Mora-Plaza, I. (2023). Task-based pronunciation teaching helps to improve L2 vowel production: Generalisation effects. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 172–187). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225325

Task-based pronunciation teaching helps to improve L2 vowel production: Generalisation effects

Ingrid Mora-Plaza
University of Barcelona

A number of task-based pronunciation teaching interventions have been shown to raise learners' awareness of relevant properties of L2 speech input during interaction (e.g., Solon et al., 2017). However, it still remains unclear whether pronunciation gains are generalisable to diverse lexical contexts (unfamiliar tokens) and elicitation modes (words vs. sentences). This study investigates whether a focus on phonetic form improves learners' production of English /iː – ɪ/ and /æ – ʌ/ and leads to generalisation effects and retention. Sixty-three L1 Catalan/Spanish EFL learners carried out 20 dyadic, problem-solving tasks over 7 weeks. Task completion required the distinction of the target lexical items (e.g., *bean – bin, cat – cut*). Gains in production and generalisation effects were assessed through delayed word and sentence repetition tasks, and Mahalanobis distances were measured between confusable vowels and between learners' and native speakers' productions. Results showed that not only did learners increase the qualitative distance between the target confusable vowels, but their L2 vowel productions also approximated those of native speakers at post-test in words produced in isolation and in sentences. In addition, gains generalised to untaught tokens and improvement was retained after 11 weeks.

Keywords: pronunciation instruction, task-based language teaching, English vowel production, generalisation effects, retention

1 Introduction

Learning foreign language (FL) phonology is an arduous task because input in formal FL contexts is generally insufficient, for instance, typically limited to a few hours of language-focused instruction per week (Muñoz, 2014) and often L1-accented. Consequently, L2 categorical perception, speech segmentation, and lexical activation and retrieval processes are inevitably affected by the already-automated L1 perceptual system. According to the Perceptual Assimilation Model (PAM-L2; Best & Tyler, 2007), L1-based perception causes difficulties in phonetic learning, especially when phonetically similar L2 sounds are perceptually mapped onto single L1 sound categories, making L2 sound contrasts confusable.

At the phonetic processing level, the English vowel contrast /iː – ɪ/ is very challenging for Catalan/Spanish speakers because they only have one single high vowel /i/ with no comparable tense–lax distinction. In PAM-L2 terms, Catalan/Spanish speakers assimilate English /iː/ and /ɪ/ (e.g., *sheep – ship*) to Catalan/Spanish /i/ via a category-goodness assimilation pattern. Thus, whilst the English vowel /iː/ has been found to be highly similar to the Catalan /i/, the English vowel /ɪ/ has been perceived as a poorer fit for the same L1 sound, being identified as Catalan /i/ or /e/ (Cebrian, 2021). The English vowel contrast /æ – ʌ/ is also difficult to perceive and produce for Catalan/Spanish speakers due to the presence of the Catalan low central vowel /a/ (Cebrian, 2021). PAM-L2 would classify it as a single-category assimilation, as English /æ/ and /ʌ/ are assimilated to Catalan /a/.

Despite the difficulty in developing learners' L2 phonological awareness in a FL context, directing learners' attention to phonetic form through various training and instructional techniques has proved effective in developing the speech perception and production of L2 learners with different proficiency levels and L2 experiences (Lee et al., 2015). Few studies have explored the role of tasks in generating a focus on phonetic form during interaction (Gurzynski-Weiss et al., 2017), as well as the extent to which L2 pronunciation gains generalise to different contexts and different speakers/voices and remain over time.

1.1 Attention to phonetic form

In order to acquire new speech sounds, it is indispensable for learners to notice and pay attention (Schmidt, 1990) to cross-linguistic differences between L1 and L2 phonologies. One way to achieve this is by explicitly instructing learners to attend to specific aspects of the speech input (Guion & Pederson, 2007) while ignoring others. Lab-based high-variability phonetic training (HVPT) can be used to raise learners' awareness of L2 phonology and its gains have been shown to be robust, generalising to new lexical items and speakers (Thomson, 2018).

Drawing attention to phonological form through explicit pronunciation instruction has also been found to be effective (Lee et al., 2015), as it helps learners notice the difference between their own productions and those of more proficient L2 speakers. However, several studies have found Focus on Form (FonF) instruction to be more effective than Focus on FormS (FonFS) instruction[1] in developing intelligibility, comprehensibility and L2 pronunciation accuracy (e.g., Darcy et al., 2021). Saito's (2012) synthesis of 15 quasi-experimental studies showed that, whereas FonFS interventions resulted in improvement only at a controlled level, FonF interventions enabled learners to improve at both controlled and spontaneous speech levels.

[1] Saito (2012) refers to FonF instruction when learners practise pronunciation form while being involved in contextualised meaning-oriented communicative activities, and FonFS instruction when learners are asked to practise the accurate use of pronunciation form via mechanical drills and choral repetition.

