Implicit feedback and lexical stress

Development of perception and production in a classroom environment

Özgür Parlak[1], Nicole Ziegler[2], and Reem Rabea[1]
[1] American University of Sharjah | [2] University of Hawai'i at Mānoa

The current study explored the effects of recasts on the production and perception of primary stress in a classroom context. Following a pretest-posttest-delayed posttest design, 28 L1 Arabic speakers were randomly assigned to intervention and control groups. Participants received four hours of instruction over a period of four days, and the lessons were recorded for stimulated recall. Teaching materials focused on argumentation, and were embedded with the target vocabulary to facilitate incidental mispronunciation. When the intervention group produced target words with misplaced primary stress, they received a recast. The control group did not receive corrective feedback. The results of linear mixed-effects analyses showed that recasts facilitated primary stress development through increased vowel duration. Stimulated recall data confirmed that participants noticed the recasts they received. However, there were no changes in participants' perceptions of stress placement. These findings suggest that incidental pronunciation errors can be addressed through implicit feedback.

Keywords: interaction, corrective feedback, recast, pronunciation, lexical stress

1. Introduction

In English, lexical stress is a prosodic feature with high communicative value. For example, research has shown that first language (L1) English speakers rely on stress cues for word recognition (Cooper et al., 2002) and that misplaced primary stress leads to misunderstandings between second language (L2) speakers (Field, 2005), thereby affecting comprehension and overall communicative success. Correct stress placement has also been shown to have a direct relationship with

https://doi.org/10.1075/jslp.24050.par | Published online: 24 February 2025
Journal of Second Language Pronunciation 11:1 (2025), pp. 76–102. ISSN 2215-1931 | E-ISSN 2215-194X

comprehensibility (Crowther et al., 2015; Kang et al., 2010), with lexical stress predicting comprehensibility and accentedness for all proficiency levels (Saito et al., 2016). Despite this demonstrated importance, only a few studies have explored the development of lexical stress using the interactionist approach to second language acquisition (SLA). The interactionist approach posits that opportunities for exposure to input, production of output, receiving corrective feedback, and engaging in negotiation for meaning are essential for L2 development (Gass, 1988; Long, 1996). Several meta-analyses and reviews covering a wide range of languages and target features have shown a clear link between corrective feedback and morphosyntactic development (e.g., Brown, 2016; Ziegler, 2016). However, these meta-analyses did not include studies with a phonological focus due to a scarcity of such studies at the time of publication. Following the call for more research in this area (Mackey et al., 2012), there has been a growing interest in exploring the impact of corrective feedback on L2 pronunciation (Bryfonski & Ma, 2020; Gooch et al., 2016; Saito, 2013). Generally, these studies have shown that corrective feedback, and particularly recasts, can promote L2 phonological development. However, few studies have focused on the effects of recasts on lexical stress development. Seeking to address this gap, the current study investigated the impact of recasts as an implicit form of corrective feedback on the development of primary stress in a face-to-face (FTF), simulated classroom environment designed to replicate real classroom interactions.

2. Background

2.1 Corrective feedback and L2 phonology

Corrective feedback triggers learner internal mechanisms and facilitates development (Long, 1996) because it can potentially direct learners to notice the linguistic differences between their non-target-like production and the feedback (Gass, 1988; Schmidt, 2001). With six main types, namely, explicit correction, metalinguistic feedback, elicitation, clarification requests, repetition, and recasts (Lyster & Ranta, 1997), corrective feedback can be placed on an implicit-explicit continuum according to the degree of overtness (Ellis & Sheen, 2011). Feedback that is more explicit is likely to interrupt the natural flow of interaction when drawing learners' attention to the error. On the other hand, implicit feedback, such as recasts, signals the error while preserving the focus on meaning and minimally disrupting the flow of communication. Recasts, defined as a target-like reformulation of the learner's non-target-like production, are a commonly occurring type of corrective feedback in the classroom (Couper, 2019; Panova & Lyster, 2002)

and naturalistic settings (Mackey et al., 2003; Oliver, 2009). Recasts are minimally intrusive and maintain focus on meaning, making them an ideal choice when addressing errors that occur incidentally (Long, 2007). Also, recasts concurrently provide negative evidence, which signals an error, and positive evidence, which provides a target-like model (Leeman, 2003). Positive evidence is particularly useful when targeting pronunciation errors, because it provides an exemplar on which learners can model their production. Although recasts do not explicitly indicate the type of error, research suggests that learners do recognize recasts targeting pronunciation errors as corrections (Carpenter et al., 2006), and they are likely to repair their errors following a recast (Lyster, 1998). Recasts can be beneficial for a range of error types; however, research suggests that they may be less effective for grammatical and lexical errors than phonological errors (Yang, 2016). Additionally, learners are generally receptive to corrective feedback provided in response to their pronunciation errors (Saito, 2021), with research suggesting that they prefer less intrusive feedback over explicit feedback (Bryfonski & Ma, 2020; Huang & Jia, 2016).

The relatively few studies conducted so far support the role of recasts in facilitating L2 phonological development. For example, among studies that focused on segmental development, Saito (2013) found that participants who received recasts during form-focused instruction (FFI) were able to generalize their gains in perception and production of /ɹ/. Similarly, Gooch et al. (2016) investigated the changes in L1 Korean learners' production of /ɹ/ after receiving recasts and prompts as part of FFI. The results revealed that recasts were useful for improving controlled production of /ɹ/. Lee and Lyster (2016) investigated the perception of the /i/-/ɪ/ phonemic contrast by L1 Korean learners and found that recasts played a facilitative role in perceptual development. As for suprasegmental features, Bryfonski and Ma (2020) investigated the effects of recasts and metalinguistic feedback on the development of Mandarin tones. Although their results demonstrated that both feedback types were beneficial, recasts were more effective in promoting productive development. Finally, in a study that focused on lexical stress development in FTF and synchronous computer-mediated environments, Parlak and Ziegler (2017) found that learners who received recasts utilized longer syllable duration to improve their primary stress placement. The results of these studies highlight the need for further research exploring the role of recasts on learners' perceptions and production of phonological features.

2.2 Lexical stress

Lexical stress is a prominent suprasegmental feature in English with direct implications for comprehensibility. Research has shown that misplaced stress dramatically

impacts L1 and L2 listeners' judgment of L2 speech for intelligibility (Field, 2005; Zielinski, 2008) as well as comprehensibility and accentedness (Kang, 2010; Isaacs & Trofimovich, 2012; Saito et al., 2016). Also, L1 listeners rely on stress cues for word recognition (Cooper et al., 2002; Jesse et al., 2017); and therefore, misplaced stress can hamper their processing of the speech signal and impair comprehension (Cutler & Clifton, 1984). Mastering lexical stress in English can also be challenging for L2 speakers (Hahn, 2004) as English words do not always follow a predictable stress pattern. Even when stress patterns are predictable, such as in noun-verb homographs, L2 speakers may still struggle with stress placement, perhaps because of being more exposed to one form over the other. For example, according to the Corpus of Contemporary American English (COCA, Davies, 2008–), the verb form of *produce* is 25.46 times more frequent than the noun form. As a result, L2 learners may make systematic errors by adopting the stress pattern of the more frequent form when producing the less frequent form, highlighting the need for more research exploring L2 primary stress development.

