



EFL LEARNERS' PERCEPTIONS, MOTIVATION, AND SELF-DIRECTED LEARNING IN STRUCTURED YOUTUBE-BASED SPEAKING PRACTICE

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 explored EFL learners' perceptions of structured YouTube-based speaking practice and the levels of motivation and self-directed learning they reported after a 12-week program at Tay Do University, Vietnam. Participants were 35 second-year non-English-major Finance and Banking undergraduates who completed the YouTube-based condition. Data came from three six-item questionnaire subscales and semi-structured interviews with 10 purposively selected participants. Questionnaire responses were summarized descriptively, and the interview findings were organized according to the thematic analysis reported in the research. Learners reported positive perceptions ($M = 4.06$, $SD = 0.45$) and motivation ($M = 3.83$, $SD = 0.50$), while self-directed learning was less strongly endorsed ($M = 3.57$, $SD = 0.47$). The three interview themes were structured self-practice routines, perceived gains in pronunciation and fluency, and growing confidence to speak. Participants described repeated viewing, caption use, shadowing, self-recording, and preparation of spoken responses. However, following a prescribed routine did not necessarily mean independently planning and evaluating learning. The findings suggest that favorable attitudes and motivation can coexist with a continuing need for support in self-directed learning. Interpretation is limited by the small, single-institution sample, post-intervention self-report design, and marginal internal consistency of the self-directed learning subscale ($\alpha = .69$). Gradual transfer of responsibility, explicit self-monitoring support, and further validation of the measure are recommended.

Keywords: YouTube; learner perceptions; motivation; self-directed learning; EFL speaking practice

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

1. Introduction

The educational value of a digital speaking activity depends not only on the availability of language input but also on how learners perceive, engage with, and manage the activity. YouTube-based practice may be viewed as useful and enjoyable without necessarily becoming an independently sustained learning routine. This article examines that distinction through the perceptions, motivation, and self-directed learning reported by non-English-major undergraduates following a structured speaking-practice program.

1.1 Background of the Study

For university learners who have limited opportunities to speak English outside class, online video can extend access to spoken models and familiar topics. It also allows repeated listening and practice at a personally convenient time. Research on extramural English and informal online learning recognizes such activities as relevant sources of language contact beyond formal lessons (Sockett, 2014; Sundqvist & Sylvén, 2016). Yet access to language and the ability to organize an effective learning process are different matters.

The research was conducted at Tay Do University in Can Tho City, Vietnam, where the researcher identified a need for supplementary speaking opportunities among non-English-major students. The intervention addressed that need by prescribing a weekly sequence of viewing, shadowing, self-recording, and guided response. Videos were selected by the researcher and linked to the General English Speaking 2 syllabus. Learners could control aspects of their practice, but the overall sequence and materials were substantially structured.

Such an arrangement raises a pedagogical question. A prescribed routine can make participation manageable by reducing uncertainty about suitable materials and expected tasks. At the same time, it may require less independent decision-making than a learner-designed program. Reinders and White (2016) and Godwin-Jones (2019) emphasize that opportunities created by technology do not automatically produce learner autonomy. It is therefore necessary to examine students' reported experiences rather than assume that out-of-class digital practice is inherently self-directed.

2. Research Problem Statements

Previous studies have reported favorable perceptions of YouTube for English learning and speaking practice. Purnamasari (2018) examined learners' perceptions of pronunciation practice, while Truong and Le (2022) investigated tertiary students' practices and perceptions of YouTube use for speaking. Ho and Le (2023) also addressed video materials, learning, motivation, and autonomy in the Vietnamese context. These findings justify attention to learner experience, but they do not remove the need to

distinguish positive evaluation of an activity from motivation to undertake it and from independent management of learning.

The research identified this distinction as particularly relevant to second-year non-English majors at a regional university. Learners were not selecting an entirely personal video-learning curriculum; they were responding to a teacher-supported program. The present article therefore treats perceptions, motivation, and self-directed learning as related but separately described dimensions. It focuses on their reported levels after participation and on interview accounts that help explain the descriptive pattern.