1.2 Task-based instruction and pronunciation

Task-based language teaching (TBLT), also known as task-based instruction (TBI), is a teaching approach that focuses on having learners complete meaningful tasks using authentic language input. It aims to draw their attention to a particular linguistic structure during interaction, with tasks specifically designed to offer opportunities for practising the target structure. It is believed that real-world interaction encourages learners to refine and restructure their inter-language by drawing their attention to linguistic code features during negotiation for meaning (Long, 2015). Following Ellis (2009), tasks may direct learners' attention to meaning while predisposing them to focus on challenging L2 phonological forms through task-essentialness (Loschky & Bley-Vroman, 1993).

Whereas most research examining the facilitative role of tasks has focused on grammar, lexical structures or pragmatics, only a few TBLT studies have investigated the effectiveness of tasks in drawing learners' attention to phonetic form during communicative task performance, by testing learners' L2 pronunciation improvement (Gurzynski-Weiss et al., 2017). Although task-based pronunciation teaching (TBPT) has been found to be beneficial for L2 pronunciation development (Mora-Plaza et al., 2018), it remains unclear to what extent TBPT helps L2 pronunciation development over time and whether gains may be transferred to new lexical contexts. Furthermore, little is known about how L2 vowel production accuracy may vary depending on the context where vowels are embedded, namely, in isolated words or in sentences (but see Mora et al., 2022).

1.3 Research questions

The present study extends this line of research by assessing the effects of form-focused instruction on the production of English high /iː, ɪ/ and low /æ, ʌ/ vowels in communicative decision-making tasks during a longitudinal intervention. Improvement in L2 vowel production was assessed in terms of elicitation mode and generalisation to untaught tokens. The present study is therefore guided by the following research questions:

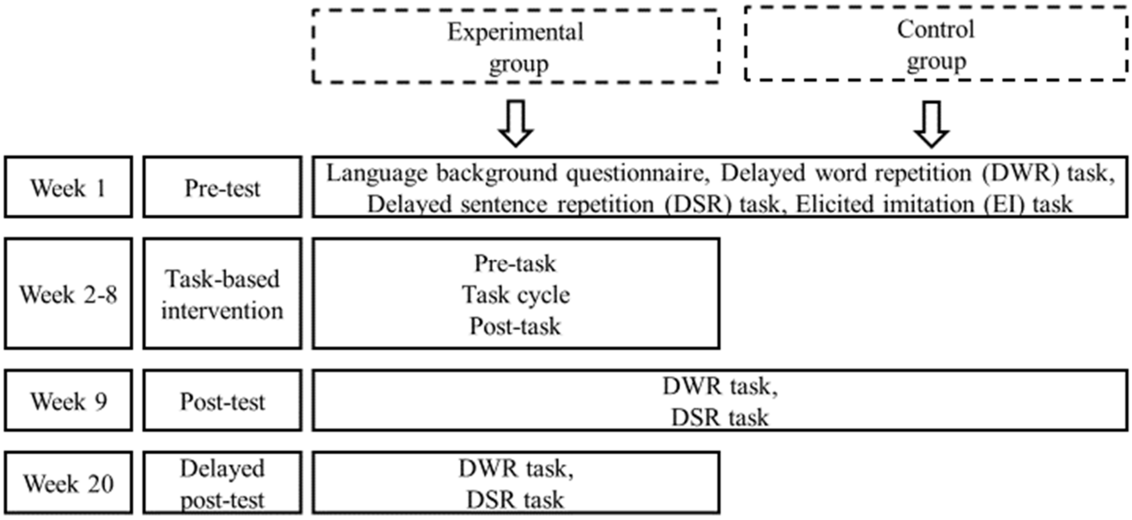
- **RQ1**: Does TBPT improve learners' production of L2 vowels embedded in words elicited in isolation?
- **RQ2**: Does TBPT improve learners' production of L2 vowels embedded in words elicited in sentences?
- **RQ3**: Are gains in L2 vowel production generalised to untaught tokens elicited in isolation and in sentences?

2 Methods

This study followed a pre-/post-test design, with a delayed post-test that learners performed 11 weeks after the task-based intervention. All participants carried out a battery of perceptual (discrimination, lexical sensitivity) and production (delayed word, sentence) tasks targeting the English vowel contrasts /iː – ɪ/ and /æ – ʌ/ individually before and after 7 weeks. A language background questionnaire and an elicited imitation (EI) task were administered before the first testing session, in order to obtain learners' biographical information and L2 proficiency level (Figure 1). In the EI task, learners were instructed to repeat 30 sentences after a beep signal. The sentences increased in grammatical and lexical complexity. Learners' productions were recorded and assessed for accuracy following Ortega et al.'s (2002) rubrics, where each sentence received a score from 0 to 4. Individual scores could thus range from 0 to 120 points.

Figure 1

Research Design



The experimental group was exposed to a series of task-based lessons three times per week. Learners dedicated the first 20–30 minutes of their English class to practising L2 reading, listening, writing and/or use of English and the last 30–40 minutes were devoted to doing pronunciation-focused oral tasks in pairs. The control group completed the pre- and post-test, and continued with their regular English classes, without taking part in any task-based intervention.