3. The current study

The current study investigated the impact of recasts on incidentally occurring primary stress errors during lessons that focused on argumentation in a simulated classroom environment. No explicit focus or instruction on pronunciation was included in these lessons. Avoiding form-focused instruction, this study sought to improve ecological validity by simulating a classroom context where the teacher may not have the time or skills to provide pronunciation instruction, an issue highlighted by previous research (Foote et al., 2011). Instead, recasts were provided in response to incidental pronunciation errors, much as would be done in a classroom. The following research questions were addressed:

RQ1: What are the effects of recasts on learners' production of primary stress?
RQ2: What are the effects of recasts on learners' perception of primary stress?

4. Method

4.1 Participants

Participants were 28 Intensive English Program students studying at a university in the Arabian Gulf. After receiving Institutional Review Board approval, they were recruited via classroom announcements and were given a payment of approximately USD 54.50 for participation. They were native speakers of various

Arabic dialects, including Egyptian ($n=3$), Gulf Arabic ($n=16$), Levantine ($n=6$), Libyan ($n=1$), Moroccan ($n=1$), and Sudanese ($n=1$). The variety of dialects was not considered to be a moderating variable as the dialects were all eastern dialects with minimal prosodic differences (Mustafawi, 2018). Also, primary stress placement is highly predictable in all Arabic dialects (Hellmuth, 2013) with duration as the strongest acoustic correlate followed by intensity and F0 as the two other acoustic correlates (Almbark et al., 2014; de Jong & Zawaydeh, 1999). Eight participants were a beginner-level speaker of another language, namely, Dutch ($n=1$), French ($n=3$), German ($n=1$), and Turkish ($n=4$). The knowledge of these languages was not considered to be a moderating variable due to participants' low proficiency level. No participant disclosed a hearing- or speaking-related problem, nor had any of the participants lived in an English L1 speaking country. Eight participants indicated that they had visited the UK or the USA for a period of three to four weeks. These participants were equally distributed between the two experimental groups. Based on the Cambridge English Placement Test administered pretreatment, participants were mainly B1- and B2-level speakers on the Common European Framework of Reference for Languages (CEFR) scale. Although there was one A1-level participant in the control group and one C1-level participant in the intervention group, the data from these participants were not statistical outliers on the pre- and posttests; and therefore, their data were maintained. Participants were randomly assigned to the intervention and control groups. At the onset of the study, there were 20 participants in each group; however, after attrition, there were 11 participants in the control group and 17 in the intervention group. Participants cited scheduling conflicts as their reason for withdrawal from the study. Table 1 provides an overview of participant demographics.

Table 1. Participant demographics

Group	Mean age	Sex		CEFR scores			
		Male	Female	C1	B2	B1	A1
Control	18 ($SD=0$)	4	7	0	6	4	1
Intervention	18.23 ($SD=0.47$)	10	7	1	7	9	0

4.2 Materials

4.2.1 *Production tests*

Production tests consisted of carrier sentences which were 8 to 12 syllables in length. The sentences were printed on paper and double-spacing was used for

ease of reading. There was a total of 90 carrier sentences, 15 embedded with the target words and 75 acting as distractors. Five carrier sentences were embedded with adjectives ending with the suffix -ic, five with noun homographs, and five with verb homographs (Table 2). The verb homographs were used in plural form, with a modal verb, or as an infinitive so that they were orthographically identical to the noun forms. Target words were embedded in sentence-medial position, and they were presented in different carrier sentences on the pretest, posttest, and the delayed posttest (Table 3). Words with the suffix -ic were selected because piloting demonstrated that learners from the same population have a tendency to misplace the primary stress for some of these words. For example, they produced the word *strategic* with stress on the initial syllable following the stress pattern of the noun form, *strategy*. This could be because the nominal forms are more frequent than the adjectival counterparts; and therefore, learners were probably more exposed to hearing the nominal forms, affecting their production of the adjectival forms. Similarly, it was observed that learners from the same population had challenges with producing noun-verb homographs with the correct stress patterns. This usually happened in the form of producing both the noun and verb with the same stress pattern. For example, they would produce the word *object* with initial stress in the verb position or the word *produce* with ultimate stress in the noun position.

Table 2. Target vocabulary

-ic words	allergic	caloric	comedic	strategic	symphonic
homographs	object	produce	progress	record	suspect

Table 3. Example carrier sentences

Test	**Carrier sentence**
Pretest	We are in a *strategic* location.
	The company made *progress* in two years.
	Japan continued to *progress* despite problems.
Posttest	We showed them *strategic* areas.
	Students made *progress* in a short time.
	You cannot *progress* without any help.
Delayed Posttest	You need to be *strategic* with her.
	I think their *progress* is very slow.
	The illness can *progress* very rapidly.

4.2.2 *Perception tests*

A forced-choice perception experiment with two parts was created on PsychoPy (Peirce et al., 2019) to measure participants' ability to identify stress placement. Participants took the experiment in a quiet computer lab. They wore Roland RH-5 model headphones to ensure that they could hear the stimuli clearly. The first part of the experiment required participants to listen to juxtaposed noun-verb pairs and decide the order in which they heard the words by choosing between two possible orderings displayed on the screen (e.g., PROgress — proGRESS vs. proGRESS — PROgress). On-screen instructions informed participants that capitalized letters indicated stress placement. Each juxtaposed stimulus played twice. There was a one-second gap between the homograph pairs, and a two-second gap between the first and the second rendering. The stress variation between noun-verb homographs provided an opportunity to explore participants' ability to distinguish between juxtaposed stress patterns, and thus this was the focus of the first part of the perception experiment.

The second part focused on words with the suffix -ic. Participants heard each individual word twice with a one-second gap between the first and second rendering. Then, they indicated the location of stress by selecting one of the orthographic representations displayed on the screen (e.g., STRAtegic, straTEgic, strateGIC). Participants were not allowed to replay the stimuli for either experiment. All stimuli were produced by a male speaker of American English, and they were presented in a randomized order for each participant. The posttest and the delayed posttest consisted of the same items as the pretest but in randomized orders. Prior to each test, participants completed a trial session with non-target vocabulary to ensure that they were familiar with the procedures.

4.2.3 *Biodata*

An online questionnaire was created to collect data on basic biographic information such as age, sex, L1, and L2(s). Additionally, participants took the Cambridge English Placement Test at the onset of the study to help determine their English proficiency level.