2.1 Research Objective

The objective was to explore how learners perceived the 12-week YouTube-based speaking-practice program and what levels of motivation and self-directed learning they reported. A further purpose was to use the interview findings to describe how participants engaged with the routine and understood their speaking-related experiences.

2.2 Research Question

- How do EFL learners perceive YouTube-based speaking practice, and what levels of motivation and self-directed learning do they report?

2.3 Scope of the Inquiry

This article uses questionnaire responses from the 35 learners in the experimental condition and the relevant interview material from a 10-participant subsample. It does not test a treatment effect on motivation or self-directed learning because these measures were collected after the intervention from the experimental group only. No directional or causal hypothesis was formulated for this descriptive question. The analysis focuses exclusively on learners' perceptions, motivation, and self-directed learning after the program; speaking-performance outcomes and reported difficulties are beyond its scope.

3. Literature Review

3.1 Learner Perceptions of YouTube-based Speaking Practice

In this study, perception referred to learners' subjective beliefs, opinions, and evaluations of the usefulness, enjoyment, and effectiveness of the practice. These evaluations are important because an activity that learners find worthwhile may be easier to incorporate into a continuing study routine. However, a favorable judgment remains a perception; it does not independently establish improvement in tested speaking ability.

YouTube videos may be positively evaluated because they present spoken language in context and provide repeatable models. Auditory information can be supported by visual context and captions. Mayer's (2009) multimedia-learning perspective and the captioned-video findings reviewed by Montero-Perez *et al.* (2013) offer a rationale for considering these supports. Within the present protocol, learners

could connect the input to shadowing and guided questions, rather than being asked simply to watch more videos.

Purnamasari (2018) reported favorable learner accounts of YouTube-supported pronunciation practice. Truong and Le (2022) likewise found positive practices and perceptions among tertiary learners using YouTube to support speaking. Saed *et al.* (2021) demonstrated the relevance of YouTube to university EFL speaking instruction. The present inquiry builds on this literature by examining perceptions within a defined, 12-week program for non-English majors rather than treating all uses of the platform as equivalent.

3.2 Motivation for Speaking Practice

Motivation concerns the orientation and effort that learners bring to language learning. Gardner (1985) connects attitudes and motivation to second-language learning, while Dörnyei and Ushioda (2011) emphasize motivation as an important area of language-learning inquiry. In the present study, the motivation items addressed willingness to improve speaking, anticipation of weekly practice, aspiration to speak like the models, the perceived importance of English, effort during practice, and willingness to recommend the activity.

These indicators overlap in their practical relevance but do not measure exactly the same experience. A learner may value English for the future while feeling less enthusiastic about a particular weekly task. Similarly, willingness to recommend an activity is not identical to sustained independent effort. Reporting the items alongside the subscale therefore helps make the scope of the motivation measure transparent.

The research proposed that engaging content, contextualized language, and manageable practice might encourage participation. Krashen's (1982) discussion of comprehensible input and affective conditions provides one lens for interpreting a positive learner response. Nevertheless, motivation was investigated as a reported experience, not assumed to improve simply because videos were used. Without a baseline motivation measure or a comparable control-group measure, the analysis can describe endorsement after the program but cannot establish motivational change caused by the program.

3.3 Self-directed Learning and Digital Autonomy

Knowles (1975) describes self-directed learning in terms of learners taking initiative and responsibility for aspects of their learning. In the source questionnaire, this construct was operationalized through goal setting, consideration of the match between assigned material and personal needs, progress monitoring, time management, searching for additional videos, and confidence in learning to speak independently. The items were adapted from the Self-Directed Learning Readiness Scale developed by Fisher *et al.* (2001).

These behaviors differ from simply completing work outside class. A learner can independently access a video while still relying on a teacher to determine the learning

goal, choose the content, sequence the tasks, and judge progress. Benson (2011) emphasizes learners' control over learning, and Reinders and White (2016) caution against equating technological opportunity with autonomy. This distinction is central to interpreting self-directed learning within a highly structured intervention.