2.1 Participants

Ninety-two Catalan/Spanish English as a Foreign Language (EFL) learners from a public high school participated in the study. They belonged to three intact classes selected for convenience, as it would have been logistically impossible to randomly assign participants to different groups. The number of males (M) and females (F) was balanced across groups (Experimental group: M = 33, F = 30; Control group: M = 14, F = 15) and their age ranged from 16 to 17 years old. Their self-estimated English proficiency level ranged from intermediate to upper-intermediate (see Table 1). The experimental and control groups were not significantly different in terms of demographic and linguistic variables ($p > .05$).

Table 1

Participants' Demographic and Linguistic Information

	M	*SD*	Range	95% Confidence Interval (CI)	
				Lower	Upper
Experimental group (*n* = 63)					
Age (years)	16.0	0.2	16-17	16.0	16.1
Age of onset (years)	5.6	1.9	3-9	5.2	6.1
L2 instruction (years)	10.3	1.9	7-13	9.8	10.7
L2 use (hours/week)	3.5	3.1	0-14	2.7	4.1
Self-estimated proficiency [a]	6.0	1.6	1-9	5.3	6.2
L2 proficiency [b]	71.2	20.0	32-113	67.6	79.1
Control group (*n* = 29)					
Age (years)	16.0	0.3	16-17	15.9	16.1
Age of onset (years)	6.1	1.6	3-10	5.5	6.7
L2 instruction (years)	9.9	1.6	6-13	9.2	10.4
L2 use (hours/week)	2.7	1.5	0-6	2.1	3.2
Self-estimated proficiency [a]	5.7	1.7	1-9	5.0	6.3
L2 proficiency [b]	71.7	21.2	35-116	63.3	80.1

[a] Averaged self-estimated ability to speak spontaneously, understand, read, write, and pronounce in English.
[b] Obtained through an elicited imitation (EI) task (Ortega et al., 2002).

2.2 Task-based intervention: Stimulus materials and procedure

Over 7 weeks, learners in the experimental group performed a sequence of 20 tasks that simulated an end-of-the-course trip to London. Tasks contained 80 consonant-vowel-consonant (CVC) words, coming from 24 minimal pairs containing the two target contrasts /iː – ɪ/ and /æ – ʌ/ as well as other words containing the four target vowels without being minimal pairs. Half of the words were monosyllabic and half were disyllabic (Appendix A). Each contrast was presented once in every task, and the stimuli consisted of six minimal pairs and eight extra words containing the target vowels. Four Southern Standard British English (SSBE) speakers recorded the aural input of the listening comprehension activity in the pre-task phase. Tasks were designed following Willis' (1996) framework for task-based learning, namely, a three-phase framework corresponding to a pre-task, task cycle, and language focus stage (see examples in Mora-Plaza, 2021; Mora-Plaza et al., 2022).

The pre-task lasted for 10 minutes. First, the teacher presented the topic area of the session through an illustration of a real-life activity (e.g., packing a suitcase) and elicited students' experiences. Relevant words and expressions were written on the whiteboard and noted down by students. Then, students listened to a conversation which replicated the task each pair was going to carry out during the task cycle. Students listened for overall comprehension, then they had to pay attention to certain words which were inserted, in order to trigger a focus on phonetic form (e.g., *cap* /kæp/ and *cup* /kʌp/). Finally, the teacher showed a picture of the target object on screen and elicited the word. Students had to guess the word and listen to it. The teacher set the goal of the main task and gave planning time to prepare for it.

The task cycle lasted approximately 15–20 minutes and consisted of 20 problem-solving tasks which were two-way, close, and convergent. Additionally, tasks were "designed to provide opportunities for communicating using some specific linguistic feature" (Ellis, 2009, p. 223); namely, learners had to be able to distinguish L2 vowel contrasts (/iː – ɪ/ or /æ – ʌ/) in order to perform the task successfully (i.e., task essentialness; Loschky & Bley-Vroman, 1993). The task cycle consisted of three different phases:

1. Task – Students performed the task in dyads and the teacher monitored students' performance. The teacher made sure students were using the L2 and promoted spontaneous talk and confidence building.
2. Planning – Students rehearsed the outcome of the task and organised their discourse before presenting it in front of their classmates.
3. Report – Students presented their reports of the task in front of their classmates. The teacher noted down any inaccuracies related to the target vowels, to comment on them during the language focus stage.

The language focus stage lasted for 5–10 minutes. During this stage, the teacher prepared some language-focused tasks to consolidate the phonological contrasts encountered through communication during the task cycle. In the analysis stage, students did consciousness-raising activities and, in the practice stage, they consolidated the target pronunciation features through communicative tasks.

2.3 Assessment

Learners' L2 phonological knowledge was assessed through perception and production tests. In this paper, we will only report the results of L2 vowel production. Learners produced the L2 target vowels through delayed word repetition (DWR) and delayed sentence repetition (DSR) tasks, which were administered in DMDX[2] (Forster & Forster, 2003) on a laptop computer.