4.2.4 *Teaching Materials*

The content of teaching materials was freshman-level argumentation strategies. There were four sets materials for four days of instruction on the following topics: (1) rhetorical analysis, (2) claim, reason, and support during argumentation, (3) facts and opinions, and (4) logical fallacies. Each set included a PowerPoint document, a handout for matching vocabulary with their meanings, and a handout for debate preparation. All PowerPoint documents included slides with carrier sen-

tences embedded with the target vocabulary. For example, a slide that asked participants to identify claims had a list of sentences embedded with target words (e.g., Instead of throwing away their *produce*, supermarkets could give them to the poor.). The vocabulary handouts were also embedded with target words as well as distractors. Each vocabulary handout introduced approximately 10 words and required participants to match words with their meanings. The part of speech was provided making it possible to present noun-verb homographs as two separate words. Finally, the handouts for debate preparation introduced the topics and required participants to work in groups to prepare arguments supporting their position. The debate handouts encouraged participants to use the target vocabulary when formulating arguments.

4.3 Procedure

Participants attended eight data collection sessions carried out either in a classroom or a computer lab. After signing the consent form during the first session, participants completed the production pretest. The written instructions asked participants to read aloud the sentences at a normal pace and informed them that if they stumble upon a sentence, they could simply repeat it. When reading aloud the sentences, participants wore a Shure WH20 XLR head-worn unidirectional microphone connected to an iPad via a Roland Duo-Capture EX USB audio interface. The microphone was positioned at 30 degrees off-axis and approximately 3 cm away from the participants' mouth. Audio data were saved as 44,100 Hz .wav files.

During the second session, participants completed the background questionnaire, took the perception pretest, and took the Cambridge English Placement Test. Then, participants were randomly assigned to the intervention and control groups. Both groups attended four lessons on argumentation techniques. The lessons were delivered by the first author, whose pronunciation is comparable to a North American English speaker. With 18 years of teaching experience and a PhD in linguistics, he had extensive experience providing instruction on the target content. Each lesson started with introducing the key concepts followed by classroom practice to reinforce participants' understanding of the content. Participants randomly took turns to read aloud statements displayed on a slide, and depending on the focus of the day, performed tasks such as identifying rhetorical elements, claims, facts and opinions, or logical fallacies. There was a total of 30 sentences, 15 embedded with target words and 15 with distractors, divided among the four lessons yielding 7–8 practice sentences per lesson. It took approximately three minutes per lesson to go through these sentences as a class. When the intervention group misplaced primary stress on a target word, they received a

recast (Excerpt 1). Next, participants completed the vocabulary handout in pairs. To check their answers, the instructor read aloud the definitions and participants provided the matching word. Once again, when participants misplaced primary stress, they received a recast. This second feedback session lasted approximately two minutes per lesson. Next, participants completed the debate handout in groups to prepare their arguments for the debate task. The instructions asked participants to use target words and distractor words when formulating arguments. The instructor monitored group work and asked each group to share their prepared arguments. Once again, when participants produced a target word with misplaced stress, they received a recast. Finally, each lesson ended with a series of mini debates where teams tried to refute the arguments by the opposing team. The total time spent on actual recast provision did not exceed two minutes per each of the 60-minute lessons, with the intervention group receiving an average of 18 recasts per lesson. The control group completed the same activities but did not receive corrective feedback. Participants took the production and perception posttests within 2–3 days of completing the last lesson based on their availability.

Excerpt 1. *Example recast provision*

Student: I have lived in this area for 20 years and I know all the STRAtegic locations.

Instructor: straTEgic [locations]

Student: [locations]

Instructor: uh huh… so what's this one?

Student: This is ethos.

Three days after the last lesson, participants attended the stimulated recall session. The rationale for conducting stimulated recall was to determine whether participants noticed the corrective force of recasts. All lessons were video recorded; however, the recording from one lesson was discarded due to sound problems. The total duration of the stimulated recall video was 30 minutes, consisting of approximately 10 minutes from each of the three lessons. The stimulated recall session was conducted as a group because of participant availability. Following recommendations by Gass and Mackey (2016), a script was developed with questions that aimed to elicit information from participants without directing them toward a specific answer (e.g., Can you tell me what you were thinking at that point?). Finally, one week after the posttest, participants took the delayed posttest (see Figure 1 for an overview).

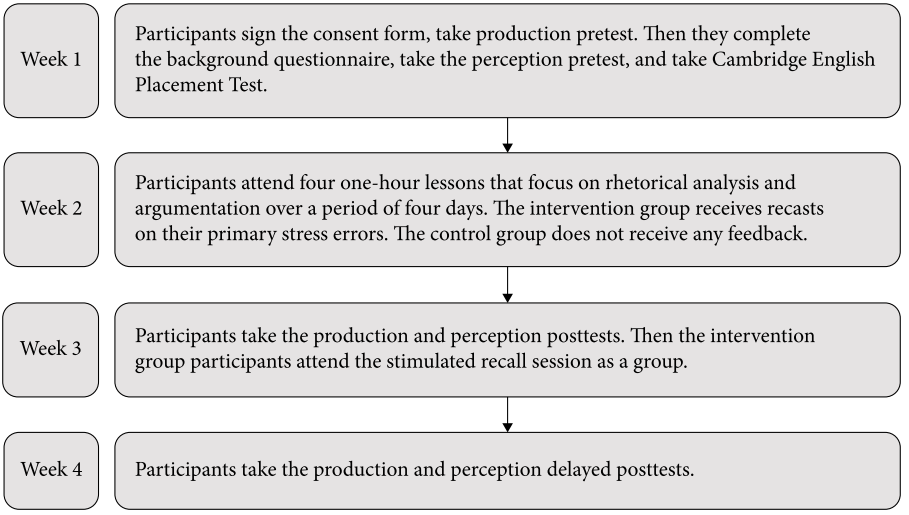

Figure 1. Overview of procedure

5. Analysis

To answer the first research question, participants' production data were analyzed acoustically. The recordings were loaded onto Praat (Boersma & Weenink, 2023) for marking word-initial and word-final boundaries. Then, target words and their corresponding TextGrid annotations were extracted from carrier sentences. Although the total number of productions including the pretest, the posttest, and the delayed posttest should have been 420 words for each of the three categories of words, there were instances of missing words due to participants skipping a carrier sentence or producing a different word instead of the target word (e.g., *calorie* instead of *caloric*). There were also a few instances of recording problems. When there was a missing word in one of the tests, the corresponding productions were also removed. Similarly, problematic productions were removed along with their corresponding pretest, posttest, or delayed posttest productions to maintain a balanced data set. At this stage there were 336 words with the suffix -ic, 396 nouns, and 390 verbs. As the current research investigates acoustic development, it was necessary to focus on words that were produced with misplaced stress during the pretest to determine whether there would be any improvement on the posttest and delayed posttest. To achieve this goal, words produced with correct stress on the pretest needed to be removed from the data set along with their corresponding posttest and delayed posttest productions. The pretest data were auditorily analyzed by the first author and words with correct stress were identified. The third author carried out the same analysis with 20% of the data, yielding a substan-

tial agreement between the two expert listeners ($\kappa = .95$, $p = .000$). The identified words then were removed along with their posttest and delayed posttest productions. Following this adjustment, 276 words with the suffix -ic, 231 nouns, and 156 verbs were ready for acoustic analysis. Next, using the Montreal Forced Aligner (McAuliffe et al., 2017), segmental boundaries were marked automatically; but also, a visual and auditory inspection was conducted and boundaries were modified if necessary. After that, duration, intensity, and F0 values were extracted from each vowel. Finally, a ratio measure was calculated prior to running inferential statistics as production of lexical stress should be examined in relation to other syllables in the same word (Ladefoged, 2006). A target syllable produced with longer duration or higher intensity on the posttest may not necessarily indicate development if the other syllables in the same word are also produced with relatively longer duration and higher intensity. Therefore, a ratio measure was calculated to capture the changes in the target syllable relative to the other syllables in the same word. For words with the suffix -ic, the value of the penultimate vowel was divided by the sum of values of the initial and ultimate vowels. For nouns, the value of the initial vowel was divided by the value of the ultimate vowel. For verbs, the value of the ultimate vowel was divided by the value of the initial vowel.