Zimmerman's (2002) account of self-regulated learning highlights forethought, performance monitoring, and self-reflection. Although self-regulation and self-directed learning are not interchangeable labels, this account helps interpret the planning and monitoring behaviors investigated in the study. Godwin-Jones (2019) similarly notes the need for learners to manage informal digital environments effectively. The present study therefore considered self-directed learning a dimension requiring direct investigation, not an automatic consequence of favorable perceptions or access to online material.

3.4 Structure, Output, and Learner Responsibility

The protocol combined opportunities for learner control with substantial teacher support. Learners could choose when and where to practice, pause or repeat segments, use captions, and listen back to recordings. Yet the researcher selected the weekly videos and prescribed the sequence. Vygotsky's (1978) sociocultural perspective provides a rationale for such mediation: support can make a difficult activity manageable before more responsibility is transferred to the learner.

Swain's (1985) Output Hypothesis also informs the distinction between engagement with content and active language practice. Shadowing and guided answers require learners to produce language, whereas viewing can remain receptive. The interview accounts are therefore useful for identifying whether learners described deliberate production, planning, and monitoring rather than simply time spent watching. These accounts illuminate participants' experiences; they do not isolate the psychological mechanisms responsible for language development.

3.5 Previous Studies and the Conceptual Research Framework

The related studies reviewed in the study indicate that positive attitudes toward video can coexist with continuing needs for support. Ho and Le (2023) examined video materials among English-major learners, while Truong and Le (2022) considered tertiary learners' practices and perceptions. Nhat (2018) identified limited out-of-class speaking opportunities in a non-English-major technical context. Differences in proficiency, degree program, and prior independent-learning experience make a context-specific inquiry necessary rather than assuming identical responses across these populations.

The conceptual framework for this article places structured YouTube-based speaking practice as the shared learning context. Perception, motivation, and self-directed learning are three dimensions of learner experience described through the questionnaire. Interview themes provide an accompanying account of practice routines, perceived speaking changes, and confidence. The dimensions are not modeled as mediators, moderators, or statistically tested predictors of speaking achievement. Their

relationship is considered at the level of interpretation, consistent with the exploratory purpose of the study.

5. Material and Methods

5.1 Research Design

A descriptive mixed-methods design was employed. Quantitative data were obtained from the post-intervention questionnaire completed by 35 learners who participated in the YouTube-based speaking-practice program. Qualitative data were obtained from semi-structured interviews with 10 purposively selected participants. The interview findings were used to elaborate on the questionnaire results.

5.2 Setting and Participants

Participants were second-year Finance and Banking undergraduates taking General English Speaking 2 at Tay Do University in Can Tho City, Vietnam. They were non-English majors with prior foundational General English coursework. The questionnaire group comprised 19 female and 16 male students. All participants belonged to the same accessible cohort, so the findings primarily concern this institutional and curricular context.

A purposive subsample of 10 learners participated in semi-structured interviews. Selection sought variation in pre-test speaking achievement and self-reported YouTube use, with approximately four higher-achieving, four average-achieving, and two lower-achieving learners. Pre-test information served this sampling purpose; the present article does not link individual test-score gains to individual questionnaire answers or interview accounts. The interview subsample was designed to provide varied perspectives rather than a statistically representative estimate of all learners' experiences.

5.3 The YouTube-based Speaking-practice Program

For 12 weeks, learners used researcher-selected videos aligned with the communicative topics of Four Corners 3. The sources were Speak English With Vanessa, Learn English with Bob the Canadian, and BBC 6 Minute English. Video selection considered difficulty, relevance, accent clarity, and audio-visual quality. Videos ranged approximately from 6 to 27 minutes, with selected segments used for shadowing longer material. The program supplemented rather than replaced the regular course.

Each weekly cycle required viewing for general understanding, repeated viewing with attention to language, shadowing, self-recording, and completion of comprehension and opinion questions. A short self-reflection checklist accompanied the worksheet. Learners submitted their work electronically, and the teacher monitored completion and a sample of recordings. The independent work was planned to take 45–60 minutes, followed by a 15–20-minute teacher-led speaking slot embedded in the regular lesson.