In the DWR task, participants heard the word (e.g., /kæt/) followed by a 1500ms pause, before a tone signal prompted them to repeat it. In order to test for generalisation effects, the testing stimuli comprised 24 taught and 24 untaught words, and were produced by two speakers that participants had not been exposed to during the intervention. The test consisted of a total of 68 trials (64 test and 4 practice trials). The testing stimuli were 32 monosyllabic and 32 disyllabic words: 48 were words from minimal pairs /iː – ɪ/ and /æ – ʌ/, and 16 words had no contrasting counterpart but contained the 4 target vowels and also had appeared during the intervention (Appendix B).

In the DSR task, learners were asked to: 1) read the sentence appearing in standard orthography on the computer screen for 3000ms; 2) listen to the sentence over the headphones; and 3) repeat the sentence from memory after a sound signal occurring 1500ms after the offset of the sentence stimulus. The DSR stimuli were identical to the DWR ones. Learners were exposed to 64 test sentences and four practice sentences, which were four words long. They were always formed by the determiner/pronoun THE/THEY + TARGET WORD containing /iː, ɪ, æ, ʌ/ + VERB + OBJECT (e.g., *The bin is empty*). All test words (DWR) and sentences (DSR) were distributed into two separate randomised blocks (1st block /iː/, /ɪ/; 2nd block /æ/, /ʌ/) with 32 stimuli each and a short break in between.

[2] DMDX (Forster & Forster, 2003) is a Win32-based software display system used in psychological and linguistic experiments for stimulus presentation (providing fast-action, dynamic gaming experience) and for measuring reaction times to visual and auditory stimuli.

2.4 Data analysis

To obtain an L2 vowel production measure, vowel frequencies were extracted in Praat. A Bark-distance normalisation procedure was used to provide speaker-independent estimates of vowel quality. The difference in Bark between F1 and *f*0 (B1-B0) estimated vowel height, whereas the difference between F2 and F1 (B2-B1) estimated vowel frontness.

Mahalanobis distances were used to calculate a measure of vowel distinctiveness (i.e., every token of vowel /æ/ and the centroid of the distribution of the tokens of the other contrasting vowel /ʌ/, and vice versa). A larger distance meant less of an overlap between the two vowels (Melnik-Leroy & Peperkamp, 2021).[3] In addition, to measure vowel nativelikeness, we calculated the distance between native speakers' and learners' productions of each target vowel produced in the same phonetic context, so a smaller distance meant a more target-like production (Kartushina et al., 2015).

In order to answer RQ1 and RQ2, linear mixed-effects models were performed in SPSS 27 with GROUP (Experimental vs. Control), TIME (pre-test vs. post-test), CONTRAST (/iː – ɪ/ vs. /æ – ʌ/) or VOWEL (/iː, ɪ, æ, ʌ/), and their interactions as fixed effects, and SUBJECT and ITEM as random intercepts. As for RQ3, we calculated gains from pre-test to delayed post-test scores for the experimental group only. In a linear-mixed effects model, we included TOKEN TYPE (taught vs. untaught) as fixed effects and a random intercept for SUBJECT. The parameter estimates are given in Appendix C.

3 Results

RQ1 queried the effects of TBPT on L2 vowel production in words elicited in isolation. On the one hand, mixed effects models revealed significant main effects of TIME ($F[1, 8680] = 7.08$, $p = .008$) and CONTRAST ($F[1, 8680] = 9.17$, $p = .002$) on Mahalanobis distances between vowels /iː – ɪ/, /æ – ʌ/, i.e., distinctiveness measure. A significant GROUP × TIME interaction ($F[1, 8680] = 6.90$, $p = .009$) revealed that, at post-test, the experimental group significantly produced a larger distance between the vowels in the contrasts (pre-test: $M = 10.24$; $SD = .37$; post-test: $M = 12.96$; $SD = .32$; $p < .001$) than the control group (pre-test: $M = 10.18$; $SD = .61$; post-test: $M = 10.16$; $SD = .80$; $p = .984$). Bonferroni-adjusted pairwise contrasts from the GROUP × TIME × CONTRAST interaction ($F[1, 8680] = 1.16$, $p = .282$) showed that, while the task-based intervention seemed to have helped learners produce a larger distance between /iː/ and /ɪ/ ($t[8680] = 2.20$, $p = .027$), and /æ/ and /ʌ/ ($t[8680] = 5.00$, $p < .001$), none of the control group's contrasting vowels distinguished significantly at post-test ($p > .05$) (see Figure 2, left panel).

