FOREIGN LANGUAGES AND VIETNAMESE

EXPLICIT PHONETIC INSTRUCTION AND THE ACQUISITION OF VOICING CONTRASTS BY VIETNAMESE ENGLISH-MAJORED STUDENTS

LUONG BA PHUONG*

ABSTRACT: This study investigates how explicit phonetic instruction affects Vietnamese English-majored students' production of English voicing contrasts. Forty undergraduate students participated in an eight-week intervention focusing on voiced and voiceless obstruents in word-initial, intervocalic, and word-final positions. A pre-test–post-test comparison showed a significant increase in overall voicing accuracy ($p < .001$), with a large effect size. Accuracy improved in all three phonological positions, with the largest gain occurring in word-final contexts, where performance was weakest before instruction. Both plosives and fricatives benefited from the intervention, although voiced fricatives in word-final position continued to present difficulties. The qualitative data further pointed to increased attention to secondary phonetic cues, especially preceding-vowel duration. The results indicate that explicit instruction can improve learners' control of English voicing contrasts by helping them use relevant phonetic cues more effectively. At the same time, the remaining difficulties point to the continuing role of first-language phonological constraints. The findings have implications for the teaching of L2 pronunciation in Vietnamese EFL settings.

KEY WORDS: Explicit phonetic instruction; English voicing contrasts; L2 pronunciation; Vietnamese EFL learners; interlanguage phonology.

RECEIVED: August, 15th, 2026. EDITED: November, 16th, 2026

1. Introduction

Second language (L2) pronunciation remains a difficult area of development for many language learners. The difficulty becomes particularly apparent when learners have to produce segmental contrasts that are weakly signalled in the speech stream or realized differently across phonological contexts. English voicing contrasts are a case in point. The distinction between voiced and voiceless consonants contributes to lexical contrasts and grammatical distinctions and is also important for intelligible speech. For L2 learners, therefore, producing these contrasts accurately involves more than simply establishing a phonological opposition between two categories.

Difficulty with English voicing contrasts may persist even when learners have had substantial exposure to the language and have reached relatively advanced levels of general proficiency. Word-final position is especially problematic. Learners may devoice final consonants, produce only partial voicing, or rely inconsistently on the phonetic information available in the signal. These patterns have been reported for learners with different first-language backgrounds and are particularly relevant when the L1 does not maintain the same voicing distinctions in coda position. Such cases provide a useful basis for examining how L1 phonology and phonological markedness shape the development of L2 contrasts.

The challenge is also partly phonetic. English voicing is not realized through a single cue that operates in the same way in every context. Vocal fold vibration interacts with other acoustic and articulatory properties, including preceding-vowel duration, closure duration, aspiration, and release characteristics. In word-final position, continuous vocal fold vibration may be difficult to maintain

*Dr; Academy of Journalism and Communication; Email: ltphuong2012@gmail.com

even in native speech, which makes the duration of the preceding vowel an important cue to the voicing distinction. Learners consequently need to develop control over several phonetic dimensions and to learn how their relative contribution changes according to phonological context.

These patterns can be understood within interlanguage phonology as part of the development of a learner's own phonological system. Non-target-like productions are not necessarily random deviations; they may reflect systematic properties of an evolving interlanguage grammar shaped by linguistic experience and phonological constraints. Final devoicing, for example, may indicate more than difficulty with the physical production of voicing. It may also reflect how learners represent the contrast and which phonetic cues they attend to when producing it. Examining voicing contrasts can therefore shed light on the relationship between phonological representation, transfer, and restructuring in L2 development.

Previous research has documented a range of voicing-related difficulties among learners of English, including final consonant devoicing, incomplete voicing in intervocalic contexts, and limited reliance on secondary phonetic cues. Much of this work, however, has concentrated on beginner or intermediate learners or has given greater attention to perception than to production. Less is known about how advanced learners develop control of these contrasts, particularly students enrolled in English-major programs who are expected to attain relatively high levels of pronunciation accuracy. Their continued segmental difficulties are important because such deviations may remain relevant to intelligibility and professional communication even when overall language proficiency is high.

Instructional intervention provides another perspective on this issue. Explicit phonetic instruction draws learners' attention to articulatory mechanisms, acoustic properties, and phonological contrasts and can be accompanied by metalinguistic explanation and focused practice. Previous research has reported improvements in segmental accuracy and intelligibility following such instruction, including among learners with intermediate or advanced proficiency. The evidence is not uniform, however. Contrasts that are position-sensitive or phonologically marked, such as word-final voicing, may respond differently to instruction from contrasts that are more strongly cued. This raises an important question: can explicit instruction lead to changes in learners' phonological control, or does it mainly produce temporary improvements in task performance?