The program thus provided a common structure but allowed some procedural choice. Learners could decide when and where to work, how often to replay a segment,

and how to use captions. Their own accounts also described personal adaptations, such as written preparation or practice with a classmate. These choices are relevant to the interpretation of self-directed learning, but the assigned videos and basic sequence remained teacher-determined.

5.4 Questionnaire and Instrument Development

The bilingual questionnaire used a five-point response scale from 1 (strongly disagree) to 5 (strongly agree). The full instrument contained four six-item subscales and an open-ended item. This article reports Perception, Motivation, and Self-Directed Learning: 18 items in total. Perception items were adapted from Purnamasari (2018) and Saed *et al.* (2021); motivation items were adapted from Gardner's (1985) Attitude/Motivation Test Battery; and the self-directed learning items were adapted from Fisher *et al.* (2001). The wording was contextualized for YouTube-based speaking practice. The Vietnamese version was checked through back-translation by an independent bilingual translator and reviewed by experienced English-language-teaching colleagues for clarity and cultural appropriateness. Revisions included clarification of self-directed learning items before administration.

Internal consistency was examined using the 35 questionnaire responses. Cronbach's alpha was .72 for Perception, .78 for Motivation, and .69 for Self-Directed Learning. The latter was retained in the study but interpreted cautiously because it was slightly below the commonly cited .70 benchmark. As Taber (2018) explains, alpha should not be treated as an absolute statement of measurement quality; its interpretation depends on the instrument and context. Nevertheless, the marginal value remains a limitation rather than evidence that self-directed learning was measured with the same confidence as the other dimensions.

5.5 Interviews and Qualitative Evidence

The interview guide included questions about learners' weekly use of the assigned videos, perceived speaking changes, difficulties, and suggestions. The present analysis focused on learners' practice routines and perceived speaking-related experiences. Interviews were conducted in Vietnamese, audio-recorded with consent, transcribed, and translated into English for reporting.

Interview data were analyzed using Braun and Clarke's (2006) six-phase thematic procedure: familiarization, initial coding, development of themes, review, definition and naming, and reporting. Short quotations are presented using anonymized participant codes P1–P10.

5.6 Data Analysis and Reporting

Questionnaire responses were summarized using means and standard deviations, consistent with the considerations for Likert-scale data discussed by Boone and Boone (2012). Overall subscale and item-level statistics are reported to two decimal places. No inferential comparison among the three subscales was conducted.

The ranking of subscale means is descriptive. No statistical test comparing the three subscale means, no relationship between questionnaire scores and speaking gains, and no causal model of motivation or self-directed learning is reported. Qualitative evidence is used to explain how participants understood the practice and to illustrate variation within the overall pattern. Neither quotations nor theme presence is converted into a numerical prevalence claim.

5.7 Ethical Considerations

Informed consent was obtained from all participants. Participation was voluntary, and students were informed of their right to withdraw without a grade penalty. Audio recording was conducted only with participants' consent. All data were anonymized using participant codes.

6. Results and Discussions

6.1 Overall Perception, Motivation, and Self-directed Learning

Table 1: Descriptive statistics and internal consistency of the three subscales

Subscale	Items	n	M	SD	Cronbach's α
Perception	6	35	4.06	0.45	.72
Motivation	6	35	3.83	0.50	.78
Self-Directed Learning	6	35	3.57	0.47	.69

Note: Responses ranged from 1 (strongly disagree) to 5 (strongly agree). Higher means indicate stronger endorsement.

Table 1 shows a positive overall response to the speaking-practice program. Perception received the highest mean ($M = 4.06$, $SD = 0.45$), followed by Motivation ($M = 3.83$, $SD = 0.50$). Self-Directed Learning was lower ($M = 3.57$, $SD = 0.47$), although still above the neutral midpoint of 3. The pattern should therefore be described as comparatively weaker endorsement of self-directed learning, not as general rejection of independent practice.