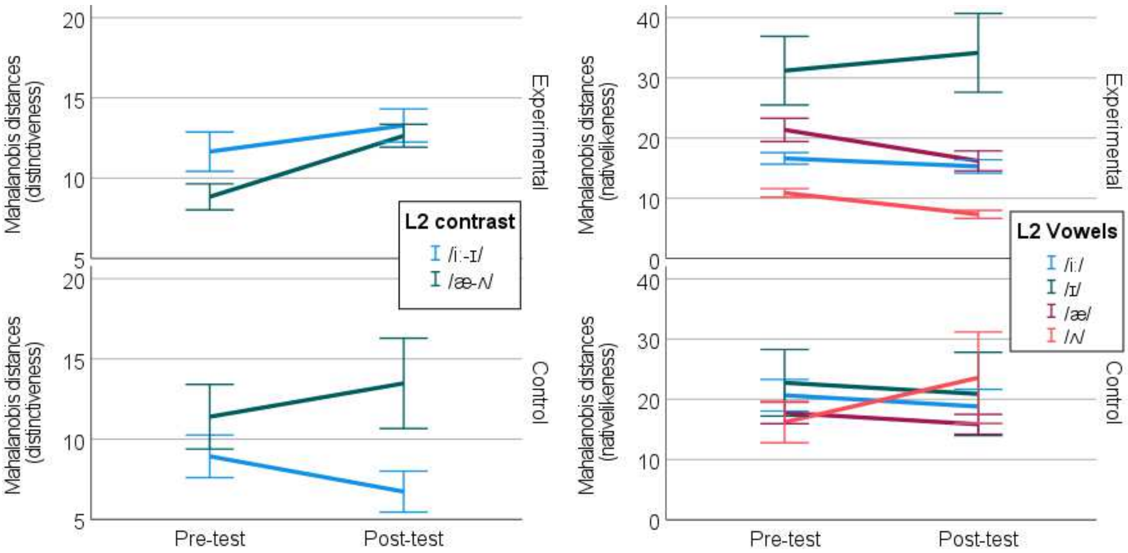
On the other hand, mixed effects models showed non-significant main effects of TIME ($F[1, 8672] = .17$, $p = .679$) and significant effects of VOWEL ($F[3, 8672] = 29.37$, $p < .001$) on Mahalanobis distances between learners' and native speakers' vowels /iː, ɪ, æ, ʌ/, i.e., nativelikeness measure. Despite the non-significant GROUP × TIME interaction ($F[1, 8672] = 2.59$, $p = .107$), Bonferroni-adjusted pairwise contrasts showed that overall the experimental group (but not the control group) had significantly shortened Mahalanobis distances compared to native speakers (i.e., the participants became more accurate) between testing times (pre-test: $M = 20.03$; $SD = .79$; post-test: $M = 18.26$; $SD = .89$; $p = .05$). As illustrated in Figure 2 (right panel), only learners' productions of /æ/ and /ʌ/ significantly approximated the values of native speakers' vowel productions. Finally, learners in the experimental group kept separating the confusing vowels in the contrasts (/iː – ɪ/ and /æ - ʌ/) as demonstrated by the vowel

[3] Mahalanobis distances are defined as the distance in standard deviations between a point and the centroid of a distribution (Melnik-Leroy & Peperkamp, 2021).

distinctiveness measures at the delayed post-test (post-test: *M* = 12.96, *SD* = .32; delayed post-test: *M* = 13.98, *SD* = .39; *p* = .05) and maintained similar Mahalanobis nativelikeness distances (post-test: *M* = 18.26, *SD* = .89; delayed post-test: *M* = 17.34, *SD* = .83; *p* = .47), suggesting that learning was retained 11 weeks after the treatment.

Figure 2

Mahalanobis Distances for Distinctiveness (on the Left) and for Nativelikeness (on the Right) Produced in Words in Isolation



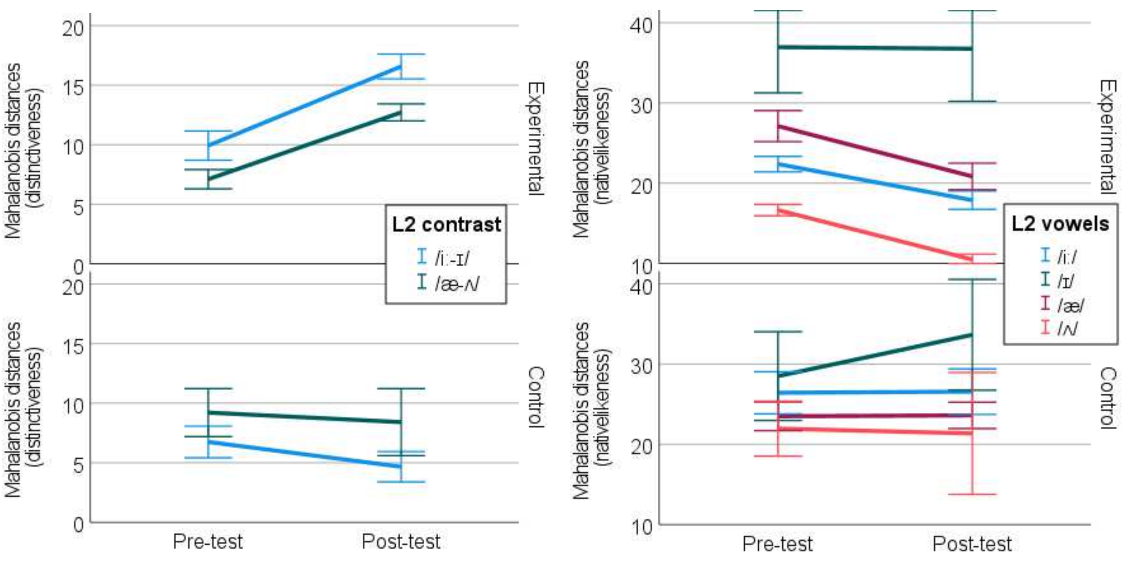
Note. The graphs are organised by GROUP (Experimental vs. Control) and TIME (pre-test vs. post-test) in the X-axis. Error bars show 95% CI.