This question is particularly relevant to Vietnamese learners of English. Vietnamese does not permit voiced obstruents in syllable-final position, and voicing distinctions are therefore neutralized in this position. Vietnamese learners of English may consequently produce final devoicing or weakened voicing contrasts in English, including after years of formal instruction. Although the contribution of L1 transfer to these patterns has been acknowledged in previous research, there is still limited evidence concerning the extent to which targeted phonetic instruction can modify such patterns in the developing interlanguage system, especially among advanced learners.

Another limitation of previous work concerns the way instructional outcomes are measured. Studies that focus mainly on overall accuracy may overlook differences associated with phonological position or consonant type. Word-initial, intervocalic, and word-final contexts do not necessarily provide learners with the same phonetic information, and plosives and fricatives may likewise differ in the cues available for voicing. Examining these dimensions separately can help identify where instructional gains are strongest and where difficulties remain. It can also provide a clearer picture of the interaction among markedness, cue salience, and L1 influence in L2 phonological development.

The present study addresses these issues by examining the effects of explicit phonetic instruction on Vietnamese English-majored students' production of English voiced and voiceless consonants. The participants were advanced EFL learners in a tertiary educational setting. The study considers both changes in production accuracy and changes in the use of phonetic cues across phonological positions and consonant types, with particular attention to word-final voicing contrasts. A pre-test–post-test design is used to examine instructional change, while the findings are interpreted in relation

to interlanguage phonology and research on instructed L2 pronunciation. The study is guided by the following research questions:

1. What voicing-related pronunciation patterns characterize Vietnamese English-majored students' production of English consonants prior to explicit phonetic instruction?
2. To what extent does explicit phonetic instruction affect learners' overall accuracy in producing English voicing contrasts, particularly in word-final position?
3. Do instructional effects differ across consonant types and phonological contexts, and what phonetic cues appear to underlie observed changes?

2. Literature review

2.1. Voicing contrastive in English phonology

Voicing is an important source of contrast in English consonants. It distinguishes lexical items such as bat–pat, zip–sip, and ride–right, and it also contributes to grammatical distinctions, including those found in plural and past-tense forms. The accurate production of voicing contrasts is therefore relevant to both phonological competence and spoken intelligibility (Ladefoged & Johnson, 2015; Munro & Derwing, 2011).

At the phonetic level, however, the distinction is more complex than a simple voiced–voiceless binary. Speakers rely on several articulatory and acoustic cues, and the relative contribution of these cues changes with phonological context. Relevant cues include vocal fold vibration, preceding-vowel duration, closure duration, aspiration, and release characteristics (Ladefoged & Johnson, 2015). In word-final position, preceding-vowel duration may provide a more reliable indication of voicing than sustained vocal fold vibration, which can be weak or absent even in native speech (Munro & Derwing, 2011; Waniek-Klimczak, 2011). English voicing is thus better understood as a context-sensitive phenomenon involving several interacting phonetic dimensions.

Word-final voicing also presents a typological challenge. Many languages neutralize voicing contrasts in coda position or do not allow voiced obstruents in that position. Learners whose L1s lack such contrasts may consequently have difficulty maintaining the English distinction in word-final contexts, even with substantial exposure to English (Waniek-Klimczak, 2011). The marked status of final voicing may therefore help explain why this aspect of pronunciation remains difficult for many L2 learners.

From a theoretical perspective, voicing illustrates the relationship between phonological categories and their phonetic realization. Although the contrast can be represented as a distinctive feature, its production depends on temporal and articulatory coordination across several dimensions. L2 learners must consequently acquire both the category itself and the phonetic means of realizing it in different environments (Munro & Derwing, 2011). This makes voicing a useful domain for investigating L2 phonological development.

2.2. Voicing in second language phonological acquisition

Voicing contrasts have been reported as a persistent source of difficulty in L2 English pronunciation. Learners from different L1 backgrounds may devoice final consonants, produce incomplete or delayed voicing, or make inconsistent use of the phonetic cues associated with the contrast (Flege, 1995; Waniek-Klimczak, 2011). The recurrence of these patterns across learners suggests that they reflect properties of developing phonological systems rather than isolated production failures.

This interpretation is consistent with the notion of interlanguage proposed by Selinker (1972). Learner phonological systems are systematic, although they may differ from the target-language system, and they can change as learners receive further input and experience. From this perspective, a voicing error is potentially informative about the learner's developing representation of the contrast. Final devoicing, for instance, may reflect not only difficulty maintaining vocal fold vibration but also the influence of an L1-based phonological organization.

The Speech Learning Model (SLM) and its later formulations provide another account of why L2 phonological contrasts can remain difficult. These models emphasize the role of perceptual similarity

and category formation in L2 speech learning (Flege, 1995; Flege & Bohn, 2021). When an L2 sound is perceived as sufficiently similar to an existing L1 category, the learner may have less incentive to establish a separate phonetic category. Related predictions have also been made in work on cross-language perception and assimilation (Best & Tyler, 2007). For voicing contrasts, these accounts suggest that learners whose L1 systems differ substantially from English may have particular difficulty establishing stable distinctions in marked contexts such as word-final codas.