The standard deviations indicate variation within each dimension, while the average scores show that favorable evaluations of the activity were more strongly expressed than confidence in directing the learning process. This ordering is descriptive and should not be interpreted as a statistically tested difference between the three subscales. The following item-level results make clear what participants were endorsing and prevent the three overall means from being treated as interchangeable indicators.

6.2 Perceptions of the Speaking-practice Activities

Table 2: Item-level perceptions of YouTube-based speaking practice (n = 35)

Item	Statement	M	SD
Per1	Watching YouTube videos helped me understand spoken English better.	3.94	0.64
Per2	Shadowing the videos helped me improve my pronunciation.	4.06	0.59
Per3	Answering questions about the videos helped me practice speaking.	4.17	0.71
Per4	I found the YouTube-based speaking practice enjoyable.	3.94	0.76
Per5	I found the YouTube-based speaking practice useful for my English course.	4.20	0.72
Per6	I would like to continue using YouTube videos for speaking practice in the future.	4.06	0.76

Note: Questionnaire responses were summarized using means and standard deviations. Overall subscale scores and item-level statistics are reported to two decimal places.

The strongest perception item concerned usefulness for the English course (Per5: $M = 4.20$, $SD = 0.72$), followed by the value of answering questions for speaking practice (Per3: $M = 4.17$, $SD = 0.71$). These findings connect positive evaluation to the course and to productive tasks rather than to video entertainment alone. Perceived pronunciation benefit from shadowing and the intention to continue using videos both received means of 4.06.

The items on understanding spoken English and enjoyment each had a mean of 3.94. They remained favorable but were not the highest-rated items in the SPSS output. Taken together, the results indicate that participants valued several aspects of the integrated sequence. The questionnaire does not establish which individual component caused that evaluation or whether each perceived benefit was accompanied by an objectively measured change.

The positive response is consistent with the perceptions reported by Purnamasari (2018) and Truong and Le (2022). It also fits the rationale that contextualized spoken models become more useful when learners have clear follow-up tasks. In the present setting, the strong ratings for course usefulness and guided questions point to the relevance of pedagogical integration, not simply the availability of the platform.

6.3 Reported Motivation

Table 3: Item-level motivation for YouTube-based speaking practice (n = 35)

Item	Statement	M	SD
Mot1	I feel motivated to improve my English speaking after using YouTube videos.	3.69	0.72
Mot2	I look forward to my weekly YouTube speaking practice.	3.80	0.63
Mot3	I want to speak English as well as the speakers I watch on YouTube.	3.91	0.66
Mot4	Learning to speak English well is important to my future.	3.80	0.72
Mot5	I try my best in every YouTube speaking practice session.	3.86	0.81
Mot6	I would recommend YouTube-based speaking practice to other students.	3.94	0.80

Note: The present analysis included three six-item subscales: Perception, Motivation, and Self-Directed Learning, comprising 18 items in total.

The highest-rated motivation item was willingness to recommend the practice to other students (Mot6: $M = 3.94$, $SD = 0.80$). Aspiration to speak as well as the video speakers was also relatively strongly endorsed (Mot3: $M = 3.91$, $SD = 0.66$). Reported effort during practice had a mean of 3.86, while anticipation of weekly practice and the importance of speaking English for the future each had a mean of 3.80.

The item explicitly stating motivation to improve speaking after using videos had the lowest mean within this subscale (Mot1: $M = 3.69$, $SD = 0.72$), although it remained above the midpoint. Thus, the positive motivation score did not mean that all facets were equally strongly endorsed. The item-level pattern distinguishes recommending the activity, identifying with a spoken model, and exerting personal effort—related experiences that should not be collapsed into a claim of uniformly high motivation.

These results describe what learners reported after participation. They cannot establish that motivation increased over the 12 weeks, because the study did not administer an equivalent pre-intervention motivation measure. They also do not show that highly motivated learners practiced more or improved more: such participant level relationships were not tested. The interpretation is therefore consistent with Gardner's (1985) motivational framing while remaining within the evidence available in this design.