RQ2 also asked about the effects of TBPT on L2 vowel production but this time in words produced in sentences. In terms of Mahalanobis distances of vowel distinctiveness, mixed effects models revealed significant main effects of TIME (F[1, 8680] = 21.31, $p < .001$) and CONTRAST (F[1, 8680] = 20.55, $p < .001$). In addition, results from the GROUP × TIME interaction (F[1, 8680] = 49.41, $p < .001$) indicated that the experimental group produced significantly greater distances between the confusing vowels for the two contrasts after the task-based intervention (pre-test: *M* = 8.52; *SD* = .37; post-test: *M* = 14.64; *SD* = .32; $p <$ .001), whereas the control group did not make any significant distinction between any of the target vowels in the contrasts (pre-test: *M* = 8.00; *SD* = .61; post-test: *M* = 6.57; *SD* = .79; *p* = .157) (see Figure 3, left panel).

Concerning Mahalanobis nativelikeness distances, linear mixed models revealed non-significant effects of TIME (F[1, 8675] = .17, *p* = .241) and significant effects of VOWEL (F[3, 8675] = 45.20, $p < .001$). Furthermore, a significant GROUP × TIME interaction (F[1, 8675] = .17, *p* = .241) arose because, according to Bonferroni-adjusted pairwise contrasts, experimental learners' vowels became more target-like (i.e., the distance to native speaker values was reduced) from pre- (*M* = 25.78, *SD* = .79) to post-test (*M* = 21.50, *SD* = .89; $p <$ .001), but the control group's vowel productions did not change significantly across times (pre-test: *M* = 25.05, *SD* = .91, post-test: *M* = 26.23, *SD* = 1.37; *p* = .268). As seen in Figure 3 (right panel), learners' vowel qualities became significantly more accurate after the intervention, except for vowel /ɪ/ (p = .798). Lastly, evidence for retention effects was found as learners in the experimental group still distinguished the confusing vowels in the contrasts at the delayed post-test (post-test: M = 14.64, SD = .32; delayed post-test: M = 13.75, SD = .39; p = .06) and did so in a native-like direction (post-test: M = 21.50, SD = .89; delayed post-test: M = 20.17, SD = .83; p = .29).

Figure 3

Mahalanobis Distances for Distinctiveness (on the Left) and for Nativelikeness (on the Right) Produced in Words in Sentences



Note. The graphs are organised by GROUP (Experimental vs. Control) and TIME (pre-test vs. post-test) in the X-axis. Error bars show 95% CI.

Interestingly, gains in distinctiveness were moderately correlated with gains in nativelikeness when vowels were embedded in words produced in isolation (r = .330, p = .008) and in sentences (r = .429, p = .001). This suggests that, overall, learners who produced more distinct vowel qualities, also produced vowels that were more target-like.

Finally, RQ3 looked into the comparison between vowel production gains for taught and untaught tokens in two different contexts: words and sentences. On the one hand, learners obtained similar gains in taught (M = 4.06; SD = .94) and untaught (M = 3.40; SD = .78) tokens produced in isolation with regards to vowel distinctiveness (F[1, 501] = .28, p = .593). In contrast, gains for vowel nativelikeness were greater (albeit non-significantly; F[1, 501] = 1.11, p = .291) for those words that had not been taught and were unfamiliar to learners than taught tokens. Likewise, Mahalanobis distance gains between vowels produced in sentences, were similar when they appeared in taught (M = 5.55; SD = .96) and untaught (M = 4.90; SD = .79) tokens (F[1, 501] = .26, p = .606). Gains in how much learners approximated native speakers' vowel qualities were larger in untaught than taught tokens, but the effects of TOKEN TYPE did not reach significance (F[1, 499] = 1.12, p = .293).

4 Discussion

Results from this investigation suggest that carefully designing and manipulating tasks induces a focus on phonetic form during meaningful interaction, and generating a linguistic focus through task-essential language raises learners' awareness about challenging L2 pronunciation features, eventually leading to more accurate vowel production (see Solon et al., 2017). Learners who took part in the task-based intervention produced L2 vowels more contrastively and more accurately than learners who did not. In word-elicitation contexts (RQ1), the overlap between confusable vowels became significantly smaller for both contrasts and L2 vowels became more target-like. Similarly, in sentence-elicitation contexts (RQ2), learners produced L2 vowels more distinctively and accurately after the task-based intervention. This was not the case for the control group, whose L2 vowel qualities remained stable. Whereas producing words in isolation may have led to more conscious reflections on form (thus emphasising the distinctiveness of vowels in terms of spectral distances), producing the target words embedded in sentences may have mirrored the occurrence of such forms during the interactive tasks, in a more realistic context, where vowel differences relied mainly on quality (Mora et al., 2022). Changes in vowel distinctiveness were significantly associated with changes in nativelikeness, meaning that TBPT helped learners align their initially unstable vowel productions with those of native speakers of English. While the goal of the TBPT intervention was not to achieve a native-like accent, approximating native vowels may have helped learners to become more intelligible, hence, to produce L2 vowels that were sufficiently distinct in order not to confuse interlocutors during communication. In addition, L2 vowel learning seems to be robust, as L2 pronunciation gains were retained 11 weeks after the intervention. Finally, as found by HVPT studies (Thomson, 2018), gains in L2 pronunciation accuracy generalised to untaught tokens in isolation, as well as to tokens embedded in sentences (RQ3) and those spoken by unfamiliar speakers/voices.