Production studies provide evidence for the influence of L1-based strategies. Learners whose languages neutralize voicing in codas, for example, commonly show final devoicing when producing English (Flege, 1995; Nguyen & Ingram, 2007). Such patterns can persist beyond the early stages of learning. Their occurrence among relatively advanced learners suggests that greater exposure to English does not necessarily lead to a complete reorganization of the relevant phonological categories (Flege & Bohn, 2021).

2.3. Vietnamese phonological constraints and first-language transfer

Vietnamese offers a particularly relevant L1 background for examining English voicing contrasts. Its syllable structure does not allow voiced obstruents in word-final position. Coda consonants are primarily voiceless stops and nasals, so the voicing distinction found in English word-final obstruents has no direct equivalent in this position in Vietnamese (Nguyen & Ingram, 2007). Vietnamese learners therefore approach English final voicing without an established L1 pattern for maintaining this contrast.

Production studies have reported final devoicing as a recurring feature of Vietnamese learners' English pronunciation. For example, /b/, /d/, and /z/ may be realized as their voiceless counterparts in word-final position, with possible consequences for intelligibility (Nguyen & Ingram, 2007). Such productions are consistent with negative transfer from Vietnamese phonotactic constraints (Flege, 1995).

L1 transfer, however, does not provide a complete explanation for persistent pronunciation difficulties. Interlanguage systems can change as learners encounter new input, receive instruction, and adapt their speech to communicative demands (Selinker, 1972; Flege & Bohn, 2021). The persistence of final voicing problems among advanced Vietnamese learners therefore makes an instructional question particularly relevant: under what conditions can learners move beyond an L1-based pattern and develop greater control of an L2 contrast?

This question has received limited empirical attention in research on Vietnamese learners. Existing work has largely described pronunciation patterns and their relation to L1 phonology rather than systematically testing the effects of targeted instruction (Nguyen & Ingram, 2007). Evidence is consequently still limited on whether explicit phonetic instruction can reduce entrenched L1-based constraints, particularly among learners who have already reached relatively advanced levels of English.

2.4. Explicit phonetic instruction in the second language pronunciation

Explicit phonetic instruction has been used to draw learners' attention to aspects of pronunciation that may otherwise remain difficult to notice or control. Instruction of this kind may include explanations of articulatory processes, attention to acoustic cues, explicit comparison of phonological contrasts, and guided production practice (Derwing & Munro, 2005; Thomson & Derwing, 2015). It differs from approaches in which pronunciation development is expected to emerge primarily through exposure and communicative use.

Studies have reported gains in segmental accuracy and intelligibility following explicit pronunciation instruction, including with learners at intermediate and advanced proficiency levels (Saito, 2012; Derwing & Munro, 2005). The size and nature of these gains, however, vary across studies. Differences in instructional focus, duration, learner characteristics, and assessment procedures may partly account for the variation reported in the literature (Thomson & Derwing, 2015; Saito & Plonsky, 2019).

The issue becomes more complicated when the target contrast depends on several phonetic cues. Instruction on English voicing commonly presents voiced and voiceless consonants as categorical opposites, whereas natural production also involves differences in vowel duration, closure duration, aspiration, and other temporal properties (Munro & Derwing, 2011). Learners may consequently improve on controlled pronunciation tasks without fully developing the ability to use these cues across different contexts (Thomson & Derwing, 2015). For this reason, examining cue use alongside accuracy may provide information that an accuracy score alone cannot capture.

A further question concerns what instructional improvement actually represents. Better performance after training does not necessarily mean that learners have reorganized their phonological categories. From an interlanguage perspective, it is useful to distinguish between a more stable change in phonological representation and an adjustment that is specific to a particular task or context (Flege & Bohn, 2021). This distinction is particularly relevant when studying advanced learners, for whom persistent L1-based patterns may coexist with considerable experience in English.

2.5. Research gaps and rationale for present study

The literature reviewed above points to several issues that remain insufficiently explored. One concerns the learners themselves. Much pronunciation research has focused on general learner populations, whereas advanced students in English-major programs have received less attention, despite the higher pronunciation demands placed on them (Saito & Plonsky, 2019). A second issue is the voicing contrast itself. Word-final voicing is both phonologically marked and strongly influenced by L1 phonotactic patterns, yet its development under explicit instruction has not been examined in sufficient detail (Waniek-Klimczak, 2011). Finally, overall accuracy does not necessarily reveal how learners produce a contrast. Differences across consonant types, phonological positions, and individual phonetic cues may be obscured when instructional outcomes are reduced to a single accuracy measure (Thomson & Derwing, 2015).