Linear mixed-effects (LME) models with restricted maximum likelihood (REML) were fitted for the analyses using lme4 package (Bates et al., 2015) in R. LME analysis was chosen due to its ability to handle small sample sizes. It is possible to fit an LME model with as few as five participants when the research interest focuses on fixed effects (Wiley & Rapp, 2019). Also, multiple observations from each participant at each time compensates for the sample size (Maas & Hox, 2005), which was the case for the current study. Finally, models fitted using REML are less likely to produce a Type 1 error due to their robustness (Luke, 2017). The LME models fitted for the analysis of data had *time* and *condition*, and the interaction between the two as the fixed effects, and *participants* as the random effect. *Words* were not entered into the model as a random effect because doing so led to a singularity error. Therefore, the model was simplified by removing the random effect for *word* so that the random effect structure was supported by the data (as recommended by Matuschek et al., 2017). After the models were fitted, the distribution of residuals was checked for assumptions of linearity and homoscedasticity (West et al., 2015). The plots did not indicate any deviations from linear form or homoscedasticity. Pairwise comparisons were obtained using the emmeans package (Lenth, 2023). Finally, following Brysbaert and Stevens's (2018) recommendation for estimating the effect size for mixed-effects analyses in psychology research, Cohen's *d* was calculated for pairwise comparisons.

The acoustic analyses mainly focused on vowel duration given that duration is a robust correlate with predictive power for stress placement (Gordon & Roettger,

2017). In addition, L1 Arabic speakers utilize duration for stress placement in their native language as well as in English (de Jong & Zawaydeh, 2002; Zuraiq & Sereno, 2021); and earlier studies have shown that L1 Arabic speakers may utilize duration more than intensity or F0 to modify their stress placement in English (e.g., Parlak, 2024). Nevertheless, intensity and F0 values were also analyzed as higher intensity or higher pitch may also indicate stress placement (Cutler, 2005).

To answer the second research question, participants' auditory judgments of stress were analyzed separately for -ic words and noun-verb homographs. There were 420 judgments for the -ic words and 420 judgments for the noun-verb homographs comprising pretest, posttest, and delayed posttest data from both groups. As there was a total of five words with the suffix -ic and five noun-verb homograph pairs, participants could potentially receive a maximum score of five on any of the tests. The auditory judgment data were also analyzed using LME models fitted with *time* and *condition* as the fixed effects and *participants* as the random effect. The residual plots did not indicate a violation of linearity or homoscedasticity.

6. Results

6.1 Research Question 1

The first research question was "What are the effects of recasts on learners' production of primary stress?" In order to examine the development of stress acoustically, it was necessary to analyze the changes in production relative to earlier productions by the same group (Ladefoged, 2006). This necessitated conducting within-groups comparisons instead of between-groups comparisons. Nevertheless, figures with confidence intervals are also provided, which will allow the reader to make between-groups comparisons visually. The descriptive statistics for the -ic words showed that the intervention group produced longer vowels on the posttest and the delayed posttest (Table 4). As for inferential statistics, the LME model yielded a conditional r^2 value of .17 and a marginal r^2 value of 0.09. The results of pairwise analyses indicated that the intervention group produced penultimate vowels with significantly longer duration on the posttest ($\beta = 0.156$, $SE = 0.041$, $df = 247$, $t = 3.799$, $p = .000$) with a moderate effect size ($d = 0.679$), as well as on the delayed posttest ($\beta = 0.134$, $SE = 0.041$, $df = 247$, $t = 3.278$, $p = .003$), again with a moderate effect size ($d = 0.653$). The control group's production of penultimate vowels did not change on the posttest ($\beta = 0.033$, $SE = 0.050$, $df = 246$, $t = 0.672$, $p = .779$) or the delayed posttest ($\beta = 0.091$, $SE = 0.050$, $df = 246$, $t = 1.834$, $p = .160$). Figure 2 provides an overview of the change in vowel duration.

Table 4. Duration ratios for -ic words

Condition	Time	*Mean*	*SD*	*Min*	*Max*
Control	T1	0.459	0.174	0.231	1.050
	T2	0.493	0.193	0.228	1.210
	T3	0.551	0.217	0.211	1.180
Intervention	T1	0.489	0.146	0.231	0.872
	T2	0.648	0.297	0.111	1.740
	T3	0.624	0.253	0.155	1.440

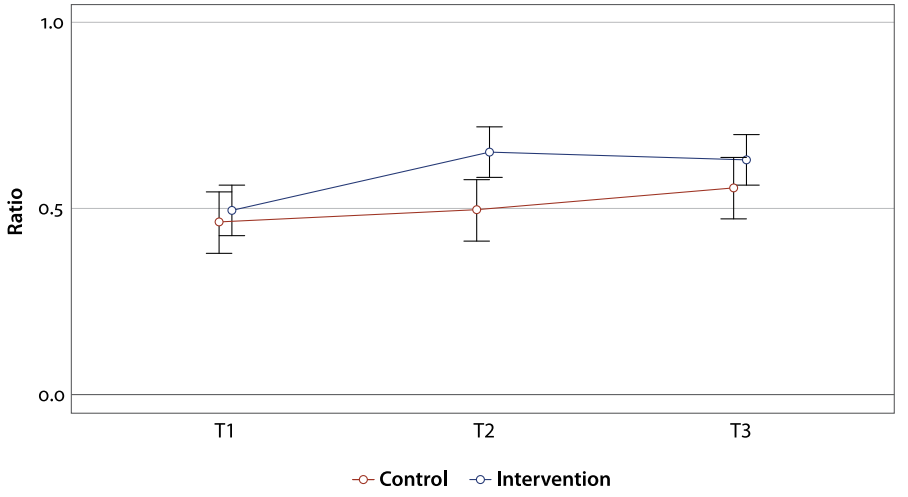

Figure 2. Vowel duration for -ic words

The descriptive statistics for nouns indicated that the intervention group produced initial vowels with longer duration on both the posttest and the delayed posttest (Table 5). The LME model yielded a conditional r^2 value of .18 and a marginal r^2 value of .12. The pairwise comparisons indicated that the intervention group produced initial vowels with significantly longer duration on the posttest ($\beta = 0.330$, $SE = 0.075$, $df = 200$, $t = 4.394$, $p = .000$) with a large effect size ($d = 0.813$) whereas they did not maintain their gains on the delayed posttest ($\beta = 0.129$, $SE = 0.075$, $df = 200$, $t = 1.721$, $p = 0.200$). The control group did not achieve gains on the posttest ($\beta = 0.016$, $SE = 0.097$, $df = 200$, $t = 0.161$, $p = .986$) or the delayed posttest ($\beta = -0.005$, $SE = 0.097$, $df = 200$, $t = -0.053$, $p = .999$).