Overall, findings from this study suggest that pronunciation instruction can be easily integrated in communicative tasks (in line with current pedagogical principles) by making L2 pronunciation features salient through task manipulation (e.g., making L2 vowel contrasts essential for task completion). Instead of teaching pronunciation in an explicit, often decontextualised manner, TBPT thus advocates for an analytic approach where learners deal with challenging L2 pronunciation features as they are communicating. In line with previous form-focused communicative studies (e.g., Darcy et al., 2021), exposing learners to L2 pronunciation features repetitively and in meaningful contexts results in L2 pronunciation gains; it also prepares learners for out-of-class conversations. Hypothetically, other L2 oral skills (e.g., fluency, prosody) may also develop along with segmentals while learners interact. Finally, L2 pronunciation improvement may be assessed in terms of successful task completion as well as objective and subjective pronunciation proficiency measures.

5 Conclusion

This TBPT study has shown that, despite the time constraints that teachers often suffer, L2 pronunciation can be part of a TBLT curriculum. Form-focused communicative instruction, which is based on tasks that are inherently repetitive yet genuinely communicative (see Sardegna, 2022), may enhance L2 pronunciation learning and lead to generalisation effects in diverse lexical contexts and elicitation modes. Further research should investigate how many segmental features can be addressed in a given task, as well as how often learners should practise the same minimal pairs for acquisition to occur. More broadly, future work should explore: the effects of task design and manipulation on L2 intelligibility and comprehensibility in face-to-face and online settings; how TBPT may apply to the teaching of suprasegmentals;

and which learner factors (experiential, affective, and/or cognitive) should be considered when designing tasks that promote second language pronunciation learning.

Acknowledgements

This research was funded through grant PID2019-107814GB-I00 from the Spanish Ministry of Science, Innovation and Universities. I would like to thank Miren Adrian, Daniela Avello and Valeria Galimberti for their help in data collection, and Dr. Joan C. Mora and Dr. Roger Gilabert for their constant support.

References

Best, C., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *The Journal of the Acoustical Society of America*, *149*(4), 2671. https://doi.org/10.1121/10.0004257

Darcy, I., Rocca, B., & Hancock, Z. (2021). A window into the classroom: How teachers integrate pronunciation instruction. *RELC Journal*, *52*(1), 110–127. https://doi.org/10.1177/0033688220964269

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 57–77). John Benjamins. https://doi.org/10.1075/lllt.17.09gui

Gurzynski-Weiss, L., Long A.Y., & Solon M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition*, *39*(2), 213–224. https://doi.org/10.1017/S0272263117000080

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Multilingual Matters.

Melnik-Leroy, G. A., & Peperkamp, S. (2021). High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition,* *24*(3), 497–506. https://doi.org/10.1017/S1366728920000644

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different training conditions: The use of non-lexical materials and masking noise. *Phonetica, 79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Mora-Plaza, I. (2021). Off to the library! TBLT Language Learning Task Bank. https://tblt.indiana.edu/tasks/details.html?id=73

Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis, C. Nagle, & E. Today (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 174–184). Iowa State University.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463–482. https://doi.org/10.1093/applin/amu024

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research*. Paper presented at the Second Language Research Forum, Toronto, Canada.

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

Sardegna, V. (2022). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching, 55*(3), 363–378. https://doi.org/10.1017/S0261444821000380

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347–380. https://doi.org/10.1017/S0272263116000425

Thomson, R. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Willis, J. (1996). *A framework for task-based learning*. Longman.