These issues motivate the present study of Vietnamese English-majored students. The study examines whether explicit phonetic instruction changes their production of English voiced and voiceless consonants and whether the effects are consistent across phonological positions and consonant types. Quantitative accuracy measures are considered together with qualitative evidence concerning error patterns and phonetic cue use. This combination makes it possible to examine instructional change while also considering how learners' developing interlanguage phonology may constrain that change.

3. Methodology

3.1. Research design

The study used a quasi-experimental pre-test–post-test design to examine whether explicit phonetic instruction affected Vietnamese English-majored students' production of English voicing contrasts. The design made it possible to compare each participant's performance before and after the instructional period within the setting of their regular pronunciation classes. A separate control group was not feasible because of curricular and organizational constraints. The participants nevertheless came from the same academic cohort, followed the same curriculum, and had comparable levels of English proficiency. To examine both changes in production accuracy and changes in learners' phonetic behavior, the study combined quantitative measures with qualitative analysis of recurring error patterns and phonetic cue use.

3.2. Participants

The participants were 40 third-year undergraduate students majoring in English at a Vietnamese university. All had previously completed formal courses in English phonetics and phonology. According to institutional benchmarks, their English proficiency ranged from upper-intermediate to advanced. The students had received at least eight years of formal English instruction. None reported a history of speech or hearing disorders. Participation was voluntary, and all students provided informed consent in accordance with the relevant institutional ethical requirements.

3.3. Target sounds and phonological context

The study examined English voiced–voiceless obstruent contrasts in two consonant classes: plosives (/b, d, g/ vs. /p, t, k/) and fricatives (/v, z, ʒ/ vs. /f, s, ʃ/). These contrasts were selected because voicing has been reported as a recurrent source of difficulty for Vietnamese learners of English. The target consonants occurred in three positions: word-initial, intervocalic, and word-final. Word-final position received particular attention because voicing contrasts in coda position are both typologically marked and phonetically variable. This distribution also made it possible to examine whether the effects of instruction differed across phonological contexts.

3.4. Data collection tasks

Two production tasks were used to elicit the target contrasts. The first was a word-list reading task. Target consonants appeared in controlled phonological environments, mainly through minimal or near-minimal pairs such as cap–cab and leaf–leave. The lexical items were familiar to the participants and were selected to reduce possible effects of unfamiliar vocabulary and orthographic complexity. The second task involved sentence reading. Target words were placed in relatively neutral sentence frames so that voicing could also be examined in a more connected speech context. Participants were recorded individually in a quiet classroom using a digital recorder and a head-mounted microphone. The same recording procedure was used for the pre-test and post-test. All recordings were saved in WAV format for subsequent phonetic analysis.

3.5. Instruction invention

The instructional intervention lasted eight weeks and was incorporated into the participants' regular pronunciation classes. It focused on both perception and production of English voicing contrasts, with particular emphasis on voiced and voiceless obstruents in word-final position. Rather than relying primarily on mechanical repetition, the instruction was designed to increase learners' awareness of the phonetic properties associated with voicing. The main activities included: (i) explicit explanation of vocal fold vibration; (ii)instruction on secondary phonetic cues, particularly preceding-vowel duration; (iii) contrastive discussion of English and Vietnamese voicing patterns; and (iv) guided perception and controlled production activities.

3.6. Data analysis

The target consonants were transcribed using the International Phonetic Alphabet (IPA) and coded according to whether their realization was target-like or non-target-like. Ambiguous tokens were evaluated using predefined phonetic criteria, including evidence of vocal fold vibration and the use of relevant secondary cues. To assess transcription reliability, approximately 20% of the recordings were independently transcribed by a second trained phonetician. Disagreements were discussed and resolved between the two transcribers. Accuracy rates were calculated for each consonant type, phonological position, and testing phase. Pre-test and post-test scores were compared using paired-samples t-tests, and Cohen's d was calculated to estimate the magnitude of the instructional effect. Qualitative analysis was then used to identify recurring error patterns and changes in learners' use of phonetic cues.

3.7. Validity considerations

The target consonants were transcribed using the International Phonetic Alphabet (IPA) and coded according to whether their realization was target-like or non-target-like. Ambiguous tokens were evaluated using predefined phonetic criteria, including evidence of vocal fold vibration and the use of relevant secondary cues. To assess transcription reliability, approximately 20% of the recordings were independently transcribed by a second trained phonetician. Disagreements were discussed and resolved between the two transcribers. Accuracy rates were calculated for each consonant type, phonological position, and testing phase. Pre-test and post-test scores were compared using paired-samples t-tests, and Cohen's d was calculated to estimate the magnitude of the instructional effect. Qualitative analysis was then used to identify recurring error patterns and changes in learners' use of phonetic cues.