6.4 Reported Self-directed Learning

Table 4: Item-level self-directed learning behaviors ($n = 35$)

Item	Statement	M	SD
SDL1	I set my own goals for what I want to learn from each video.	3.57	0.74
SDL2	I think about how well the assigned video for each week matches my own speaking needs and level.	3.43	0.78
SDL3	I keep track of my own progress in speaking practice.	3.57	0.78
SDL4	I manage my own time to complete the weekly viewing and speaking tasks.	3.69	0.63
SDL5	I look for extra videos on my own, beyond what is required.	3.57	0.78
SDL6	I am confident in my ability to learn speaking skills on my own using YouTube.	3.57	0.78

Note: Responses ranged from 1 (strongly disagree) to 5 (strongly agree). Higher means indicate stronger endorsement. Statistics are reported to two decimal places. The scale represents self-reported behavior rather than independently observed practice.

Self-directed learning items ranged from 3.43 to 3.69. Time management for completing weekly tasks received the highest mean (SDL4: $M = 3.69$, $SD = 0.63$). Considering how well the assigned video matched personal speaking needs and level received the lowest mean (SDL2: $M = 3.43$, $SD = 0.78$). Goal setting, progress monitoring, searching for additional videos, and confidence in learning independently each had a mean of 3.57.

The pattern is compatible with a routine in which learners take responsibility for completing assigned work while exercising less control over its design and evaluation. Managing time for a prescribed task is one aspect of self-directed learning, but it does not establish independent selection of learning goals or materials. Likewise, thinking about the suitability of an assigned video differs from choosing the weekly video oneself. These

distinctions are preserved in the item wording and are important when interpreting the lower subscale mean.

The findings echo the technology-and-autonomy literature: opportunity and positive attitudes do not automatically establish the capacity to plan, monitor, and evaluate learning (Reinders & White, 2016; Godwin-Jones, 2019). However, the questionnaire alone cannot demonstrate that teacher structuring caused lower self-directed learning. That explanation remains plausible in this context rather than experimentally tested. The subscale's α of .69 adds a further reason to treat the finding as an indicative group pattern requiring further validation.

6.5 Interview Findings

The thematic analysis identified three themes: Structured Self-practice Routines, Perceived Gains in Pronunciation and Fluency, and Growing Confidence to Speak. These themes complement the questionnaire without turning perceived changes into independently measured outcomes. The examples below are selected from the existing transcript section and retain the original participant identifiers.

6.5.1 Structured Self-practice Routines

Participants described deliberate sequences of viewing and production. P1 watched repeatedly, used English captions, shadowed the model, and prepared answers to the guided questions. P8 described using different subtitle or caption options across viewings and focusing shadowing on difficult segments. P9 planned answers in writing before saying them aloud. These accounts show how learners adapted parts of an assigned sequence to make the task manageable.

"Twice a week I watched, shadowed, and then recorded myself with my phone to compare with the original." (P5)

P5's account illustrates self-recording as a monitoring activity rather than passive exposure. P7 described using a phone timer to keep the routine within an hour, while P6 practiced with a classmate and took turns during shadowing. These examples indicate procedural choices and some self-management. They do not show that every participant independently designed or consistently followed a complete learning plan.

"Honestly, I did not always follow the schedule. Some weeks I watched the video three or four times; other weeks I only did it once, quickly, right before the deadline." (P3)

This variation is important. The theme identifies purposeful practice, but it does not imply identical adherence. Following a teacher-provided sequence and adapting parts of that sequence can coexist with incomplete participation. The interview accounts therefore help explain why a generally useful routine could receive stronger perception ratings than endorsements of fully self-directed behavior.

6.5.2 Perceived Gains in Pronunciation and Fluency

The second theme focused primarily on perceived gains in pronunciation and fluency, although participants also described perceived vocabulary and grammar development. Participants described clearer sounds, fewer pauses, and more readily available language. P1 connected repeated practice with clearer word endings and a larger vocabulary. P2 reported being able to speak longer without stopping to search for words. P9 attributed perceived improvement in tense use to organizing ideas before speaking.