Finally, according to the descriptive statistics for verbs, the intervention group produced ultimate vowels with longer duration on both the posttest and delayed posttest (Table 6). The LME model yielded a conditional r^2 value of .20 and a

Table 5. Duration ratios for nouns

Condition	Time	*Mean*	*SD*	*Min*	*Max*
Control	T1	0.654	0.251	0.210	1.480
	T2	0.669	0.328	0.136	1.450
	T3	0.649	0.335	0.227	1.500
Intervention	T1	0.700	0.297	0.235	1.330
	T2	1.030	0.491	0.283	1.970
	T3	0.830	0.443	0.291	1.960

Table 6. Duration ratios for verbs

Condition	Time	*Mean*	*SD*	*Min*	*Max*
Control	T1	1.100	0.482	0.411	1.960
	T2	1.020	0.421	0.392	1.750
	T3	1.090	0.508	0.269	1.930
Intervention	T1	1.020	0.408	0.331	1.790
	T2	1.370	0.519	0.517	2.200
	T3	1.400	0.474	0.330	1.990

marginal r^2 value of .11; and the pairwise comparisons indicated that the intervention group produced ultimate vowels with significantly longer duration on the posttest ($\beta = 0.345$, $SE = 0.109$, $df = 129$, $t = 3.168$, $p = .005$) with a moderate effect size ($d = .749$) and the delayed posttest ($\beta = 0.376$, $SE = 0.109$, $df = 129$, $t = 3.452$, $p = .002$) with a large effect size ($d = .859$). In contrast, there were no gains for the control group on the posttest ($\beta = -0.078$, $SE = 0.150$, $df = 129$, $t = -0.518$, $p = .863$) nor the delayed posttest ($\beta = -0.005$, $SE = 0.150$, $df = 129$, $t = -0.035$, $p = .999$). Figure 3 provides an overview of the change over time based on the fitted data. As mentioned earlier, statistical analyses were also conducted for intensity and F0. However, no significant differences were found between pretest and posttest productions of intensity or F0 for any group of words (Table 7 and Table 8).

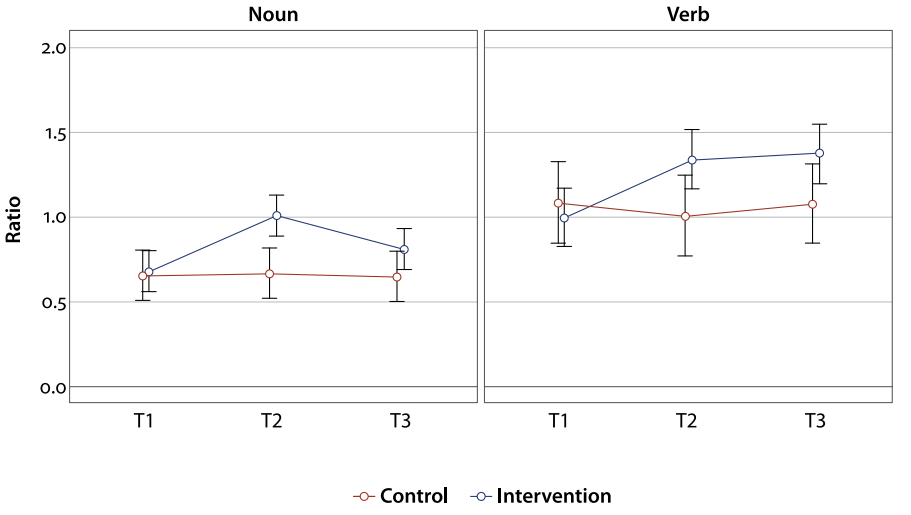

Figure 3. Vowel duration for homographs

Table 7. LME analysis of intensity and Fo ratios for homographs

Type	Correlate	Condition	Time	*β*	*SE*	*df*	*Lower CL*	*Upper CL*
Nouns	Intensity	Control	T1	0.964	0.014	97	0.936	0.993
			T2	0.967	0.014	97	0.939	0.996
			T3	0.956	0.014	97	0.927	0.984
		Intervention	T1	0.976	0.011	86	0.954	0.999
			T2	0.999	0.011	86	0.976	1.021
			T3	0.992	0.011	86	0.969	1.014
	Fo	Control	T1	0.987	0.028	64.4	0.930	1.040
			T2	0.988	0.028	64.4	0.931	1.040
			T3	0.988	0.028	64.4	0.931	1.050
		Intervention	T1	0.966	0.022	60.3	0.920	1.010
			T2	1.010	0.022	60.3	0.965	1.060
			T3	0.976	0.022	60.3	0.931	1.020
Verbs	Intensity	Control	T1	0.983	0.019	87.3	0.945	1.020
			T2	1.010	0.019	87.3	0.971	1.050
			T3	1.006	0.019	87.3	0.968	1.040
		Intervention	T1	0.994	0.014	69.6	0.965	1.020
			T2	1.005	0.014	69.6	0.977	1.030
			T3	1.028	0.014	69.6	1.000	1.060

Table 7. *(continued)*

Type	Correlate	Condition	Time	*β*	*SE*	*df*	*Lower CL*	*Upper CL*
	Fo	Control	T1	0.983	0.019	87.3	0.945	1.020
			T2	1.010	0.019	87.3	0.971	1.050
			T3	1.006	0.019	87.3	0.968	1.040
		Intervention	T1	0.994	0.014	69.6	0.965	1.020
			T2	1.005	0.014	69.6	0.977	1.030
			T3	1.028	0.014	69.6	1.000	1.060

Table 8. LME analysis of intensity and Fo ratios for -ic words

Correlate	Condition	Time	*β*	*SE*	*df*	*Lower CL*	*Upper CL*
Intensity	Control	T1	0.501	0.005	93.8	0.491	0.512
		T2	0.509	0.005	93.8	0.499	0.519
		T3	0.511	0.005	93.8	0.501	0.521
	Intervention	T1	0.502	0.004	92.5	0.494	0.511
		T2	0.510	0.004	92.5	0.501	0.518
		T3	0.510	0.004	92.5	0.501	0.518
Fo	Control	T1	0.491	0.010	60.1	0.470	0.512
		T2	0.501	0.010	60.1	0.480	0.522
		T3	0.501	0.010	60.1	0.480	0.522
	Intervention	T1	0.501	0.008	60.3	0.484	0.518
		T2	0.519	0.008	60.3	0.502	0.536
		T3	0.516	0.008	60.3	0.499	0.533

6.2 Research Question 2

The second research question asked "What are the effects of recasts on learners' perception of primary stress?" The LME analyses indicated that the intervention group's perception of stress placement did not change significantly on the posttest ($\beta = 0.412$, $SE = 0.401$, $df = 52$, $t = 1.027$, $p = 0.563$) or the delayed posttest ($\beta = 0.176$, $SE = 0.401$, $df = 52$, $t = 0.440$, $p = 0.898$) for the -ic words. In addition, the control group's perceptions on the posttest ($\beta = 0.545$, $SE = 0.498$, $df = 52$, $t = 1.094$, $p = 0.521$) or the delayed posttest ($\beta = 0.000$, $SE = 0.498$, $df = 52$, $t = 0.000$, $p = 1.000$) did not improve. Table 9 provides a summary of descriptive statistics and Figure 4 provides an overview of the confidence intervals.