3.8. Methodological rationale and analysis scope

Several procedures were used to maintain consistency across the two testing phases. The same types of production tasks, recording conditions, and target contrasts were used in the pre-test and post-test. The combination of word-list and sentence-reading tasks also provided two levels of production data, ranging from relatively controlled to more contextualized speech. Because the study did not include a separate control group, changes between the two testing phases cannot be attributed exclusively to the instructional intervention. The pre-test–post-test design nevertheless allowed for a direct examination of within-participant change under the conditions of the instructional setting.

4. Results and Discussion

4.1. Results

4.1.1. Overall changes in voicing accuracy

Learners' overall accuracy in producing English voicing contrasts increased from the pre-test to the post-test. As shown in table 1, mean accuracy rose from 62.4% (SD = 8.7) before instruction to 78.9% (SD = 7.9) after the eight-week intervention. A paired-samples t-test showed that this increase was statistically significant ($p < .001$), with a large effect size. Although post-test performance was substantially higher than pre-test performance, the mean accuracy remained below a ceiling level, indicating that some difficulties with English voicing contrasts persisted after the intervention.

Table 1. *Overall voicing accuracy before and after instruction*

Test phase	Mean accuracy (%)	Standard deviation (SD)
Pre-test	62.4	8.7
Post-test	78.9	7.9

4.1.2. Positional effects on voicing realization

Accuracy was examined separately across word-initial, intervocalic, and word-final positions (Table 2). Performance improved in all three phonological contexts following instruction, although the size of the gain varied across positions.

In word-initial position, mean accuracy increased from 75.1% to 83.4%, representing an 8.3-percentage-point gain. Accuracy in the intervocalic position increased from 69.8% to 80.2%, a gain of 10.4 percentage points. The largest change occurred in word-final position, where accuracy rose from 48.3% to 73.1%, corresponding to a 24.8-percentage-point gain. Thus, word-final position showed the greatest improvement among the three contexts. Despite this substantial increase, post-test accuracy in word-final position remained lower than accuracy in word-initial and intervocalic positions.

Table 2. *Voicing accuracy by phonological position*

Position	Pre-test (%)	Post-test (%)	Gain (percentage points)
Word-initial	75.1	83.4	+8.3
Intervocalic	69.8	80.2	+10.4
Word-final	48.3	73.1	+24.8

4.1.3. Differences between consonant types

Accuracy was also compared between plosives and fricatives (Table 3). Both consonant types showed improvement after the intervention. For plosives, mean accuracy increased from 65.7% in the pre-test to 82.6% in the post-test, corresponding to a gain of 16.9 percentage points. Fricative accuracy increased from 58.9% to 74.2%, representing a gain of 15.3 percentage points.

Both consonant groups showed statistically significant improvement following instruction ($p < .001$). Plosives nevertheless maintained higher accuracy than fricatives at both testing stages. Examination of the post-test responses further showed that voiced fricatives in word-final position accounted for a substantial proportion of the remaining errors.

Table 3. *Voicing accuracy by consonant type*

Consonant type	Pre-test (%)	Post-test (%)	Gain(percentage points)
Plosives	65.7	82.6	+16.9
Fricatives	58.9	74.2	+15.3

4.1.4. Voicing error patterns before and after instruction

The qualitative analysis identified three recurring patterns in the pre-test data: final devoicing, partial voicing, and inconsistent use of secondary phonetic cues, particularly preceding-vowel duration. Final devoicing was the most frequently observed pattern and occurred especially often with fricatives and velar plosives.

The post-test data showed a reduction in final devoicing, although the pattern was not completely eliminated. There were also changes in learners' use of secondary phonetic cues. In a number of productions, vowels preceding intended voiced consonants were longer even when vocal fold vibration was not consistently maintained.

These observations suggest that some learners showed changes in phonetic implementation that were not fully captured by the binary accuracy measure. In particular, the post-test responses indicated greater use of secondary phonetic information alongside the improvement in segmental voicing accuracy.

4.2. Discussion

The results point to a clear improvement in learners' production of English voicing contrasts after the eight-week intervention, but the improvement was not uniform across phonological contexts or consonant types. The strongest change occurred in word-final position, which was also the least accurate context before instruction. Plosives were produced more accurately than fricatives at both testing stages, and the qualitative analysis showed a reduction in final devoicing together with greater use of secondary phonetic cues. These patterns suggest that explicit instruction helped learners attend to aspects of the voicing contrast that had previously received limited attention. At the same time, the remaining errors indicate that improvement did not amount to complete restructuring of the learners' phonological system.