"Recording myself really showed me my weak points in fluency - lots of 'um' and long pauses." (P5)

P5 subsequently described fewer pauses in later recordings, while also stating that grammar remained a struggle. This account is useful because it does not portray perceived development as uniform. Learners could notice improvement in one aspect while remaining uncertain about another. The theme supports the questionnaire's positive appraisal of the activity but remains grounded in participants' own reports, not an additional speaking assessment.

The role of active production is visible in these accounts. Shadowing and spoken answers required learners to use language rather than merely recognize it, a distinction consistent with Swain's (1985) account of output. Repeated listening to an accessible model also fits Krashen's (1982) input perspective. These connections provide interpretations of learner experience, not evidence that the theories' mechanisms were separately demonstrated.

6.5.3 Growing Confidence to Speak

Several accounts emphasized willingness to participate. P1 described feeling less afraid to speak in class, and P3 reported a small increase in confidence when answering the teacher. P6 associated practice with a classmate with greater comfort in initiating and responding. P10 described sustaining classroom conversation longer without switching to Vietnamese. These experiences were coded as confidence-related changes rather than being treated as proof of general interactional mastery.

"My biggest change is confidence - I used to avoid speaking in front of classmates; now I volunteer more." (P4)

The theme adds an affective dimension to the positive questionnaire response. Learners described the activity as relevant not only to pronunciation or vocabulary but also to their readiness to speak. However, confidence was not tested with a separate pre-test/post-test confidence scale in this article. It is therefore reported as an interview theme, preserving the distinction between subjective experience and an experimentally established change.

6.6 Integrated Discussion: Positive Evaluation and Developing Independence

The combined evidence gives a coherent but qualified answer to the research question. Learners generally valued the program and reported motivation to participate. Their accounts described active use of the videos, including repeated viewing, caption decisions, shadowing, preparation, and listening back to recordings. These descriptions help explain why course usefulness and productive follow-up tasks were strongly endorsed in the questionnaire.

At the same time, the program supplied many of the decisions that fully self-directed learners would otherwise need to make. The researcher selected the videos, provided the weekly procedure, and set submission requirements. Learners could show initiative within that structure without independently determining the whole process. Vygotsky's (1978) perspective helps explain the supportive function of such mediation, while Benson (2011) and Zimmerman (2002) draw attention to the further development of learner control and monitoring.

The findings therefore suggest a distinction between accepting structured practice and independently sustaining learning. They do not show that scaffolding was undesirable; for these learners it may have made participation feasible. Nor do they demonstrate that the participants were incapable of autonomy. The more defensible interpretation is that favorable perceptions and motivation coexisted with less strongly endorsed self-directed learning, and that responsibility may need to be developed deliberately rather than presumed to emerge from digital access.

7. Recommendations

Teachers should retain clear links between videos and speaking goals. The positive ratings for course usefulness and guided questions suggest that the instructional purpose of the material should remain visible. Viewing can be organized as preparation for spoken production, with manageable segments, opportunities for replay, and a clear sequence that distinguishes listening, shadowing, and independent response.

To support self-directed learning, responsibility can be transferred gradually. For example, teachers may initially provide a small set of videos of comparable relevance and difficulty, then ask learners to explain their selection in relation to a personal goal. Learners may also record a specific target, compare successive recordings, and reflect on whether the chosen practice addressed that target. These are recommendations derived from the interpretation of the findings, not procedures whose independent effects were tested in the study.

Self-monitoring should be made concrete. A brief log or simple self-check rubric could connect goal setting, practice completion, and reflection rather than measuring participation only through submission. Such support is compatible with the planning and monitoring processes discussed by Zimmerman (2002). It should supplement, not be confused with, a claim that all participants already practiced autonomously.

Future studies should validate and refine the self-directed learning measure and supplement self-reports with behavioral evidence, such as practice logs or task-completion records. Repeated measurements would help distinguish a positive post-program response from change over time. Larger samples from different disciplines and institutions would strengthen transferability, while careful participant-level integration could examine how particular practice accounts relate to questionnaire responses. These extensions would address questions that the present descriptive analysis leaves open.

