1996-Ling%20Env%20L2%20Acq.pdf>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Montero-Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- Muslem, A., Fata, I. A., & Saputri, W. I. (2022). The effect of English authentic YouTube videos on students' speaking skills in a rural school. *Journal of Education Technology*, 6(3), 423–430. <https://doi.org/10.23887/jet.v6i3.45412>

- Nguyen, T. N. H. (2026). *The effect of YouTube videos on EFL learners' speaking skills* [Master's thesis, Can Tho University].
- Nhat, T. T. M. (2018). Thực trạng kỹ năng nói tiếng Anh và đề xuất một số hoạt động tự rèn luyện nói tiếng Anh ngoài lớp học cho sinh viên chuyên ngành kỹ thuật Trường Đại học Công nghiệp Thành phố Hồ Chí Minh. *Tạp chí Giáo dục*, 435(1), 54–59.
- Oanh, H. T. K., & Van, N. T. H. (2020). Steps to improve English speaking skills for students of Electric Power University through vlog-based practice. *Journal of Science, Hanoi Open University*, 67, 21–30.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Purnamasari, A. (2018). What EFL learners say about YouTube use to improve pronunciation in a blended learning environment. *Lingua Cultura*, 12(3), 233–238.
- Saed, H. A., Haider, A. S., Al-Salman, S., & Hussein, R. F. (2021). The use of YouTube in developing the speaking skills of Jordanian EFL university students. *Heliyon*, 7(7), Article e07543. <https://doi.org/10.1016/j.heliyon.2021.e07543>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin. Retrieved from <https://psycnet.apa.org/record/2002-17373-000>
- Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan. Retrieved from <https://link.springer.com/book/10.1057/978-1-137-46048-6>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House. Retrieved from <https://www.taylorfrancis.com/chapters/mono/10.4324/9781315835877-13/communicative-competence-roles-comprehensible-input-comprehensible-output-development-jim-cummins-merrill-swain>
- Thornbury, S. (2005). *How to teach speaking*. Longman.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. *Research Synthesis in Applied Linguistics*, 1(2), 239–269. <https://doi.org/10.1080/29984475.2025.2546827>