Similarly, the intervention group's perception scores for the noun-verb data did not indicate significant differences on the posttest ($\beta = -0.650$, $SE = 0.370$, $df = 53.5$, $t = -1.757$, $p = 0.193$) or the delayed posttest ($\beta = -0.118$, $SE = 0.370$, $df = 53.5$, $t = -0.319$, $p = 0.9454$). Finally, the control group scored the same on the posttest ($\beta = -0.273$, $SE = 0.458$, $df = 51.4$, $t = 0.595$, $p = 0.823$) and the delayed posttest ($\beta = -0.273$, $SE = 0.458$, $df = 51.4$, $t = 0.595$, $p = 0.823$) with no significant gains compared to the pretest. Table 10 provides a summary of descriptive statistics and Figure 5 provides an overview of the confidence intervals.

Table 9. Descriptive statistics for the perception of primary stress for -ic words

Condition	Time	*Mean*	*SD*	*Min*	*Max*
Control	T1	3.270	1.100	2	5
	T2	3.820	1.080	2	5
	T3	3.270	1.350	1	5
Intervention	T1	2.820	1.380	1	5
	T2	3.240	1.440	1	5
	T3	3.000	1.370	1	5

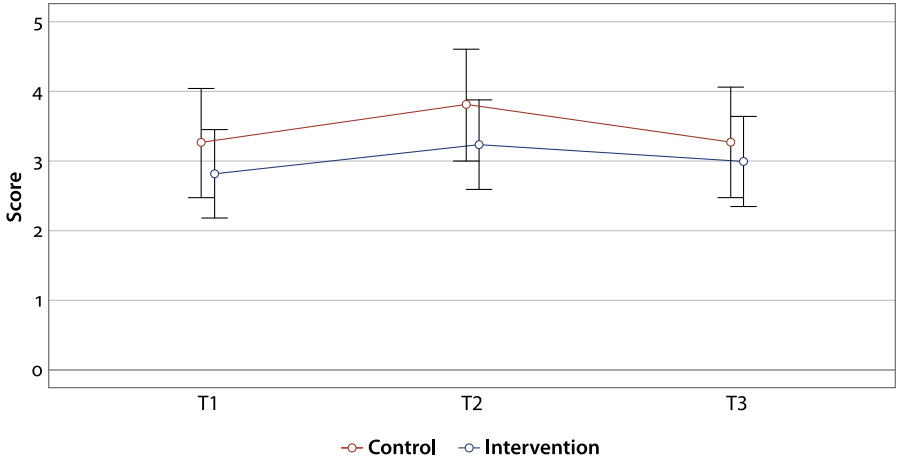

Figure 4. Perception scores for -ic words

Table 10. Descriptive statistics for the perception of primary stress for homographs

Condition	Time	*Mean*	*SD*	*Min*	*Max*
Control	T1	3.910	0.944	2	5
	T2	3.640	1.030	2	5
	T3	3.640	1.120	2	5
Intervention	T1	3.24	0.970	2	5
	T2	2.59	1.120	1	4
	T3	3.12	1.450	0	5

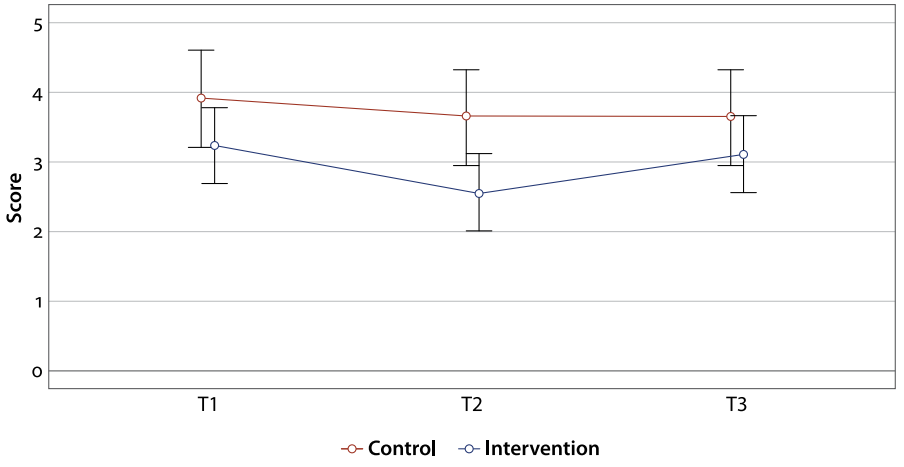

Figure 5. Perception scores for homographs

7. Discussion

The current study explored the effects of recasts on learners' production and perception of primary stress during classroom interaction where the main focus of instruction was argumentation rather than pronunciation. The results indicated a positive impact of recasts on primary stress development for the three categories of target words. The intervention group produced stressed syllables with statistically longer duration on the posttest and delayed posttest for the -ic words. Their gains for nouns, however, were less durable and were only significant on the posttest, suggesting these differences may have been less salient to learners than the other target words. Participants' gains for verbs were significant on both the posttest and the delayed posttest. Taken together, the observed increase in vowel duration across all three categories of target words is an indica-

tion of development, as duration is the most robust correlate of stress in English (Gordon & Roettger, 2017). Furthermore, duration is a strong correlate of stress in Arabic (de Jong & Zawaydeh, 2002; Zuraiq & Sereno, 2021), and earlier studies have shown that L1 Arabic speakers utilize duration to adjust their primary stress placement in English (Parlak, 2024). Importantly, pronunciation was not the focus of instruction, with learners instead focusing on argumentation strategies. Nonetheless, findings suggest that even if learners are directing attention towards other features or content, they may still benefit from implicit feedback on pronunciation.

The productive gains were moderate to large in terms of effect sizes, and there was no significant gain for nouns on the delayed posttest. This positive but complex developmental pattern could be due to a few factors. The variety of stress patterns may have increased processing demands, which may have impacted performance (O'Grady, 2015). Yet, despite the complex patterns and differential rates of improvement, implicit feedback led to increased duration indicating adjustment of stress placement. Also, pronunciation development usually requires more time than other aspects of language (Derwing & Munro, 2015). Therefore, the participants probably need more opportunities for input, output, and feedback to further improve their primary stress production. Nevertheless, the amount of improvement observed after only four lessons is noteworthy, particularly because there was no explicit feedback or instruction targeting pronunciation and recasts were provided in a rather short portion of the actual classroom time. Thus, these results suggest that even short periods of feedback provision on primary stress errors can have a positive impact on production.