8. Conclusions

The learners reported positive perceptions and motivation after participating in the structured YouTube-based speaking program, while self-directed learning received comparatively lower ratings. The interview findings elaborated this pattern through three themes: deliberate practice routines, perceived gains in pronunciation and fluency, and growing confidence to speak. Participants described active engagement with the assigned material, but their accounts did not imply that all learners independently planned, selected, monitored, and evaluated the entire learning process.

The study's main contribution is the distinction between valuing a learning activity and directing it independently. Within this context, teacher-selected materials and prescribed steps may support participation while leaving important self-directed learning decisions under external control. Access to a video platform, positive perceptions, and reported motivation should therefore not be treated as sufficient evidence of developed learner autonomy.

The findings are limited to 35 learners from one accessible cohort at one university, with interviews from a purposive subsample of 10. Questionnaire data were collected only after the program and only from the YouTube group. As a result, they do not establish causal effects or pre-program-to-post-program changes in perceptions, motivation, or self-directed learning. Responses may also be affected by recall and socially desirable reporting in a teacher-associated research context.

The marginal internal consistency of the self-directed learning subscale further limits the precision of its interpretation. No systematic participant-level linkage between speaking gains, questionnaire responses, and interview accounts was reported in the study. The conclusions should consequently remain descriptive and context-specific. A practical next step is to preserve the supportive structure that learners valued while deliberately teaching and gradually extending their responsibility for learning decisions.

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 of 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

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Benson, P. (2011). *Teaching and researching autonomy* (2nd ed.). Pearson Education. <https://doi.org/10.4324/9781315833767>
- Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, 50(2), Article 48. <https://doi.org/10.34068/joe.50.02.48>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications. Retrieved from https://books.google.ro/books/about/Research_Design.html?id=4uB76IC_pOQC&redir_esc=y
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Pearson Education. <https://doi.org/10.4324/9781315833750>
- Fisher, M., King, J., & Tague, G. (2001). Development of a self-directed learning readiness scale for nursing education. *Nurse Education Today*, 21(7), 516–525. <https://doi.org/10.1054/nedt.2001.0589>

- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold. Retrieved from https://books.google.ro/books/about/Social_Psychology_and_Second_Language_Le.html?id=VcBitAEACAAJ&redir_esc=y
- Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning & Technology*, 23(1), 8–25. <https://doi.org/10.10125/44667>
- Ho, T. V., & Le, T. T. (2023). The effects of video materials on English-major students' learning: A test of hypotheses in the Vietnamese context. *FOSTER Journal of English Language Teaching*, 3(4), 185–196. <https://doi.org/10.24256/foster-jelt.v3i4.114>
- Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. Association Press. Retrieved from https://books.google.ro/books/about/Self_directed_Learning.html?id=ljVZAAAAYAAJ&redir_esc=y
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press. Retrieved from https://books.google.ro/books/about/Principles_and_Practice_in_Second_Langua.html?id=p9toAAAIAAJ&redir_esc=y
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>
- 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>
- 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.
- Purnamasari, A. (2018). What EFL learners say about YouTube use to improve pronunciation in a blended learning environment. *Lingua Cultura*, 12(3), 233–238. Retrieved from https://www.researchgate.net/publication/358043542_What_EFL_Learners_Say_about_YouTube_Use_to_Improve_Pronunciation_in_a_Blended_Learning_Environment
- Reinders, H., & White, C. (2016). Twenty years of autonomy and technology: How far have we come and where to next? *Language Learning & Technology*, 20(2), 143–154.
- 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>
- Socket, G. (2014). *The online informal learning of English*. Palgrave Macmillan. <https://doi.org/10.1057/9781137414885>

- Sundqvist, P., & Sylvé, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan. <https://doi.org/10.1093/elt/ccy051>
- 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>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Truong, N. K. V., & Le, Q. T. (2022). Utilizing YouTube to enhance English speaking skills: EFL tertiary students' practices and perceptions. *AsiaCALL Online Journal*, 13(4), 7–31. <https://doi.org/10.54855/acoj.221342>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2