4.2.1. The effect of explicit phonetic instruction

The increase in overall accuracy from 62.4% to 78.9% provides evidence that the instructional intervention was associated with substantial improvement in learners' production of English voicing contrasts. This result is in line with previous work showing that pronunciation instruction can produce measurable changes in L2 segmental accuracy, including among learners beyond the earliest stages of acquisition (Derwing & Munro, 2005; Saito, 2012; Thomson & Derwing, 2015).

What is particularly relevant here is that the participants were English majors who had already received formal exposure to English phonetics and phonology. Their relatively low pre-test performance therefore suggests that knowledge of phonetic terminology or articulatory descriptions does not necessarily translate into reliable production of a difficult contrast. The improvement following targeted instruction indicates that learners may benefit from returning to such contrasts with more focused attention to how they are actually realized in speech.

The results can be viewed from an interlanguage perspective (Selinker, 1972). Learner phonological systems are not fixed once a certain level of proficiency has been reached; they can continue to change as learners receive new input and experience different forms of instruction. The present findings do not, however, establish that the intervention produced a complete restructuring of phonological categories. The post-test mean remained well below ceiling performance, and several error patterns persisted. It is therefore more appropriate to regard the improvement as evidence of ongoing phonetic and phonological development rather than complete acquisition of the target contrast.

4.2.2. Why word-final position remained difficult

The positional results are especially revealing. Accuracy increased in all three contexts, but the gain in word-final position was much larger than those observed word-initially or intervocalically: 24.8 percentage points, compared with 8.3 and 10.4 points, respectively. Yet word-final accuracy remained the lowest after instruction. The same context was therefore both the greatest source of improvement and the most persistent source of difficulty.

This pattern is understandable in light of the phonological differences between English and Vietnamese. Vietnamese does not maintain a voicing contrast among word-final obstruents in the way English does, and final consonants are subject to substantial phonological restrictions (Nguyen & Ingram, 2007). Vietnamese learners consequently have limited L1 experience on which to draw when producing English voiced obstruents in coda position. Previous research has likewise reported final devoicing among Vietnamese learners of English and has related this pattern to L1 phonotactic constraints (Nguyen & Ingram, 2007).

The present results add an instructional dimension to this earlier observation. The large increase in final-position accuracy shows that the difficulty was not completely resistant to instruction. At the same time, the remaining gap between word-final and the other two positions suggests that the L1 constraint continued to influence performance. In this sense, the results fit better with a gradient account of L1 influence than with either of two extreme interpretations: that learners simply reproduce the L1 pattern, or that explicit instruction can readily replace it.

The findings also have a connection with markedness-based accounts of L2 acquisition. Word-final voicing is typologically less stable than voicing contrasts in many other positions, and learners whose first languages lack such contrasts are expected to experience greater difficulty in acquiring it (Eckman, 1977, 2004; Broselow & Kang, 2013). The present data are compatible with this prediction. Instruction reduced the size of the difficulty, but did not remove the positional asymmetry.

4.2.3. Plosives, fricatives, and the availability of phonetic cues

The difference between plosives and fricatives provides a second indication that improvement was not uniform across the voicing contrast. Plosive accuracy increased from 65.7% to 82.6%, while fricative accuracy rose from 58.9% to 74.2%. Thus, although both groups improved, plosives remained more accurate at both stages of testing.

One possible explanation lies in the phonetic information available to learners. The realization of voicing in plosives can be accompanied by several cues, including vocal-fold vibration, closure characteristics, release properties, and the duration of the preceding vowel (Ladefoged & Johnson, 2015; Munro & Derwing, 2011). Fricatives present a somewhat different challenge because maintaining voicing during sustained oral constriction requires continuous coordination between airflow and vocal-fold vibration. The relevant acoustic differences can consequently be less straightforward for learners to control.

The qualitative results support this interpretation to some extent. Voiced fricatives in word-final position remained a noticeable source of post-test errors. This suggests that instruction did not affect all segment types in exactly the same way. Learners appear to have made greater progress where the relevant cues were easier to identify and implement, while more demanding combinations of segment type and phonological position remained problematic.

This pattern also cautions against treating “voicing accuracy” as a single, uniform phonetic skill. The same phonological contrast may place different demands on learners depending on the consonant and its position in the word.

4.2.4. Error reduction and the changing use of secondary cues

The qualitative analysis provides a more detailed picture of what changed behind the overall accuracy scores. Before instruction, three recurring patterns were identified: final devoicing, partial voicing, and inconsistent use of secondary phonetic cues, particularly preceding-vowel duration. Final devoicing was the most frequent pattern, especially with fricatives and velar plosives.

After instruction, final devoicing occurred less often, although it was still present. More importantly, some productions showed a change in the use of secondary cues. In several cases, vowels preceding intended voiced consonants were longer even when vocal-fold vibration was not consistently maintained. This pattern is noteworthy because it suggests that learners did not necessarily move directly from an inaccurate production to a fully target-like one. Some productions appear to occupy an intermediate stage in which learners make greater use of a relevant phonetic cue without yet maintaining the complete voicing pattern.