Another noteworthy point is the amount of variance explained by the random effect *participant*. The differences between the conditional and marginal r^2 values indicate that the random effect *participant* adds to the explanatory power for each of the three LME models (8% for -ic words, 6% for nouns, and 9% for verbs), suggesting that a meaningful amount of the observed gains in duration is explained by individual differences which potentially impact the extent to which participants can benefit from recasts. Although the current study did not investigate the effects of individual differences on the outcome, this statistical finding suggests that individual differences may mediate the impact of recasts on L2 pronunciation development, highlighting the need for more research in this particular area.

On the other hand, the results of the perception experiments did not yield significant changes on the posttest or the delayed posttest. It is possible that participants may not have fully attended to the task at hand, instead prioritizing task completion over attending to input in order to be done quickly (Toth et al., 2013). To explore this possibility, post hoc analyses exploring reaction time data were conducted. The intervention group's mean reaction time in seconds for the -ic

words was 6.03 (*SD* = 2.89) on the pretest, 5.46 (*SD* = 2.26) on the posttest, and 4.75 (*SD* = 3.61) on the delayed posttest. Similarly, the control group's reaction time decreased over time, with a mean reaction time of 5.79 (*SD* = 1.92) on the pretest, and 4.79 (*SD* = 1.96) on the posttest, and 4.27 (*SD* = 1.80) on the delayed posttest. The duration of each stimulus for the -ic words was two to three seconds. Therefore, even at a mean of 4.75 or 4.27 seconds, participants must have heard the complete stimulus once before making a judgment. Reaction times for noun-verb judgments followed a similar pattern, suggesting that learners heard the stimuli once and made a judgment without hearing the second rendering. This single rather than double exposure may have contributed to the lack of improvement in perception scores, as learners may have been relying on incomplete implicit knowledge to make rash decisions. Research suggests that when learners react quickly, they may be drawing on implicit rather than explicit knowledge (e.g., Rebuschat, 2013). Given that previous research indicates that learners may need longer periods of time to develop implicit knowledge (Ellis et al., 2009), the findings suggest learners may need extended exposure to stress patterns to develop their implicit knowledge over time.

Current perception results seem to provide support for the view that perception is not a prerequisite for production (de Leeuw et al., 2021) and contradict with speech acquisition models which suggest that perception develops before production, such as Flege's (1995) Speech Learning Model (SLM) or Best and Tyler's (2007) Perceptual Assimilation Model (PAM), as learners' gains in production were independent of gains in perception. However, this may be due to a difference in focus, as the SLM and the PAM focus on segmental features and do not address lexical stress directly. In addition, judgment of lexical stress can be challenging for untrained listeners, such as the participants in this study, as it has been shown that limited language proficiency or linguistic awareness may negatively impact performance on a range of listener judgment tasks (Yan & Ginther, 2017). In a similar vein, Saito and Plonsky (2019) suggest that expert listeners trained in phonology and phonetics may be better able to judge lexical stress when conducting auditory analyses. Interestingly, the participants' responses during the stimulated recall session showed that they were aware of the stress differences between their production and the recast. However, learners' awareness of differences between their production and the recast may be different from identifying the location of stress in a forced-choice task. For example, during a forced-choice task, learners experience time pressure and thus may direct attention to task completion, rather than to noticing differences in stimuli. It may also be possible that perception is not a prerequisite for production in the case of lexical stress development, underscoring the need for further research to investigate this particular issue.

Another noteworthy point comes from the qualitative stimulated recall data which showed that participants were able to recognize recasts as corrective feedback. Multiple participants explicitly described the differences in pronunciation between their own production and the stress placement by their interlocutor, which suggests noticing of the recasts (Excerpts 2).

Excerpts 2. *SR comments*

⇒ "When you said straTEgic, the word just got different for us because we were saying STRAtegic,"

⇒ "The difference between PROduce and proDUCE was new to me."

⇒ "You said the word differently and I was like what?"

These comments demonstrate participants' noticing of the recasts as corrections, particularly the differences in lexical stress. Some participants also commented on how they actively directed their attention to the target items (Excerpts 3).

Excerpts 3. *SR comments*

⇒ "Whenever I am nodding my head, I recognize the pronunciation."

⇒ "I read the sentence, then you said REcord or reCORD, then I repeat the word. I was thinking about the difference between them."

⇒ "I was explaining to him how there were two words like progRESS or PROGress."

This last excerpt demonstrates noticing and subsequent understanding of the target items, as this remark referred to a language-related episode (LRE, Swain & Lapkin, 1995) in which one participant was sharing their explicit knowledge with the other participant. Comments from the stimulated recall align with previous research demonstrating that learners are able to notice the corrective intent of recasts when the target is a phonological feature (Carpenter et al., 2006).

Overall, results are supportive of previous interactionist studies which argue for the positive role of implicit corrective feedback in facilitating L2 phonological development (Parlak, 2024; Saito, 2013). Findings also provide important information regarding the suitability of recasts depending on proficiency level, with results lending support to Bryfonski and Ma (2020) by showing that recasts could potentially benefit intermediate-level learners. In contrast to Ammar and Spada (2006), who argued that recasts may be more suited for advanced-level learners, the current findings suggest that not only are intermediate-level learners able to notice the corrective nature of recasts, their L2 pronunciation may also benefit. However, it is important to highlight that proficiency level may have differential effects on the usefulness of corrective feedback depending on the target language feature, as previous research did not examine lexical stress (Ammar & Spada, 2006, possessive determiners; Bryfonski & Ma, 2020, Mandarin tones).

In addition, recasts may be more effective on salient forms which have direct implications on meaning (Long, 2007). Stress as a phonetic feature is inherently salient and can also be utilized to enhance the salience of other target features (Gass, 2018), suggesting that type of target feature may play a role in the effectiveness of recasts. Future research is necessary to better understand the relationship between recasts, proficiency level, and the development of L2 phonology.

Importantly, this study also demonstrated how recasts can facilitate phonological development when pronunciation is not the focus of instruction. Long (2007) explains that recasts can be used to address various production errors that arise incidentally, with the current findings supporting this view by demonstrating phonological improvement following instruction on argumentation. The results suggest that recasts can be an effective pedagogical tool for incidental pronunciation errors; an important finding given that in many language teaching contexts, teachers may lack the skills or resources that would allow them to explicitly teach pronunciation (Foote et al., 2011). Rather, instructors may instead respond to students' individual pronunciation needs by providing recasts as errors occur incidentally.