Appendix A

Intervention Stimuli by Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Type (Minimal Pairs vs. Extra Words also Containing the Target L2 Vowels) and Syllable (One- vs. Two-Syllable Words)

	/iː/	/ɪ/	/æ/	/ʌ/
Minimal pairs				
1 syllable	bean	bin	bag	bug
	cheek	chick	bat	butt
	feast	fist	cap	cup
	peel	pill	cat	cut
	sheep	ship	mag	mug
	teen	tin	ram	rum
2 syllables	heating	hitting	amber	umber
	keeper	kipper	ankle	uncle
	lever	liver	babble	bubble
	sleeper	slipper	batter	butter
	sneakers	Snickers	carry	curry
	weeping	whipping	natty	nutty
Extra words				
1 syllable	leave	kill	act	run
	weed	fish	hat	drum
	tea	chips	ham	bun
	jeans	pin	jam	gun
2 syllables	illegal	bitter	jacket	public
	kiwi	whiskey	baggy	nugget
	Peter	Jimmy	Patrick	Luster
	Sheila	Lily	Cathy	Sunset

Appendix B

Testing Stimuli by Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Token Type (Taught vs. Untaught, and Practice Items) and Syllable (One- vs. Two-Syllable Words)

	/iː/	/ɪ/	/æ/	/ʌ/
Taught				
1 syllable	bean	bin	bag	bug
	cheek	chick	cat	cut
	sheep	ship	ram	rum
	leave	fish	hat	drum
	tea	chips	jam	bun
2 syllables	keeper	kipper	amber	umber
	lever	liver	batter	butter
	sneakers	Snickers	carry	curry
	illegal	bitter	jacket	public

	/iː/	/ɪ/	/æ/	/ʌ/
	kiwi	whiskey	baggy	nugget
Untaught				
1 syllable	beef	biff	crash	crush
	feel	fill	lag	lug
	seal	sill	stab	stub
2 syllables	greeting	gritting	attar	utter
	litre	litter	bagger	bugger
	weaner	winner	clatter	clutter
Practice items				
	feet	hill	rat	sun

Appendix C

Parameter estimates of linear mixed-effects models for the measures of distinctiveness and of nativelikeness

RQ1: Words in Isolation

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	13.633	1.1654	11.698	0.000	11.348	15.917
Group	-0.858	1.4082	-0.609	0.542	-3.619	1.902
Time	-2.157	1.2021	-1.795	0.073	-4.514	0.199
Contrast	-6.838	1.1556	-5.917	0.000	-9.103	-4.573
Group x Time	-1.741	1.4328	-1.215	0.224	-4.550	1.068
Group x Time x Contrast	-2.102	1.9523	-1.077	0.282	-5.929	1.725

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	24.079	3.1393	7.670	0.000	17.925	30.232
Group	-17.648	3.7929	-4.653	0.000	-25.082	-10.213
Time	-8.243	3.4130	-2.415	0.016	-14.934	-1.553
Vowel	-4.969	3.3487	-1.484	0.138	-11.533	1.595
Group x Time	12.267	4.0817	3.005	0.003	4.265	20.268
Group x Time x Vowel	-11.446	5.6656	-2.020	0.043	-22.552	-0.340

RQ2: Sentences

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	15.810	1.9139	8.261	0.000	12.059	19.562
Group	1.978	2.3216	0.852	0.394	-2.573	6.529
Time	0.527	1.2041	0.437	0.662	-1.834	2.887
Contrast	-6.830	1.1772	-5.802	0.000	-9.137	-4.522
Group x Time	-6.187	1.4285	-4.331	0.000	-8.987	-3.387
Group x Time x Contrast	-2.564	1.9207	-1.335	0.182	-6.329	1.201

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	21.844	3.1414	6.953	0.000	15.686	28.002
Group	-12.259	3.7954	-3.230	0.001	-19.698	-4.819
Time	-0.259	3.4145	-0.076	0.939	-6.953	6.434
Vowel	5.016	3.3501	1.497	0.134	-1.551	11.583
Group x Time	6.804	4.0835	1.666	0.096	-1.201	14.808
Group x Time x Vowel	-0.905	5.6680	-0.160	0.873	-12.016	10.205

RQ3: Words in Isolation

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	3.566313	1.064224	501	3.351	0.001	1.475421	5.657205
Token Type	0.657299	1.228860	501	0.535	0.593	-1.757055	3.071653
Contrast	-0.320170	1.228860	501	-0.261	0.795	-2.734525	2.094184

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	5.244276	4.671135	499	1.123	0.262	-3.933241	14.421792
Token Type	-4.414901	4.177990	499	-1.057	0.291	-12.623522	3.793720
Vowel	-1.363169	5.908571	499	-0.231	0.818	-12.971912	10.245573

RQ3: Sentences

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	4.307979	1.083649	501	3.975	0.000	2.178924	6.437035
Token Type	0.645394	1.251290	501	0.516	0.606	-1.813027	3.103816
Contrast	1.198348	1.251290	501	0.958	0.339	-1.260073	3.656770

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	8.799831	4.684930	499	1.878	0.061	-0.404789	18.004451
Token Type	-4.414901	4.190329	499	-1.054	0.293	-12.647763	3.817961
Vowel	-1.363169	5.926020	499	-0.230	0.818	-13.006195	10.279856

About the author

Ingrid Mora-Plaza is a Lecturer in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona, Spain, where she teaches English phonetics and phonology courses. She is part of the Barcelona L2 speech research group and GRAL research group on the acquisition of languages. Her research focuses on L2 pronunciation learning and teaching in foreign language contexts and the effects of individual differences on L2 speech acquisition. She is especially interested in the potential benefits of task-based pronunciation teaching and high-variability phonetic training on L2 speech development in instructional settings.

E-mail: imoraplaza@ub.edu