Such findings fit the broader view that English voicing is realized through a combination of phonetic cues rather than through vocal-fold vibration alone (Munro & Derwing, 2011; Waniek-Klimczak, 2011). They also provide a more nuanced interpretation of the post-test results. A binary target-like/non-target-like measure can capture whether a production meets the coding criterion, but it may not fully represent changes in how learners implement the contrast.

At the same time, this interpretation should remain cautious. The present study identified changes in cue use qualitatively; it did not establish a quantitative change in preceding-vowel duration. The evidence therefore supports the claim that learners showed greater attention to or use of secondary cues, rather than demonstrating a statistically verified shift in vowel duration.

4.2.5. L1 influence and phonological plasticity

The persistence of final devoicing after instruction is important for understanding the role of first-language influence. The results support earlier observations that Vietnamese phonological constraints can shape learners' production of English final obstruents (Nguyen & Ingram, 2007). However, the substantial improvement after instruction also shows that L1 influence should not be treated as an immutable constraint.

This combination of persistence and change is compatible with an interlanguage account in which learner phonology develops through interaction between existing L1 patterns, L2 input, and learning experience (Selinker, 1972). It is also broadly consistent with the Speech Learning Model and its later formulations, which emphasize the importance of phonetic similarity and learners' ability to establish sufficiently differentiated L2 representations (Flege, 1995; Flege & Bohn, 2021). From this perspective, learners may initially interpret unfamiliar L2 contrasts through categories and patterns already available in the L1 system. Focused instruction can draw attention to distinctions that have previously been weakly represented or inconsistently implemented.

The present data do not allow us to determine whether learners developed entirely new phonological categories. They do, however, show that performance remained responsive to focused intervention even among English majors with previous phonetic training. This is an important point because persistent pronunciation errors are sometimes interpreted as evidence that advanced learners have reached a relatively fixed state. The present findings suggest a more flexible picture: some aspects of an established interlanguage system can remain open to modification, although the extent of change may differ across phonological environments.

4.2.6. Implications for pronunciation teaching

The findings have a straightforward pedagogical implication for Vietnamese EFL contexts. Teaching English voicing contrasts solely as a binary distinction between "voiced" and "voiceless" consonants may not be sufficient, particularly when learners are dealing with word-final obstruents. Instruction that makes the relevant phonetic cues explicit can give learners additional resources for monitoring and modifying their productions.

The intervention used in this study combined explanation of vocal-fold vibration, attention to secondary cues such as preceding-vowel duration, contrastive discussion of English and Vietnamese patterns, and guided perception and production activities. The improvement observed after the intervention suggests that this combination was useful for the learners in the present context. It also points to the value of linking phonological knowledge to actual speech production rather than treating phonetics and pronunciation as primarily descriptive subjects.

For Vietnamese learners in particular, word-final voiced obstruents may warrant sustained attention because they represent a contrast that is not supported by the L1 phonological system. The aim should not necessarily be to eliminate every trace of L1 influence in a short period of instruction. A more realistic goal is to help learners recognize the contrast, identify the cues that signal it, and gradually gain greater control over its realization.

5. Conclusion

This study examined the effect of explicit phonetic instruction on Vietnamese English-majored students' production of English voicing contrasts. The results showed a significant improvement in overall accuracy, from 62.4% in the pre-test to 78.9% in the post-test. Improvement occurred across word-initial, intervocalic, and word-final positions, with the largest gain found in word-final position (+24.8 percentage points). Nevertheless, word-final accuracy remained the lowest after instruction. Plosives were also produced more accurately than fricatives at both testing stages, while voiced fricatives in word-final position remained a notable source of difficulty. The qualitative analysis further indicated a reduction in final devoicing and greater use of secondary phonetic cues.

The findings suggest that explicit phonetic instruction can improve learners' production of English voicing contrasts even after several years of English study. At the same time, the persistence of final devoicing indicates that improvement was incomplete and that L1 phonological constraints continued to influence performance. The results are therefore consistent with an interlanguage perspective in which established pronunciation patterns remain open to modification but may not disappear through short-term instruction.

Several pedagogical implications follow. First, pronunciation teaching should move beyond a simple voiced–voiceless distinction and draw learners' attention to relevant phonetic cues, including vocal-fold vibration and preceding-vowel duration. Second, word-final voiced obstruents deserve sustained practice because they remained particularly difficult despite showing the greatest improvement. Practice can combine focused perception, controlled production, and communicative activities. Finally, persistent pronunciation problems should not be assumed to disappear at higher levels of English study. Learners with previous phonetic training may still benefit from targeted instruction and feedback.