8. Limitations and future directions

Although this study provided essential information about the role of implicit feedback in L2 phonological development, it has a number of limitations. Importantly, this study measured the impact of recasts on trained vocabulary; therefore, future research should explore to what extent implicit feedback may support learners' ability to generalize gains to untrained words with the same stress pattern. It should also be noted that even if learners are able to make generalizations, lexical stress in English is highly complex and the stress pattern for many words needs to be learned individually. The predictable stress patterns for noun-verb homographs and derived words form a category that is different from the non-predictable stress pattern for many non-derived words. In addition, the production tests elicited controlled speech rather than spontaneous speech. This methodological decision was appropriate as it allowed elicitation of the target forms consistently, which would have been difficult to capture using a spontaneous speech task. Although there are arguments in support of making generalizations based on controlled speech tasks (e.g., Xu, 2010), it is important to acknowledge that L2 speech can be hampered by processing demands of spontaneous speech. Therefore, the findings should be interpreted with caution as learners may not be able to sustain their gains consistently during spontaneous production. Furthermore, although it was possible to observe meaningful pro-

ductive gains on the delayed posttest, it should be noted that due to logistical constraints the delayed posttest was given only one week after the posttest. Future studies would benefit from longer periods of time between the posttest and the delayed posttest. Finally, although it was possible to observe meaningful productive gains after four lessons, L2 learners need longer periods of time to improve their pronunciation through extensive amounts of input, corrective feedback, and output. As such, future studies with longitudinal designs would provide a more comprehensive understanding of the relationship between corrective feedback and L2 phonological development.

Funding

Research funded by American University of Sharjah (FRG19-M-S76) to Özgür Parlak. The work in this paper was supported, in part, by the Open Access Program from the American University of Sharjah, and made Open Access under a CC BY 4.0 license through payment of an APC on behalf of the author. This paper represents the opinions of the author(s) and does not mean to represent the position or opinions of the American University of Sharjah.

References

Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014). Acquiring the phonetics and phonology of English word stress: Comparing learners from different L1 backgrounds. *Concordia Working Papers in Applied Linguistics*, *5*, 19–35.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*(4), 543–574.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O. S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–24). John Benjamins.

Boersma, P. & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer program]. Version 6.3.10. Available at http://www.praat.org/

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*(4), 436–458.

Bryfonski, L., & Ma, X. (2020). Effects of implicit versus explicit corrective feedback on Mandarin tone acquisition in a SCMC learning environment. *Studies in Second Language Acquisition*, *42*, 61–88.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1).

Carpenter, H., Jeon, K. S., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition*, *28*, 209–236.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*(3), 207–228.

Couper, G. (2019). Teachers' cognitions of corrective feedback on pronunciation: Their beliefs, perceptions and practices. *System*, *84*, 41–52.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal*, *99*(1), 80–95.

Cutler, A. (2005). Lexical stress. In D.B. Pisoni, & R.E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Blackwell.

Cutler, A., & Clifton Jr, C. (1984). The use of prosodic information in word recognition. In H. Bouma, & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Lawrence Erlbaum.

Davies, M. (2008–). *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/

de Jong, K., & Zawaydeh, B.A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, *27*(1), 3–22.

de Jong, K., & Zawaydeh, B.A. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*, 53–75.

de Leeuw, E., Stockall, L., Lazaridou-Chatzigoga, D., & Gorba Masip, C. (2021). Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research*, *37*(4), 587–618.

Derwing, T.M., & Munro, M.J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.

Ellis, R. & Sheen, Y. (2011). Corrective feedback in language teaching. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning* (pp. 593–610). Routledge.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423.

Flege, J.E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425–442.

Foote, J.A., Holtby, A.K., & Derwing, T.M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22.

Gass, S.M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, *9*(2), 198–217.

Gass, S.M. (2018). *Input, interaction, and the second language learner*. Routledge.

Gass, S.M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Taylor & Francis.

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System*, *60*, 117–127.

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1), 1–21.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223.

Hellmuth, S. (2013). Phonology. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 45–70). Oxford University Press.

Huang, X., & Jia, X. O. (2016). Corrective feedback on pronunciation: Students' and teachers' perceptions. *International Journal of English Linguistics*, *6*(6), 245–254.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, *34*(3), 475–505.

Jesse, A., Poellmann, K., & Kong, Y. Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of Speech, Language, and Hearing Research*, *60*(1), 190–198.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566.

Ladefoged, P. (2006). *A course in phonetics*. Thomson Wadsworth.

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*(1), 35–64.

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition*, *25*(1), 37–63.

Lenth, R. (2023). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.8.6. https://CRAN.R-project.org/package=emmeans

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Long, M. H. (2007). *Problems in SLA*. Lawrence Erlbaum.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, *48*(2), 183–218.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *20*(1), 37–66.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

Mackey, A., Abbuhl, R., & Gass, S. M. (2012). Interactionist approach. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 7–23). Routledge.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS–NNS and NNS–NNS adult and child dyads. *Language Learning*, *53*(1), 35–66.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Interspeech* (Vol. *2017*, pp. 498–502).

Mustafawi, E. (2018). Arabic phonology. In E. Benmamoun, & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 11–31). Routledge.

O'Grady, W. (2015). Processing determinism. *Language Learning*, *65*(1), 6–32.

Oliver, R. (2009). How young is too young? Investigating negotiation of meaning and feedback in children aged five to seven years. In A. Mackey, & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 135–156). Routledge.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, *36*(4), 573–595.

Parlak, Ö. (2024). The effects of implicit corrective feedback on production of lexical stress in L2 English. *Studies in Second Language Learning and Teaching*, *14*(4), 661–686.

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition*, *39*(2), 257–285.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E. & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626.

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language learning*, *63*(3), 499–529.

Saito, K. (2021). Corrective feedback and the development of L2 pronunciation. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge University Press.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217–240.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, *16*(3), 371–391.

Toth, P. D., Wagner, E., & Moranski, K. (2013). 'Co-constructing' explicit L2 knowledge with high school Spanish learners through guided induction. *Applied Linguistics*, *34*(3), 279–303.

West, B. T., Welch, K. B., & Galecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software*. CRC Press.

Wiley, R. W., & Rapp, B. (2019). Statistical analysis in small-n designs: Using linear mixed-effects modeling for evaluating intervention effectiveness. *Aphasiology*, *33*(1), 1–30.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336.

Yan, X., & Ginther, A. (2017). Listeners and raters: Similarities and differences in evaluation of accented speech. In O. Kang, & A. Ginther (Eds.), *Assessment in Second language pronunciation* (pp. 67–88). Routledge.

Yang, J. (2016). Learners' oral corrective feedback preferences in relation to their cultural background, proficiency level and types of error. *System*, *61*, 75–86.

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, *38*(3), 553–586.

Zielinski, B.W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*, 69–84.

Zuraiq, W., & Sereno, J. (2021). Production of English lexical stress by Arabic speakers. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 290–311). Cambridge University Press.

Address for correspondence

Özgür Parlak
Department of English
American University of Sharjah
26666, Sharjah
United Arab Emirates
oparlak@aus.edu
https://orcid.org/0000-0003-3174-8700

Co-author information

Nicole Ziegler
Department of Second Language Studies
University of Hawai'i at Mānoa
nziegler@hawaii.edu

Reem Rabea
Department of English
American University of Sharjah
rrabea@aus.edu

Publication history

Date received: 4 October 2024
Date accepted: 1 February 2025
Published online: 24 February 2025