The study has several limitations. The eight-week intervention does not show whether the gains were maintained over time, and the qualitative treatment of secondary cues was not supported by detailed acoustic measurements. In addition, the participants came from a relatively homogeneous group of English-majored students. Future research could therefore include delayed post-tests, acoustic analysis, and learners from different proficiency levels and instructional contexts.

The findings support the value of sustained, cue-oriented phonetic instruction for helping Vietnamese learners develop greater control over English voicing contrasts while recognizing the continuing influence of their first-language phonology.

REFERENCES

1. Derwing, T. M., & Munro, M. J. (2005), Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397. https://doi.org/10.2307/3588486.
2. Munro, M. J., & Derwing, T. M. (2011), The foundations of accent and intelligibility in pronunciation research. *Language Teaching, 44*(3), 316–327. https://doi.org/10.1017/S0261444811000103.
3. Nguyen, T. T. T., & Ingram, J. (2007), A phonetic study of Vietnamese learners' production of English final consonants. *Australian Journal of Linguistics, 27*(2), 175–195. https://doi.org/10.1080/07268600701359056.
4. Saito, K. (2012), Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842–854. https://doi.org/10.1002/tesq.67.

5. Saito, K. (2015), Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*(3), 563–595. https://doi.org/10.1111/lang.12120.
6. Saito, K., & Plonsky, L. (2019), Effects of second language pronunciation teaching revisited: A meta-analysis of controlled studies. *Language Learning, 69*(3), 652–708.
7. Selinker, L. (1972), Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*(3), 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209.
8. Thomson, R. I., & Derwing, T. M. (2015), The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344. https://doi.org/10.1093/applin/amu076.
9. Best, C. T., & Tyler, M. D. (2007), Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins.
10. Flege, J. E. (1995), Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
11. Flege, J. E., & Bohn, O.-S. (2021), The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002.
12. Derwing, T. M., & Munro, M. J. (2015), *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
13. Hansen Edwards, J. G., & Zampini, M. L. (2008), *Phonology and second language acquisition*. John Benjamins.
14. Ladefoged, P., & Johnson, K. (2015), *A course in phonetics* (7th ed.). Cengage Learning.
15. Lyster, R. (2007), *Learning and teaching languages through content: A counterbalanced approach*. John Benjamins.
16. Boersma, P., & Weenink, D. (2023), *Praat: Doing phonetics by computer* (Version 6.3) [Computer software]. http://www.praat.org.

Giảng dạy ngữ âm tường minh và quá trình lĩnh hội các đối lập hữu thanh của sinh viên chuyên ngành tiếng Anh ở Việt Nam

Tóm tắt: Nghiên cứu này khảo sát tác động của việc giảng dạy ngữ âm tường minh đối với khả năng phát âm các đối lập hữu thanh-vô thanh trong tiếng Anh của sinh viên Việt Nam chuyên ngành tiếng Anh. Bốn mươi sinh viên đại học tham gia một chương trình can thiệp kéo dài tám tuần, tập trung vào các âm tắc và âm xát hữu thanh và vô thanh ở vị trí đầu từ, giữa từ và cuối từ. So sánh kết quả trước và sau can thiệp cho thấy độ chính xác về đối lập hữu thanh-vô thanh tăng đáng kể ($p < .001$), với kích thước hiệu ứng lớn. Độ chính xác được cải thiện ở cả ba vị trí âm vị học, trong đó mức tăng lớn nhất xuất hiện ở vị trí cuối từ, vốn là vị trí có kết quả thấp nhất trước khi giảng dạy. Cả âm tắc và âm xát đều có sự tiến bộ, mặc dù các âm xát hữu thanh ở cuối từ vẫn tiếp tục gây khó khăn cho người học. Dữ liệu định tính cũng cho thấy người học chú ý nhiều hơn đến các dấu hiệu ngữ âm thứ cấp, đặc biệt là độ dài nguyên âm đứng trước. Kết quả cho thấy việc giảng dạy ngữ âm tường minh có thể cải thiện khả năng kiểm soát các đối lập hữu thanh-vô thanh trong tiếng Anh của người học bằng cách giúp họ sử dụng các dấu hiệu ngữ âm liên quan một cách hiệu quả hơn. Đồng thời, những khó khăn còn tồn tại cho thấy vai trò tiếp diễn của các ràng buộc âm vị học từ tiếng mẹ đẻ. Những phát hiện này có ý nghĩa đối với việc giảng dạy phát âm ngôn ngữ thứ hai trong bối cảnh EFL tại Việt Nam.

Từ khóa: Giảng dạy ngữ âm tường minh; các đối lập hữu thanh-vô thanh trong tiếng Anh; phát âm ngôn ngữ thứ hai; người học tiếng Anh như một ngoại ngữ tại Việt Nam; âm vị học liên ngôn ngữ.